# Methods to Estimate Dynamic Stochastic General Equilibrium Models

Francisco J. Ruge-Murcia*

First Draft: August 2002
This Version: December 2005

**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 Bayesian priors, 2) Generalized Method of Moments, 3) Simulated Method of Moments, and 4) Indirect Inference. It is shown that stochastic singularity limits the number of variables and moments that can be used for the estimation of DSGE models and, in general, affects more severely Maximum Likelihood than other estimation procedures. Monte Carlo analysis indicates that all methods deliver reasonably good estimates under the null, but that moment-based procedures are substantially more robust to misspecification than Maximum Likelihood.

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

*Départemente de sciences économiques and CIREQ, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec) H3C 3J7, Canada, E-mail: francisco.ruge-murcia@umontreal.ca. I have benefitted from conversations with Rui Castro, Wouter den Haan, Jesús Fernandez-Villaverde, Jim Hamilton, Frank Schorfheide, and Chris Sims. 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 of Canada is gratefully acknowledged.

# 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 those of actual data. The model is usually evaluated in terms of the distance between these two set of moments. This comparison may 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 may 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 Indirect Inference procedure proposed by Smith (1993). All these procedures are standard and their asymptotic properties are well known. 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 examine their robustness to misspecification, to compare their computational costs, and to discuss fully the implications of the stochastic 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.

A feature of linearized DSGE models that has important implications for all estimation methods is their stochastic singularity. Linearized DSGE models are singular because they generate predictions about more observable variables than exogenous shocks are specified in the model.[2] Thus, these models imply that certain linear combinations of observable variables should hold without noise. This implication is not satisfied by actual data and arises from a particular form of misspecification, namely that models assume a smaller number of shocks than are present in the real world.

Stochastic singularity limits the number of variables and moments that can be used for the estimation of the model, and imposes restrictions on both the order and the number of variables in the VAR representation of artificial data generated by a DSGE model. In general, singularity affects more severely Maximum Likelihood than the other estimation methods. ML estimation is limited by the number of linearly independent variables, while Method of Moments and Indirect Inference are limited by the number of linearly independent moments. That latter is a weaker restriction because it is possible to find independent moments that incorporate information about more variables than those that are linearly independent. For example, the RBC model studied here cannot be estimated by Maximum Likelihood using more than one variable, unless measurement errors are added, but it can be estimated by the Method of Moments or Indirect Inference using moments that involve two variables. This observation has two implications. First, the efficiency gain of using ML rather than Method of Moments or Indirect Inference may be modest. Second, identification may be sharper under either of the latter procedures than under ML because information on a larger set of variables can be employed.

Monte Carlo analysis is used to study the small-sample properties of the estimators

[2]Although strictly-speaking nonlinear DSGE models are not singular, the discussion in Section 3.4 makes clear that their estimation may also be affected depending on the extent to which they differ from their linearized counterparts.

under the null, and to examine their robustness to misspecification. Results indicate that all procedures deliver reasonably good estimates under the null, but that statistical inference may be subject to size distortions because asymptotic standard errors are a poor measure of the small-sample variability of the estimates. Results also show that the Method Method of Moments (and to a lesser extent, Indirect Inference) are substantially more robust to misspecification than Maximum Likelihood.

Finally, this paper also studies the effect of adding measurement errors to sidestep the singularity of the model and the use of Bayesian priors that incorporate information from 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 classical Maximum Likelihood 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 one-sector Real Business Cycle model with indivisible labor (Hansen, 1985).[3] The representative agent maximizes expected lifetime utility

$$U_t = E_t \sum_{i=t}^{\infty} \beta^{i-t} \left( \ln(c_i) + \psi(1 - n_i) \right),$$

where $\beta \in (0,1)$ is the discount factor, $c_t$ is consumption, $n_t$ is hours worked and $\psi$ is a utility weight. There is no population growth. The population size and time endowment are normalized to one. 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 units of the consumption good.

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

Investment increases the capital stock 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

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

The single, perishable good in this economy is produced by perfectly competitive firms using a constant returns to scale technology. 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 $\epsilon_t$ is an innovation assumed to be independently, identically, and Normally distributed with zero mean and variance $\sigma^2$. 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, aggregate output equals aggregate demand

$$y_t = c_t + x_t.$$

A common 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 deterministic steady state. The linearized equations for this model are shown in the Appendix. These equations form a dynamic system that determines the path of consumption, capital, output, investment, hours worked, and the technology shock.[4] Using the circumflex to denote percentage deviation from steady state and after some manipulations, it is possible to

[4]For convenience, I have substituted out the wage and rental rate by equating them to the marginal products of labor and capital, respectively.

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

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

$\varsigma = \alpha\beta(k/n)^{\alpha-1}$, $k/n = ((1/\beta+\delta-1)/\alpha)^{1/(\alpha-1)}$ 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 standard methods. Here, I use the approach in Blanchard and Kahn (1980) 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 $\phi_{ck}$ and $\phi_{cz}$ are combinations of the eigenvectors and eigenvalues of the matrix $\mathbf{A}$ and, consequently, depend nonlinearly on the structural parameters.

Define the $2\times 1$ vector $\boldsymbol{\xi}_t = (\hat{k}_t, \hat{z}_t)'$ with the state variables of the system, and the $3\times 1$ vector $\mathbf{s}_t = (\hat{n}_t, \hat{y}_t, \hat{c}_t)'$ with the observable variables that the researcher will use in the estimation of the model. Using the linearized equations of the model, write the components of $\mathbf{s}_t$ as functions of the capital stock and technology shock alone

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

where the elements of the $3\times 2$ matrix $\boldsymbol{\Phi}$ are nonlinear functions of the structural parameters of the model, and the last row of $\mathbf{s}_t$ reproduces exactly equation (2).

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. Hence, the model is stochastically singular. 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,

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

Combining (4) and (5), it is easy to see 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 variance-covariance matrices of these systems are singular. It is shown below that the stochastic singularity of DSGE models has different and nontrivial implications for the actual application of each of the estimation methods studied in this paper.

# 3 Estimation Methods

## 3.1 Maximum Likelihood

The Maximum Likelihood (ML) estimation of DSGE models requires the construction and evaluation of the likelihood function of the data given the parameters. This paper considers the case where the model involves unobservable state variables and, consequently, filtering techniques are required to evaluate the likelihood function. In particular, I follow Ireland (2004) in treating the capital stock as unobservable.[5] With unobservable state variables, one can exploit the recursive nature of the model and its fully-specified laws of motion to apply filters like the one proposed by Kalman (1960).[6] The Kalman filter allows the construction of inferences about the unobserved state vector and the evaluation of the joint likelihood function of observable endogenous variables. In turn, the maximization of this function yields efficient, consistent and asymptotically Normal estimates of the model parameters.

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. Then,

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

where

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

[5]This assumption is made throughout the paper. Thus, the difference in Monte Carlo results across estimation methods is not due to the way in which the capital stock is treated. The case where the capital stock is assumed to be observable is a straightforward simplification of the setup considered here. See, for example, Christiano (1988), Altug (1989) and Bencivenga (1992), who employ the Maximum Likelihood procedure proposed by Hansen and Sargent (1980).

[6]As an alternative, Fernández-Villaverde and Rubio (2004) propose the use of a nonlinear Sequential Monte Carlo filter to evaluate the log likelihood function of non-linearized DSGE models.

is a $2 \times 2$ matrix and $\mathbf{v}_t = (0, \boldsymbol{\epsilon}_t)'$ is a $2 \times 1$ 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 $1 \times 3$ selection vector. For example, in the case where the model is estimated using output alone, $\mathbf{h} = (1, 0, 0)$ and $\mathbf{x}_t = \hat{y}_t$. The reason why $\mathbf{x}_t$ contains only one variable will become clear below.

Collect the parameters of the model in the $q \times 1$ vector $\boldsymbol{\theta}$. Denote by $\aleph_{t-1} = (\mathbf{x}_{t-1}, \ldots, \mathbf{x}_1)$ the set of past observations of $\mathbf{x}_t$, by $\tilde{\boldsymbol{\xi}}_{t|t-1}$ the time $t-1$ forecast of $\boldsymbol{\xi}_t$ constructed on the basis of $\aleph_{t-1}$, and by $\mathbf{P}_{t|t-1}$ the Mean Square Error (MSE) of this forecast. Then, under the assumption that the technology innovation, $\epsilon_t$, 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, Griffiths, Hill, Lütkepohl and Lee, 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.

Due to their stochastic singularity, DSGE models cannot be estimated by Maximum Likelihood using more observable variables than structural shocks are specified in the model. To illustrate this point, consider the (one-shock) model in Section 2 and write the innovation to $\mathbf{x}_t$

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

Note that when $\mathbf{x}_t$ contains more than one variable (for example, consumption and output), the matrix $Var(\mathbf{x}_t - E(\mathbf{x}_t|\aleph_{t-1})) = \mathbf{HP}_{t|t-1}\mathbf{H}'$ is singular because the innovations to the variables in $\mathbf{x}_t$ are all proportional to the technology innovation and, consequently, perfectly correlated. Thus, the likelihood function is not well defined. While the model studied here cannot be estimated by Maximum Likelihood using more than one variable, we will see below that it can be estimated by Method of Moments or by Indirect Inference using moments from more than one variable. In this sense, stochastic singularity affects Maximum Likelihood more severely than the other estimation methods.

