JOURNAL OF APPLIED ECONOMETRICS, VOL. 8, S63–S84 (1993)

ESTIMATING NONLINEAR TIME-SERIES MODELS USING SIMULATED VECTOR AUTOREGRESSIONS

A. A. SMITH, JR
Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA

and

Department of Economies, Queen's University, Kingston, Ontario, Canada, K7L 3N6

SUMMARY

This paper develops two new methods for conducting formal statistical inference in nonlinear dynamic economic models. The two methods require very little analytical tractability, relying instead on numerical simulation of the model's dynamic behaviour. Although one of the estimators is asymptotically more efficient than the other, a Monte Carlo study shows that, for a specific application, the less efficient estimator has smaller mean squared error in samples of the size typically encountered in macroeconomics. The estimator with superior small sample performance is used to estimate the parameters of a real business cycle model using observed US time-series data.

1. INTRODUCTION

This paper develops and implements two new methods for estimating the parameters of fully specified structural dynamic economic models, such as, for example, nonlinear stochastic equilibrium models of the business cycle. These models are typically difficult to estimate using standard methods because of analytically intractable likelihood functions and/or the presence of unobservable variables. A key feature of the two methods developed in this paper is that no analytical tractability is required: one need only be able to simulate numerically the behaviour of the structural model for different values of the structural parameters. This feature of the two methods obviates the need either for an analytically tractable likelihood function or for analytical expressions of population moments as functions of structural parameters. Moreover, these methods circumvent the problem of unobserved or poorly measured time series by allowing one to focus on the marginal distribution of well-measured, observed time series.

This paper first develops the asymptotic properties of the two methods. It shows that both methods yield consistent and asymptotically normal estimates of the true structural parameters. In addition, both methods produce test statistics that can be used to evaluate the goodness-of-fit of the structural model. Next, this paper uses Monte Carlo methods to compare the small-sample performance of the two estimators in a specific application involving the estimation of the parameters of a real business cycle model. Although one of the methods is asymptotically more efficient than the other, the Monte Carlo study shows that, in samples of the size typically encountered in macroeconomics, the mean squared error of the more efficient estimator is larger than that of the less efficient estimator. Finally, this paper uses the

0883–7252/93/0S0S63–22$16.00

Received September 1992
Revised May 1993

estimator with a smaller mean squared error to estimate the parameters of a real business cycle model using observed data.

Vector autoregressions have proven to be a useful tool for exploring the dynamic interaction of multiple time series. Accordingly, both the estimation methods considered in this paper focus on the parameters of a potentially misspecified vector autoregression (VAR) that is used to summarize the statistical properties of both observed and simulated time series. The key idea of both methods is to estimate the structural parameters by matching as closely as possible estimated VAR parameters calculated from, respectively, observed and simulated time series.[1]

The two methods differ in the choice of a metric for measuring the 'distance' between observed and simulated VAR parameters. The first method (called 'extended method of simulated moments', or EMSM) is a generalization of the method developed in Lee and Ingram (1991) and Duffie and Singleton (1988). Following Hansen (1982), the EMSM approach measures the distance between the 'observed' and 'simulated' VAR parameters by forming a quadratic form in a vector of differences between the two sets of parameters. The EMSM estimator of the structural parameters minimizes this quadratic form. The optimal (variance minimizing) weighting matrix is the inverse of the asymptotic covariance matrix of the VAR parameters. The consistent estimation of the optimal weighting matrix must therefore take into account the potential misspecification of the VAR as a representation of the true data-generating process.

The second method (called 'simulated quasi-maximum likelihood', or SQML) uses the likelihood function associated with the VAR as a quasi-likelihood function for the structural model. The SQML estimator of the structural parameters maximizes this quasi-likelihood function, subject to the 'cross-equation' restrictions that the structural model places on the VAR parameters. Although the SQML estimator is consistent despite the misspecification of the quasi-likelihood function, the SQML estimator is, except in special cases, asymptotically less efficient than the EMSM estimator.

Since the laws of motion of many non-linear structural time series models (e.g. real business cycle models) are often well approximated by linear laws of motion, the loss of efficiency associated with an SQML estimator that is defined in terms of a linear model such as a VAR is likely to be small in many circumstances. Moreover, the SQML estimator does not require the estimation of an optimal weighting matrix prior to estimation of structural parameters, suggesting that the SQML method might have better finite sample properties than the EMSM method. A Monte Carlo study consisting of 1000 replications of each estimator in samples of size 150 bears out this intuition. In particular, in a real business cycle model with six unknown structural parameters, the SQML estimates of these parameters have smaller mean squared error than their EMSM counterparts.

This paper is organized as follows. Section 2 develops the asymptotic properties of the SQML estimator and Section 3 the asymptotic properties of the EMSM estimator. Section 4 compares the finite sample performance of the two methods in a Monte Carlo study. Section 5 uses the SQML estimator to estimate the parameters of a real business cycle model using US time-series data. Section 6 concludes. Proofs of all propositions are gathered in Technical Appendices 1 and 2.

[1] More generally, any analytically tractable parametric econometric model can serve as a 'window' through which to view the observed and simulated time series. The structural estimation methods developed in this paper estimate the structural parameters by matching as closely as possible estimated 'window' parameters computed using, respectively, observed and simulated time series.

2. SIMULATED QUASI-MAXIMUM LIKELIHOOD

2.1. Preliminaries

Let the $k \times 1$ vector $\beta \in C$, where C is a compact subset of $\mathscr{R}^k$, consist of the parameters of a fully specified dynamic economic model. Given β, the economic model generates an $m \times 1$ vector stochastic process $y(\beta) \equiv \{y_s(\beta), s \geqslant 1\}$. The vector $y_s(\beta)$ need not incorporate all the variables encompassed by the model. For example, unobserved or poorly measured variables can be omitted from $y_s(\beta)$. It is assumed that, given a set of structural parameters β, the investigator can generate numerically a finite realization $\{y_s(\beta)\}_{s=-(p-1)}^{S}$ of the $y(\beta)$ process ($p \geqslant 0$ a fixed constant). Corresponding to $y(\beta)$ is an observed $m \times 1$ vector time series $x \equiv \{x_t, t \geqslant 1\}$. In practice, the investigator observes a finite realization $\{x_t\}_{t=-(p-1)}^{T}$ of the x process. Assumption 1 requires the processes x and $y(\beta)$ to be stationary and ergodic.

Assumption 1 (1) The observed process x is stationary and ergodic. (2) For all $\beta \in C$, the process $y(\beta)$ is stationary and ergodic.

Under the null hypothesis, there exists a unique set of structural parameters β_0 such that the observed process x and the simulated process $y(\beta_0)$ are drawn from the same distribution. Assumption 2 formalizes the null hypothesis.

Assumption 2 There exists a unique $\beta_0 \in C$ (β_0 an interior point of C) such that the random vectors $[x_t' \ldots x_{t-l}']'$ and $[y_s(\beta_0)' \ldots y_{s-l}(\beta_0)']'$ have identical (stationary) distributions for all $l \geqslant 0$.

Let H_l denote the stationary joint density of the random vector $[x_t' \ldots x_{t-l}']'$ and let G_l^{β} denote the stationary joint density of the random vector $[y_s(\beta)' \ldots y_{s-l}(\beta)']'$. Under the null hypothesis, the densities H_l and $G_l^{\beta_0}$ are identical.

2.2. Definition of the SQML Estimator

To implement the SQML approach, the investigator must choose a conditional density function $f(y_s(\beta), \ldots, y_{s-p}(\beta); \theta)$ characterized by an $n \times 1$ vector of parameters $\theta \in \Theta$, where Θ is a compact subset of $\mathscr{R}^n$. This density function specifies the density of $y_s(\beta)$ conditional on p lags $y_{s-1}(\beta), \ldots, y_{s-p}(\beta)$. In general, the conditional density function f is misspecified in the sense that the true conditional density of $y_s(\beta)$ given $y_{s-1}(\beta), \ldots, y_{s-p}(\beta)$ does not belong to the set of conditional densities $\{f(y_s(\beta), \ldots, y_{s-p}(\beta); \theta): \theta \in \Theta\}$. It is assumed that $n \geqslant k$, i.e. the dimensionality of the space of econometric, or 'shallow', parameters θ is at least as large as the dimensionality of the space of structural, or 'deep', parameters β.

