Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER Journal of Economic Dynamics & Control 31 (2007) 2599–2636
JOURNAL OF Economic Dynamics & Control
www.elsevier.com/locate/jedc

Methods to estimate dynamic stochastic general equilibrium models

Francisco J. Ruge-Murcia*

Départemente de sciences économiques and CIREQ, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Qué., Canada H3C 3J7

Received 9 December 2005; accepted 10 July 2006
Available online 15 November 2006

Abstract

This paper employs the one-sector real business cycle model as a testing ground for four different procedures to estimate dynamic stochastic general equilibrium (DSGE) models. The procedures are: (1) maximum likelihood, with and without measurement errors and incorporating priors, (2) generalized method of moments, (3) simulated method of moments, and (4) indirect inference. Monte carlo analysis is used to study the small-sample properties of these estimators and to examine the implications of misspecification, stochastic singularity, and weak identification.
© 2006 Elsevier B.V. All rights reserved.

JEL classification: E13; C11; C13; C15; C32

Keywords: Estimation methods; DSGE models; Stochastic singularity; Priors

1. Introduction

This paper employs the one-sector real business cycle (RBC) model as a testing ground for four different methods to estimate dynamic stochastic general equilibrium (DSGE) models. The estimation methods are maximum likelihood (ML), generalized method of moments (GMM), simulated method of moments

*Tel.: +1 514 343 2396; fax: +1 514 343 5831.
E-mail address: francisco.ruge-murcia@umontreal.ca.

0165-1889/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jedc.2006.09.005

(SMM), and the indirect inference procedure proposed by Smith (1993). All these methods 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 compare their computational costs, and to examine the implications of weak identification and misspecification.

Monte Carlo experiments are carried out under the null hypothesis and under three possible alternatives using samples of the size typically found in empirical work. Under the null, the data generating process (DGP) and the estimated model are the same. Although all methods deliver consistent parameter estimates, weak identification, stochastic singularity, and small-sample distortion are (or should be) important considerations in their practical application. Weak identification may arise intrinsically from the model solution[1] and/or from an unfortunate choice of variables or moments to estimate the model. For example, we will see that the log likelihood function of output is flatter with respect to the discount factor than that of consumption or hours worked, and that the objective function used in indirect inference may be less locally convex than that of GMM because they focus on different moments of the data.

Stochastic singularity imposes restrictions on the variables and moments that may be used for model estimation, and on the VAR representation of artificial data generated by a DSGE model. DSGE models are singular because they use a small number of structural shocks to generate predictions about a large number of observable variables. Hence, these models predict that linear combinations of observable variables should hold without noise.[2] This prediction is not satisfied by the data and is only the result of a particular misspecification, namely that the model assumes a smaller number of shocks than are present in the real world. This paper shows that singularity affects more severely ML than moment-based methods: ML estimation is limited by the number of linearly independent variables while moment-based estimation is limited by the number of linearly independent moments. The 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. The use of measurement errors to sidestep stochastic singularity in the ML framework is studied here as well.

The small-sample distortion in statistical inference is primarily due to fact that the asymptotic distributions of test statistics may be different from their small-sample analogues. For example, we will see that the empirical size of the t test that the parameter takes its true value may be quite different from the nominal size because asymptotic standard errors are not always a good measure of the small-sample variability of the estimates.

[1] See Canova and Sala (2005) for an example.

[2] Strictly speaking, stochastic singularity is a feature of linearized DSGE models, but it may also have implications for the estimation of nonlinear models depending on the extent to which they differ from their linearized counterparts. For some of the econometric issues that arise in the estimation of nonlinear DSGE models, see An and Schorfheide (2005) and Fernández-Villaverde and Rubio-Ramirez (2006) in the context of ML; and Kim and Ruge-Murcia (2006) in the context of method of moments.

Under the alternative, the data are not generated by the model of interest but instead by an alternative model. Three alternative models are considered. First, the DGP is the linearized RBC model, but with multiple structural shocks. Second, the DGP is a linearized RBC model with habit formation in consumption. Third, the DGP is the nonlinear version of the one-shock RBC model. In all cases, the estimated model is the economically interesting but (now) misspecified RBC model with time separable preferences and only one technology shock. The goal of these experiments is to study the robustness of each method to misspecification. The main results of this analysis are that GMM and SMM are generally more robust to misspecification than ML, but that adding measurement errors and using informative priors are helpful in limiting the effects of misspecification in the ML framework.

The paper is organized as follows: Section 2 outlines the DSGE model that will be used as backdrop for the analysis, Section 3 describes the estimation methods and their application to DSGE models, Section 4 presents the Monte Carlo design and report the results, and 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 RBC model with indivisible labor (Hansen, 1985). The representative agent maximizes expected lifetime utility

$$\mathrm{E}_t \sum_{i=t}^{\infty} \beta^{i-t}(\ln(c_i) + \psi(1 - n_i)),$$

where $\beta \in (0, 1)$ is the discount factor, c_t is consumption, n_t is hours worked, and ψ 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. 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

$$1/c_t = \beta \mathrm{E}_t((1/c_{t+1})(1 + r_{t+1} - \delta)),$$
$$\psi c_t = w_t.$$

The single, perishable good in this economy is produced by perfectly competitive firms. The representative firm rents labor and capital from the agent and combines them using the constant returns to scale technology

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

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

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

where $\rho \in (-1,1)$ and ε_t is an innovation assumed to be independently, identically, and normally distributed (*i.i.d.N.*) with zero mean and variance σ^2. In every period, the firm chooses input levels to maximize profits and equates the marginal product of labor (capital) to the real wage (rental rate). 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.

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. Using the circumflex to denote percentage deviation from steady state and after some manipulations write

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

where

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

$$\mathbf{B} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \delta/(\alpha-\alpha\gamma) \\ \varsigma\rho/(\varsigma+\alpha-\alpha\varsigma) \end{bmatrix},$$

$\varsigma = \alpha\beta(k/n)^{\alpha-1}$, $k/n = ((1/\beta+\delta-1)/\alpha)^{1/(\alpha-1)}$ 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 ϕ_{ck} and ϕ_{cz} are combinations of the eigenvectors and eigenvalues of the matrix $\mathbf{A}$ and, consequently, depend nonlinearly on the structural parameters.

Define the 2×1 vector $\boldsymbol{\xi}_t = (\hat{k}_t, \hat{z}_t)'$ with the state variables of the system, and the 3×1 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

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

where the elements of the 3×2 matrix $\mathbf{\Phi}$ are nonlinear functions of structural parameters and the last row exactly reproduces Eq. (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 observable endogenous variables, the model predicts that there are linear-combinations of these variables that hold without noise. Thus, the model is stochastically singular. Two examples of linear-dependent relations are

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

and

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

Eq. (4) is obtained by substituting out $\hat{k}_t$ and $\hat{z}_t$ from (3). Eq. (5) is obtained by manipulating the equations for $\hat{y}_t$ and $\hat{c}_t$ in (3) and the linearized law of motion for capital.[3] Under the model, the variance–covariance matrices of the systems $(\hat{n}_t, \hat{y}_t, \hat{c}_t)$ and $(\hat{c}_t, \hat{c}_{t-1}, \hat{y}_{t-1}, \hat{y}_t)$ are singular for any sample size and parameter values.

3. Estimation methods

3.1. Maximum likelihood

The 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, and use the Kalman filter to generate optimal inferences about the unobserved state vector and to evaluate the joint likelihood function of observable endogenous variables. In turn, the maximization of this function yields consistent and asymptotically normal estimates of the model parameters.[4]

[3] Similar manipulations deliver relations like (5) for the systems $(\hat{y}_t, \hat{n}_t, \hat{y}_{t-1}, \hat{n}_{t-1})$ and $(\hat{n}_t, \hat{c}_t, \hat{n}_{t-1}, \hat{c}_{t-1})$.

[4] In the case where disturbances are not normal, this procedure still delivers consistent and asymptotically normal (Quasi-) ML estimates, but with the variance–covariance matrix derived by White (1982). For the case of nonlinear DSGE models, Fernández-Villaverde and Rubio-Ramirez (2006) propose the use of a sequential Monte Carlo filter.

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

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

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

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

where $\mathbf{h}$ is a 1×3 selection vector. 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 the set of past observations of $\mathbf{x}_t$ by $\aleph_{t-1}$, the time $t-1$ forecast of $\boldsymbol{\xi}_t$ constructed on the basis of $\aleph_{t-1}$ by $\tilde{\boldsymbol{\xi}}_{t|t-1}$, and the mean square error (MSE) of this forecast by $\mathbf{P}_{t|t-1}$. Then, under the assumption that the technology innovation is normally distributed, the density of $\mathbf{x}_t$ conditional on $\aleph_{t-1}$ is

$$f(\mathbf{x}_t|\aleph_{t-1}; \boldsymbol{\theta}) = \mathrm{N}(\mathbf{H}\tilde{\boldsymbol{\xi}}_{t|t-1}, \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}'). \tag{6}$$

The ML estimator of $\boldsymbol{\theta}$ is

$$\tilde{\boldsymbol{\theta}}_{\mathrm{ml}} = \underset{\{\boldsymbol{\theta}\}}{\operatorname{argmax}}\ L(\boldsymbol{\theta}), \tag{7}$$

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 may be started with the unconditional moments $\tilde{\boldsymbol{\xi}}_{1|0} = \mathrm{E}(\boldsymbol{\xi}_t) = (0, 0)'$ and $\mathbf{P}_{1|0} = \mathrm{E}(\boldsymbol{\xi}_t\boldsymbol{\xi}_t')$. 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, Chapter 13). Under standard regularity conditions (see Judge et al., 1985, p. 178), the ML estimator is consistent and asymptotically normal

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

where $\Im = -\mathrm{E}(\partial^2 L(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}')$ 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 ML using more observable variables than structural shocks are specified in the model. To see this, consider the one-shock model in Section 2 and write the innovation $\mathbf{x}_t - \mathrm{E}(\mathbf{x}_t|\aleph_{t-1}) = \mathbf{H}(\boldsymbol{\xi}_t - \tilde{\boldsymbol{\xi}}_{t|t-1})$. Note that when $\mathbf{x}_t$ contains more than one variable (for example, consumption and output), the matrix $Var(\mathbf{x}_t - \mathrm{E}(\mathbf{x}_t|\aleph_{t-1})) = \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}'$ 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.