There are at least two strategies to deal with singular DSGE models in the Maximum Likelihood framework. First, one can estimate the model using at most as many observable variables as structural shocks. This approach is followed by Kim (2000), Ireland (2001) and Bouakez, Cardia and Ruge-Murcia (2005). 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 (2004).[7]

### 3.1.1 Adding Measurement Errors

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

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

where $\mathbf{x}_t$ is now a $d \times 1$ vector, $d$ 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 $E(\mathbf{u}_t\mathbf{u}_t') = \mathbf{R}$. Following Sargent (1989), it is common 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 structural shocks. The extension to serially correlated errors is straightforward and is discussed in Hansen and Sargent (1998, ch. 8).

[7] Another alternative is to extend the model to permit additional structural shocks. This strategy is attractive because it increases the realism of the model and allows the use of more observable variables in the estimation of the model. Early advocates of this strategy include Leeper and Sims (1994) and Ingram, Kocherlakota and Savin (1994). In general, adding structural shocks postpones, but does not necessarily solve, the stochastic singularity of DGSE models. For example, incorporating a government expenditure shock into the RBC model would add a second structural disturbance, but also would add another observable variable (*i.e.,* government expenditure). However, adding structural shocks can solve the singularity of a model if a sufficient number of these shocks are latent. See Bergin (2003) for an application of this idea. Another issue is that while adding measurement errors preserves the original economic model, adding structural errors does not. The reason is that decision rules in the extended model would depend on a larger set of state variables than in the original model.

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

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

but now the log likelihood function is

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

Since $\mathbf{R}$ is full rank by assumption, $\mathbf{H}\mathbf{P}'_{t|t-1}\mathbf{H}' + \mathbf{R}$ will not be singular when the number of observable variables included in $\mathbf{x}_t$ is larger than the number of structural shocks.

Adding extra error terms to the observation equation is a simple way to sidestep the stochastic singularity of DSGE models and may provide a less stringent platform to assess the theory. Relationships like (4) and (5) may hold approximately in the data, even if not without noise as predicted by the model. On the other hand, 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 shocks than those assumed in the model. This is the main specification error that is captured by $\mathbf{u}_t$. When $\mathbf{u}_t$ is modeled as serially and/or contemporaneously correlated, other forms of misspecification may be captured by this term as well. Ireland (2004) interprets measurement error as a mean to capture movements and comovements in the data that the model cannot explain. Finally, appending measurement errors may complicate the identification of the parameters of the process of the structural shocks because the disturbances in the observation equation are linear combinations of structural shocks and measurement errors (see Boivin and Giannoni, 2005).

#### 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 $\aleph_T = (\mathbf{x}_T, \mathbf{x}_{T-1}, \ldots, \mathbf{x}_1)$ to obtain an estimate of $\boldsymbol{\theta}$ as in

$$\tilde{\boldsymbol{\theta}}_{qb} = \max_{\{\boldsymbol{\theta}\}} \left(L(\boldsymbol{\theta}) + L(\boldsymbol{\mu}(\boldsymbol{\theta}))\right), \tag{9}$$

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 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 is equal to the mean and, consequently, the point estimate of $\boldsymbol{\theta}$ is the same as that that would be 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.

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 $\tau T$. The SMM estimator, $\tilde{\boldsymbol{\theta}}_{smm}$, is the value of $\boldsymbol{\theta}$ that solves

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

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

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

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 $\tau T$ data points. $\mathbf{W}$ is computed using the Newey-West estimator with a Barlett kernel. When the model is overidentified (that is,

$p > q$), a general specification test of the model may be constructed using the chi-square statistic proposed by Hansen (1982). In the case of SMM

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

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.

The stochastic singularity of DSGE models also affects their estimation by SMM. In particular, singularity restricts the variables that can be included in $\mathbf{m}_i$ or, equivalently, the moments that can be exploited for SMM estimation. For example, multiply (4) by $\hat{y}_t$ and note that the variables $\hat{n}_t\hat{y}_t$, $\hat{y}_t^2$, and $\hat{c}_t\hat{y}_t$ are linearly dependent. Since this is true for all $t$ and $\boldsymbol{\theta}$, it follows that (under the model) the variance of output, the covariance of output with hours, and the covariance of output with consumption are linearly dependent. If one were to include $\hat{n}_t\hat{y}_t$, $\hat{y}_t^2$, and $\hat{c}_t\hat{y}_t$ in $\mathbf{m}_i$, the Jacobian matrix $\mathbf{D}$ would not be of full rank, Assumption 6 in Duffie and Singleton (1993, p. 944) would not be satisfied, and the asymptotic distribution of SMM estimates would not be well defined.[8]

Singularity has different implications for SMM and Maximum Likelihood. SMM estimation is limited by the number of linearly independent moments. ML estimation is limited by the number of linearly independent variables. The former is a weaker restriction because it is possible to find independent moments that incorporate information about more variables than those that are linearly independent. For example, for the SMM estimation of the RBC model in Section 2, it is easy to find independent moments that involve up to two variables, even though (conditional on the predetermined stock of capital) no two variables are linearly independent. The fact that singularity imposes milder restrictions on SMM than ML has two implications. First, the efficiency gain associated with using ML rather than SMM may be modest. Second, identification may be sharper under SMM because information on a larger number of variables can be used.[9]

## 3.3 Generalized Method of Moments

Consider now the case where it is possible to compute analytical expressions for the unconditional moments as a function of the parameters. Then, the simulation-based estimate $(1/\tau T) \sum_{i=1}^{\tau T} \mathbf{m}_i(\boldsymbol{\theta})$ may be replaced by its analytical counterpart $E(\mathbf{m}(\boldsymbol{\theta}))$ in the objective

[8] Note that singularity does not prevent the evaluation of the objective function in (10). The reason is that the weighting matrix depends on the data alone, which is not singular. Singularity matters only for the Jacobian matrix $\mathbf{D}$ and, hence, for the computation of standard errors and statistical inference.

[9] On the other hand, since SMM requires at least as many moments as parameters are estimated but only linearly independent objects can be included in $\mathbf{m}_i$, singularity may impose a limit on the number of parameters that can be estimated.

function (10), and a Generalized Method of Moments (GMM) estimator of $\boldsymbol{\theta}$ can be computed 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 Burnside, Eichenbaum and Rebelo (1993).

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 the regularity conditions in Hansen (1982)

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

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. 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. The optimal weighting matrix, $\mathbf{W}$, is computed using the Newey-West estimator with a Barlett kernel. A global specification test for overidentified DSGE models may be performed using the usual chi-square statistic (see Hansen, 1982)

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

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.

Comparing the asymptotic variance-covariance matrices of SMM and GMM estimates, one can see 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 $\tau T \rightarrow \infty$, the difference in the standard errors of both estimates is primarily due to the term $(1+1/\tau)$. This term can be thought of as a measure of the increase in sample uncertainty due to the use of simulation to compute the population moments. Note, however, that this term decreases rapidly as $\tau$ increases. For example, for $\tau = 5$, 10 and 20, the standard errors of SMM estimates are, respectively, $1.10, 1.05$ and $1.025$ times larger than those of GMM estimates. 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.

The implications of stochastic singularity for GMM are the same as for SMM. That is, singularity restricts the moments that can be employed to those that form a linearly independent set. Because this is a milder restriction than the one singularity imposes on Maximum Likelihood, the efficiency gain associated with using ML rather than GMM may be small, but identification may be sharper under GMM because it can exploit information on a larger number of variables.

### 3.4 Indirect Inference

This section examines the Indirect Inference procedure proposed by Smith (1993). For reasons to be made clear below, Smith refers to this procedure as Extended Method of Simulated Moments (EMSM). To understand the mechanics of EMSM, it is useful to recall that SMM constructs an estimate of $\boldsymbol{\theta}$ by minimizing the distance between the unconditional moments of the data and those of an artificial series simulated using given parameter values. In contrast, EMSM constructs an estimate of $\boldsymbol{\theta}$ 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. This approach has a Method of Moments interpretation because the coefficients of the VAR are proportional to the covariances and autocovariances of the variables in the VAR. This approach also has an Indirect Inference interpretation (Gouriéroux, Monfort and Renault, 1993) because the auxiliary model (the VAR) is a misspecified version of the true state-space representation of the model.[10]

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 $\tau T$. Then, the Indirect Inference estimator of $\boldsymbol{\theta}$, $\tilde{\boldsymbol{\theta}}_{im}$, is the value that solves

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

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),

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

[10] Rather than matching VAR parameters, some authors match instead the impulse-responses of the model. Since the impulse-responses are nonlinear transformations of the VAR parameters, this strategy is similar to the one studied here. For an application of this idea, see Christiano, Eichenbaum and Evans (2005). An advantage of matching impulse responses is that the researcher can focus on specific horizons by a choosing an appropriate weighting matrix.

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 $\tau T$ data points. Smith suggests a test based on Hansen's (1982) chi-square statistic as specification test for overidentified DSGE models

$$T(1+1/\tau)\left(\left(\boldsymbol{\eta}-\boldsymbol{\eta}(\tilde{\boldsymbol{\theta}}_{im})\right)'\mathbf{V}\left(\boldsymbol{\eta}-\boldsymbol{\eta}(\tilde{\boldsymbol{\theta}}_{im})\right)\right) \to \chi^2(p-q),$$

where $\left(\boldsymbol{\eta}-\boldsymbol{\eta}(\tilde{\boldsymbol{\theta}}_{im})\right)'\mathbf{V}\left(\boldsymbol{\eta}-\boldsymbol{\eta}(\tilde{\boldsymbol{\theta}}_{im})\right)$ 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 auxiliary VAR. More precisely, the artificial data generated by the DSGE model does not have an unrestricted VAR representation. In the case of the RBC model studied here, recall that the systems $(\hat{n}_t, \hat{y}_t, \hat{c}_t)$, $(\hat{y}_t,\ \hat{c}_t,\ \hat{y}_{t-1},\ \hat{c}_{t-1})$, $(\hat{y}_t,\ \hat{n}_t,\ \hat{y}_{t-1},\ \hat{n}_{t-1})$, and $(\hat{n}_t,\ \hat{c}_t,\ \hat{n}_{t-1},\ \hat{c}_{t-1})$ are singular, meaning that their variance-covariance matrices are singular. This implies that the data generated by this DSGE model has only a bivariate VAR representation of order one.[11] 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.