Subject to the regularity conditions described below, the investigator is free, in principle, to choose any conditional density function f. From the viewpoint of asymptotic efficiency, it is desirable to choose a conditional density function f which, given the proper choice of θ, can provide a 'close' approximation to the true but unknown density of $y_s(\beta)$ conditional on $y_{s-1}(\beta), \ldots, y_{s-p}(\beta)$. From the viewpoint of computational ease, it is desirable to choose a conditional density function f for which quasi-maximum likelihood estimates of θ, given a data set $\{y_s(\beta)\}_{s=-(p-1)}^{S}$, can be computed relatively easily. For the Monte Carlo study in Section 4 and the empirical application in Section 5, the conditional density function f corresponds to a vector autoregression (VAR) with i.i.d. normal errors. In this case, the vector of econometric parameters θ consists of the coefficients on lagged endogenous variables as well as the elements of the error covariance matrix. Although the structural model that is estimated in Sections 4 and 5 is non-linear, the laws of motion implied by this model can be well approximated by linear laws of motion such as those provided by a VAR. Moreover, estimates of the VAR parameters can be computed easily using ordinary least squares.

Given a simulated time series $\{y_s(\beta)\}_{s=-(p-1)}^{S}$, let

$$L_S(\{y_s(\beta)\};\,\theta) \equiv \sum_{s=1}^{S} \log f(y_s(\beta), \ldots, y_{s-p}(\beta);\,\theta) \tag{1}$$

be the quasi-log-likelihood function (conditional on $y_0(\beta), \ldots, y_{1-p}(\beta)$) associated with the conditional density function f. L_S is not the true conditional log-likelihood because f is, in general, misspecified. For a given β, maximizing the quasi-log-likelihood function with respect to θ induces a mapping from structural parameters β to econometric parameters θ. Formally, define

$$\hat{\theta}_S^{\beta} \equiv \arg\max_{\theta\in\Theta} L_S(\{y_s(\beta)\};\,\theta) \tag{2}$$

Under a set of regularity conditions discussed in Section 2.3, it can be shown that $\hat{\theta}_S^{\beta}$ converges in probability (as S grows large) to a vector of 'pseudo' true values θ_β. In general, it is not possible to find a closed-form expression for θ_β in terms of the structural parameters β. Using simulation methods, however, one can obtain an arbitrarily accurate estimate $\hat{\theta}_S^{\beta}$ of θ_β by choosing the simulation sample size S suitably. To emphasize the functional dependence of both $\hat{\theta}_S^{\beta}$ and θ_β on the structural parameters β, define $h_S(\beta) \equiv \hat{\theta}_S^{\beta}$ and $h(\beta) \equiv \theta_\beta = \text{plim}_{S\to\infty} \hat{\theta}_S^{\beta}$.

The (conditional) quasi-log-likelihood function can also be evaluated using the observed time series $\{x_t\}_{t=-(p-1)}^{T}$. Define

$$L_T(\{x_t\};\,\theta) \equiv \sum_{t=1}^{T} \log f(x_t, \ldots, x_{t-p};\,\theta) \tag{3}$$

and

$$\hat{\theta}_T \equiv \arg\max_{\theta\in\Theta} L_T(\{x_t\};\,\theta) \tag{4}$$

Under a set of regularity conditions discussed in Section 2.3, $\hat{\theta}_T$ converges in probability (as T grows large) to a vector of pseudo-true values θ_0. Under the null hypothesis (Assumption (2)), $\theta_0 = h(\beta_0) \equiv \theta_{\beta_0}$.

We can now define the SQML estimator of the true structural parameter vector β_0. It is assumed that, for each β, the investigator generates a simulated time series $\{y_s(\beta)\}_{s=-(p-1)}^{S}$ of length $S+p$.[2] By construction, $\{y_s(\beta)\}$ is independent of $\{x_t\}$ for all β. In addition, it is assumed that $S=\tau T$, there $\tau>0$ is a fixed constant. Thus as the observed sample size T tends to infinity, the simulated sample size S also tends to infinity. Simulation error can be controlled by a suitable choice of τ.

Definition 1 The SQML estimator $\hat{\beta}_T$ of β_0 solves the following maximization problem:

$$\hat{\beta}_T \equiv \arg\max_{\beta\in C} L_T(\{x_t\};\,\hat{\theta}_S^{\beta}) \tag{5}$$

where $\hat{\theta}_S^{\beta}$ is defined by equation (2).

Under the regularity conditions set forth in Section 2.3 it can be shown that $\hat{\beta}_T$ converges

[2] As discussed in Section 2.3, in order to ensure that $\{y_s(\beta)\}$ is a 'smooth' function of β, it is necessary to use the *same* random numbers across values of β when generating the simulated time series $\{y_s(\beta)\}$.

in probability (as T grows large) to β_0 and that $T^{1/2}(\hat{\beta}_T - \beta_0)$ has a well-defined asymptotic normal distribution.

As an example of the steps involved in the computation of $\hat{\beta}_T$, consider the case where the quasi-log-likelihood function corresponds to a vector autoregression with, say, two lags (i.e. a VAR(2)). The investigator chooses a set of structural parameters β, simulates a vector time series $\{y_s(\beta)\}$, and fits a VAR(2) to these data using ordinary least squares, yielding parameter estimates $\hat{\theta}_S^\beta$. Next, the investigator inserts $\hat{\beta}_S^\beta$ into the quasi-log-likelihood function defined by equation (3), obtaining a value $L_T(\{x_t\};\hat{\theta}_S^\beta)$. Finally, the investigator searches across values of β to find that value of β (i.e. $\hat{\beta}_T$) which maximizes $L_T(\{x_t\};\hat{\theta}_S^\beta)$.[3]

Note that the structural model places a set of restrictions across the parameters θ of the VAR(2). In particular, under the null hypothesis, the n-dimensional vector θ can be expressed (via the function $h(\beta)$) in terms of the k-dimensional vector β, where $n \geqslant k$. In effect, the SQML estimator maximizes the quasi-likelihood function subject to the constraints that the structural model places across the parameters of the quasi-likelihood function. Since these constraints do not, in general, possess closed-form expressions, the constraints are approximated by means of simulation using equation (2). For the case $n > k$, the model imposes $n - k$ overidentifying restrictions on the parameters of the quasi-likelihood function. As discussed in Section 2.4, these overidentifying restrictions can form the basis of tests of the goodness-of-fit of the structural model.

2.3. Asymptotic Properties of the SQML Estimator

In order to characterize the asymptotic behaviour of the SQML estimator $\hat{\beta}_T$, it is necessary, first, to characterize the asymptotic behaviour of $\hat{\theta}_T$ and $\hat{\theta}_S^\beta$, and, second, to place some structure on the mapping from structural parameters β to econometric parameters θ defined by equation (2). Throughout Assumptions 3–10, the E operator means to compute the mathematical expectation with respect to the appropriate stationary density H_l (for observed data) or G_l^β (for simulated data).

Assumption 3 For all $\beta \in C$, $\log f(y_s(\beta),\dots,y_{s-p}(\beta);\theta)$ is twice continuously differentiable in θ for all $(y_s(\beta),\dots,y_{s-p}(\beta))$.

Assumption 4 For all $\beta \in C$, the functions $\log f(\cdot;\theta)$, $\partial \log f(\cdot;\theta)/\partial\theta_i$, $i = 1,\dots,n$, $\partial^2 \log f(\cdot;\theta)/\partial\theta_i\,\partial\theta_j$, $i,j = 1,\dots,n$, and $\partial \log f(\cdot;\theta)/\partial\theta_i \cdot \partial \log f(\cdot;\theta)/\partial\theta_j$, $i,j = 1,\dots,n$, are measurable for all $\theta \in \Theta$, are separable (see Definition 1 of Tauchen, 1985), and are dominated.[4]

Assumption 5 For all $\beta \in C$, the non-stochastic function $E\log f(y_s(\beta),\dots,y_{s-p}(\beta);\theta)$ is uniquely maximized at θ_β, an interior point of Θ.

Under Assumptions 1 and 3–5 it can be shown that $\hat{\theta}_S^\beta$ converges in probability (as S grows large) to the pseudo-true value θ_β.[5]

---

[3] In practice, gradient hillclimbing methods can be used to locate $\hat{\beta}_T$.

[4] A real-valued function $r(y_s(\beta),\dots,y_{s-p}(\beta);\theta)$ is said to be 'dominated' if there exists a measurable function $b(y_s(\beta),\dots,y_{s-p}(\beta))$ such that $Eb(y_s(\beta),\dots,y_{s-p}(\beta))$ exists and $|r(y_s(\beta),\dots,y_{s-p}(\beta);\theta)| \leqslant b(y_s(\beta),\dots,y_{s-p}(\beta))$ for all $\theta \in \Theta$.

[5] This result and the result in equation (8), whose proofs are contained in the Technical Appendix 1, are not new to this paper. See, for example, Domowitz and White (1982).