To address the singularity of DSGE models, researchers either (1) estimate the model using at most as many observable variables as structural shocks (see, for example, Kim, 2000; Ireland, 2001; Bouakez et al., 2005) or (2) add error terms to the observation equation of the state-space representation of the model (see, among others, McGrattan, 1994; Hall, 1996; Ireland, 2004). 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 approach include Leeper and Sims (1994) and Ingram et al. (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 it would also add another observable variable (i.e., government expenditure). Still, adding structural shocks may 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.

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 *i.i.d.N.* with zero mean and variance–covariance matrix $\mathrm{E}(\mathbf{u}_t\mathbf{u}_t') = \mathbf{R}$. Following Sargent (1989), it is common to interpret $\mathbf{u}_t$ as measurement error. It is further assumed that $\mathrm{E}(\mathbf{v}_i\mathbf{u}_j') = \mathbf{0}$ for all $i \geqslant j$, 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, Chapter 8).

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

$$\tilde{\boldsymbol{\theta}}_{\mathrm{ml}} = \underset{\{\boldsymbol{\theta}\}}{\operatorname{argmax}}\ L(\boldsymbol{\theta}), \tag{8}$$

but now the log likelihood function is

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

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 provides a less stringent platform to assess the theory. For example, relationships like (4) and (5) may approximately hold in the data, even if not without noise as predicted by the model. On the other hand, measurement error lacks a structural interpretation and it essentially represents misspecification error. The uncertainty about the true value of $\mathbf{x}_t$ which arises from 'measurement' error is taken into account by the macroeconomist for the purpose of estimating the model, but not by agents for the purpose of solving their optimization problems. Thus (in contrast to the technology shock, z_t), $\mathbf{u}_t$ is not a structural disturbance. A more natural interpretation of $\mathbf{u}_t$ is that of misspecification error. The stochastic singularity of DSGE models arises from a particular misspecification: there are more shocks in the real world than those assumed in the model. This is the main misspecification that is captured by $\mathbf{u}_t$. In addition, when $\mathbf{u}_t$ is modeled as serially and/or contemporaneously correlated, this term also captures dynamics and/or cross-correlations in the data that the model cannot explain.[5]

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 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. A simple way to incorporate priors into the ML 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 optimally incorporates prior information regarding the parameter values. This GLS estimator is given a ML interpretation by Stone (1954). It is easy to show that the mean and variance of the optimal GLS estimator corresponds exactly to mean and variance of the Bayesian posterior distribution (see Hamilton, 1994, p. 359).

[5]This point was made earlier by Ireland (2004). Boivin and Giannoni (2005) show that adding 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. To improve identification, Boivin and Giannoni propose a ML method that exploits the cross-sectional variation of multiple indicators (for example, multiple measures of output).

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

$$\boldsymbol{\mu} = \mathbf{K}\boldsymbol{\theta} + \mathbf{e}, \tag{9}$$

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 $f(\boldsymbol{\theta}) = \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

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

$$\tilde{\boldsymbol{\theta}}_{\mathrm{qb}} = \underset{\{\boldsymbol{\theta}\}}{\operatorname{argmax}} \left(L(\boldsymbol{\theta}) + L(\boldsymbol{\mu}(\boldsymbol{\theta}))\right), \tag{10}$$

where $L(\boldsymbol{\mu}(\boldsymbol{\theta}))$ is the log of the density of $\boldsymbol{\mu}$ in (9) 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 (10) 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 would be obtained using the approach in DeJong et al. (2000). In the special case where the prior is diffuse or uninformative, the estimator $\tilde{\boldsymbol{\theta}}_{\mathrm{qb}}$ converges to the classical ML estimator. When the prior is proper, the quasi-Bayesian estimate of $\boldsymbol{\theta}$ may 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.

The ML estimation of DSGE models using proper Bayesian techniques is a very active area of research. Among early contributions are Schorfheide (2000), DeJong et al. (2000), Chang et al. (2002), and Smets and Wouters (2003). Typically researchers use either the Metropolis–Hasting algorithm or Importance Sampling to generate draws from the posterior distribution of the parameters. The former method uses a path-dependent algorithm to construct draws of proposed realizations of $\boldsymbol{\theta}$ and the Metropolis criterion to determine whether a realization is accepted into the sample. This sample is then used to estimate the moments of the posterior distribution. The latter method constructs directly an approximation to the moments of the posterior distribution by sampling from a multivariate distribution of $\boldsymbol{\theta}$ (for example, the Gaussian one obtained above) and re-weighting by the importance ratio. A more detailed presentation of Bayesian methods and an exhaustive review of this literature can be found in An and Schorfheide (2005).

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

$$\min_{\{\boldsymbol{\theta}\}} \mathbf{G}(\boldsymbol{\theta})'\mathbf{W}\mathbf{G}(\boldsymbol{\theta}), \tag{11}$$

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

$$\mathbf{W} = \lim_{T\to\infty} Var\left((1/\sqrt{T})\sum_{t=1}^{T}\mathbf{m}_t\right)^{-1}. \tag{12}$$

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

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

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

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

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

Singularity restricts the variables that may be included in $\mathbf{m}_i$ or, equivalently, the moments that can be exploited for their 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 null) 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. Note that singularity does not prevent the evaluation of the objective function in (11) because 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.

Singularity has different implications for SMM and ML. 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 model in Section 2, it is easy to find independent moments that involve two variables, even though (conditional on the predetermined capital stock) no two variables are linearly independent.

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 $\mathrm{E}(\mathbf{m}(\boldsymbol{\theta}))$ in the objective function (11), and a GMM estimator of $\boldsymbol{\theta}$ may be computed by minimizing the distance between the empirical moments of the data and the theoretical moments predicted by the model.[6] This approach has been followed, among others, by Christiano and Eichenbaum (1992) and Burnside et al. (1993).

The GMM estimator is defined by

$$\tilde{\boldsymbol{\theta}}_{\mathrm{gmm}} = \underset{\{\boldsymbol{\theta}\}}{\operatorname{argmin}}\ \mathbf{G}(\boldsymbol{\theta})'\mathbf{W}\mathbf{G}(\boldsymbol{\theta}), \tag{13}$$

where

$$\mathbf{G}(\boldsymbol{\theta}) = (1/T)\sum_{t=1}^{T}\mathbf{m}_t - \mathrm{E}(\mathbf{m}(\boldsymbol{\theta})), \tag{14}$$

and $\mathbf{W}$ is the $q \times q$ optimal weighting matrix defined in (12). Under the regularity conditions in Hansen (1982)

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

where $\mathbf{D} = \partial\mathrm{E}(\mathbf{m}(\boldsymbol{\theta}))/\partial\boldsymbol{\theta}$ is a $q \times p$ matrix assumed to be finite and of full rank. In principle, one could obtain $\partial\mathrm{E}(\mathbf{m}(\boldsymbol{\theta}))/\partial\boldsymbol{\theta}$ analytically using the expressions for the unconditional moments $\mathrm{E}(\mathbf{m}(\boldsymbol{\theta}))$. 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 $\partial\mathrm{E}(\mathbf{m}(\boldsymbol{\theta}))/\partial\boldsymbol{\theta}$ numerically prior to the calculation of the standard errors. The optimal weighting matrix is computed

[6]This formulation of GMM is closely related to the minimum distance estimator in Malinvaud (1970).

using the Newey–West estimator with a Barlett kernel. A global specification test for overidentified DSGE models may be performed using the chi-square statistic (see Hansen, 1982)

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

where $\mathbf{G}(\tilde{\boldsymbol{\theta}}_{\text{gmm}})'\mathbf{W}\mathbf{G}(\tilde{\boldsymbol{\theta}}_{\text{gmm}})$ 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 $\tau T \to \infty$, 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 quickly as τ increases. For example, for $\tau = 5$, 10 and 20, the standard errors of SMM estimates should be, respectively, 1.10, 1.05 and 1.025 times larger than those of GMM estimates. Since $(1+1/\tau) \to 1$ as $\tau \to \infty$, the efficiency of SMM converges to that of GMM as the length of the simulated series increases.[7]

Some authors (see, for example, Braun, 1994) construct estimates of the parameters of DSGE models by applying GMM to a subset of the model equations (for example, the first-order conditions). This single-equation GMM estimation does not involve the solution of the model and, consequently, does not exploit its cross-equation restrictions. As a result, single-equation GMM estimates will be less efficient than those obtained using ML and may be subject to more severe identification problems. In addition, the use of instruments to deal with the endogeneity of right-hand regressors may lead to biased parameter estimates and poor small-sample properties when instruments are weakly correlated with the variables they are meant to instrument for. See Fuhrer et al. (1995) for a comparison of ML and GMM in the context of the linear-quadratic inventory model, and Stock et al. (2002) for a survey on the effects of weak instruments for GMM estimation and inference.

The GMM estimator defined in (13) does not involve the use of instruments and requires the solution of the DSGE model in every step of the minimization routine. Thus, compared with single-equation GMM estimation, it delivers more efficient estimates and is less subject to identification problems, but it may suffer from a problem analogous to the weak-instrument problem when the moments included in the objective function are not very informative about the structural parameters. Compared with ML, GMM may be less efficient because it focuses on limited aspects of the data. To see this it is useful to recall that under certain conditions ML is a special case of GMM applied to the scores. In terms of our notation, $\mathbf{m}_t$ may be the $q \times 1$ vector $\partial \log f(\mathbf{x}_t|\aleph_{t-1};\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ where $f(\mathbf{x}_t|\aleph_{t-1};\boldsymbol{\theta})$ was defined in (6). Then

$$(1/T)\sum_{t=1}^{T}\mathbf{m}_t = (1/T)\sum_{t=1}^{T}\partial \log f(\mathbf{x}_t|\aleph_{t-1};\boldsymbol{\theta})/\partial\boldsymbol{\theta}$$

[7]The implications of stochastic singularity for GMM are the same as for SMM. That is, singularity restricts the moments that may be employed to those that form a linearly independent set.

is just the sample average of the scores. With $E(\mathbf{m}(\boldsymbol{\theta})) = 0$, the solution to the (now) exactly identified system in (13) delivers a GMM estimator that also satisfies the first-order conditions for the maximization of the log likelihood function (since $\mathbf{G}(\tilde{\boldsymbol{\theta}}_{\text{gmm}}) = 0$ in the exactly identified case). However, this estimator is constructed by matching all information contained in the density while the typical GMM estimator matches only some features of the distribution like, for example, some second moments. Thus, a general concern is that identification by limited-information methods is bound to be poorer than by full-information methods.[8] On the other hand, as was explained in Section 3.1, when $\mathbf{x}_t$ consists of more variables than structural shocks, the log likelihood function and, consequently, the scores are not well defined. Hence, when the model is singular, it is not clear that matching moments of, say, two variables would produce less precise estimates than matching the sample average of the scores of one variable.