Although this paper focuses on linearized DSGE models, the above discussion allow us to see that stochastic singularity also has implications for the estimation of nonlinear models, which strictly speaking are not singular. Note that in the case of the nonlinear RBC model studied here, a system like $(\hat{y}_t,\ \hat{c}_t,\ \hat{y}_{t-1},\ \hat{c}_{t-1})$, with linear terms in output and consumption, is not singular. Thus, in principle, artificial data from the nonlinear version of this model may be used to estimate a VAR of order higher than one. However, up the extent that the model is close to linear, the explanatory variables will be close to being collinear. Then, as in the standard linear regression model with collinear regressors, standard errors of the VAR parameter estimates will be unduly large. Even though this observation is suggestive, the more exhaustive analysis of estimation issues concerning nonlinear DSGE models is left for future research.

[11] 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 investment. 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.

# 4 Monte-Carlo Experiments

## 4.1 Design

The small-sample properties of all estimators are studied here using Monte Carlo analysis. First, I study the small-sample properties under the null hypothesis. That is, the Data Generating Process (DGP) is the linearized RBC model with one structural (technology) shock that was described in Section 2. Second, I study the small-sample properties in the case where the DSGE model is misspecified. 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.

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. To reduce the computational burden in the Monte Carlo experiments, I concentrate on three (of the five) model parameters, namely the subjective discount factor ($\beta$), the autocorrelation coefficient of the technology shock ($\rho$), and the standard deviation of the technology innovation ($\sigma$). Thus, $\boldsymbol{\theta} = (\beta, \rho, \sigma)'$ is a $3 \times 1$ vector. The data were generated using parameter values $(\beta, \rho, \sigma) = (.95, .85, .04)$. The share of capital in production ($\alpha$) and the depreciation rate ($\delta$) 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.[12]

For the Monte Carlo experiments with priors, I consider independent prior densities for $\beta$ and $\rho$: $\beta \sim N(.95, .025^2)$, $\rho \sim N(.85, .07^2)$, and a diffuse prior for $\sigma$. In terms of the prior representation in (8), they correspond to

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

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

For SMM and Indirect Inference, 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

[12] A slightly larger value of $\beta$, 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 $\beta$ 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.

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 necessarily 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.

## 4.2 Results under the Null Hypothesis

Monte Carlo results are reported in Tables 1 to 3. In all tables, Mean is the average of the estimated parameter values, and 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 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 per cent. In other words, Size is 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 per cent confidence interval around the empirical size would contain the nominal size of 5 per cent. 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 2 and 3, OI is the empirical size of the $\chi^2$ test of the overidentification restrictions.

Panel A in Table 1 reports results using Maximum Likelihood without adding measurement errors or priors. Recall that singularity implies that the RBC model cannot be estimated by Maximum Likelihood using more than one observable variable. Experiments 1 through 3 refer to experiments using output, consumption, and hours, 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 may matter. We will see below that this is also true for the other estimation methods. In particular, the average estimate and A.S.E. of $\beta$ vary substantially depending on whether one uses output, consumption, or hours. When the model is estimated using output alone, the average estimate of $\beta$ is well below its true value and the A.S.E. is very large. When the model is estimated using either consumption or hours, the average estimate is 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 is obtained using consumption or hours, 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 $\rho$ seem to vary with the variable employed. For example, the A.S.E. of $\rho$ is much 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. 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 in the $t$ test.

Panel B in Table 1 reports results using Maximum Likelihood incorporating the priors specified in (14). Experiments 4 through 6 refer, respectively, to the experiments using output, consumption, and hours. Because the prior about $\sigma$ was uninformative, results regarding this parameter are basically the same as those reported in Panel A. Regarding $\beta$ and $\rho$, 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 Maximum Likelihood with priors tend to be numerically closer to their true values than ML estimates without priors. An interesting case is the estimate of $\beta$ obtained using output. Results from Panel A suggest that output is not very informative regarding the discount factor. Consequently when the prior and 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 overestimate the variability of the parameter estimates in small samples. As a result, the $t$ test has a smaller empirical than nominal size and tends to underreject the null hypothesis.[13]

Table 2 report results using the Simulated and Generalized Methods of Moments. In all cases, the mean and median of the estimated parameters are close to their true values. Standard errors are reasonably low and vary with $\tau$ as predicted by the discussion in Section 3.4.[14] 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. For example, in Panel A, the standard deviation of the estimate of $\beta$ obtained using the moments of output and hours is roughly 30 per cent larger than the one obtained using the moments of consumption and hours, for the same value of $\tau$.

There are some differences between the A.S.E.'s 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 Table 2 report the empirical size of the $\chi^2$ 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 per cent. This means that a researcher comparing the $\chi^2$ statistic with the 5 per cent 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 per cent probability of rejecting a true model, the researcher actually is actually taking (approximately) a 1 per cent probability. In some cases (see, for example, Experiment 8 in Panel C), the model's overidentification restrictions were never rejected by the $\chi^2$ test in the 500 replications. The fact that Hansen's $\chi^2$ test easily fails to detect a

[13]Notice that when the DGP is the RBC model with one shock, it is not possible to estimate the model by Maximum Likelihood with measurement errors added. The reason is that under the null, the mean and variance of measurement errors is zero. A previous version of this paper reports results where the DGP is the one-shock model with extra errors added (see, Ruge-Murcia, 2002). Those results are not comparable with the ones reported here because the DGP is not the same.

[14]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.

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 $\chi^2$ test also arise in the context of fully-specified DSGE models.

Finally, Table 3 reports results using Indirect Inference. In all experiments, the mean and median of the estimated parameters are close to their true values. The standard errors of the estimates of $\beta$ and $\rho$ are large, but the standard deviation of the innovation to the technology shock, $\sigma$, is estimated more precisely by Indirect Inference than by any other method. As in Maximum Likelihood, SMM and GMM, the choice of variables and moments to use in estimation has some effect on the precision of the estimates. This can be traced back to the fact that not all variables and moments are equally informative about all structural parameters. For example, $\sigma$ can only be identified under Indirect Inference 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 innovations. In line with the finding that under Maximum Likelihood, consumption and hours are more informative than output about the subjective discount rate, the standard deviation of the empirical distribution of $\beta$ is smallest when the 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 $\chi^2$ test of the overidentification restrictions. As in the Method of Moments, 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 $\beta, \rho$, and $\sigma$ are plotted in Figures 1 to 3, respectively. They correspond to results in Experiments 3 and 6 in Table 1, Experiments 5, 6, 11, and 12 in Table 2, and Experiments 5 and 6 in Table 3. The reason I focus on these experiments, is because they help illustrate more general results uncovered by the Monte Carlo analysis.

Four conclusion can be drawn from these Figures. First, the Method of Moments can deliver sharper estimates than Maximum Likelihood. This result is due to the fact that singularity affects more severely the latter than the former. Thus, the RBC model studied here can be estimated by Maximum Likelihood using only one observable variable, but using moments that involve up to two variables by the Method of Moments. Second, informative priors can be combined with sample data to improve 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 $\beta$ and $\rho$ obtained using Maximum Likelihood with and without priors. Third, 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 $\tau$. Fourth, the choice of variables and moments employed can have some effect on the precision of the estimators. This can be seen in Figures 1 and 2 by comparing the empirical distribution of the estimators of $\beta$ and $\rho$ obtained by SMM, GMM, and Indirect Inference using the moments of output and hours, and consumption and hours. Indirect Inference estimates based on the moments of output and hours 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 reasons 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 may 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. Figure 4 plots the dynamic responses of consumption, output and hours following a technology shock. The dotted lines are the 90 per cent 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 supports four conclusions. First, the mean response does not depend on the estimation method because all methods deliver unbiased parameter estimates. However, there are some small differences in the coverage probabilities of the estimated confidence intervals because not all estimation procedures are equally efficient. Second, confidence intervals for the response based on Indirect Inference estimates obtained using the moments of output and hours are the widest. This reflects the diffuse empirical distributions of $\beta$ and $\rho$ reported in Figures 1 and 2. Third, aside from the latter case, 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 necessarily translate into substantial differences in the coverage probabilities of their impulse responses.

### 4.3 Robustness to Misspecification

This section reports the results of Monte Carlo experiments where the DSGE model is misspecified meaning that the true model that generates the data is not the linearized one-shock RBC model. The goal of this exercise is to study the robustness of each estimation method to misspecification. Two forms of misspecification are considered. First, the DGP is the linearized RBC model, but with three (rather than only one) structural shocks. Thus, a researcher looking at the artificial data on output, consumption, and hours would find them to be nonsingular. Second, the DGP is the nonlinear one-shock RBC model. In both cases, the estimated model is the economically-interesting but misspecified and singular RBC model with only one technology shock.

### 4.3.1 True Model has Multiple Structural Shocks

In this experiment, the artificial data is generated by the linearized RBC model in Section 2, but expanded to include, in addition to the technology shock, a preference shock and a shock to the investment function. The preference shock is a disturbance to the weight of leisure in the utility function, $\psi$. This weight follows the stochastic process

$$\ln \psi_{t+1} = 0.6 \ln \psi_t + \zeta_t,$$

where $\zeta_t$ is an innovation assumed to be independently, identically, and Normally distributed with zero mean and variance $0.03^2$. The shock to the investment function affects the transformation of investment into new capital. The law of motion of the capital stock is generalized to

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

where $\nu_t$ is a shock that follows the process

$$\ln \nu_{t+1} = 0.6 \ln \nu_t + \varepsilon_t,$$

and $\varepsilon_t$ is an innovation assumed to be independently, identically, and Normally distributed with zero mean and variance $0.04^2$. Notice that the mean of $\nu_t$ equals 1 so that on average one unit invested becomes one unit of capital.