Define the matrices $A_\beta(\theta) \equiv E\,\nabla^2 \log f(y_s(\beta), \dots, y_{s-p}(\beta); \theta)$ and

$$B_\beta(\theta) \equiv \Gamma_0^\beta(\theta) + \sum_{k=1}^{\infty} (\Gamma_k^\beta(\theta) + \Gamma_k^\beta(\theta)')$$

where $\Gamma_k^\beta(\theta) \equiv E(\nabla \log f(y_s(\beta), \dots, y_{s-p}(\beta); \theta)\,\nabla \log f(y_{s-k}(\beta), \dots, y_{s-k-p}(\beta); \theta)')$. When the conditional density f is evaluated using observed data, the counterparts to $A_\beta(\theta)$ and $B_\beta(\theta)$ are given by:

$$A(\theta) \equiv E\,\nabla^2 \log f(x_t, \dots, x_{t-p}; \theta) \tag{6}$$

and

$$B(\theta) \equiv \Gamma_0(\theta) + \sum_{k=1}^{\infty} (\Gamma_k(\theta) + \Gamma_k(\theta)') \tag{7}$$

where $\Gamma_k(\theta) \equiv E(\nabla \log f(x_t, \dots, x_{t-p}; \theta)\,\nabla \log f(x_{t-k}, \dots, x_{t-k-p}; \theta)')$. Note that under the null hypothesis, $A_{\beta_0}(\theta) = A(\theta)$ and $B_{\beta_0}(\theta) = B(\theta)$.

Assumption 6 For all $\beta \in C$, $S^{-1/2}\,\nabla L_S(\{y_s(\beta)\}; \theta_\beta) \to N(0, B_\beta(\theta_\beta))$.[6]

Assumption 7 For all $\beta \in C$, the matrices $A_\beta(\theta_\beta)$ and $B_\beta(\theta_\beta)$ are invertible.

Under Assumptions 1 and 3–7, it can be shown that

$$S^{1/2}(\hat{\theta}_S^\beta - \theta_\beta) \to N(0, A_\beta(\theta_\beta)^{-1} B_\beta(\theta_\beta) A_\beta(\theta_\beta)^{-1}) \tag{8}$$

Assumption 8 states that versions of Assumptions 3–7 hold for the functions $L_T(\{x_t\}; \theta)$ and $\log f(x_t, \dots, x_{t-p}; \theta)$.[7] Under Assumption 8, $\text{plim}_{T\to\infty}\,\hat{\theta}_T = \theta_0$ and

$$T^{1/2}(\hat{\theta}_T - \theta_0) \to N(0, A(\theta_0)^{-1} B(\theta_0) A(\theta_0)^{-1}) \tag{9}$$

Assumption 8 The functions $\log f(x_t, \dots, x_{t-p}; \theta)$ and $L_T(\{x_t\}; \theta)$ satisfy Assumptions 3–7, with $\log f(x_t, \dots, x_{t-p}; \theta)$ taking the place of $\log f(y_s(\beta), \dots, y_{s-p}(\beta); \theta)$, $L_T(\{x_t\}; \theta)$ taking the place of $L_S(\{y_s(\beta)\}; \theta)$, $T^{-1/2}$ taking the place of $S^{-1/2}$ (in Assumption 6), θ_0 taking the place of θ_β, $A(\cdot)$ taking the place of $A_\beta(\cdot)$, and $B(\cdot)$ taking the place of $B_\beta(\cdot)$.

Assumption 9 imposes regularity conditions on the mapping from structural parameters β to econometric parameters θ defined by equation (2). Part (1) of Assumption 9 ensures smoothness of $h_S(\beta)$ 'near' β_0 and part (2) ensures local identifiability of β_0.

Assumption 9 There exists an open neighbourhood $N(\beta_0)$ of β_0 such that:

(1) $h_S(\beta)$ is twice continuously differentiable in β for $\beta \in N(\beta_0)$.
(2) $h(\beta)$ is continuously differentiable in β for $\beta \in N(\beta_0)$ and $J(\beta_0) \equiv \nabla h(\beta_0)$ has full-column rank k.

Given a simulated sample size S, Assumption 9(1) requires that, for all β, the *same* seed be used for the (pseudo) random number generator that is used to generate the simulated series $\{y_s(\beta)\}$. Since the random errors used to create $\{y_s(\beta)\}$ are held fixed and since the observed

[6] Assumption 6 can be derived from more primitive assumptions. See, for example, Hansen (1982). The fact that $\text{plim}_{S\to\infty}\,S^{-1}\,\nabla L_S(\{y_s(\beta)\}; \theta_\beta) = 0$ follows from Assumptions 4 and 5.

[7] If the null hypothesis (Assumption 2) holds, then Assumption 8 is redundant.

sample is (of course) also fixed, the optimization problem (5) that defines the SQML estimator is a well-defined deterministic problem.

Finally, Assumption 10 gathers together an additional set of regularity conditions required for the proofs of Propositions 1, 2, 3, and 5.

Assumption 10 Let $\{\beta_T^*\}$ be any sequence of random vectors converging in probability to β_0. Recall that $S = \tau T$, where $\tau > 0$ is a fixed constant.

(1) There exists an open neighbourhood $N(\beta_0)$ of β_0 such that $T^{-1} L_T(\{x_t\};\ \hat{\theta}_S^{\beta})$ converges in probability to $E \log f(x_t, \ldots, x_{t-p};\ \theta_\beta)$ uniformly in $\beta \in N(\beta_0)$.
(2) $\operatorname{plim}_{T\to\infty} T^{-1}\ \nabla_\theta L_T(\{x_t\};\ h_S(\beta_T^*)) = E\ \nabla_\theta \log f(x_t, \ldots, x_{t-p};\ h(\beta_0))$.
(3) $\operatorname{plim}_{T\to\infty} T^{-1}\ \nabla_\theta^2 L_T(\{x_t\};\ h_S(\beta_T^*)) = E\ \nabla_\theta^2 \log f(x_t, \ldots, x_{t-p};\ h(\beta_0))$.
(4) $\operatorname{plim}_{T\to\infty} \nabla h_S(\beta_T^*) = \nabla h(\beta_0) \equiv J(\beta_0)$.
(5) $\operatorname{plim}_{T\to\infty} \partial^2 h_S(\beta_T^*)/\partial\beta_i\, \partial\beta_j = \partial^2 h(\beta_0)/\partial\beta_i\, \partial\beta_j$ for $i,j = 1, \ldots, k$.
(6) $h_S(\beta)$ converges in probability to $h(\beta)$ uniformly in $\beta \in C$.

Proposition 1 characterizes the asymptotic behaviour of $\hat{\beta}_T$.

Proposition 1 Assumptions 1–7 and 9–10 imply the following results for the SQML estimator $\hat{\beta}_T$ defined by equation (5):

$$\operatorname*{plim}_{T\to\infty} \hat{\beta}_T = \beta_0 \tag{10}$$

$$T^{1/2}(\hat{\beta}_T - \beta_0) \to N((1+\tau^{-1})\Sigma(\beta_0)) \tag{11}$$

where $\Sigma(\beta_0) \equiv (J(\beta_0)' A(\theta_0) J(\beta_0))^{-1} J(\beta_0)' B(\theta_0) J(\beta_0) (J(\beta_0)' A(\theta_0) J(\beta_0))^{-1}$, $A(\theta_0)$ is defined by equation (6), $B(\theta_0)$ by equation (7), and $J(\beta_0)$ by Assumption 9.

If the function $h(\beta)$ were known, then there would be no need to simulate the behaviour of the structural model, in which case the asymptotic covariance matrix of $T^{1/2}(\hat{\beta}_T - \beta_0)$ would be simply $\Sigma(\beta_0)$. The use of simulations to evaluate the mapping from β to θ therefore inflates the asymptotic covariance matrix of $T^{1/2}(\hat{\beta}_T - \beta_0)$ by the factor $(1+\tau^{-1})$ (recall that the simulated sample size $S = \tau T$, where $\tau > 0$ is a fixed constant).[8] By choosing an appropriate value for τ, the investigator can control the extent to which the use of simulations increases sampling uncertainty. If $\tau = 10$, for example, so that the simulated sample is ten times as large as the observed sample, then asymptotic standard errors are only approximately 5% larger than in the case where h is known.

The asymptotic covariance matrix $\Sigma(\beta_0)$ can be estimated using standard methods. For example, $A_T(\theta) \equiv T^{-1} \nabla^2 L_T(\{x_t\};\ \theta)$ consistently estimates $A(\theta)$, $A_T(\hat{\theta}_T)$ consistently estimates $A(\theta_0)$, and $\nabla h_S(\hat{\beta}_T)$ consistently estimates $J(\beta_0)$. The required first and second partial derivatives can be computed numerically. The matrix $B(\theta_0)$ can be consistently estimated using a heteoscedasticity and autocorrelation consistent covariance matrix estimator, such as the Newey–West (1987) estimator. See equation (26) in Section 4 for further details.