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. Instead, 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 et al., 1993) because the auxiliary model (the finite VAR) is a misspecified version of the true state-space representation of the model.

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

$$\min_{\{\boldsymbol{\theta}\}} (\boldsymbol{\eta} - \boldsymbol{\eta}(\boldsymbol{\theta}))' \mathbf{V} (\boldsymbol{\eta} - \boldsymbol{\eta}(\boldsymbol{\theta})), \tag{15}$$

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),

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

[8]For example, for a model similar to one used here, Canova and Sala (2005, p. 13) report that weak identification problems appear to be more acute when the distance between impulse responses, rather than the likelihood function, is used as objective function.

where $\mathbf{J} = \mathrm{E}(\partial \boldsymbol{\eta}(\boldsymbol{\theta})/\partial \boldsymbol{\theta})$ is a $q \times p$ matrix assumed to be finite and of full rank. In the Monte Carlo, the derivatives $\partial \boldsymbol{\eta}_t(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ are computed numerically and the expectation approximated by the average over the simulated τT data points. Smith suggests a test based on Hansen's (1982) chi-square statistic as specification test for overidentified DSGE models

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

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

The stochastic singularity of DSGE models has implications for the order of, and the number of variables in, the auxiliary VAR. This means that the artificial data generated by a DSGE model does not have an unrestricted VAR representation. In the case of the RBC model in Section 2, the variance–covariance matrices of 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. Hence, data generated by this model have only a bivariate VAR representation of order one. 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 totwo of the three observable endogenous variables and to only one of their lags, I use VARs of order one for $(\hat{y}_t, \hat{n}_t)$, $(\hat{y}_t, \hat{c}_t)$, and $(\hat{n}_t, \hat{c}_t)$ in the Monte Carlo experiments to examine the sensitivity of the results to the variables used.

Rather than matching VAR parameters, some authors match instead the impulse responses of the model (see, for example, Rotemberg and Woodford, 1997; Christiano et al., 2005). Since the impulse responses are nonlinear transformations of VAR parameters, this strategy is conceptually similar to the one studied here. Matching impulse responses is attractive because they intuitively summarize the dynamics of the model and data, and allow the researcher to focus on specific horizons by choosing an appropriate weighting matrix. On the other hand, Canova and Sala (2005) argue that identical impulse responses may be generated by different economic models, that under-identification is possible because the shock size is normalized in impulse-response analysis, and that weak identification problems may arise because responses at long horizons are noisy and may contain little information about structural parameters. Canova and Sala suggest that some of these issues may apply to matching VAR parameters despite the fact that shock identification and normalization are not required in this approach.

4. Monte Carlo experiments

4.1. Design

In this Section, Monte Carlo analysis is used to study the small-sample properties of the estimators and to examine the implications of misspecification, stochastic singularity, and weak identification. The analysis is carried out under the null and under three possible alternatives. In the former case, the DGP and the estimated model are both the linearized RBC model in Section 2. In the latter case, the DGPs

are a model with multiple structural shocks, a model with habit formation in consumption, and a nonlinear model, but the macroeconomists estimates the (now misspecified) linearized RBC model. All experiments are based on 500 replications using a sample size of 200 observations. This sample size corresponds to, say quarterly observations of series for a period of 50 years.

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. For the Monte Carlo experiments, I concentrate on three (of the five) model parameters, namely the subjective discount factor (β), the autocorrelation coefficient of the technology shock (ρ), and the standard deviation of the technology innovation (σ), and fix the values of the capital share (α) and the depreciation rate (δ) in all experiments. This reduces the computational burden in the Monte Carlo and sidesteps the weak identification of α and δ intrinsic to this model.[9] Hence, in what follows, $\boldsymbol{\theta} = (\beta, \rho, \sigma)'$ is a 3×1 vector. The parameter values used to generate the data are $\beta = .95$, $\alpha = .36$, $\delta = .025$, $\rho = .85$, and $\sigma = .04$. This Monte Carlo design is realist in the sense that the true parameter values correspond to the ones typically found in empirical analysis. For the Monte Carlo experiments with priors, I consider the following independent prior densities for β and ρ: $\beta \sim \mathrm{N}(.95, .025^2)$ and $\rho \sim \mathrm{N}(.85, .07^2)$, and a diffuse prior for σ. These priors mean that, before observing the data, the researcher believes that with a 95% probability their true values lay in the intervals $(.901, .999)$ and $(.713, .987)$, respectively. Notice that prior distributions are centered around the true parameter values. For SMM and indirect inference, I use three different values of $\tau = 5$, 10, 20, meaning that the simulated series are, respectively, 5, 10, and 20 times larger than the sample size of 200 observations.

4.2. Results under the null hypothesis

This section reports the results of Monte Carlos experiments where the data are generated by the same economic model that the macroeconomist estimates, namely the linearized RBC model in Section 2. The intention here is to study the small-sample properties of the estimators and of the tests used in statistical inference, and to shed some light on identification issues concerning DSGE models.

Results are reported in Tables 1–3. In all tables, mean is the average of the estimated parameter values, and ASE is the average asymptotic standard error. Both averages are taken over the 500 replications in each experiment. Median and SD are the median and standard deviation of the empirical parameter distribution.

[9] Canova and Sala (2005) show that the rational-expectations solution of a RBC model with inelastic labor supply implies that the dynamics of the capital stock are only weakly influenced by α and ρ, and insensitive to proportional changes in δ and β. For the model here (where labor supply is elastic), I found that all loss function were fairly flat when one tries to simultaneously estimate the five parameters. This observation underscores the fact that weak identification may arise despite the fact that DSGE models are fully specified.

Table 1
Maximum likelihood under the null hypothesis

Experiment no.	*Var.*	β			ρ			σ		
		Mean (median)	ASE (SD)	Size (SE)	Mean (median)	ASE (SD)	Size (SE)	Mean (median)	ASE (SD)	Size (SE)
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 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; ASE is the average asymptotic standard error; median and SD 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% of the null hypothesis that the parameter takes its true value; and SE is the standard error of the actual test size. The experiments were based on 500 replications. The priors used are $\beta \sim N(.95, .025^2)$ and $\rho \sim N(.85, .07^2)$, with a diffuse prior for σ.

Comparing median and mean with the true parameter value, and SD with ASE, is informative about the small-sample distribution of the estimates. For example, if Mean is well below the true parameter value, this indicates a 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 SD is larger than ASE, this indicates that using the asymptotic formula to compute the standard error may 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%. In other words, size is the empirical size of this t test. SE 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, or, more formally, the 95% confidence interval around the empirical size would contain the nominal size of 5%. 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-square test of the overidentification restrictions.

Table 2
Method of moments under the null hypothesis

Experiment no.	*Var.*	β			ρ			σ			OI (SE)
		Mean (median)	ASE (SD)	Size (SE)	Mean (median)	ASE (SD)	Size (SE)	Mean (median)	ASE (SD)	Size (SE)	
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.

Panel A in Table 1 reports results using ML without adding measurement errors or priors.[10] Recall that in this case, the RBC model cannot be estimated by ML using

[10]Notice that when the DGP is the linearized RBC model with one shock, it is not possible to estimate the model by ML with measurement errors added. The reason is that under the null, the mean and variance of measurement errors are 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.

Table 3
Indirect inference under the null hypothesis

Experiment no.	*Var*	β			ρ			σ			OI (SE)
		Mean (median)	ASE (SD)	Size (SE)	Mean (median)	ASE (SD)	Size (SE)	Mean (median)	ASE (SD)	Size (SE)	
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.

more than one observable variable. Experiments 1–3 refer to experiments using output, consumption, and hours, respectively. When the model is estimated using output alone, the average estimate of β is well below its true value and the ASE is large. Instead, when the model is estimated using either consumption or hours, the estimate is close to the true value and the ASE is small. Thus, for the RBC model studied here, a sharper estimate of the subjective discount factor is obtained using consumption or hours, rather than output. The variable used has a limited effect on the point estimates of the other structural parameters, but plays a role in the standard error of the empirical distribution and the ASE of ρ. For example, the ASE of ρ is 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 reasonably well the small-sample standard deviations of the estimates, but there are size distortions in the t test. In particular, the test tends to overreject the null hypothesis that the parameter value is the true value.

From the above discussion it follows that, despite the fact that the model is a general equilibrium one, not all variables may be equally informative about all structural parameters. This means that the choice of data series to estimate the model may matter. The reason is that identification may depend on the variable or set of variables used to estimate the model. (Although this result is illustrated here in the context of ML, we will see below that it applies to other estimation methods as well.) General analytical results are difficult to obtain but, in order to develop some intuition, Fig. 1 plots the objective function of each method for a range of values of a parameter holding the other parameters fixed to their true value.[11] Consider first the top panel of Column 1 and notice that the log likelihood function of output is flatter in the dimension of β than the functions of consumption or hours. This indicates that output is less informative about the discount factor than either consumption or hours, and explains why the uncertainty regarding the estimate of β (measured, for example, by the ASE) is larger when output rather than either consumption or hours is used. On the other hand, note that the functions are similarly concave in the dimensions of ρ and σ. This explains why the variable used to estimate the model has a limited effect on the estimates of these two parameters.

Panel B in Table 1 reports results using ML with priors. Regarding β and ρ, these estimates tend to be numerically closer to their true values than ML estimates without priors. This is due to the fact that prior distributions are centered around the true values and, consequently, this statement will not be generally true in actual empirical applications. It is interesting to note that although the prior about σ was uninformative, estimates of σ are closer to the true value in Panel B than in Panel A. Thus, in the context of DSGE models, imposing priors on a set of parameters may be helpful in the estimation of remaining parameters. In most cases, the ASE is larger than the standard deviation of the parameter estimates. Since asymptotic standard errors overestimate the variability of the estimates in small samples, the t test has a smaller empirical than nominal size and tends to underreject the null hypothesis.

Results above indicate that β is weakly identified when estimation is carried out using output data alone. Thus, when the prior distribution for β and output data are combined, the resulting marginal posterior distribution looks similar to the marginal prior density.[12] Comparing the top panels of Columns 1 and 2 in Fig. 1 shows that information extraneous to the model and summarized in the form of a prior may be helpful in introducing curvature and making easier the numerical maximization of the log likelihood function. However, a possible concern is that priors, naively used, may conceal weak identification. In order to guard against this potential pitfall, An and Schorfheide (2005) suggests the direct comparison of priors and posteriors and Canova and Sala (2005) propose the use of progressively less informative priors as a diagnostic tool.