Results are reported in Tables 4 to 6. In all tables, Bias is the average difference between the estimated parameter values and the true value used to generate the artificial data. Remaining headings are as previously described. Table 4 reports results obtained using Maximum Likelihood. Panel A considers the case where neither priors nor measurement errors are used, and the researcher is limited to using at most as many variables as structural shocks. Estimates of $\sigma$ are biased upwards, while estimates of $\beta$ and $\rho$ are biased downwards.[15] The biases are more severe when data on hours or output, rather than consumption, is used. The reason $\sigma$ is biased upwards is that all the conditional variance of the

[15] A possible concern is that these biases simply reflect the existence of multiple local maxima in the log likelihood function. In order to address this issue, the ML estimation of the misspecified model were carried out using a genetic algorithm. Genetic algorithms do not completely eliminate the possibility that the maximization routine will converge to a local, rather than the global, maximum, but they make it much less likely.

series is attributed to the technology shock. The asymptotic standard errors usually underestimate the standard errors of the empirical small-sample distribution of the estimates. Consequently, $t$ tests tend to reject the null hypothesis more often that the nominal size of 5 per cent.

To understand why different series lead to different estimates, recall that Maximum Likelihood minimizes the one-step-ahead predictions of the model. The variance decomposition of the true model with multiple shocks indicates that the technology shock explains 89, 78, and 44 per cent of the conditional variance of the one-step ahead forecasts errors of consumption, output, and hours, respectively. Because the only shock in the estimated RBC model (that is, the technology shock) is more important in explaining the variation of consumption in the short-run than the one of hours or output, the misspecification affects ML estimates less severely when one uses data on consumption than hours or output.

Panel B in Table 4 reports results using Maximum Likelihood but incorporating Bayesian priors. In general, using priors reduces the bias of the estimates and the variance of their small-sample distributions. The reason is simply that the priors are centered around the true parameter value. More generally, incorporating prior information will "pull" ML estimates towards more economically reasonable values, even if the estimated model is misspecified. The asymptotic standard errors tend to overestimate the standard error of the small-sample distribution and, as a result, $t$ tests tend to reject the null hypothesis less often than 5 per cent of the times.

Panel C reports the results using Maximum Likelihood but adding measurement errors. Because measurement errors are added, more variables than the number of structural shocks may be used in Maximum Likelihood estimation. In general, the bias of the estimates are smaller in this case than the ones reported in Panel A. However, because the technology shock is not equally important in explaining the short-run behavior of all series, results depend on the data used. For the example considered here, biases are less severe for combinations of variables that include consumption. Finally, note that, as in Panel A, the asymptotic standard errors underestimate the standard errors of the small-sample distribution of the estimates.

Table 5 report results for the Simulated and Generalized Method of Moments. Estimate of $\rho$ are downward biased in all cases but, in general, the bias of the Method of Moments estimates is smaller than that of ML estimates. Because the asymptotic standard errors differ from the small-sample standard errors, the $t$ test is subject to size distortions, but these distortions are smaller than the ones observed in Table 4 for Maximum Likelihood. Hence, the Method of Moments appear more robust than Maximum Likelihood to this form of misspecification, in the sense that biases are smaller and statistical inference is less distorted.

As in Table 2, standard errors vary with $\tau$, but this variation in efficiency is much smaller than that observed across moments employed in estimation. Thus, when the model is misspecified, the choice of moments to match might be more important than the length of the simulated series in SMM.

Table 6 reports results using Indirect Inference. In general, results are similar to those obtained using the Method of Moments. The reason is that the version of Indirect Inference studied here matches coefficients on a VAR representation of the data and these coefficients are proportional to covariances and autocovariances of the variables in the VAR. These explains why the bias in the parameter estimates are usually smaller than those observed under Maximum Likelihood and roughly of the same magnitude as those obtained under the Method of Moments. However, the size distortions of the $t$ test are usually larger than those observed under the Method of Moments and there are instances where the bias of $\tilde{\beta}$ and $\tilde{\rho}$ is severe.

Figure 5 compares the biases in all experiments under each estimation method. In particular, I focus on the bias of the estimates of $\beta$ and $\rho$ by plotting them in a plane with the bias of $\tilde{\beta}$ in the horizontal axis and that of $\tilde{\rho}$ in the vertical axis. Notice that for the Method of Moments estimators, biases are closely packed near the $(0, 0)$ point, meaning that, although not exactly zero, these biases are comparatively small. For Indirect Inference, that is also true in most cases, but there are some outliers where the biases are large. By far the most diffuse plot is that of Maximum Likelihood. The points closest to $(0, 0)$ are those where Bayesian priors used. Biases tend to be generally larger than under the Method of Moments, and there are cases where they can be very severe.

Figure 6 examines the effects of misspecification and bias on impulse response analysis. In particular, I compare the response to a technology shock for the misspecified model (continuous line) with the response predicted by the true model (dotted line). These responses were constructed using estimates from Experiments 3, 6, and 9 in Table 3, Experiments 5, 6, and 12 in Table 4, and Experiments 5 and 6 in Table 6. In order not to saturate the picture, I have abstained from plotting confidence intervals for these responses. Consider first the responses predicted by GMM and SMM estimates in columns 4 to 6. Quantitatively, the (average) estimated responses are not too far from the true response and the predicted dynamics are roughly similar. Statistically, the true response lies just above the 90 per cent confidence bound (not shown). Note that because, GMM and SMM results in Table 5 are similar, these plots are representative of the general effect of misspecification on impulse response analysis with GMM and SMM estimates. Columns 1 to 3 plot the responses obtained using ML estimates. When no priors are added, the (average) estimated response can be quite different from the true response, both in terms of numerical and statistical distance and in terms of the dynamics. Notice, however, that these results are not representative of all ML results. For example, responses using results from Maximum Likelihood estimation with consumption data (not shown) yield responses whose 90 per cent confidence interval includes the true response. Columns 7 and 8 plot the responses obtained using Indirect Inference estimates. These plots illustrate the previous observation that matching VAR coefficient will lead to similar results that matching moments directly. However, the moments coded in the VAR might not always be the most informative. In this case, biases may be severe and predicted impulse response might differ substantially from the true response.

#### 4.3.2 True Model is Nonlinear

In this experiment, the artificial data is generated by the nonlinear version of the one-shock RBC model in Section 2. The nonlinear model was solved using the Parameterized Expectation Algorithm (Williams and Wright, 1991, ch. 3) with the marginal utility of consumption in the Euler equation parameterized as a Chebychev polynomial in the two state variables, namely the capital stock and the technology shock. Since the model is close to linear in logs, the misspecification is not too severe in this case. Nonetheless, we will see below that there is a difference in robustness across estimation methods, with moment-based procedures substantially more robust than Maximum Likelihood.

Results are reported in Tables 7 to 9. Table 7 reports results obtained using Maximum Likelihood.[16] Estimates of $\sigma$ ($\beta$ and $\rho$) tend to be biased upwards (downwards), and by more when hours or output, rather than consumption, are used. In most instances, biases are large despite the fact that, as noted above, misspecification is minor. When no priors are added, the asymptotic standard errors of $\beta$ and $\rho$ tend to underestimate the standard error of the small-sample distribution and, consequently, $t$ tests tend to reject the null hypothesis more often than 5 per cent of the times. Biases are usually smaller when priors are imposed, but this is primarily due to fact that priors are centered around the true parameter values.

Tables 8 and 9 report results for the Simulated and Generalized Methods of Moments, and Indirect Inference, respectively. It is remarkable that results in these Tables are very close to those in Tables 2 and 3, where the model was estimated under the null. Biases are small and asymptotic standard errors differ from the small-sample standard errors so that the $t$ test is subject to size distortions. However, it is clear that the latter result is due to the small sample, rather than to misspecification because very similar results were found under the null.

[16]This Table does not include the case where measurement errors are added because ML was very poorly behaved and too many replications had to be thrown out because either estimates were at the boundary of the parameter space or stability conditions were violated.

To understand the reason why estimators based on unconditional second moments are more robust to this form of misspecification than Maximum Likelihood, recall that ML minimizes the one-step-ahead predictions of the model. By estimating the linear model when the true DGP is nonlinear, ML ignores the contribution of terms proportional to all moments of order two and higher. In contrast, second moments computed using the linear model ignore terms proportional to moments of order four, six, eight, etc. Hence, the quantitative magnitude of the neglected terms is larger under ML than under the Method of Moments. This result is important because solving and estimating nonlinear DSGE models is much more computationally demanding than their linearized counterparts. Hence, there may be instances where a reasonable empirical strategy may be to linearized the DSGE model and then estimate the (misspecified) approximation using a method that is robust to this form of misspecification.

A graphical comparison of biases under each estimation method is provided in Figure 7. This is the analog of Figure 5 reported above, but for a different form of misspecification. The bias of $\tilde{\beta}$ is plotted in the horizontal axis and that of $\tilde{\rho}$ in the vertical axis. As in Figure 5, the biases of the Method of Moments and Indirect Inference estimators are generally small and packed close to the point $(0, 0)$. In contrast, the plot is more diffuse for the Maximum Likelihood estimates and, in certain cases, biases can be severe.

## 4.4 Comparison in Terms of Computing Time

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 computing time.