As a final point, note that if the quasi-likelihood function is actually correctly specified, then the information matrix equality holds: $A(\theta_0) + B(\theta_0) = 0$. In this case, the asymptotic covariance matrix of $T^{1/2}(\hat{\beta}_T - \beta_0)$ reduces to $-(1+\tau^{-1})(J(\beta_0) A(\theta_0) J(\beta_0))^{-1}$. Under the null hypothesis, as τ grows large, this covariance matrix approaches the Cramér–Rao lower bound.

[8] Duffie and Singleton (1988), McFadden (1989), Pakes and Pollard (1989), McFadden and Ruud (1990), Gourieroux and Monfort (1991), and Lee and Ingram (1991) obtain similar results in a variety of different settings.

2.4. Testing the Overidentifying Restrictions

When the number of 'econometric' parameters θ exceeds the number of 'structural' parameters β (i.e. when $n-k>0$), the structural model places $n-k$ overidentifying restrictions on the parameters θ of the quasi-likelihood function $L_T(\{x_t\};\theta)$. These restrictions can form the basis of tests of the goodness-of-fit of the structural model.

Proposition 2 characterizes the asymptotic behaviour of the test statistic:

$$Q_T \equiv T(1+\tau^{-1})^{-1}(\hat{\theta}_T - h_S(\hat{\beta}_T))' A_T(\hat{\theta}_T) B_T(\hat{\theta}_T)^{-1} A_T(\hat{\theta}_T)(\hat{\theta}_T - h_S(\hat{\beta}_T)) \tag{12}$$

where $\hat{\theta}_T$ is defined by equation (4), $\hat{\beta}_T$ is the SQML estimator of β_0, the function h_S is defined by equation (2), and $A_T(\hat{\theta}_T)$ and $B_T(\hat{\theta}_T)$ are consistent estimates of, respectively, $A(\theta_0)$ and $B(\theta_0)$. This test statistic is a quadratic form in the vector $(\hat{\theta}_T - h_S(\hat{\beta}_T))$ of differences between the 'econometric' parameters calculated using the observed data and the 'econometric' parameters calculated using the simulated data (given the consistent estimate $\hat{\beta}_T$ of the structural parameters).

Proposition 2 Suppose $n>k$. Let A, B, and J serve as shorthand for, respectively, $A(\theta_0)$, $B(\theta_0)$ and $J(\beta_0)$ and define V by $B=VV'$. Next, define

$$\begin{aligned}\Lambda \equiv I_n - V'J(J'AJ)^{-1}J'AV'^{-1} - V^{-1}AJ(J'AJ)^{-1}J'V \\ + V^{-1}AJ(J'AJ)^{-1}J'BJ(J'AJ)^{-1}J'AV'^{-1}\end{aligned}$$

where I_n is the $n\times n$ identity matrix. Finally, let λ_i, $i=1,\ldots,n-k$, be the $n-k$ non-zero eigenvalues of Λ[9]. Then, under Assumptions 1–7, 9, and 10, the asymptotic distribution of the statistic Q_T defined by equation (12) is identical to the distribution of $\Sigma_{i=1}^{n-k}\lambda_i c_i^2$, where $c_i \sim$ i.i.d. $N(0,1)$, $i=1,\ldots,n-k$.

It is easy to see that if the information matrix equality $A(\theta_0)+B(\theta_0)=0$ holds, then the matrix Λ defined in Proposition 2 is idempotent with rank $n-k$. In this case, the $n-k$ non-zero eigenvalues of Λ are all equal to 1, so that Q_T converges in distribution to $\chi^2(n-k)$. If the information matrix equality does *not* hold, then the non-zero eigenvalues of Λ can be estimated by computing the non-zero eigenvalues of a consistent estimate of Λ (see the end of Section 2.3 for further details). Accurate estimates of critical values corresponding to the distribution of $\Sigma_{i=1}^{n-k}\lambda_i c_i^2$ can then be easily computed by means of simulation.

3. EXTENDED METHOD OF SIMULATED MOMENTS

This section develops an alternative approach to estimating the true structural parameters β_0. This approach can be viewed as a generalization of the estimation strategy (which we will refer to as 'method of simulated moments', or MSM) developed by Lee and Ingram (1991) and Duffie and Singleton (1988) for the estimation of structural time-series models. The 'extended method of simulated moments', or EMSM, approach estimates β_0 by minimizing the 'distance' between $\hat{\theta}_T$ and $\hat{\theta}_S^\beta$, where this distance is measured by forming a quadratic form in the vector $\hat{\theta}_T - \hat{\theta}_S^\beta$.[10] Since the statistics $\hat{\theta}_T$ and $\hat{\theta}_S^\beta$ cannot, in general, be expressed as simple time averages

[9] Note that $\Lambda = DD'$, where $D \equiv I_n - V^{-1}AJ(J'AJ)^{-1}J'V$. D is idempotent with rank $n-k$, implying that Λ is positive semi-definite with rank $n-k$. Λ therefore has $n-k$ eigenvalues greater than zero and k eigenvalues equal to zero.

[10] This approach is related to the literature on 'minimum distance estimation'. See the discussion and references in Section 4.4 of Ogaki (1992). The EMSM approach can be generalized to include minimizing the distance between non-linear functions of the two sets of 'econometric' parameters, such as, for example, impulse response functions. See Chapter 2 of Smith (1990).

of functions of, respectively, the observed and simulated data, the EMSM approach is not a special case of the MSM approach.[11]

Let $\{W_T\}$ be a sequence of $n \times n$ positive definite 'weighting' matrices that converges in probability to a non-stochastic positive definite matrix W. In practice, W_T can depend on the observed sample $\{x_t\}$. As for the SQML estimator, the simulated sample size $S = \tau T$, where $\tau > 0$ is a fixed constant.

Definition 2 The EMSM estimator $\tilde{\beta}_T$ of β_0 solves the following minimization problem:

$$\tilde{\beta}_T \equiv \arg \min_{\beta \in C} (\hat{\theta}_T - \hat{\theta}_S^\beta)' W_T (\hat{\theta}_T - \hat{\theta}_S^\beta) \tag{13}$$

where $\hat{\theta}_T$ is defined by equation (4) and $\hat{\theta}_S^\beta$ by equation (2).

Proposition 3 characterizes the asymptotic behaviour of $\tilde{\beta}_T$.

Proposition 3 Assumptions 1–7 and 9–10 imply the following results for the EMSM estimator $\tilde{\beta}_T$ defined by equation (13):

$$\operatorname*{plim}_{T \to \infty} \tilde{\beta}_T = \beta_0 \tag{14}$$

$$T^{1/2}(\tilde{\beta}_T - \beta_0) \to N(0, (1 + \tau^{-1}) K(\beta_0)^{-1} J(\beta_0)' W \Omega(\theta_0) W J(\beta_0) K(\beta_0)^{-1}) \tag{15}$$

where $K(\beta_0) \equiv J(\beta_0)' W J(\beta_0)$, $\Omega(\theta_0) \equiv A(\theta_0)^{-1} B(\theta_0) A(\theta_0)^{-1}$, $A(\theta_0)$ is defined by equation (6), $B(\theta_0)$ by equation (7), and $J(\beta_0)$ by Assumption 9.

Definition 2 defines a class of estimators indexed by the (asymptotic) weighting matrix W. By choosing W appropriately, one can select the 'best' member of this class in the sense that the asymptotic covariance matrix of any other estimator in the class exceeds the asymptotic covariance matrix of the 'best' estimator by a positive semi-definite matrix. Using standard arguments (see, for example, the proof of Proposition 4), it can be shown that the optimal (asymptotic) weighting matrix $W^* = \Omega(\theta_0)^{-1}$. Note that the optimal weighting matrix is the inverse of the asymptotic covariance matrix of $T^{1/2}(\hat{\theta}_T - \theta_0)$. The data-dependent matrix

$$W_T^* \equiv A_T(\hat{\theta}_T) B_T(\hat{\theta}_T)^{-1} A_T(\hat{\theta}_T) \tag{16}$$

consistently estimates the optimal weighting matrix, where $A_T(\theta)$ and $B_T(\theta)$ are consistent estimates of $A(\theta)$ and $B(\theta)$, respectively. Let $\tilde{\beta}_T^*$ be the optimal EMSM estimator (i.e. the EMSM estimator that results from using W_T^* as the weighting matrix). The asymptotic covariance matrix of $T^{1/2}(\tilde{\beta}_T^* - \beta_0)$ is

$$(1 + \tau^{-1})(J(\beta_0)' \Omega(\theta_0)^{-1} J(\beta_0))^{-1} \tag{17}$$

Clearly, if the quasi-log-likelihood function is correctly specified in the sense that the information matrix equality $A(\theta_0) + B(\theta_0) = 0$ holds, then the asymptotic covariance matrices of $T^{1/2}(\hat{\beta}_T - \beta_0)$ and $T^{1/2}(\tilde{\beta}_T^* - \beta_0)$ are both equal to $-(1 + \tau^{-1})(J(\beta_0)' A(\theta_0) J(\beta_0))^{-1}$. In this case, therefore, the SQML and optimal EMSM estimators have equal asymptotic efficiency.