[11]These objective functions were constructed for a sample of 200 observations generated under the null. Since the sample is random and small, the value that optimizes these functions may not necessarily coincide with the true value used to generate the data. To facilitate the visual comparison of these functions, they were scaled so that their maxima (in the case of ML) or minima (in the case of SMM, GMM and indirect inference) are the same.

[12]For a complete discussion of the Bayesian analysis of nonidentified models, see Poirier (1998).

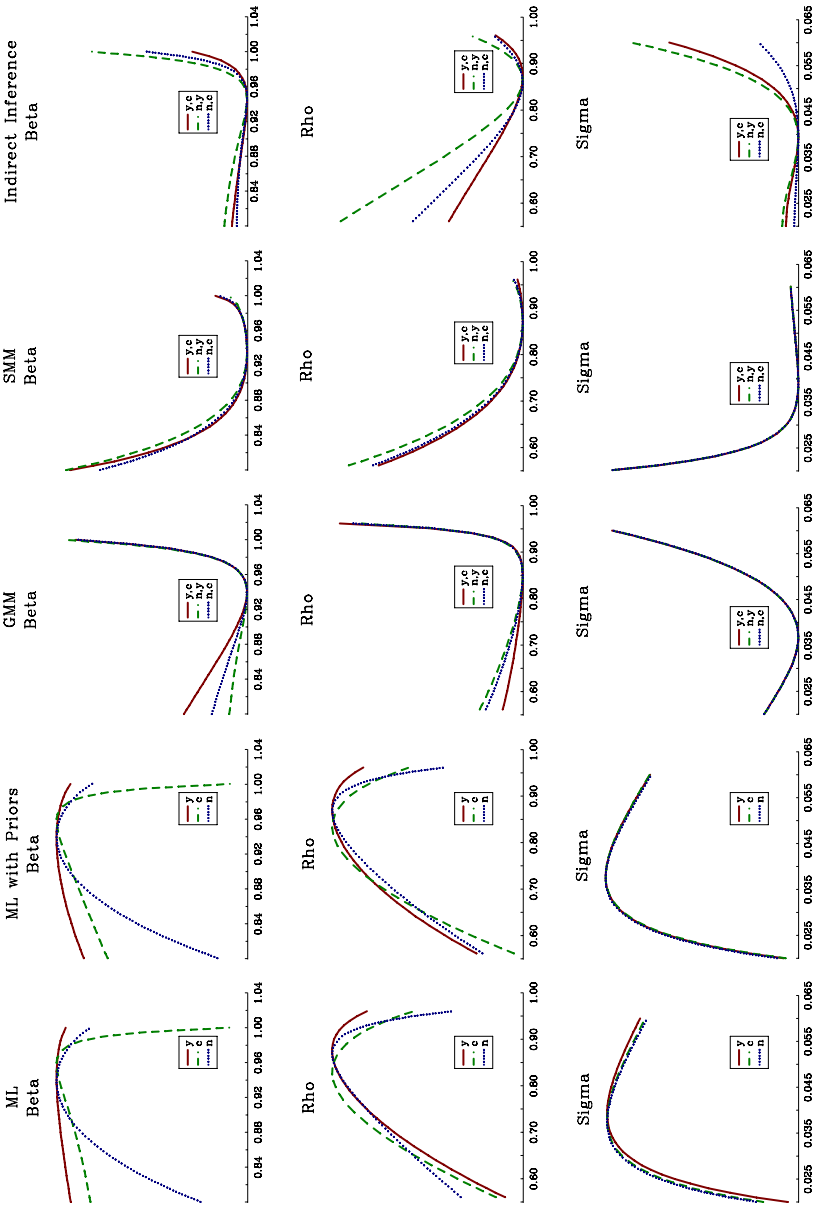

Fig. 1. Objective functions.

Table 2 reports results using the SMM and GMM. In all cases, the mean and median of the estimated parameters are close to their true values. Standard errors are smaller than those obtained using ML without priors because SMM and GMM are less severely affected by the stochastic singularity of the model and can exploit information on additional variables. Standard errors vary with τ as predicted by the discussion in Section 3.4. That is, SMM standard errors based on simulated series with $\tau = 5, 10$, and 20 are approximately 1.1, 1.05, and 1.025 times larger than those obtained using GMM, respectively. However, this variation in statistical efficiency is of the same order of magnitude as that observed across moments used in estimation. This means that the choice of the length of the simulated series in SMM may be as important as the choice of moments to match. For example, in Panel A, the standard deviation of the estimate of β obtained using the moments of output and hours is larger than the one obtained using those of consumption and hours, for the same value of τ. The observation that different moments may be differently informative about the model parameters is illustrated in Columns 3 and 4 of Fig. 1, which plots the objective functions of GMM and SMM with $\tau = 10$ along each of the three possible dimensions. Notice that the convexity of the objective functions depends to some extent on the choice of moments, especially in the case of GMM.

Returning to Table 2, note that there are some differences between the ASEs and the standard deviations of the empirical distribution of the parameters. Since the t statistics are typically computed using asymptotic standard errors, this translates into fairly large size distortions for the t test. There is no obvious pattern for these 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 reports the empirical size of the chi-square 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%. This means that a researcher comparing the chi-square statistic with the 5% 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% probability of rejecting a true model, the researcher actually is taking (approximately) a 1% probability. In some cases (see, for example, Experiment 8), the model's restrictions were never rejected by the chi-square test in the 500 replications. The fact that Hansen's chi-square test easily fails to detect a misspecified model is well known in the literature (see among others, Newey, 1985). The results in this paper suggest that the severe size distortions of the chi-square test also arise in the context of fully specified DSGE models.

Table 3 reports results using indirect inference. In all experiments, the mean and median of the estimated parameters are close to their true values, but the standard errors of the estimates of β and ρ are fairly large. As in the other methods, the choice of variables and moments used has some effect on the precision of the estimates because they are not equally informative about the structural parameters. For example, σ 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 to be informative regarding the standard deviation of technology innovations. Plots of

the objective functions in the last column of Fig. 1 show that convexity, and hence identification, varies substantially across VARs.

Comparing the last three columns of Fig. 1 provides some clues as to why identification may be sharper under either GMM or SMM than under the indirect inference procedure studied here. The moments in the objective function used in indirect inference are those encoded in a VAR, but they may be independently chosen by the macroeconomist in GMM or SMM. The fact that objective functions are usually less (locally) convex in the former than in the latter indicates that (as a whole) the moments in the VAR may not be as informative regarding the parameters as those used here in GMM and SMM estimation. In turn, this suggests that some of the identification problem highlighted by Canova and Sala (2005) may be more easily addressed in the context of GMM or SMM through a judicious choice of moments.

Returning to Table 3, note that there are very large size distortions for both the t test that the parameters take their true value and the chi-square test of the overidentification restrictions. As before, there is no clear pattern for these size distortions, but they seem to vary more with the moments matched than with the length of the simulated series.

It is enlightening to go beyond the summary statistics in these tables and plot the empirical distribution of the parameter estimates obtained under the different methods. The frequency histograms for estimators of β, ρ and σ are, respectively, plotted in the top, middle and bottom panels of Fig. 2. 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 may be drawn from this figure. First, the method of moments (either in its GMM or SMM formulation) may deliver sharper estimates than ML under the null hypothesis. This is due to the fact that stochastic singularity affects less severely the former than the latter method. While the RBC model studied here cannot be estimated by ML using more than one variable, it can be estimated by the method of moments or indirect inference using moments that involve two variables. Second, informative priors may be combined with sample data to improve the researcher's inferences regarding the structural parameters of DSGE model. This can be seen by comparing the empirical distribution of the estimators of β and ρ obtained using ML 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 τ. Fourth, the choice of variables and moments employed can have some effect on the precision of the estimators. This can be seen by comparing the empirical distribution of the estimators of β and ρ 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 diffuse empirical distribution and a substantial number of outliers, though their mean is close to the true value used to generate the sample.

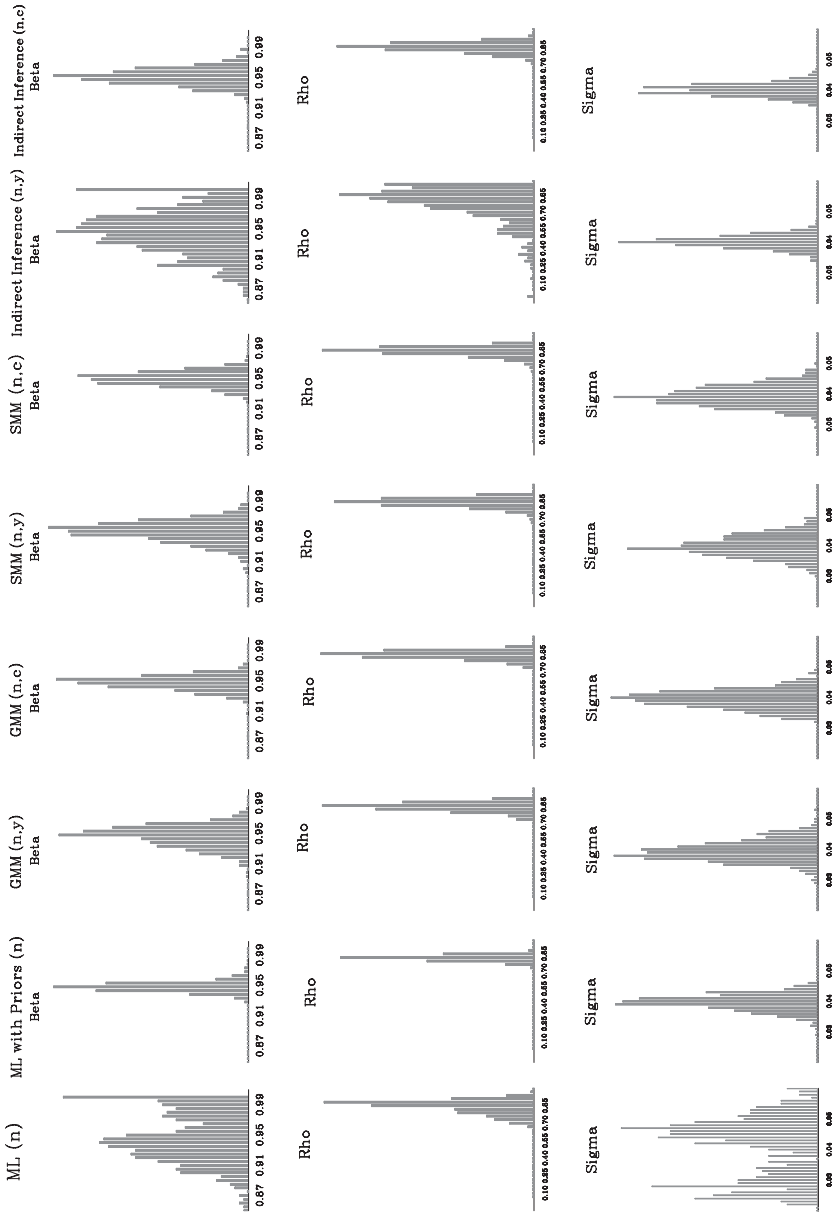

Fig. 2. Empirical distributions under the null.