Table 7 reports the average number of seconds taken to complete a replication, including the computation of standard errors, using each estimation method. The average is taken over all replications for all experiments in Tables 1 to 3 that employ the same estimation method. From this Table is clear that GMM is by far the most computationally efficient procedure, followed by Maximum Likelihood and SMM. There is a fairly large difference in computational efficiency between simulation-based methods and GMM, and the time per iteration increases proportionally with $\tau$. The reason is that the simulation-based methods require the solution of the DSGE model and computation of the gradients using $\tau$ times more observations than GMM.[17]

[17]Results regarding Indirect Inference 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 Indirect Inference, the algorithm would

# 5 Conclusions

The paper studies the application of standard econometric techniques for the estimation of DSGE models. A basic RBC model is used to illustrate the fact that the stochastic singularity of linearized DSGE models imposes restrictions on the variables and moments that may be used for estimation purposes. Singularity affects more severely Maximum Likelihood than the Method of Moments or the version of Indirect Inference examined here. The former requires linearly independent variables while the latter two require linearly independent moments. In general, it is possible to find independent moments that incorporate information about more variables than those that are linearly independent. This implies that 1) the efficiency gain of using ML may be modest, but 2) identification may be sharper under moment-based procedures.

Monte-Carlo analysis shows that all procedures deliver reasonably good estimates under the null hypothesis, but that the Method of Moments (and to a lesser extent, Indirect Inference) are substantially more robust to misspecification than Maximum Likelihood, even when measurement errors are added to sidestep the singularity of the model. Results show that GMM is computationally more efficient than ML, which in turn is more efficient than simulation-based estimators. However, differences in computational cost are not large enough to condition our choice of estimation method. Instead, other considerations like robustness to misspecification and to stochastic singularity suggest that moment-based procedures are an attractive alternative to the more widely used method of Maximum Likelihood.

frequently blow up if the routine was started at the true value of $\sigma$. Hence, the minimization routine was started using a value for $\sigma$ much larger than the one used to generate the sample. Just for this reason alone Indirect Inference would take longer to converge than the other estimation methods. This means that the numbers in Table 10 most likely overstate the computating time required by this method.

**Table 1. Maximum Likelihood**
**Under the Null Hypothesis**
**Artificial Data is Stochastically Singular**

| Experiment # | Var. | $\beta$ Mean / Median | $\beta$ A.S.E. / S.D. | $\beta$ Size / S.E. | $\rho$ Mean / Median | $\rho$ A.S.E. / S.D. | $\rho$ Size / S.E. | $\sigma$ Mean / Median | $\sigma$ A.S.E. / S.D. | $\sigma$ Size / S.E. |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | *A. Using as Many Variables as Shocks* | | | | | | |
| 1 | $\hat{y}_t$ | .7896 | .2205 | .0900 | .8080 | .0600 | .0620 | .0495 | .0173 | .2300 |
| | | .8202 | .1721 | .0128 | .8206 | .0583 | .0108 | .0505 | .0104 | .0188 |
| 2 | $\hat{c}_t$ | .9156 | .0847 | .1660 | .8127 | .1201 | .0980 | .0534 | .0375 | .1040 |
| | | .9489 | .1359 | .0166 | .8585 | .1370 | .0133 | .0376 | .0371 | .0137 |
| 3 | $\hat{n}_t$ | .9403 | .0623 | .1280 | .8406 | .0890 | .0580 | .0433 | .0225 | .1440 |
| | | .9425 | .0477 | .0149 | .8575 | .0609 | .0105 | .0441 | .0157 | .0157 |
| | | | | *B. Incorporating Bayesian Priors* | | | | | | |
| 4 | $\hat{y}_t$ | .9501 | .0250 | .0000 | .8434 | .0334 | .0280 | .0398 | .0037 | .0000 |
| | | .9502 | .0010 | .0000 | .8464 | .0295 | .0074 | .0397 | .0023 | .0000 |
| 5 | $\hat{c}_t$ | .9494 | .0230 | .0020 | .8482 | .0522 | .0060 | .0397 | .0088 | .0000 |
| | | .9499 | .0067 | .0020 | .8500 | .0286 | .0035 | .0396 | .0027 | .0000 |
| 6 | $\hat{n}_t$ | .9498 | .0225 | .0000 | .8450 | .0440 | .0020 | .0397 | .0082 | .0000 |
| | | .9495 | .0068 | .0000 | .8473 | .0257 | .0020 | .0399 | .0026 | .0000 |

*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 per cent 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. The priors used are: $\beta \sim N(.95, .025^2)$, $\rho \sim N(.85, .07^2)$. The prior for $\sigma$ is diffuse.

**Table 2. Method of Moments**
**Under the Null Hypothesis**
**Artificial Data is Stochastically Singular**

| | | | $\beta$ | | | $\rho$ | | | $\sigma$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Experiment | | Mean | A.S.E. | Size | Mean | A.S.E. | Size | Mean | A.S.E. | Size | OI |
| # | Var. | Median | S.D. | S.E. | Median | S.D. | S.E. | Median | S.D. | S.E. | S.E. |
| | | | | *A. Simulated Method of Moments with $\tau = 5$* | | | | | | | |
| 1 | $\hat{y}_t, \hat{c}_t$ | .9503 | .0128 | .0300 | .8344 | .0410 | .0920 | .0396 | .0041 | .0280 | .0060 |
| | | .9504 | .0094 | .0076 | .8426 | .0489 | .0129 | .0396 | .0031 | .0074 | .0035 |
| 2 | $\hat{n}_t, \hat{y}_t$ | .9489 | .0140 | .0980 | .8415 | .0314 | .1520 | .0399 | .0037 | .0640 | .0000 |
| | | .9497 | .0159 | .0133 | .8455 | .0440 | .0161 | .0398 | .0038 | .0109 | .0000 |
| 3 | $\hat{n}_t, \hat{c}_t$ | .9506 | .0107 | .0420 | .8403 | .0297 | .1320 | .0395 | .0032 | .0940 | .0020 |
| | | .9508 | .0098 | .0090 | .8460 | .0413 | .0151 | .0393 | .0033 | .0131 | .0020 |
| | | | | *B. Simulated Method of Moments with $\tau = 10$* | | | | | | | |
| 4 | $\hat{y}_t, \hat{c}_t$ | .9504 | .0124 | .0280 | .8342 | .0400 | .1100 | .0393 | .0040 | .0340 | .0160 |
| | | .9511 | .0097 | .0074 | .8426 | .0499 | .0140 | .0393 | .0030 | .0081 | .0056 |
| 5 | $\hat{n}_t, \hat{y}_t$ | .9495 | .0134 | .0680 | .8425 | .0298 | .1700 | .0397 | .0036 | .0700 | .0000 |
| | | .9507 | .0140 | .0113 | .8477 | .0423 | .0168 | .0395 | .0035 | .0114 | .0000 |
| 6 | $\hat{n}_t, \hat{c}_t$ | .9508 | .0102 | .0400 | .8410 | .0281 | .1580 | .0393 | .0030 | .0960 | .0000 |
| | | .9510 | .0094 | .0088 | .8468 | .0400 | .0163 | .0391 | .0033 | .0132 | .0000 |
| | | | | *C. Simulated Method of Moments with $\tau = 20$* | | | | | | | |
| 7 | $\hat{y}_t, \hat{c}_t$ | .9497 | .0122 | .0160 | .8351 | .0388 | .1240 | .0395 | .0039 | .0300 | .0080 |
| | | .9503 | .0084 | .0056 | .8462 | .0508 | .0147 | .0394 | .0028 | .0076 | .0040 |
| 8 | $\hat{n}_t, \hat{y}_t$ | .9485 | .0131 | .0800 | .8395 | .0295 | .1480 | .0395 | .0034 | .0680 | .0000 |
| | | .9497 | .0145 | .0121 | .8450 | .0407 | .0159 | .0394 | .0034 | .0113 | .0000 |
| 9 | $\hat{n}_t, \hat{c}_t$ | .9506 | .0100 | .0600 | .8432 | .0272 | .1280 | .0395 | .0030 | .0940 | .0020 |
| | | .9509 | .0098 | .0106 | .8479 | .0368 | .0149 | .0394 | .0032 | .0131 | .0020 |
| | | | | *D. Generalized Method of Moments* | | | | | | | |
| 10 | $\hat{y}_t, \hat{c}_t$ | .9502 | .0117 | .0080 | .8366 | .0375 | .0940 | .0395 | .0038 | .0200 | .0080 |
| | | .9505 | .0082 | .0040 | .8423 | .0446 | .0131 | .0394 | .0027 | .0063 | .0040 |
| 11 | $\hat{n}_t, \hat{y}_t$ | .9485 | .0127 | .0700 | .8383 | .0288 | .1520 | .0394 | .0033 | .0520 | .0000 |
| | | .9506 | .0141 | .0114 | .8436 | .0393 | .0161 | .0392 | .0034 | .0099 | .0000 |
| 12 | $\hat{n}_t, \hat{c}_t$ | .9503 | .0098 | .0400 | .8390 | .0270 | .1380 | .0394 | .0029 | .0980 | .0020 |
| | | .9506 | .0093 | .0088 | .8420 | .0362 | .0154 | .0395 | .0030 | .0133 | .0020 |

*Notes*: For Experiments 1, 4, 7, and 10, $\mathbf{m}_t = (\hat{y}_t^2,\ \hat{c}_t^2,\ \hat{c}_t\hat{y}_t,\ \hat{c}_t\hat{c}_{t-1},\ \hat{y}_t\hat{y}_{t-1})'$; for Experiments 2, 5, 8, and 11, $\mathbf{m}_t = (\hat{y}_t^2,\ \hat{n}_t^2,\ \hat{n}_t\hat{y}_t,\ \hat{n}_t\hat{n}_{t-1},\ \hat{y}_t\hat{y}_{t-1})'$; and for Experiments 3, 6, 9, and 12, $\mathbf{m}_t = (\hat{n}_t^2,\ \hat{c}_t^2,\ \hat{c}_t\hat{n}_t,\ \hat{c}_t\hat{c}_{t-1},\ \hat{n}_t\hat{n}_{t-1})'$. See the notes to Table 1.