Now suppose that $n = k$, i.e. the true structural parameter vector β_0 is exactly identified, but the information matrix equality does not hold. Since $n = k$, the Jacobian matrix $J(\beta_0)$ is square and, by Assumption 9(2), is invertible. It is easy to see that, in this case, the asymptotic

[11] Gourieroux *et al.* (1992) show how the EMSM approach can be modified to incorporate exogenous variables.

covariance matrix of $T^{1/2}(\hat{\beta}_T - \beta_0)$ simplifies to the expression in equation (17). Thus if no overidentifying restrictions are imposed, the SQML and optimal EMSM estimators once again have equal asymptotic efficiency.

Proposition 4 compares the efficiency of the SQML estimator and the optimal EMSM estimator in the case where $n > k$ and the information matrix equality is violated. This proposition states that the difference between the asymptotic covariance matrix of the SQML estimator and the asymptotic covariance matrix of the optimal EMSM estimator is positive semi-definite if $k < n < 2k$ and is positive definite if $n \geqslant 2k$.

Proposition 4 Fix a quasi-log-likelihood function $L_T(\{x_t\}; \theta)$ and suppose that $n > k$ and $A(\theta_0) + B(\theta_0) \neq 0$. Let V_1 be the asymptotic covariance matrix of $T^{1/2}(\hat{\beta}_T - \beta_0)$ and let V_2 be the asymptotic covariance matrix of $T^{1/2}(\tilde{\beta}_T^* - \beta_0)$, where $\hat{\beta}_T$ is the SQML estimator of β_0 and $\tilde{\beta}_T^*$ is the optimal EMSM estimator of β_0. Then $V_1 - V_2$ is positive definite if $n \geqslant 2k$ and is positive semi-definite (with rank $n - k$) if $k < n < 2k$.

Proposition 5 states that if an estimate of the optimal weighting matrix is used to define the EMSM estimator, then the minimized value of the criterion function (13), normalized by the factor $T(1 + \tau^{-1})^{-1}$, can be used as a test of the $n - k$ overidentifying restrictions.

Proposition 5 Suppose $n > k$. Define the statistic

$$Z_T \equiv T(1 + \tau^{-1})^{-1}(\hat{\theta}_T - h_S(\tilde{\beta}_T^*))' W_T^*(\hat{\theta}_T - h_S(\tilde{\beta}_T^*)) \tag{18}$$

where $\hat{\theta}_T$ is defined by equation (4), $\tilde{\beta}_T^*$ is the optimal EMSM estimator, W_T^* is defined by equation (16), and the function h_S is defined by equation (2). Under Assumptions 1–7, 9, and 10, Z_T converges in distribution to $\chi^2(n - k)$.

4. A MONTE CARLO STUDY

This section uses the SQML and EMSM estimation strategies to estimate the parameters of a real business cycle model using repeated samples drawn from the data-generating process associated with the real business cycle model. The goal of this Monte Carlo study is to compare the performance of the two estimators in a realistic application using 'observed' samples of the size typically encountered in macroeconomics.

Recall that the optimal EMSM estimator requires that an estimate W_T^* of the optimal weighting matrix W^* be computed (using observed data) prior to estimation. Generally, estimates of W^* do not perform well in small samples. The difficulty of estimating W^* suggests that the EMSM estimator may not perform as well as the SQML estimator in small samples, despite the fact that the EMSM estimator is asymptotically more efficient than the SQML estimator. Moreover, for structural models such as the real business cycle model described below, the loss of efficiency associated with the SQML estimator is likely to be small even in large samples, since, for this class of models, one can generally find a quasi-likelihood function that provides a good approximation to the true but unknown likelihood function.

The real business cycle to be estimated takes the form of the following social planner's problem:

$$\max_{\{c_t\}_{t=0}^{\infty}, \{i_t\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \omega^t \gamma^{-1}(c_t^{\gamma} - 1), \text{ given } k_0, \lambda_0, \text{ and } z_0 \tag{19}$$

subject to the following constraints for all $t \geqslant 0$:

$$c_t + i_t = Ak_t^{\alpha}\lambda_t \tag{20}$$

$$k_{t+1} = (1-\delta)k_t + z_t i_t \tag{21}$$

$$\lambda_t = \rho_1\lambda_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \text{i.i.d. } N(1-\rho_1, \sigma_\varepsilon^2) \tag{22}$$

$$z_t = \rho_2 z_{t-1} + \nu_t, \qquad \nu_t \sim \text{i.i.d. } N(1-\rho_2, \sigma_\nu^2) \tag{23}$$

where c_t is period t consumption, i_t is period t investment, k_t is the period t capital stock, λ_t is the period t 'technology' shock, and z_t is the period t shock to the productivity of new investment goods.[12] The innovations ε_t and ν_t are mutually uncorrelated at all leads and lags. The nine structural parameters are:

α: capital's share of income ($0 < \alpha < 1$)
ω: discount factor ($0 < \omega < 1$)
A: scaling factor in production function ($A > 0$)
δ: rate of depreciation of capital stock ($0 \leqslant \delta \leqslant 1$)
γ: risk-aversion parameter (coefficient of relative risk aversion $= 1 - \gamma > 0$)
ρ_1: persistence parameter in 'technology' shock process ($|\rho_1| < 1$)
σ_ε: standard deviation of innovation in 'technology' shock process
ρ_2: persistence parameter in 'investment' shock process ($|\rho_2| < 1$)
σ_ν: standard deviation of innovation in 'investment' shock process

This real business cycle model is similar to the one studied in Greenwood *et al.* (1988).

The solution to this real business cycle model consists of a decision rule expressing the optimal choice for i_t as a function of the period t state variables k_t, λ_t, and z_t. Since the decision rule for this problem does not possess a known closed-form expression, a linear approximation to the decision rule is computed using the linear-quadratic methodology introduced by Kydland and Prescott (1982) (see also Christiano, 1990; McGrattan, 1990). The linear approximation to the decision rule takes the form:

$$i_t = b_0 + b_1 k_t + b_2\lambda_t + b_3 z_t \tag{24}$$

where the decision rule coefficients b_0, b_1, b_2, and b_3 are complicated non-linear functions of the structural parameters.[13] Given initial conditions k_0, λ_0, and z_0 and sequences of innovations $\{\varepsilon_t\}_{t=1}^{T}$ and $\{\nu_t\}_{t=1}^{T}$, recursive iterations on the nonlinear laws of motion (20)–(24) can be used to yield time series for output $w_t \equiv Ak_t^{\alpha}\lambda_t$, investment i_t and other variables of interest.

For the Monte Carlo study whose results are reported below, the observed sample size T is set at 150. The structural model is simulated 1000 times, yielding 1000 'observed' time series $\{x_t\}_{t=0}^{T}$, where $x_t = [\log w_t \ \log i_t]'$. To reduce computational burden in the Monte Carlo experiments, three of the nine structural parameters (in particular, ρ_2, σ_ν, and γ) are considered known, while the remaining six structural parameters (in particular, $\beta = [A \ \delta \ \rho_1 \ \sigma_\varepsilon \ \alpha \ \omega]'$) are considered unknown. For each of the observed series $\{x_t\}$, the six unknown structural

[12] The laws of motion (22) and (23) do not rule out the possibility of negative realizations for the shocks λ_t and z_t. For empirically plausible values of ρ_1, σ_ε, ρ_2, and σ_ν, however, the probability of a negative realization for either λ_t or z_t is essentially zero.

[13] The estimation methods developed in this paper do not require that the decision rule be linear. For the structural model studied here, however, linear decision rules are both highly accurate and easy to compute using the 'doubling' algorithm described in McGrattan (1990).

parameters are estimated using both the SQML and EMSM estimation strategies. The simulated sample size is set at 1500, so that the constant $\tau = S/T = 10$. To minimize the effects of initial conditions, observed samples of length $T + 200$ and simulated samples of length $S + 200$ are simulated using initial conditions set at deterministic steady state values; the first 200 data points of each series are subsequently discarded.[14]

The (misspecified) conditional density f which forms the basis of the quasi-log-likelihood function corresponds to a bivariate vector autoregression with one lag for the vector $x_t = [\log w_t \ \log i_t]'$: $x_t = C[1 \ \ x_{t-1}']' + \eta_t$, where C is a 2×3 matrix and the i.i.d. vector of innovations η_t is assumed to be normally distributed with covariance matrix DD', where D is lower triangular. In particular,

$$f(x_t, x_{t-1}; \theta) \equiv -1/2 \log(\det(DD')) - 1/2\eta_t'(DD')^{-1}\eta_t \tag{25}$$

The 9×1 vector of econometric parameters θ consists of the six elements of C as well as the three non-zero elements of D.