One of the reasons DSGE models are interesting is because they allow the macroeconomist to examine the response of a model economy to shocks. The estimation of DSGE models delivers a measure of parameter uncertainty in the form of standard errors and bootstraps techniques may then be used to transform this uncertainty into, for example, confidence intervals for the model's 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. Fig. 3 plots the dynamic responses of consumption, output, and hours following a technology shock. The dotted lines are the 90% confidence intervals for the response. The parameter estimates used to construct these figures come from the same experiments used to obtain Fig. 2.

Fig. 3 supports three conclusions. First, since all methods deliver unbiased parameter estimates, the mean response does not depend on the method used. However, there are some differences in the coverage probabilities of the estimated confidence intervals because not all methods 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 β and ρ reported in Fig. 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. That is, the data are not generated by the linearized one-shock RBC model but instead by some alternative model. The intention here is to study the robustness of each estimation method to misspecification. Three possible forms of misspecification are examined. First, the DGP is the linearized RBC model, but with three structural shocks. Thus, a researcher looking at the artificial data on output, consumption, and hours would find them to be linearly independent. Second, the DGP is a linearized RBC model with habit formation in consumption. Third, the DGP is the nonlinear version of the one-shock RBC model. In all cases, the estimated model is the economically interesting but misspecified RBC model with time separable preferences and only one technology shock.

4.3.1. True model has multiple structural shocks

In this experiment, the artificial data are generated by the linearized RBC model in Section 2, but expanded to include 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, ψ. Thus, the weight follows the process $\ln \psi_{t+1} = 0.6 \ln \psi_t + \zeta_t$, where ζ_t is an innovation assumed $i.i.d.N.(0, 0.03^2)$. The shock to the investment function affects the transformation of investment into new capital. The law of motion of the

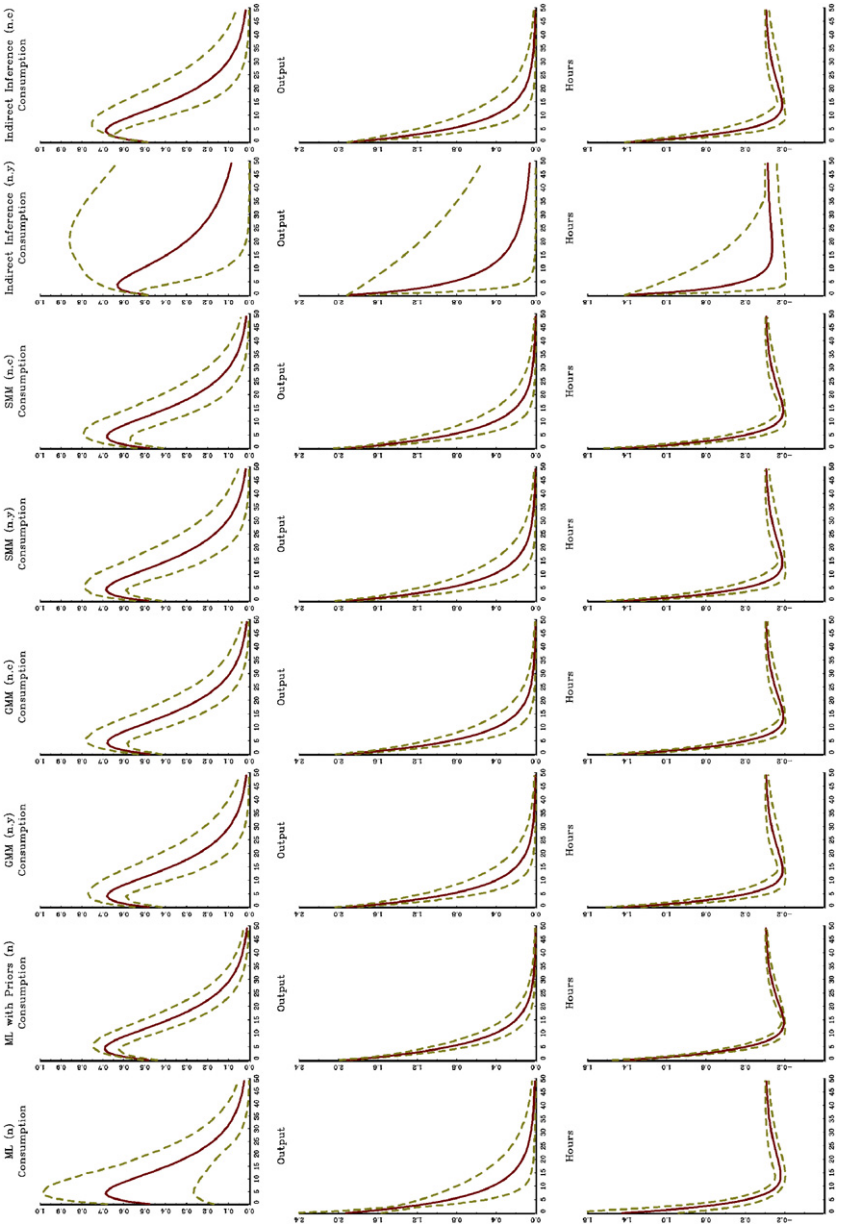

Fig. 3. Response to technology shock under the null.

capital stock is generalized to $k_{t+1} = (1-\delta)k_t + v_t x_t$, where v_t is a shock that follows the process $\ln v_{t+1} = 0.6 \ln v_t + \varepsilon_t$, and ε_t is an innovation assumed $i.i.d.N.(0, 0.04^2)$.

Results are reported in Tables 4–6 under the heading 'multiple shocks'. In all tables, mean is the average of the estimated parameter values and bias is the average difference between the estimated parameter values and the true value used to generate the artificial data. Table 4 reports results obtained using ML. Panel A considers the case where neither priors nor measurement errors are used. Estimates of σ are biased upwards because the conditional variance of the series is attributed to the one technology shock. Estimates of β and ρ are usually biased downwards, with significant differences across series. To understand why different series lead to different estimates and biases, recall that ML minimizes the one-step-ahead prediction errors of the model. The variance decomposition of the true model with multiple shocks shows that the technology shock explains 89%, 78%, and 44% 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

Table 4
Maximum likelihood (DSGE model is misspecified)

Experiment no.	*Var.*	Multiple shocks			Habit formation			Nonlinear		
		β Mean (bias)	ρ Mean (bias)	σ Mean (bias)	β Mean (bias)	ρ Mean (bias)	σ Mean (bias)	β Mean (bias)	ρ Mean (bias)	σ Mean (bias)
A. Using as many variables as shocks										
1	$\hat{y}_t$	.8220	.7822	.0535	.7380	.7751	.0455	.7775	.8070	.0498
		−.1280	−.0678	.0135	−.2120	−.0749	.0055	−.1725	−.0430	.0098
2	$\hat{c}_t$	.9089	.8020	.0554	.5238	.9313	.0157	.9242	.7967	.0570
		−.0411	−.0480	.0154	−.4262	.0813	−.0243	−0.0258	−.0533	.0170
3	$\hat{n}_t$	.9703	.7208	.0481	.9332	.9064	.0438	.9713	.8085	.0332
		.0203	−.1292	.0081	−.0168	.0564	.0038	.0213	−.0415	−.0068
B. Incorporating priors										
4	$\hat{y}_t$	.9506	.8268	.0448	.9503	.8307	.0346	.9501	.8458	.0396
		.0006	−.0232	.0048	.0003	−.0193	−.0054	.0001	−.0042	−.0004
5	$\hat{c}_t$	.9480	.8425	.0422	.9515	.8090	.0272	.9487	.8457	.0404
		−.0020	−.0075	.0022	.0015	−.0410	−.0128	−.0013	−.0043	.0004
6	$\hat{n}_t$	.9500	.7995	.0564	.9652	.8924	.0320	.9553	.8517	.0386
		.0000	−.0505	.0164	.0152	.0424	−.0080	.0053	.0017	−.0014
C. Adding measurement errors to										
7	$\hat{n}_t$, $\hat{y}_t$	.9493	.8212	.0445	.9511	.6983	.0308	.9497	.8585	.0400
		−.0007	−.0288	.0045	.0011	−.1517	−.0092	−.0003	.0085	.0000
8	$\hat{y}_t$,$\hat{c}_t$	.9679	.7234	.0475	.9906	.8057	.0222	.9654	.8627	.0390
		.0179	−.1266	.0075	.0406	−.0443	−.0178	.0154	.0127	−.0010
9	$\hat{n}_t$, $\hat{c}_t$	.9491	.8189	.0449	.9601	.8012	.0328	.9499	.8566	.0400
		−.0009	−.0311	.0049	.0101	−.0488	−.0072	−.0001	.0066	.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 (DSGE model is misspecified)