**Table 3. Indirect Inference**
**Under the Null Hypothesis**
**Artificial Data is Stochastically Singular**

| Experiment # | Var. | $\beta$ Mean / Median | $\beta$ A.S.E. / S.D. | $\beta$ Size / S.E. | $\rho$ Mean / Median | $\rho$ A.S.E. / S.D. | $\rho$ Size / S.E. | $\sigma$ Mean / Median | $\sigma$ A.S.E. / S.D. | $\sigma$ Size / S.E. | OI / S.E. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | *A.* $\tau = 5$ | | | | | | |
| 1 | $\hat{y}_t, \hat{c}_t$ | .9514 | .0249 | .3360 | .8182 | .0795 | .3580 | .0399 | .0004 | .7360 | .0800 |
| | | .9506 | .0325 | .0211 | .8521 | .1425 | .0214 | .0399 | .0022 | .0197 | .0121 |
| 2 | $\hat{n}_t, \hat{y}_t$ | .9489 | .0250 | .2300 | .8119 | .0920 | .2380 | .0402 | .0002 | .8520 | .1740 |
| | | .9502 | .0298 | .0188 | .8508 | .1466 | .0190 | .0401 | .0023 | .0159 | .0170 |
| 3 | $\hat{n}_t, \hat{c}_t$ | .9492 | .0267 | .0020 | .8428 | .0982 | .0020 | .0398 | .0006 | .6240 | .0020 |
| | | .9489 | .0111 | .0020 | .8458 | .0425 | .0020 | .0399 | .0023 | .0217 | .0020 |
| | | | | | *B.* $\tau = 10$ | | | | | | |
| 4 | $\hat{y}_t, \hat{c}_t$ | .9535 | .0235 | .3240 | .8266 | .0708 | .3740 | .0401 | .0003 | .7180 | .0820 |
| | | .9513 | .0329 | .0209 | .8548 | .1333 | .0216 | .0401 | .0020 | .0201 | .0123 |
| 5 | $\hat{n}_t, \hat{y}_t$ | .9463 | .0251 | .2100 | .7995 | .0955 | .2120 | .0401 | .0002 | .8260 | .1760 |
| | | .9472 | .0302 | .0182 | .8393 | .1543 | .0182 | .0401 | .0020 | .0170 | .0170 |
| 6 | $\hat{n}_t, \hat{c}_t$ | .9488 | .0256 | .0060 | .8414 | .0944 | .0000 | .0401 | .0005 | .6380 | .0000 |
| | | .9487 | .0111 | .0035 | .8452 | .0423 | .0000 | .0401 | .0021 | .0215 | .0000 |
| | | | | | *C.* $\tau = 20$ | | | | | | |
| 7 | $\hat{y}_t, \hat{c}_t$ | .9566 | .0246 | .2920 | .8377 | .0691 | .3700 | .0399 | .0003 | .7460 | .1040 |
| | | .9558 | .0334 | .0203 | .8704 | .1308 | .0216 | .0397 | .0020 | .0195 | .0137 |
| 8 | $\hat{n}_t, \hat{y}_t$ | .9399 | .0237 | .2800 | .7672 | .0982 | .2480 | .0398 | .0002 | .8520 | .2180 |
| | | .9416 | .0307 | .0201 | .8167 | .1726 | .0193 | .0398 | .0019 | .0159 | .0185 |
| 9 | $\hat{n}_t, \hat{c}_t$ | .9480 | .0248 | .0040 | .8391 | .0928 | .0020 | .0400 | .0005 | .5940 | .0000 |
| | | .9475 | .0101 | .0028 | .8406 | .0396 | .0020 | .0400 | .0020 | .0220 | .0000 |

*Notes*: For Experiments 1, 4, and 7, the VAR consists of $\hat{y}_t$ and $\hat{c}_t$; for Experiments 2, 5, and 8, the VAR consists of $\hat{y}_t$ and $\hat{n}_t$; and for Experiments 3, 6, and 9, the VAR consists of $\hat{n}_t$, and $\hat{c}_t$. In all cases a VAR of order one is used. See the notes to Table 1.

**Table 4. Maximum Likelihood**
**Artificial Data Appears to be Nonsingular Because**
**True Model has Multiple Structural Shocks**

| | | | $\beta$ | | | $\rho$ | | | $\sigma$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| Experiment | | Mean | A.S.E. | Size | Mean | A.S.E. | Size | Mean | A.S.E. | Size |
| # | Var. | Bias | S.D. | S.E. | Bias | S.D. | S.E. | Bias | S.D. | S.E. |
| | | | | | *A. Using as Many Variables as Shocks* | | | | | |
| 1 | $\hat{y}_t$ | .8764 | .0440 | .0400 | −.0151 | .1230 | .9920 | .2430 | .0706 | .9900 |
| | | −.0736 | .1580 | .0088 | −.8651 | .1667 | .0040 | .2030 | .0796 | .0044 |
| 2 | $\hat{c}_t$ | .9089 | .1036 | .1440 | .8020 | .1276 | .1060 | .0554 | .0390 | .1100 |
| | | −0.0411 | .1249 | .0157 | −.0480 | .1374 | .0138 | .0154 | .0373 | .0140 |
| 3 | $\hat{n}_t$ | .6276 | .1664 | .4920 | .0479 | .1494 | .9880 | .1211 | .0240 | .9900 |
| | | −.3224 | .2154 | .0224 | −.8021 | .1523 | .0049 | .0811 | .0233 | .0044 |
| | | | | | *B. Incorporating Bayesian Priors* | | | | | |
| 4 | $\hat{y}_t$ | .9304 | .0288 | .0000 | .7661 | .0877 | .0720 | .1890 | .0302 | .9860 |
| | | −.0196 | .0162 | .0000 | −.0839 | .0589 | .0116 | .1490 | .0156 | .0053 |
| 5 | $\hat{c}_t$ | .9480 | .0229 | .0000 | .8425 | .0091 | .0040 | .0422 | .0091 | .0000 |
| | | −.0020 | .0069 | .0000 | −.0075 | .0023 | .0028 | .0022 | .0023 | .0000 |
| 6 | $\hat{n}_t$ | .9230 | .0278 | .0000 | .7329 | .0982 | .1280 | .2005 | .0284 | .9900 |
| | | −0.0270 | .0131 | .0000 | −0.1171 | .0506 | .0149 | .1605 | .0147 | .0044 |
| | | | | | *C. Adding Measurement Errors* | | | | | |
| 7 | $\hat{y}_t, \hat{c}_t$ | .9999 | .0050 | .9840 | .9766 | .0088 | .9900 | .0546 | .0046 | .8880 |
| | | .0499 | .0005 | .0056 | .0266 | .0048 | .0044 | .0146 | .0037 | .0141 |
| 8 | $\hat{n}_t, \hat{y}_t$ | .0294 | .0827 | .9420 | .7819 | .0910 | .0160 | .0302 | .0079 | 1.000 |
| | | −.9206 | .1189 | .0105 | −.0681 | .0572 | .0056 | −.0098 | .0059 | .0000 |
| 9 | $\hat{n}_t, \hat{c}_t$ | .9550 | .0020 | .6400 | .7747 | .0246 | .8220 | .0507 | .0028 | 1.000 |
| | | .0050 | .0052 | .0215 | −.0753 | .0381 | .0171 | .0107 | .0031 | .0000 |

*Notes*: Bias is the arithmetic average of the difference between the estimated and true parameter values. See the notes to Table 1.

**Table 5. Method of Moments**
**Artificial Data Appears to be Nonsingular Because**
**True Model has Multiple Structural Shocks**