The 'true' values of the structural parameters to be used in the Monte Carlo study are listed in Table I. These parameters were selected by estimating the parameters of the real business cycle using observed US time series, with the quasi-log-likelihood function defined in terms of the conditional density f given by equation (25). Section 5, to which the reader is referred for further details, estimates the parameters of the real business cycle model using a conditional density f with an additional lag. As mentioned previously, the parameters ρ_2, σ_ν, and γ are considered known in the Monte Carlo study, while the remaining six parameters are regarded as unknown.

Implementing the optimal EMSM estimator requires the choice of an estimator for the optimal weighting matrix $W^* = A(\theta_0)B(\theta_0)^{-1}A(\theta_0)$. As discussed in Section 2.3, $A_T(\hat{\theta}_T) \equiv T^{-1}\nabla^2 L_T(\{x_t\}; \hat{\theta}_T)$ consistently estimates $A(\theta_0)$. $B(\theta_0)$ can be consistently estimated in a variety of ways (see, for example, Andrews, 1991). The present Monte Carlo study uses the estimator suggested by Newey and West (1987). Define

$$B_T(\theta) \equiv \hat{\Gamma}_0(\theta) + \sum_{k=1}^{M} (1 - k/(M+1))(\hat{\Gamma}_k(\theta) + \hat{\Gamma}_k(\theta)') \tag{26}$$

where $\hat{\Gamma}_k(\theta) \equiv T^{-1}\sum_{t=k+1}^{T} \nabla \log f(x_t, \ldots, x_{t-p}; \theta) \nabla \log f(x_{t-k}, \ldots, x_{t-k-p}; \theta)'$. Then $B_T(\hat{\theta}_T)$ consistently estimates $B(\theta_0)$. The number of lags, M, used to construct the estimate of $B(\theta_0)$ is set at 10.

Table II summarizes the results of the Monte Carlo study. The EMSM estimator failed to converge for two of the 1000 'observed' time series. These two draws were discarded, leaving a sample of size 998 for both the SQML estimator and the EMSM estimator. Table II gives

Table I. 'True' values of structural parameters for the Monte Carlo study

α	ω	A	δ	γ	ρ_1	σ_ε	ρ_2	σ_ν
0·5579	0·8456	1·9269	0·07234	0·2123	0·9182	0·01096	0·8363	0·04624

[14] Duffie and Singleton (1988) argue that this procedure does not remove dependence on initial conditions, so that simulated time series are non-stationary, thereby violating Assumption 1. Duffie and Singleton (1988) provide a more general set of conditions under which simulated time series are asymptotically stationary, thereby guaranteeing consistency and asymptotic normality of simulation estimators.

the sample mean and sample standard deviation of each set of estimates, as well as the estimated bias and the square root of the estimated mean squared error. The table shows that, for both methods, estimates of δ, σ_ε, α, and ω display very little bias. Estimates of A, on the other hand, are biased upwards (by approximately 8% for SQML and by approximately 11% for EMSM), while estimates of ρ_1 are biased downwards (by approximately 3% for both SQML and EMSM).

Table II also confirms that the EMSM estimator does not perform as well as the SQML estimator in samples of the size typically encountered in macroeconomics. In particular, for all six parameters, the estimated standard errors of the EMSM estimates are larger than the estimated standard errors of the SQML estimates. The ratio of root mean squared error for the EMSM estimates to root mean squared error for the SQML estimates ranges from 1·063 (for ω) to 1·129 (for σ_ε). Although the improvement of the SQML estimates over the EMSM estimates in terms of mean squared error is modest, the SQML estimates nonetheless do have greater precision (in the 'mean squared error' sense) in observed samples of size 150.[15]

5. AN EMPIRICAL APPLICATION OF SQML

This section uses the SQML estimation strategy to estimate and test the real business cycle model described by equations (20)–(24) in Section 4. The conditional density function f which underlies the quasi-log-likelihood function is chosen to be a bivariate vector autoregression with two lags for the vector $x_t = [\log w_t \ \log i_t]'$, where $\log w_t$ is the detrended log of output

Table II. Monte Carlo results: summary statistics for SQML and EMSM estimates

	A	δ	ρ_1	σ_ε	α	ω
'True' values of structural parameters						
	1·9269	0·07234	0·9182	0·01096	0·5579	0·8456
SQML estimates of structural parameters						
Mean	2·0797	0·07290	0·8942	0·01094	0·5636	0·8455
Bias	0·1528	0·00056	−0·0240	−0·00002	0·0057	−0·0001
Std dev.	0·9149	0·01140	0·0649	0·00072	0·1324	0·0488
$MSE^{1/2}$	0·9276	0·01141	0·0692	0·00072	0·1325	0·0488
EMSM estimates of structural parameters						
Mean	2·1322	0·07199	0·8888	0·01088	0·5576	0·8502
Bias	0·2053	−0·00035	−0·0294	−0·00008	−0·0003	0·0046
Std dev.	1·0184	0·01256	0·0719	0·00081	0·1453	0·0516
$MSE^{1/2}$	1·0389	0·01257	0·0777	0·00082	0·1453	0·0518
$MSE^{1/2}_{EMSM}/MSE^{1/2}_{SQML}$						
	1·120	1·102	1·122	1·129	1·096	1·063

Notes: These results are based on 998 draws in observed samples of size 150. The first row gives the values of the structural parameters used to generate the 'observed' series. Rows labelled 'Mean' give the sample mean of the 998 point estimates. Rows labelled 'Bias' give the estimated bias, where est. bias ≡ sample mean − true value. Rows labelled 'Std Dev.' give the sample standard deviation of the 998 point estimates. Rows labelled '$MSE^{1/2}$' give the square root of the estimated mean squared error (MSE), where estimated MSE ≡ (est. bias)^2 + (std dev.)^2. The last row of the table gives the ratio of $MSE^{1/2}$ for the SQML estimator to $MSE^{1/2}$ for the EMSM estimator.

[15] Clearly, these results are specific to the application studied here. Comparisons of mean squared error for SQML and EMSM in other applications must await further Monte Carlo studies.

and log i_t is the detrended log of investment: $x_t = C[1\ \ x'_{t-1}\ \ x'_{t-2}]' + \eta_t$, where C is a 2×5 matrix and the i.i.d. vector of innovations η_t is assumed to be normally distributed with covariance matrix DD', where D is lower triangular. In particular,

$$f(x_t, x_{t-1}, x_{t-2};\ \theta) \equiv -1/2\ \log(\det(DD')) - 1/2\eta'_t(DD')^{-1}\eta_t$$

The 13×1 vector of econometric parameters θ consists of the ten elements of C as well as the three non-zero elements of D.

The observed data consists of US time series for log per capita GNP and log per capita investment for the time period 1947:1 to 1988:4.[16] Following Perron (1989), the time series are detrended by fitting a deterministic linear time trend to the logged series, with a structural break in the slope coefficient in the first quarter of 1973.

Since two lags are used as initial conditions in the VAR(2), there are 162 observations on the vector x_t. The simulation sample size, S, is set at 2000, so that $\tau = 2000/162 \approx 12 \cdot 3$.[17] The use of simulations in the estimation process therefore leads to an increase in (asymptotic) standard errors of about 4% ($\approx (1 + \tau^{-1})^{1/2}$). For the calculation of (1) the estimated standard errors of the SQML estimates, (2) the test statistic Q_T defined by equation (12), and (3) an estimate of the matrix Λ defined in Proposition 2, the matrices $A(\theta_0)$ and $B(\theta_0)$ are consistently estimated as described in Section 4. The number of lags used in the estimation of $B(\theta_0)$ is set at 25. The matrix $\nabla h_S(\hat{\beta}_T)$ serves as a consistent estimate of $J(\beta_0)$.

Table III tabulates the point estimates for the structural parameters, together with estimated standard errors. Unlike other empirical studies of real business cycle models (see, for example, Altug, 1989; Christiano, 1988), the present empirical study estimates all nine of the model's structural parameters, including the discount rate ω and the rate of depreciation δ. The estimates of α (capital's share of income), ω, and δ (respectively, 0·76, 0·79, and 0·07) differ substantially from values that are typically considered 'reasonable'. Hansen (1985), for example, fixes α at 0·36, ω at 0·99, and δ at 0·025. The standard errors of the estimates for α, ω, and δ, however, are quite large. At the 1% significance level, one would not be able to reject any of the null hypotheses H_0: $\alpha = 0 \cdot 36$, H_0: $\omega = 0 \cdot 99$, or H_0: $\delta = 0 \cdot 025$. A Wald test of the null hypothesis H_0: $\alpha = 0 \cdot 36$, $\omega = 0 \cdot 99$, $\delta = 0 \cdot 025$, however, leads to a rejection at the 1% significance level. The null hypothesis H_0: $\alpha = 0 \cdot 36$, $\delta = 0 \cdot 025$ is also rejected at the 1% level. The statistically significant divergence between the estimated values and the 'reasonable' values for these parameters probably reflects the poor fit of the structural model with observed data, as discussed below.