Experiment no.	*Var.*	Multiple shocks			Habit formation			Nonlinear		
		β Mean (bias)	ρ Mean (bias)	σ Mean (bias)	β Mean (bias)	ρ Mean (bias)	σ Mean (bias)	β Mean (bias)	ρ Mean (bias)	σ Mean (bias)
A. Simulated method of moments with $\tau = 5$										
1	$\hat{y}_t$, $\hat{c}_t$	.9478	.8088	.0451	.9702	.8192	.0321	.9505	.8408	.0390
		−.0022	−.0412	.0051	.0202	−.0308	−.0079	.0005	−.0092	−.0010
2	$\hat{n}_t$, $\hat{y}_t$	.9708	.7897	.0457	.8903	.8352	.0399	.9477	.8386	.0393
		.0208	−.0603	.0057	−.0597	−.0148	−.0001	−.0023	−.0114	−.0007
3	$\hat{n}_t$, $\hat{c}_t$	.9523	.7532	.0525	.9342	.7159	.0262	.9506	.8410	.0393
		.0023	−.0968	.0125	−.0158	−.1341	−.0138	.0006	−.0090	−.0007
B. Simulated method of moments with $\tau = 10$										
4	$\hat{y}_t$, $\hat{c}_t$	.9479	.8099	.0448	.9701	.8203	.0319	.9507	.8401	.0390
		−.0021	−.0401	.0048	.0201	−.0297	−.0081	.0007	−.0099	−.0010
5	$\hat{n}_t$, $\hat{y}_t$	.9701	.7822	.0457	.8896	.8370	.0402	.9486	.8410	.0392
		.0201	−.0678	.0057	−.0604	−.0130	.0002	−.0014	−.0090	−.0008
6	$\hat{n}_t$, $\hat{c}_t$	.9520	.7556	.0527	.9400	.7170	.0255	.9493	.8410	.0391
		.0020	−0.0944	.0127	−.0100	−.1330	−.0145	−.0007	−0.0090	−.0009
C. Simulated method of moments with $\tau = 20$										
7	$\hat{y}_t$, $\hat{c}_t$	.9483	.8072	.0445	.9702	.8146	.0318	.9495	.8351	.0389
		−.0017	−.0428	.0045	.0202	−.0354	−.0082	−.0005	−.0149	−.0011
8	$\hat{n}_t$, $\hat{y}_t$	.9709	.7915	.0455	.8860	.8355	.0405	.9494	.8419	.0390
		.0209	−.0585	.0055	−.0640	−.0145	.0005	−.0006	−.0081	−.0010
9	$\hat{n}_t$, $\hat{c}_t$	.9511	.7523	.0529	.9339	.7129	.0259	.9499	.8406	.0389
		.0011	−.0977	.0129	−.0161	−.1371	−.0141	−.0001	−.0094	−.0011
D. Generalized method of moments										
10	$\hat{y}_t$, $\hat{c}_t$	.9473	.8075	.0447	.9701	.8163	.0318	.9509	.8446	.0387
		−.0027	−.0425	.0047	.0201	−.0337	−.0082	.0009	−.0054	−.0013
11	$\hat{n}_t$, $\hat{y}_t$	.9710	.7909	.0457	.8886	.8333	.0402	.9490	.8426	.0393
		.0210	−.0591	.0057	−.0614	−.0167	.0002	−.0010	−.0074	−.0007
12	$\hat{n}_t$, $\hat{c}_t$	.9512	.7541	.0529	.9367	.7079	.0265	.9502	.8411	.0390
		.0012	−.0959	.0129	−.0133	−.1421	−.0135	.0002	−.0089	−.0010

Notes: See the notes to Tables 1 and 2.

(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 ML with priors. Using priors reduces the bias of the estimates because they are centered around the true parameter values. While this may overstate the usefulness of priors in overcoming misspecification, it is still the case that, in general, informative priors will 'pull' ML estimates towards more economically reasonable values. Panel C reports results using ML with measurement errors. Estimation is carried out using the three observable variables

Table 6
Indirect inference (DSGE model is misspecified)

Experiment no.	*Var.*	Multiple shocks			Habit formation			Nonlinear		
		β Mean (bias)	ρ Mean (bias)	σ Mean (bias)	β Mean (bias)	ρ Mean (bias)	σ Mean (bias)	β Mean (bias)	ρ Mean (bias)	σ Mean (bias)
A. $\tau = 5$										
1	$\hat{y}_t$, $\hat{c}_t$	.9423	.7389	.0448	.9777	.8193	.0317	.9501	.8171	.0400
		−.0077	−.1111	.0048	.0277	−.0307	−.0083	.0001	−.0329	.0000
2	$\hat{n}_t$, $\hat{y}_t$	.9999	.1050	.0344	.9079	.8508	.0400	.9416	.7746	.0399
		.0499	−.7450	−.0056	−.0421	.0008	.0000	.0084	−.0754	−.0001
3	$\hat{n}_t$, $\hat{c}_t$	.9572	.7649	.0523	.9378	.7453	.0267	.9467	.8420	.0398
		.0072	−.0851	.0123	−.0122	−.1047	−.0133	−.0033	−.0080	−.0002
B. $\tau = 10$										
4	$\hat{y}_t$, $\hat{c}_t$	.9437	.7491	.0447	.9798	.8311	.0314	.9492	.8166	.0400
		−.0063	−.1009	.0047	.0298	−.0189	−.0086	−.0008	−.0334	.0000
5	$\hat{n}_t$, $\hat{y}_t$	.9999	.0713	.0342	.9053	.8409	.0400	.9406	.7740	.0398
		.0499	−.7787	−.0058	−.0447	−.0091	.0000	−.0094	−.0760	−.0002
6	$\hat{n}_t$, $\hat{c}_t$	.9572	.7657	.0524	.9471	.7526	.0270	.9469	.8432	.0401
		.0072	−.0843	.0124	−.0029	−.0974	−.0130	−.0031	−.0068	.0001
C. $\tau = 20$										
7	$\hat{y}_t$, $\hat{c}_t$	.9463	.7614	.0448	.9806	.8354	.0316	.9534	.8315	.0400
		−.0037	−.0886	.0048	.0306	−.0146	−.0084	.0034	−.0185	.0000
8	$\hat{n}_t$, $\hat{y}_t$	.9999	.0722	.0341	.9049	.8509	.0399	.9405	.7738	.0399
		.0499	−.7778	−.0059	−.0451	.0009	−.0001	−.0095	−.0762	−.0001
9	$\hat{n}_t$, $\hat{c}_t$	.9570	.7636	.0520	.9470	.7522	.0268	.9459	.8402	.0399
		.0070	−.0864	.0120	−.0030	−.0978	−.0132	−.0041	−.0098	−.0001

Notes: See the notes to Tables 1 and 3.

but with measurement errors added to two of the variables.[13] Since the number of variables is no larger than the number of shocks, singularity is sidestepped and the log likelihood function for the three-variable system is well defined. To examine whether results depend on the variables that are treated as noisy, results are reported for each possible variable pair. With only one exception, biases are smaller than those reported in Panel A. Thus, adding measurement errors appears to be very helpful in limiting the effect of misspecification on parameter estimates. Results depend on the variables that are assumed to be measured with noise and, in this example, biases are smaller when errors are added to output and hours. The reason is simply that these are the two variables most affected by the form of misspecification studied in these experiments.

Table 5 reports results using SMM and GMM. Estimates of $\rho(\sigma)$ are downward (upward) biased in all cases. In general, the bias of these estimates is smaller than (comparable to) that of ML estimates without (with) measurement errors. In the case

[13]I am indebted to an anonymous referee for suggesting the configuration of these experiments.

of SMM, the variation in standard errors (not shown) across τ is smaller than across moments. Hence, when the model is misspecified, the choice of moments to match might be more important than the length of the simulated series. Table 6 reports results using indirect inference. Results are similar to those obtained using the GMM and SMM, but there are instances where biases are large. In particular, biases are larger when the VAR consists of hours and output because misspecification affects these two variables more severely than it does consumption.

In order to get a sense of the robustness of each method, Column 1 in Fig. 4 graphs the biases of the estimates of β and ρ in a plane, with the bias of the former in the horizontal axis and of the latter in the vertical axis.[14] The biases of GMM and SMM estimates are close to the $(0, 0)$ point, meaning that, although not exactly zero, they are small. The plots for ML and indirect inference are relatively more diffuse but they reflect two consistent patterns. First, biases for ML with measurement errors added or with priors are small and outliers correspond to experiments from Panel A. Thus, adding measurement errors and using informative priors help ML overcome its general lack of robustness to misspecification. Second, biases for indirect inference when the VAR includes consumption are also small and outliers correspond to the case where the VAR consists of output and hours. Thus, a careful choice of moments to match, based on a good understanding of the strengths and weaknesses of the economic model, may be beneficial in actual empirical analysis.

Fig. 5 examines the effects of misspecification 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). In order not to saturate the figure, I have abstained from plotting confidence intervals for these responses. The variables in parenthesis in each heading are those used to estimate the model. Consider first the responses predicted by GMM and SMM estimates in Columns 4–6. Quantitatively, the (average) estimated responses are close to the true response and the predicted dynamics are similar. Statistically, the true response lies just above the 90% 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 for these estimation methods. Columns 1–3 plot the responses obtained using ML estimates. Responses are also similar to the true response but whether the 90% confidence interval includes the true response depends on the variables assumed to be measured with noise. Columns 7 and 8 plot the responses obtained using indirect inference estimates. These plots illustrate the previous observations that matching VAR coefficient may lead to similar results than matching moments directly, but that moments coded in the VAR might not as informative as those independently chosen by the macroeconomist to estimate the model *via* GMM or SMM. Note in Column 7 that when the VAR consists of output and hours, the predicted impulses responses differ significantly from the true response.

[14] In preliminary work, I also plotted graphs in the planes (β, σ) and (ρ, σ), with similar conclusions to the ones reported here.

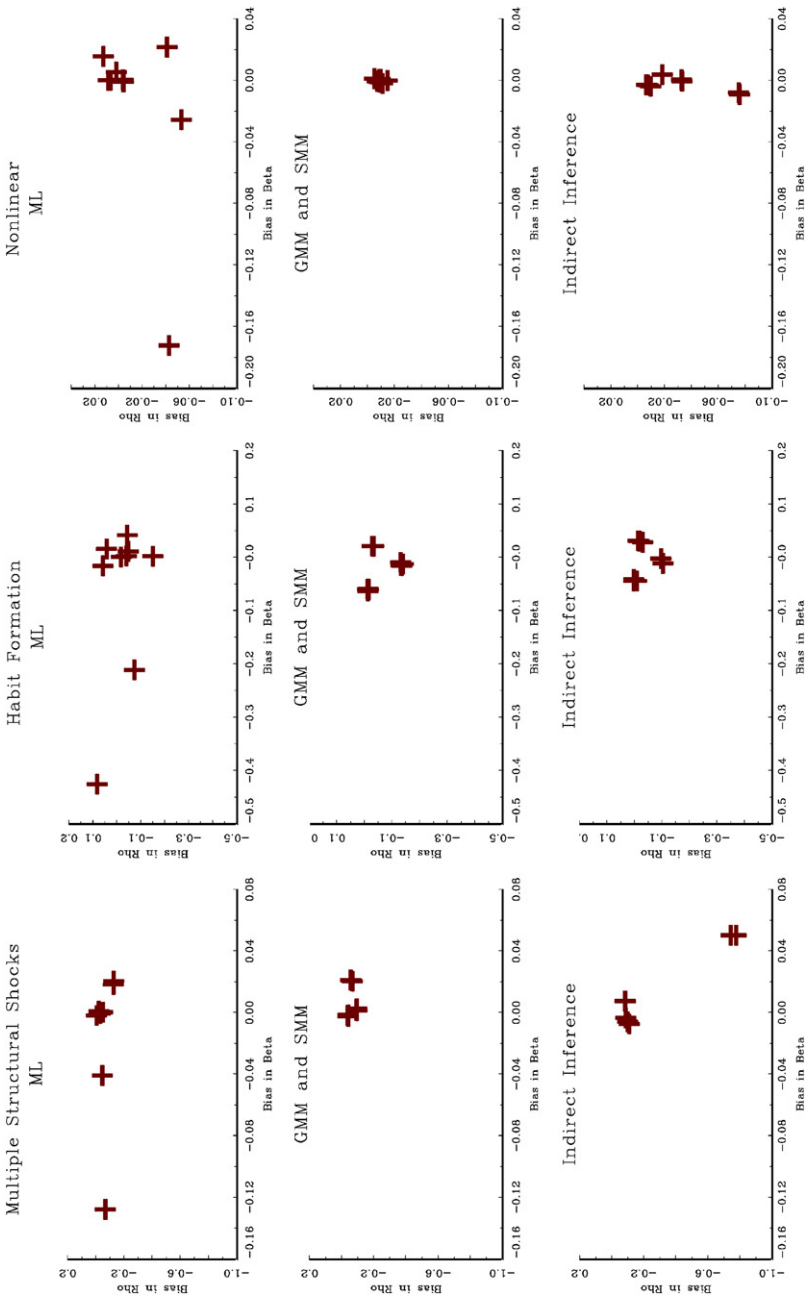

Fig. 4. Bias comparison.