| Experiment | | $\beta$ | | | $\rho$ | | | $\sigma$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | A.S.E. | Size | Mean | A.S.E. | Size | Mean | A.S.E. | Size |
| # | Var. | Bias | S.D. | S.E. | Bias | S.D. | S.E. | Bias | S.D. | S.E. |
| | | | | | *A. Simulated Method of Moments with $\tau = 5$* | | | | | |
| 1 | $\hat{y}_t, \hat{c}_t$ | .9478 | .0116 | .0320 | .8088 | .0431 | .1800 | .0451 | .0042 | .1220 |
| | | −.0022 | .0092 | .0079 | −.0412 | .0562 | .0172 | .0051 | .0031 | .0146 |
| 2 | $\hat{n}_t, \hat{y}_t$ | .9708 | .0094 | .6080 | .7897 | .0359 | .3640 | .0457 | .0036 | .3420 |
| | | .0208 | .0124 | .0218 | −.0654 | .0603 | .0215 | .0057 | .0045 | .0212 |
| 3 | $\hat{n}_t, \hat{c}_t$ | .9523 | .0106 | .1200 | .7532 | .0381 | .6540 | .0525 | .0038 | .9000 |
| | | .0023 | .0118 | .0145 | −.0968 | .0573 | .0213 | .0125 | .0043 | .0134 |
| | | | | | *B. Simulated Method of Moments with $\tau = 10$* | | | | | |
| 4 | $\hat{y}_t, \hat{c}_t$ | .9479 | .0112 | .0036 | .8099 | .0417 | .1560 | .0448 | .0041 | .1340 |
| | | −.0021 | .0091 | .0083 | −.0401 | .0541 | .0162 | .0048 | .0031 | .0152 |
| 5 | $\hat{n}_t, \hat{y}_t$ | .9701 | .0092 | .5920 | .7822 | .0350 | .4300 | .0457 | .0034 | .3920 |
| | | .0201 | .0117 | .0220 | −.0678 | .0533 | .0221 | .0057 | .0043 | .0218 |
| 6 | $\hat{n}_t, \hat{c}_t$ | .9520 | .0102 | .0800 | .7556 | .0360 | .7040 | .0527 | .0037 | .9300 |
| | | .0020 | .0109 | .0121 | −0.944 | .0494 | .0204 | .0127 | .0042 | .0114 |
| | | | | | *C. Simulated Method of Moments with $\tau = 20$* | | | | | |
| 7 | $\hat{y}_t, \hat{c}_t$ | .9483 | .0110 | .0280 | .8072 | .0409 | .2040 | .0445 | .0039 | .0940 |
| | | −.0017 | .0087 | .0074 | −.0428 | .0541 | .0180 | .0045 | .0029 | .0131 |
| 8 | $\hat{n}_t, \hat{y}_t$ | .9709 | .0088 | .6280 | .7915 | .0335 | .3660 | .0455 | .0033 | .3920 |
| | | .0209 | .0117 | .0216 | −.0585 | .0437 | .0215 | .0055 | .0043 | .0218 |
| 9 | $\hat{n}_t, \hat{c}_t$ | .9511 | .0100 | .0940 | .7523 | .0356 | .7320 | .0529 | .0036 | .9460 |
| | | .0011 | .0114 | .0131 | −.0977 | .0508 | .0198 | .0129 | .0041 | .0101 |
| | | | | | *D. Generalized Method of Moments* | | | | | |
| 10 | $\hat{y}_t, \hat{c}_t$ | .9375 | .0108 | .0300 | .8075 | .0398 | .1840 | .0447 | .0038 | .1340 |
| | | −.0125 | .0087 | .0076 | −.0425 | .0518 | .0173 | .0047 | .0029 | .0152 |
| 11 | $\hat{n}_t, \hat{y}_t$ | .9710 | .0087 | .6480 | .7909 | .0328 | .3840 | .0457 | .0033 | .3880 |
| | | .0210 | .0104 | .0214 | −.0591 | .0475 | .0218 | .0057 | .0038 | .0218 |
| 12 | $\hat{n}_t, \hat{c}_t$ | .9512 | .0097 | .0920 | .7541 | .0346 | .7040 | .0529 | .0035 | .9660 |
| | | .0012 | .0109 | .0129 | −.0959 | .0506 | .0204 | .0129 | .0040 | .0081 |

*Notes*: See the notes to Tables 1, 2 and 4.

**Table 6. Indirect Inference**
**Artificial Data Appears to be Nonsingular Because**
**True Model has Multiple Structural Shocks**

| Experiment | | $\beta$ | | | $\rho$ | | | $\sigma$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| # | Var. | Mean<br>Bias | A.S.E.<br>S.D. | Size<br>S.E. | Mean<br>Bias | A.S.E.<br>S.D. | Size<br>S.E. | Mean<br>Bias | A.S.E.<br>S.D. | Size<br>S.E. |
| | | | | | *A.* $\tau = 5$ | | | | | |
| 1 | $\hat{y}_t, \hat{c}_t$ | .9423 | .0156 | .5000 | .7389 | .0699 | .5280 | .0448 | .0003 | .9540 |
| | | −.0077 | .0301 | .0224 | −.1111 | .1624 | .0223 | .0048 | .0025 | .0094 |
| 2 | $\hat{n}_t, \hat{y}_t$ | .9999 | .0013 | 1.000 | .1050 | .0864 | .9880 | .0344 | .0001 | 1.000 |
| | | .0499 | .0001 | .0000 | −.7450 | .3734 | .0049 | −.0056 | .0018 | .0000 |
| 3 | $\hat{n}_t, \hat{c}_t$ | .9572 | .0133 | .0040 | .7649 | .0847 | .0720 | .0523 | .0005 | 1.000 |
| | | .0072 | .0073 | .0028 | −.0851 | .0480 | .0116 | .0123 | .0027 | .0000 |
| | | | | | *B.* $\tau = 10$ | | | | | |
| 4 | $\hat{y}_t, \hat{c}_t$ | .9437 | .0157 | .4480 | .7491 | .0682 | .4800 | .0447 | .0003 | .9600 |
| | | −.0063 | .0293 | .0222 | −.1009 | .1571 | .0223 | .0047 | .0024 | .0088 |
| 5 | $\hat{n}_t, \hat{y}_t$ | .9999 | .0013 | 1.000 | .0713 | .0828 | .9880 | .0342 | .0001 | 1.000 |
| | | .0499 | .0000 | .0000 | −.7785 | .3355 | .0049 | −.0058 | .0017 | .0000 |
| 6 | $\hat{n}_t, \hat{c}_t$ | .9572 | .0127 | .0040 | .7657 | .0808 | .0860 | .0524 | .0005 | 1.000 |
| | | .0072 | .0064 | .0028 | −.0843 | .0418 | .0125 | .0124 | .0026 | .0000 |
| | | | | | *C.* $\tau = 20$ | | | | | |
| 7 | $\hat{y}_t, \hat{c}_t$ | .9463 | .0164 | .4180 | .7614 | .0663 | .4720 | .0448 | .0003 | .9720 |
| | | −.0037 | .0303 | .0221 | −.0886 | .1544 | .0223 | .0048 | .0022 | .0074 |
| 8 | $\hat{n}_t, \hat{y}_t$ | .9999 | .0012 | 1.000 | .0722 | .0804 | .9980 | .0341 | .0001 | 1.000 |
| | | .0499 | .0000 | .0000 | −.7778 | .2894 | .0020 | −.0059 | .0017 | .0000 |
| 9 | $\hat{n}_t, \hat{c}_t$ | .9570 | .0124 | .0040 | .7636 | .0795 | .0780 | .0520 | .0005 | 1.000 |
| | | .0070 | .0063 | .0028 | −.0864 | .0419 | .0120 | .0120 | .0025 | .0000 |

*Notes*: See the notes to Tables 1, 3 and 4.

**Table 7. Maximum Likelihood**
**Artificial Data is Generated by Nonlinear Model**

| Experiment # | Var. | $\beta$ Mean Bias | $\beta$ A.S.E. S.D. | $\beta$ Size S.E. | $\rho$ Mean Bias | $\rho$ A.S.E. S.D. | $\rho$ Size S.E. | $\sigma$ Mean Bias | $\sigma$ A.S.E. S.D. | $\sigma$ Size S.E. |
|---|---|---|---|---|---|---|---|---|---|---|
| | | *A. Using as Many Variables as Shocks* | | | | | | | | |
| 1 | $\hat{y}_t$ | .8564 | .0409 | .0740 | .0463 | .1225 | .9880 | .2281 | .0702 | .9900 |
| | | −.0936 | .2347 | .0117 | −.8037 | .2040 | .0049 | .1881 | .0823 | .0044 |
| 2 | $\hat{c}_t$ | .9242 | .0844 | .1680 | .7967 | .1267 | .1000 | .0570 | .0398 | .0940 |
| | | −0.0258 | .1162 | .0167 | −.0533 | .1503 | .0134 | .0170 | .0399 | .0131 |
| 3 | $\hat{n}_t$ | .7540 | .1152 | .1660 | .0829 | .1474 | .9880 | .0963 | .0225 | .8600 |
| | | −.1960 | .1652 | .0166 | −.7671 | .1592 | .0049 | .0563 | .0213 | .0155 |
| | | *B. Incorporating Bayesian Priors* | | | | | | | | |
| 4 | $\hat{y}_t$ | .9297 | .0252 | .0000 | .7651 | .0844 | .0600 | .1624 | .0243 | .9920 |
| | | −.0203 | .0124 | .0000 | −.0849 | .0479 | .0106 | .1224 | .0107 | .0040 |
| 5 | $\hat{c}_t$ | .9487 | .0229 | .0000 | .8457 | .0531 | .0040 | .0404 | .0089 | .0000 |
| | | −.0013 | .0072 | .0000 | −.0043 | .0288 | .0028 | .0004 | .0026 | .0000 |
| 6 | $\hat{n}_t$ | .9215 | .0244 | .0000 | .7300 | .0850 | .1620 | .1267 | .0174 | .9980 |
| | | −0.0285 | .0074 | .0000 | −0.2200 | .0321 | .0165 | .0867 | .0081 | .0020 |

*Notes*: See the notes to Tables 1 and 4.