Table III. SQML estimates of structural parameters using VAR(2)

$\hat{\alpha}$	$\hat{\omega}$	$\hat{A}$	$\hat{\delta}$	$\hat{\gamma}$	$\hat{\rho}_1$	$\hat{\sigma}_\varepsilon$	$\hat{\rho}_2$	$\hat{\sigma}_\nu$
0·7585	0·7945	0·9620	0·06965	−0·06920	0·8961	0·01023	0·7851	0·05240
(0·3798)	(0·0681)	(1·4569)	(0·02961)	(0·55375)	(0·0849)	(0·00078)	(0·0689)	(0·00889)

Notes: The first row of the table contains point estimates. The second row of the table contains estimated standard errors.

[16] Output per capita is defined as US GNP in 1982 dollars (Citibase variable *GNP82*) divided by the US population over the age of 15 (the sum of Citibase variables *MPOP* and *FPOP*). Investment per capita is defined as US gross private domestic investment in 1982 dollars (Citibase variable *GI82*) divided by the US population over the age of 15. All series are seasonally adjusted and detrended as described in the text.

[17] As in the Monte Carlo study reported in Section 4, to minimize the effects of initial conditions, simulated time series of length $S + 200$ are generated using initial conditions set at deterministic steady-state values; the first 200 data points are subsequently discarded.

Table IV. Estimated VAR(2) parameters using observed and simulated data

Parameter	Output equation Observed	Output equation Simulated	Investment equation Observed	Investment equation Simulated
Constant	0·0951	0·0476	−0·1404	−0·0948
	0·0564	0·0092	0·2816	0·0511
Once-lagged output	1·4751	1·3037	1·7045	0·7444
	0·0907	0·0326	0·5015	0·1874
Twice-lagged output	−0·5108	−0·3231	−1·5808	−0·6561
	0·0964	0·0324	0·4457	0·1847
Once-lagged investment	−0·0244	−0·0365	0·8678	0·7386
	0·0178	0·0066	0·0682	0·0383
Twice-lagged investment	0·0117	0·0369	−0·1285	0·0673
	0·0146	0·0063	0·0633	0·0386

Parameter	Observed	Simulated
d_{11}	0.0099	0.0104
	(0.0007)	(0.0002)
d_{21}	0.0388	0.0425
	(0.0040)	(0.0011)
d_{22}	0.0329	0.0340
	(0.0036)	(0.0006)

Notes: The top and bottom numbers in each cell are, respectively, the point estimate and the estimated standard error. Misspecification-robust standard errors are computed using the asymptotic result in equation (9). The number of lags M used to compute an estimate of $B(\theta_0)$ is 25 for the observed data and 100 for the simulated data. The parameters d_{11}, d_{21}, and d_{22} are the non-zero elements of the Choleski decomposition D of the error covariance matrix (see Section 5). The observed sample size is 162 (see Section 5 and footnote 14 for a description of the data). The simulated sample size is 2000; the structural parameter estimates contained in Table III are used to generate the simulated data.

Table IV tabulates two sets of estimated VAR(2) parameters, one set for the observed data and one set for the simulated data (given the estimated structural parameters in Table III). The vector of differences between the 'observed' and 'simulated' VAR(2) parameters is a key component of the statistic Q_T that is used to test the goodness-of-fit of the real business cycle model. For this estimation problem, $Q_T = 119{\cdot}4$. Since four $(=n-k=13-9)$ overidentifying restrictions are imposed in estimating the structural parameters, the matrix Λ defined in Proposition 2 has four non-zero eigenvalues. The estimated eigenvalues of Λ (i.e. the eigenvalues of a consistent estimate of Λ) are: 3·908, 3·438, 1·914, and 1·398. As discussed following Proposition 2, these estimated eigenvalues can be used to obtain an estimated p-value corresponding to the computed value of Q_T. In particular, the distribution of $z \equiv \Sigma_{i=1}^{4} \hat{\lambda}_i c_i^2$, where $\hat{\lambda}_i$, $i = 1, 2, 3, 4$, are the estimated eigenvalues of Λ and $c_i \sim i.i.d.N(0, 1)$, $i = 1, 2, 3, 4$, is approximated by generating 150,000 i.i.d. draws for z. The largest of these draws is 107·5, which indicates that the p-value associated with the computed value of Q_T is essentially zero. In other words, the structural model is strongly rejected as an adequate representation of the data-generating process of the observed time series.[18]

[18] This conclusion hinges on the assumption that the asymptotic distribution provides a reliable approximation to the distribution of the test statistic Q_T in small samples. Deviations between the asymptotic and small-sample distributions could lead to different conclusions.

6. CONCLUSION

This paper develops two new methods for conducting formal statistical inference in fully specified dynamic structural economic models. The methods yield consistent and asymptotically normal estimates of structural parameters and can be used to test the goodness-of-fit of the structural model. They require very little analytical tractability, relying instead on numerical simulation of the structural model for different values of the structural parameters. The two methods can therefore be applied in a wide variety of interesting but analytically intractable dynamic economic models.

This paper compares the small-sample performance of the two methods by implementing them in a Monte Carlo study involving the estimation of the parameters of a simple real business cycle model. This Monte Carlo study shows that, in samples of the size typically encountered in macroeconomics, the simulated quasi-maximum likelihood (SQML) estimator has smaller mean squared error than the extended method of simulated moments (EMSM) estimator, despite the fact that the EMSM estimator is, except in special cases, asymptotically more efficient than the SQML estimator. This finding suggests that, in cases where the econometrician can choose a quasi-likelihood function that provides a good approximation to the true but unknown likelihood function, the SQML estimator can lead to an improvement in small sample performance.

Finally, this paper uses the SQML estimation strategy to estimate the parameters of a real business cycle model similar to the one studied by Greenwood *et al.* (1988). All nine of the model's parameters are estimated, including the time discount parameter and the rate of depreciation. The empirical analysis strongly rejects the null hypothesis that the real business cycle model considered here is the data-generating process for observed post-war US time series for output and investment.

TECHNICAL APPENDIX 1

This appendix demonstrates the consistency of $\hat{\theta}_S^\beta$ for θ_β and proves the asymptotic result in equation (8). In the light of Assumption 8, similar arguments can be used to demonstrate the consistency of $\hat{\theta}_T$ for θ_0 and to prove the asymptotic result in equation (9). These results are not new to this paper. Similar results (under slightly different conditions) have appeared, for example, in Domowitz and White (1982).

Throughout this Appendix $l_s(\beta;\theta)$ denotes $\log f(y_s(\beta),\ldots,y_{s-p}(\beta);\theta)$ and $L_S(\beta;\theta)$ denotes $\Sigma_{s=1}^S l_s(\beta;\theta)$. The following preliminary results will prove useful. By straightforward modifications of the arguments in Tauchen (1985), it can be shown that Assumption 4 implies that $El_s(\beta;\theta)$ and $E\,\nabla_\theta^2 l_s(\beta;\theta)$ both exist and are continuous in θ. Moreover, $S^{-1}L_S(\beta;\theta)$ and $S^{-1}\,\nabla_\theta^2 L_S(\beta;\theta)$ converge almost surely uniformly in θ to, respectively, $El_s(\beta;\theta)$ and $E\,\nabla_\theta^2 l_s(\beta;\theta)$. Let $\{\theta_S^*\}$ be any sequence satisfying plim $\theta_S^*=\theta_\beta$. Then, given the above results, Theorem 4.1.5 in Amemiya (1985) implies that plim $S^{-1}\,\nabla_\theta^2 L_S(\beta;\theta_S^*)=E\,\nabla_\theta^2 l_s(\beta;\theta_\beta)\equiv A_\beta(\theta_\beta)$.

First, we show consistency of $\hat{\theta}_S^\beta$ for θ_β. By assumption, the parameter space C is compact. By Assumption 3, the measurable function $L_S(\beta;\theta)$ is continuous in θ. As shown above, $S^{-1}L_S(\beta;\theta)$ converges in probability uniformly in $\theta\in\Theta$ to $El_s(\beta;\theta)$. Furthermore, $El_s(\beta;\theta)$ is continuous in θ. Finally, by Assumption 5, $El_s(\beta;\theta)$ is uniquely maximized at θ_β. Thus the conditions of Theorem 4.1.1. in Amemiya (1985) are satisfied, implying that $\hat{\theta}_S^\beta=\arg\max_\theta L_S(\beta;\theta)$ converges in probability to θ_β.