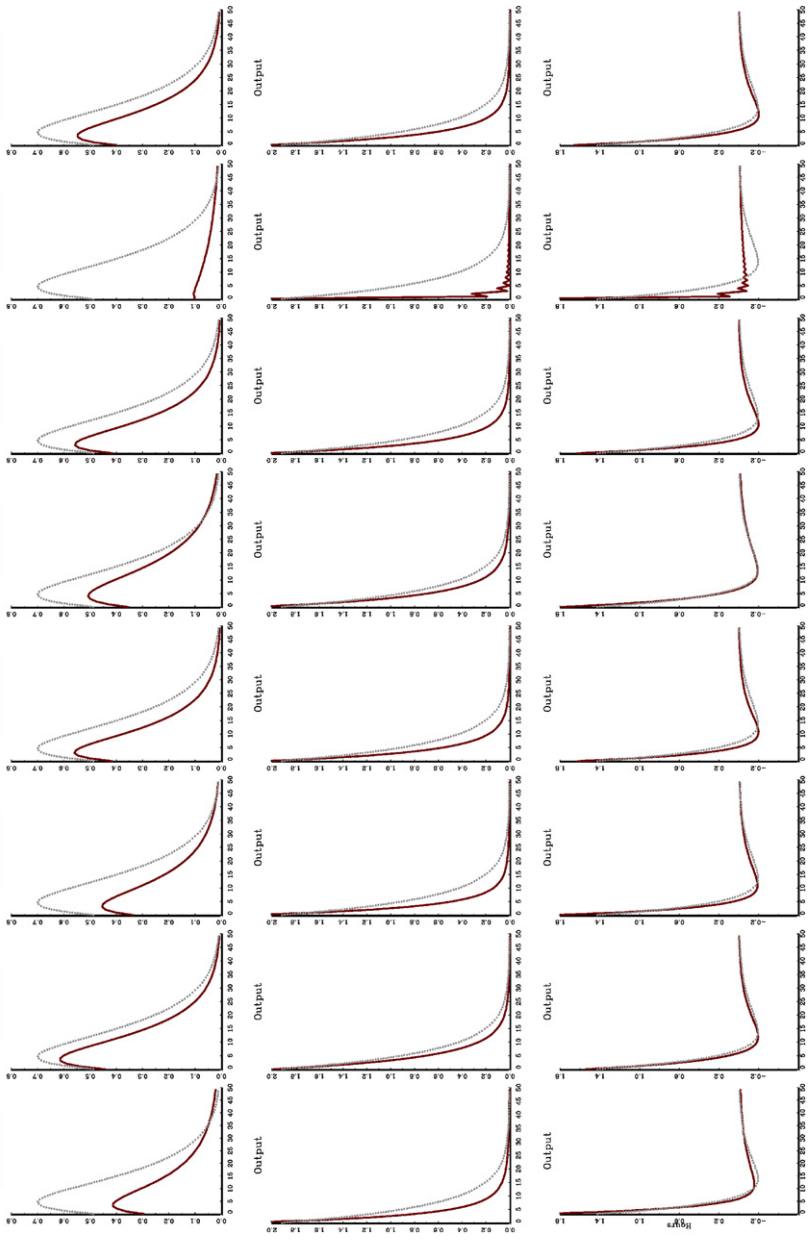

Fig. 5. Response to technology shock. True model has multiple shocks.

4.3.2. True model has habit formation

In this experiment, the data are generated by a linearized RBC model identical to the one in Section 2 except that the instantaneous utility is given by

$$\frac{(c_t/c_{t-1}^{\varkappa})^{1-\lambda}-1}{1-\lambda}+\psi(1-n_t),$$

with $\lambda, \varkappa > 0$. In particular, for the Monte Carlo experiments below I use parameter values $\lambda = 1.5$ and $\varkappa = 0.8$. This specification exhibits internal habit formation whereby the agent's own consumption in two adjacent periods are complements, and it nests the utility function in Section 2 in the special case where $\varkappa \to 0$ and $\lambda \to 1$. With one-period habit formation the linearized Euler equation for consumption is an AR(2) process rather than the usual AR(1) when preferences are time-separable. Thus, consumption dynamics are severely misspecified in this experiment and, consequently, output and hour dynamics will be misspecified as well.

Results are reported in Tables 4–6 under the heading 'habit formation.' Table 4 reports results using ML. When neither priors nor measurement errors are used, biases are fairly large specially when consumption, rather than output or hours, is used to estimate the model. Estimates of β are downward biased but there is no pattern for the biases of ρ and σ. As was reported in the previous section, incorporating extra information in the form of priors and adding measurement errors (see Panels B and C) are useful ways to address misspecification in the ML setup though, in the latter case, results may depend on the variables that are assumed to measured with noise.

Tables 5 and 6 report results for the SMM and GMM, and indirect inference, respectively. Results are similar across these three methods because they all minimize a weighted distance between moments of the data and those predicted by the model. However, within each method, results are different depending on the variables whose moments are matched. This is why in Fig. 4 one usually observes three clusters for each of the three possible combinations of observable variables. Column 2 of Fig. 4 plots the biases in all experiments according to the estimation method employed. As before, the plot is more scattered for ML than for the other methods. However, when one abstracts from the experiments in Panel A (which incorporate neither priors nor measurement errors) biases are smaller than those obtained using SMM, GMM, or indirect inference.

Fig. 6 examines the effects of misspecification on impulse-response analysis. The response to a technology shock under the misspecified model is the continuous line and under the true model is the dotted line. Except for the first column, the predicted dynamics are similar and the (average) estimated response is close to the true response in the sense that the latter lies inside the 90% confidence bound (not shown). Although results for ML in Column 1 are quantitatively different from those implied by the true model, they are not representative of ML results because, as seen in Panel A of Table 4, they are extremely sensitive to the variable used to estimate the model.

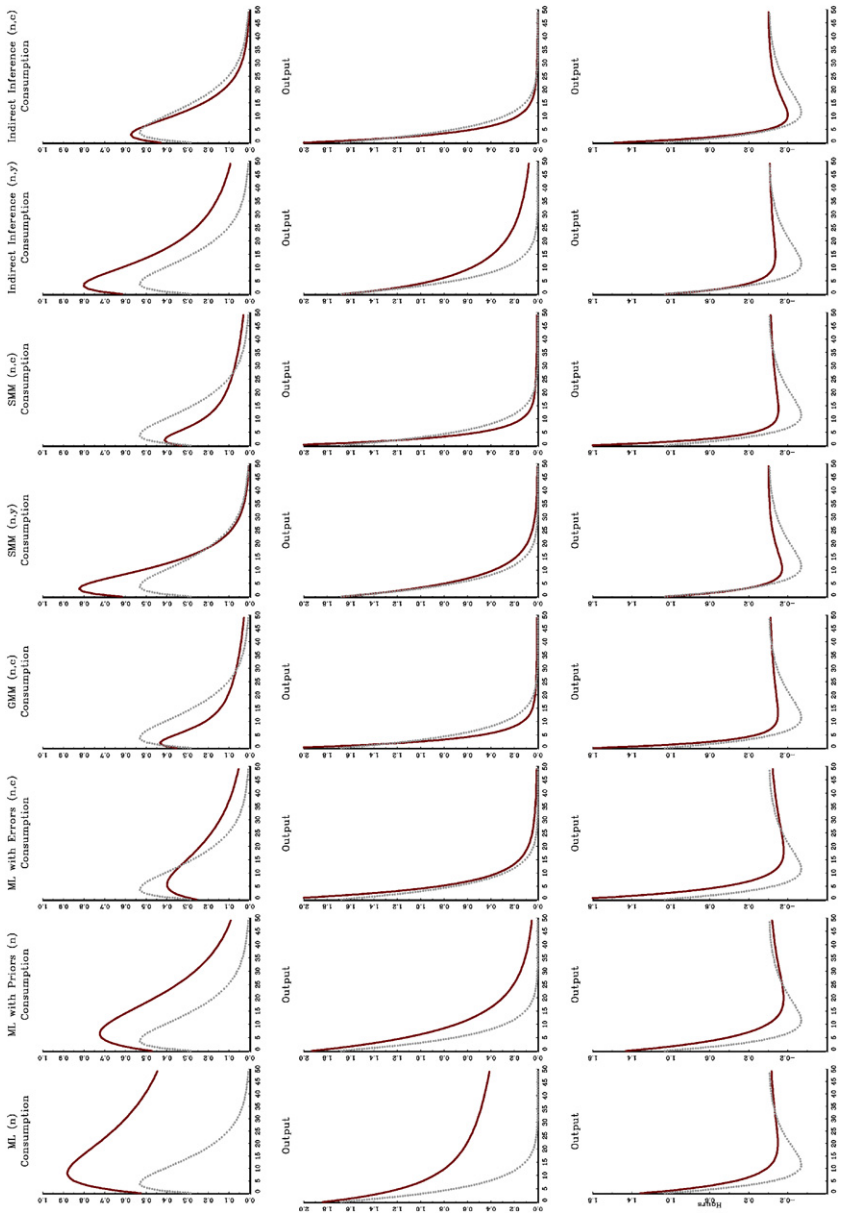

Fig. 6. Response to technology shock. True model has habit formation.