**Table 8. Method of Moments**
**Artificial Data is Generated by Nonlinear Model**

| | | | $\beta$ | | | $\rho$ | | | $\sigma$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| Experiment # | Var. | Mean Bias | A.S.E. S.D. | Size S.E. | Mean Bias | A.S.E. S.D. | Size S.E. | Mean Bias | A.S.E. S.D. | Size S.E. |
| | | | | | *A. Simulated Method of Moments with $\tau = 5$* | | | | | |
| 1 | $\hat{y}_t, \hat{c}_t$ | .9505 | .0134 | .0360 | .8408 | .0409 | .1100 | .0390 | .0041 | .1100 |
| | | .0005 | .0117 | .0083 | −.0092 | .0501 | .0140 | −.0010 | .0032 | .0140 |
| 2 | $\hat{n}_t, \hat{y}_t$ | .9477 | .0144 | .0740 | .8386 | .0319 | .1480 | .0393 | .0037 | .0780 |
| | | −.0023 | .0157 | .0117 | −.0114 | .0450 | .0159 | −.0007 | .0037 | .0120 |
| 3 | $\hat{n}_t, \hat{c}_t$ | .9506 | .0115 | .0660 | .8410 | .0302 | .1380 | .0393 | .0033 | .1020 |
| | | .0006 | .0117 | .0111 | −.0090 | .0415 | .0154 | .0007 | .0036 | .0135 |
| | | | | | *B. Simulated Method of Moments with $\tau = 10$* | | | | | |
| 4 | $\hat{y}_t, \hat{c}_t$ | .9507 | .0126 | .0070 | .8401 | .0386 | .1240 | .0390 | .0040 | .0280 |
| | | .0007 | .0122 | .0114 | −.0099 | .0503 | .0147 | −.0010 | .0031 | .0074 |
| 5 | $\hat{n}_t, \hat{y}_t$ | .9486 | .0138 | .0700 | .8410 | .0306 | .1340 | .0392 | .0036 | .0680 |
| | | −.0014 | .0148 | .0114 | −.0090 | .0400 | .0152 | −.0008 | .0035 | .0113 |
| 6 | $\hat{n}_t, \hat{c}_t$ | .9493 | .0110 | .0720 | .8410 | .0285 | .1760 | .0391 | .0031 | .0660 |
| | | −.0007 | .0115 | .0116 | −0.0090 | .0404 | .0170 | −.0009 | .0031 | .0111 |
| | | | | | *C. Simulated Method of Moments with $\tau = 20$* | | | | | |
| 7 | $\hat{y}_t, \hat{c}_t$ | .9495 | .0127 | .0380 | .8351 | .0388 | .1180 | .0389 | .0038 | .0440 |
| | | −.0005 | .0116 | .0086 | −.0149 | .0486 | .0144 | −.0011 | .0030 | .0092 |
| 8 | $\hat{n}_t, \hat{y}_t$ | .9494 | .0133 | .0760 | .8419 | .0295 | .1480 | .0390 | .0035 | .1480 |
| | | −.0006 | .0144 | .0119 | −.0081 | .0427 | .0159 | −.0010 | .0033 | .0159 |
| 9 | $\hat{n}_t, \hat{c}_t$ | .9499 | .0108 | .0800 | .8406 | .0281 | .1380 | .0389 | .0030 | .0960 |
| | | −.0001 | .0115 | .0121 | −.0094 | .0398 | .0154 | −.0011 | .0030 | .0132 |
| | | | | | *D. Generalized Method of Moments* | | | | | |
| 10 | $\hat{y}_t, \hat{c}_t$ | .9509 | .0124 | .0460 | .8446 | .0367 | .1040 | .0387 | .0038 | .0340 |
| | | .0009 | .0113 | .0094 | −.0054 | .0442 | .0137 | −.0013 | .0028 | .0081 |
| 11 | $\hat{n}_t, \hat{y}_t$ | .9490 | .0130 | .0520 | .8426 | .0284 | .0780 | .0393 | .0034 | .0780 |
| | | −.0010 | .0136 | .0099 | −.0074 | .0379 | .0120 | −.0007 | .0035 | .0120 |
| 12 | $\hat{n}_t, \hat{c}_t$ | .9502 | .0106 | .0760 | .8411 | .0276 | .0980 | .0390 | .0030 | .1040 |
| | | .0002 | .0112 | .0119 | −.0089 | .0339 | .0133 | −.0010 | .0031 | .0137 |

*Notes*: See the notes to Tables 1, 2 and 4.

**Table 9. Indirect Inference**
**Artificial Data is Generated by Nonlinear Model**

| Experiment # | Var. | $\beta$ Mean Bias | $\beta$ A.S.E. S.D. | $\beta$ Size S.E. | $\rho$ Mean Bias | $\rho$ A.S.E. S.D. | $\rho$ Size S.E. | $\sigma$ Mean Bias | $\sigma$ A.S.E. S.D. | $\sigma$ Size S.E. |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | *A.* $\tau = 5$ | | | | | |
| 1 | $\hat{y}_t, \hat{c}_t$ | .9501 | .0246 | .3380 | .8171 | .0744 | .3840 | .0400 | .0004 | .7060 |
| | | .0001 | .0349 | .0212 | −.1111 | −.0329 | .0218 | .0000 | .0021 | .0204 |
| 2 | $\hat{n}_t, \hat{y}_t$ | .9416 | .0234 | .2820 | .7746 | .0945 | .2860 | .0399 | .0002 | .8600 |
| | | .0084 | .0323 | .0201 | −.0754 | .1897 | .0202 | −.0001 | .0022 | .0155 |
| 3 | $\hat{n}_t, \hat{c}_t$ | .9467 | .0274 | .0080 | .8420 | .0985 | .0020 | .0398 | .0006 | .5960 |
| | | −.0033 | .0118 | .0040 | −.0080 | .0435 | .0020 | −.0002 | .0021 | .0219 |
| | | | | | *B.* $\tau = 10$ | | | | | |
| 4 | $\hat{y}_t, \hat{c}_t$ | .9492 | .0237 | .3400 | .8166 | .0721 | .3680 | .0400 | .0003 | .7540 |
| | | −.0008 | .0345 | .0212 | −.0334 | .1434 | .0216 | .0000 | .0021 | .0193 |
| 5 | $\hat{n}_t, \hat{y}_t$ | .9406 | .0236 | .2620 | .7740 | .0949 | .2440 | .0398 | .0002 | .8560 |
| | | −.0094 | .0320 | .0197 | −.0760 | .1769 | .0192 | −.0002 | .0021 | .0157 |
| 6 | $\hat{n}_t, \hat{c}_t$ | .9469 | .0265 | .0120 | .8432 | .0944 | .0020 | .0401 | .0005 | .6260 |
| | | −.00312 | .0118 | .0049 | −.0568 | .0439 | .0020 | .0001 | .0022 | .0216 |
| | | | | | *C.* $\tau = 20$ | | | | | |
| 7 | $\hat{y}_t, \hat{c}_t$ | .9534 | .0250 | .2940 | .8315 | .0693 | .3620 | .0400 | .0003 | .7580 |
| | | .0034 | .0346 | .0204 | −.0185 | .1396 | .0215 | .0000 | .0021 | .0192 |
| 8 | $\hat{n}_t, \hat{y}_t$ | .9405 | .0245 | .2540 | .7738 | .0967 | .2560 | .0399 | .0002 | .8460 |
| | | −.0095 | .0317 | .0195 | −.0762 | .1813 | .0195 | −.0001 | .0019 | .0161 |
| 9 | $\hat{n}_t, \hat{c}_t$ | .9459 | .0253 | .0060 | .8402 | .0923 | .0000 | .0399 | .0005 | .6420 |
| | | −.0041 | .0105 | .0035 | −.0098 | .0390 | .0000 | −.0001 | .0021 | .0214 |

*Notes*: See the notes to Tables 1, 3 and 4.

**Table 10. Comparison in Terms of Computing Time**

| Method | $\tau$ | Seconds per Replication |
|---|---|---|
| Maximum Likelihood | – | 0.60 |
| Simulated Method of Moments | 5 | 1.76 |
| | 10 | 2.66 |
| | 20 | 5.22 |
| Indirect Inference (EMSM) | 5 | 4.59 |
| | 10 | 8.34 |
| | 20 | 16.40 |
| Generalized Method of Moments | – | 0.02 |

*Notes*: The Monte Carlo was performed using GAUSS for Windows running in a Dell Extreme Edition with Pentium 4 processor.

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

[3] Bergin, P. R., (2003), "Putting the 'New Open Economy Macroeconomics' to a Test," *Journal of International Economics* 60, 3-34.

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

[5] Boivin, J. and Giannoni, M. P., (2005), "DSGE Models in a Data-Rich Environment," Columbia University, *Mimeo.*

[6] Bouakez, H., Cardia, E., and Ruge-Murcia, F. J., (2005), "Habit Formation and the Persistence of Monetary Shocks," *Journal of Monetary Economics* 52, 1073-1088.

[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] Burnside, C., Eichenbaum, M., and Rebelo, S., (1993), "Labor Hoarding and the Business Cycle," *Journal of Political Economy* 101, 245-273.

[9] Chang, Y., Gomes, J. F., and Schofheide, F., (2002), "Learning-by-Doing as a Propagation Mechanism," *American Economic Review* 92, 1498-1520.

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

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

[12] Christiano, L. J., Eichenbaum, M., and Evans, C. L., (2005), "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy* 113, 1-45.

[13] 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.

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

[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., (2004), "Estimating Macroeconomic Models: 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] Gouriéroux, C., Monfort, A., and Renault, E., (1993), "Indirect Inference," *Journal of Applied Econometrics* 8, S85-S118.

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

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

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

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

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

[24] 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.

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

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

[27] Ingram, B. F., Kocherlakota, N., and Savin, N. E., (1994), "Explaining Business Cycles: A Multiple-Shock Approach," *Journal of Monetary Economics* 34, 415-428.

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

[29] Ireland, P., (2004), "A Method for Taking Models to the Data," *Journal of Economic Dynamics and Control* 28, 1205-1226.

[30] 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.

[31] 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.

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

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

[34] 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.

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

[36] Leeper, E. M. and Sims, C. A., (1994), "Toward a Modern Macroeconomic Model Usable for Policy Analysis" *NBER Macroeconomics Annual 1994,* 81-118.

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

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

[39] 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.

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

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

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

[43] Ruge-Murcia, F. J., (2002), "Methods to Estimate Dynamic Stochastic General Equilibrium Models," University of California at San Diego, *Economics Discussion Paper 2002-18.*

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

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

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

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

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

[49] Williams, J. C. and Wright, B. D., (1991), *Storage and Commodity Markets*, Cambridge University Press: Cambridge.