4.3.3. True model is nonlinear

In this experiment, the artificial data are generated by the nonlinear version of the one-shock RBC model in Section 2. The nonlinear model was solved using the Parameterized Expectation Algorithm with the marginal utility of consumption in the Euler equation parameterized as a Chebychev polynomial in the capital stock and the technology shock. Results are reported in Tables 4–6 under the heading 'nonlinear.' Table 4 reports results obtained using ML. In Panel A, estimates of σ (β and ρ) tend to be biased upwards (downwards), and by more when output, rather than hours or consumption, is used. In some cases biases are large despite the fact that misspecification is relatively minor (since the model is close to linear in logs). As in previous experiments, biases are smaller when priors are imposed or measurement errors are added.

Tables 5 and 6 report results for the SMM and GMM, and indirect inference, respectively. Estimates in these tables are very close to those in Tables 2 and 3, where the model was estimated under the null, and biases are uniformly small. To understand the reason why estimators based on unconditional second moments are more robust to this form of misspecification than ML, 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 three and higher. 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 Column 3 of Fig. 3. The biases of the method of moment estimators are generally small and packed close to the point $(0, 0)$. In contrast, the plot is more diffuse for the ML and indirect inference and, in certain cases, biases are large.[15]

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 procedure that optimizes 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 using each estimation method. The average is taken over all replications for all

[15]I have abstained from comparing impulse responses here because in the case of nonlinear models responses usually depend on the sign and size of the shock and, consequently, cannot be summarized by means of a normalization.

Table 7
Comparison in terms of computing time

Method	τ	Seconds per replication
ML	—	0.60
SMM	5	1.76
	10	2.66
	20	5.22
Indirect inference	5	4.59
	10	8.34
	20	16.40
GMM	—	0.02

Notes: The Monte Carlo was performed using GAUSS for Windows running in a Dell Extreme Edition with Pentium 4 processor. Results regarding indirect inference need to be interpreted with caution. For all estimation methods, the optimization routines were started at the true parameter values. However, I found that for indirect inference, the algorithm would frequently blow up if the routine was started at the true value of σ. Hence, the routine was started using a value larger than the one used to generate the sample. For this reason alone indirect inference would take longer to converge than the other estimation methods. Thus, results in this Table most likely overstate the computating time required by this method.

experiments in Tables 1–3 that employ the same estimation method. From this table is clear that GMM is by far the most computationally efficient procedure, followed by ML. There is a fairly large difference in computational efficiency between simulation-based methods and GMM, and the time per iteration increases proportionally with τ. The reason is that the simulation-based methods require the solution of the DSGE model and computation of the gradients using τ times more observations than GMM.

5. Conclusions

The paper studies the application of standard econometric techniques for the estimation of DSGE models. Monte Carlo analysis shows that all procedures deliver reasonably good estimates under the null but that important practical considerations are weak identification, stochastic singularity, small-sample distortions in statistical inference, robustness to misspecification, and computational convenience. Although ML is more severely affected by singularity and generally less robust to misspecification than moment-based methods, adding measurement errors and incorporating priors may be helpful in empirical applications. Overall results indicate that moment-based methods (specially in their GMM and SMM formulations) compare very favorably to the more widely used method of ML and are an attractive addition to our toolbox to estimate DSGE models.

A general issue uncovered here is that the choice of variables or moments used to estimate the model is important for at least two reasons. First, identification depends on the shape of the objective function which, in turn, depends on the data series or

moments used to estimate the model. Second, a misspecification may affect more severely some variables than others. Thus, a selection based on a good understanding of the strengths and weaknesses of the economic model and on the diagnostics mentioned above may be beneficial in actual empirical analysis.

Acknowledgment

I have benefited from comments by Rui Castro, Wouter den Haan, Jesús Fernández-Villaverde, Jim Hamilton, Frank Schorfheide, Chris Sims, and anonymous referees. This project was completed while I was Visiting Professor at the University of California at San Diego in the Fall of 2002. 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.

Appendix A. The log-linearized model

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

$$\mathrm{E}_t\hat{c}_{t+1} = \hat{c}_t + \varsigma(\alpha - 1)\mathrm{E}_t\hat{k}_{t+1} + \varsigma(1 - \alpha)\mathrm{E}_t\hat{n}_{t+1} + \varsigma\mathrm{E}_t\hat{z}_{t+1},$$
$$\hat{n}_t = -(1/\alpha)\hat{c}_t + \hat{k}_t + (1/\alpha)\hat{z}_t,$$

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

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

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

$$\hat{k}_{t+1} = (1 - \delta)\hat{k}_t + \delta\hat{x}_t,$$
$$\hat{z}_{t+1} = \rho\hat{z}_t + \varepsilon_t.$$

References

An, S., Schorfheide, F., 2005. Bayesian analysis of DSGE models. Econometric Reviews, forthcoming.

Bergin, P.R., 2003. Putting the 'new open economy macroeconomics' to a test. Journal of International Economics 60, 3–34.

Blanchard, O.J., Kahn, C.M., 1980. The solution of linear difference models under rational expectations. Econometrica 48, 1305–1311.

Boivin, J., Giannoni, M.P., 2005. DSGE models in a data-rich environment. Columbia University, Manuscript.

Bouakez, H., Cardia, E., Ruge-Murcia, F.J., 2005. Habit formation and the persistence of monetary shocks. Journal of Monetary Economics 52, 1073–1088.

Braun, R.A., 1994. Tax disturbances and real economic activity in the postwar United States. Journal of Monetary Economics 33, 441–462.

Burnside, C., Eichenbaum, M., Rebelo, S., 1993. Labor hoarding and the business cycle. Journal of Political Economy 101, 245–273.

Canova, S., Sala, L., 2005. Back to square one: identification issues in DSGE models. Universitat Pompeu Fabra, Manuscript.

Chang, Y., Gomes, J.F., Schorfheide, F., 2002. Learning-by-doing as a propagation mechanism. American Economic Review 92, 1498–1520.

Christiano, L.J., Eichenbaum, M., 1992. Current real-business cycle theories and aggregate labor-market fluctuations. American Economic Review 82, 430–450.

Christiano, L.J., Eichenbaum, M., Evans, C.L., 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. Journal of Political Economy 113, 1–45.

DeJong, D.N., Ingram, B.F., Whiteman, C.H., 2000. A Bayesian approach to dynamic macroeconomics. Journal of Econometrics 98, 203–223.

Duffie, D., Singleton, K.J., 1993. Simulated moments estimation of Markov models of asset prices. Econometrica 61, 929–952.

Fernández-Villaverde, J., Rubio-Ramirez, J., 2006. Estimating macroeconomic models: a likelihood approach. NBER Technical Working Paper No. 321.

Fuhrer, J.C., Moore, G.R., Schuh, S.D., 1995. Estimating the linear-quadratic inventory model: maximum likelihood versus generalized method of moments. Journal of Monetary Economics 35, 115–157.

Gouriéroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics 8, S85–S118.

Hall, G.J., 1996. Overtime effort and the propagation of business cycle shocks. Journal of Monetary Economics 38, 139–160.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton.

Hansen, G.D., 1985. Indivisible labor and the business cycle. Journal of Monetary Economics 16, 309–327.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1929–1954.

Hansen, L.P., Sargent, T.J., 1998. Recursive linear models of dynamic economies. University of Chicago, Manuscript.

Ingram, B.F., Kocherlakota, N., Savin, N.E., 1994. Explaining business cycles: a multiple-shock approach. Journal of Monetary Economics 34, 415–428.

Ireland, P., 2001. Sticky price models of the business cycle: specification and stability. Journal of Monetary Economics 47, 3–18.

Ireland, P., 2004. A method for taking models to the data. Journal of Economic Dynamics and Control 28, 1205–1226.

Judge, G.G., Griffiths, W.E., Carter Hill, R., Lütkepohl, H., Lee, T.-C., 1985. The Theory and Practice of Econometrics. Wiley, New York.

Kim, J., 2000. Constructing and estimating a realistic optimizing model of monetary policy. Journal of Monetary Economics 45, 329–359.

Kim, J., Ruge-Murcia, F.J., 2006. How much inflation is necessary to grease the wheels? University of Montréal, Manuscript, in preparation.

Lee, B.-S., Ingram, B.F., 1991. Simulation estimation of time-series models. Journal of Econometrics 47, 195–205.

Leeper, E.M., Sims, C.A., 1994. Toward a modern macroeconomic model usable for policy analysis. NBER Macroeconomics Annual, pp. 81–118.

Malinvaud, E., 1970. Statistical Methods of Econometrics. North-Holland, Amsterdam.

McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 57, 995–1026.

McGrattan, E.R., 1994. The macroeconomic effects of distortionary taxation. Journal of Monetary Economics 33, 573–601.

Newey, W.K., 1985. Generalized method of moments specification testing. Journal of Econometrics 29, 229–256.

Pakes, A., Pollard, D., 1989. The asymptotic distribution of simulation experiments. Econometrica 57, 1027–1057.

Poirier, D., 1998. Revising beliefs in nonidentified models. Econometric Theory 14, 483–509.

Rotemberg, J., Woodford, M., 1997. An optimization-based econometric framework for the evaluation of monetary policy. NBER Macroeconomics Annual, pp. 297–346.

Ruge-Murcia, F.J., 2002. Methods to estimate dynamic stochastic general equilibrium models. University of California at San Diego, Economics Discussion Paper 2002–18.

Sargent, T., 1989. Two models of measurement and the investment accelerator. Journal of Political Economy 97, 251–287.

Schorfheide, F., 2000. Loss function-based evaluation of DSGE models. Journal of Applied Econometrics 15, 645–670.

Smets, F., Wouters, R., 2003. An estimated stochastic dynamic general equilibrium model for the euro area. Journal of the European Economic Association 1, 1123–1175.

Smith, A.A., 1993. Estimating nonlinear time-series models using simulated vector autoregressions. Journal of Applied Econometrics 8, 63–84.

Stock, J.H., Wright, J.H., Yogo, M., 2002. A survey of weak instruments and weak identification in generalized method of moments. Journal of Business and Economic Statistics 20, 518–529.

Stone, J.R.N., 1954. The Measurement of Consumers' Expenditure and Behaviour in the United Kingdom. Cambridge University Press, Cambridge.

Theil, H., Goldberger, A.S., 1961. On pure and mixed statistical estimation in economics. International Economic Review 2, 65–78.

White, H., 1982. Maximum likelihood estimation of misspecified model. Econometrica 50, 1–25.