Now we will show asymptotic normality. By definition, $\nabla_\theta L_S(\beta;\hat{\theta}_S^\beta)=0$. A first-order Taylor series expansion about θ_β yields: $\nabla_\theta L_S(\beta;\ \theta_\beta) + \nabla^2_\theta L_S(\beta;\ \theta^*_S)(\hat{\theta}^\beta_S - \theta_\beta) = 0$, where θ^*_S lies on the line joining θ_β and $\hat{\theta}^\beta_S$. (To be precise, θ^*_S should vary from row to row of $\nabla^2_\theta L_S(\beta;\ \cdot)$, but this subtlety makes no difference asymptotically.) Rearranging yields:

$$S^{1/2}(\hat{\theta}^\beta_S - \theta_\beta) = -(S^{-1}\ \nabla^2_\theta L_S(\beta;\ \theta^*_S))^{-1} S^{-1/2}\ \nabla_\theta L_S(\beta;\ \theta_\beta) \tag{27}$$

Since plim $\hat{\theta}^\beta_S = \theta_\beta$ implies that plim $\theta^*_S = \theta_\beta$, plim $S^{-1}\ \nabla^2_\theta L_S(\beta;\ \theta^*_S) = A_\beta(\theta_\beta)$. By Assumption 6, $S^{-1/2}\ \nabla_\theta L_S(\beta;\ \theta_\beta) \to N(0, B_\beta(\theta_\beta))$. Applying Slutsky's theorem to the right-hand side of equation (27) yields the asymptotic result in equation (8) (note that $A_\beta(\theta_\beta)$ is invertible by Assumption 7).

TECHNICAL APPENDIX 2

This appendix provides proofs of Propositions 1–5. Throughout this Appendix $l_t(\theta)$ denotes $\log f(y_t, \ldots, y_{t-p};\ \theta)$ and $L_T(\theta)$ denotes $\Sigma^T_{t=1}\, l_t(\theta)$ Also, as in Appendix 1, $l_s(\beta;\ \theta)$ denotes $\log f(y_s(\beta), \ldots, y_{s-p}(\beta);\ \theta)$ and $L_S(\beta;\ \theta)$ denotes $\Sigma^S_{s=1}\, l_s(\beta;\ \theta)$. Recall, too, that, by virtue of Assumption 8, versions of Assumptions 3–7 hold for the functions $L_T(\theta)$ and $l_t(\theta)$ as well as for the functions $L_S(\beta;\ \theta)$ and $l_s(\beta;\ \theta)$.

Proof of Proposition 1

Let $\beta_0 \in E$, an open subset of C (the existence of such an E follows from the fact that β_0 is an interior point of C). Since $S = \tau T$, $\tau > 0$ fixed, it follows from the consistency of $\hat{\theta}^\beta_S$ for θ_β that $\text{plim}_{T\to\infty}\, \hat{\theta}^\beta_S = \theta_\beta = h(\beta)$. Using arguments analogous to those in Appendix 1, it can be shown that $T^{-1}L_T(\theta)$ converges in probability uniformly in θ to $El_t(\theta) \equiv G(\theta)$, where $G(\theta)$ is continuous in θ. By Theorem 4.1.5 in Amemiya (1985), it follows that $\text{plim}_{T\to\infty}\, T^{-1}L_T(\hat{\theta}^\beta_S) = El_t(h(\beta)) = G(h(\beta))$. By Assumption 5, $h(\beta) = \arg\max_\theta F(\beta, \theta)$, where $F(\beta, \theta) \equiv El_s(\beta;\ \theta)$. Under the null hypothesis (Assumption 2),

$$G(h(\beta)) = El_s(\beta_0;\ h(\beta)) = F(\beta_0, h(\beta))$$

Let $\beta^* = \arg\max_\beta G(h(\beta)) = \arg\max_\beta F(\beta_0, \arg\max_\theta F(\beta, \theta))$. Since $h(\beta)$ is invertible in a neighbourhood of β_0 (by Assumption 9(a)), $\beta^* = \beta_0$.

The preceding argument shows that the function $T^{-1}L_T(h_S(\beta))$ (which is measurable by Assumption 4) converges in probability to a non-stochastic function $G(h(\beta))$ which is maximized at β_0 under the null. By Assumption 10(1), this convergence is uniform in β in a neighbourhood of β_0. By Assumption 3, $L_T(\theta)$ is continuously differentiable in θ; by Assumption 9(1), $h_S(\beta)$ is continuously differentiable in β for $\beta \in N(\beta_0)$, an open neighbourhood of β_0. It follows that $L_T(h_S(\beta))$ is continuously differentiable in β for $\beta \in N(\beta_0)$. The conditions of Theorem 4.1.2 in Amemiya (1985) are therefore met, implying that there exists a consistent root $\hat{\beta}_T$ of the equation $\nabla_\beta L_T(h_S(\beta))$.

Now we will establish the asymptotic normality of $T^{1/2}(\hat{\beta}_T - \beta_0)$. By definition, $\hat{\beta}_T$ satisfies $J_S(\hat{\beta}_T)'\ \nabla_\theta L_T(h_S(\hat{\beta}_T)) = 0$, where $J_S(\beta) \equiv \nabla h_S(\beta)$. (By Assumption 9(1), $J_S(\beta)$ exists in a neighbourhood $N(\beta_0)$ of β_0. Since plim $\beta_T = \beta_0$, the probability that $\hat{\beta}_T \notin N(\beta_0)$ goes to 0 as $T \to \infty$. Hence $J_S(\hat{\beta}_T)$ exists for sufficiently large T.) A first-order Taylor series expansion about β_0 yields:

$$J_S(\beta_0)'\ \nabla_\theta L_T(h_S(\beta_0)) + D_T(\beta^*_T)(\hat{\beta}_T - \beta_0) = 0 \tag{28}$$

where $D_T(\beta^*_T) = C_T(\beta^*_T) + J_S(\beta^*_T)'\ \nabla^2_\theta L_T(h_S(\beta^*_T))J_S(\beta^*_T)$, $C_T(\beta)$ is a $k \times k$ matrix with i,jth element $c_{ij}(\beta) = \nabla_\theta L_T(h_S(\beta))' \, \partial^2 h_S(\beta)/\partial\beta_i \, \partial\beta_j$, and β_T^* lies on the line joining β_0 and $\hat{\beta}_T$ (more precisely, each row of $D_T(\cdot)$ is evaluated at a *different* β_T^* each of which lies on the line joining β_0 and $\hat{\beta}_T$). Rearranging equation (28) yields:

$$T^{1/2}(\hat{\beta}_T - \beta_0) = -(T^{-1}D_T(\beta_T^*))^{-1} J_S(\beta_0)' T^{-1/2} \, \nabla_\theta L_T(h_S(\beta_0)) \tag{29}$$

By Assumptions 10(2) and 10(5), plim $T^{-1}c_{ij}(\beta_T^*) = E[\nabla_\theta l_t(h(\beta_0))] \, \partial^2 h(\beta_0)/\partial\beta_i \, \partial\beta_j$, since plim $\hat{\beta}_T = \beta_0$ implies that plim $\beta_T^* = \beta_0$. Recalling that $\theta_0 = h(\beta_0)$ under the null, $E \, \nabla_\theta l_t(h(\beta_0)) = \nabla_\theta E l_t(\theta_0) = 0$ since, by Assumption 5, θ_0 maximizes $El_t(\theta)$. (By Corollary 5.9 of Bartle, 1966, Assumption 4 guarantees that interchanging the integration and differentiation operators in the preceding derivation is legitimate.) Thus plim $T^{-1}C_T(\beta_T^*) = 0$.

Since plim $\beta_T^* = \beta_0$, Assumptions 10(3) and 10(4) imply that

$$\text{plim } J_S(\beta_T^*)' \, [T^{-1} \, \nabla_\theta^2 L_T(h_S(\beta_T^*))] \, J_S(\beta_T^*) = J(\beta_0)' \, [E \, \nabla_\theta^2 l_t(h(\beta_0))] \, J(\beta_0) = J(\beta_0)' A(\theta_0) J(\beta_0)$$

since $\theta_0 = h(\beta_0)$ under the null. Thus it has been established that

$$\text{plim}(T^{-1}D_T(\beta_T^*))^{-1} J_S(\beta_0)' = (J(\beta_0)' A(\theta_0) J(\beta_0))^{-1} J(\beta_0)'$$

(Note that plim $J_S(\beta_0) = J(\beta_0)$ by Assumption 10(4) and that $J(\beta_0)' A(\theta_0) J(\beta_0)$ is invertible since (1), $A(\theta_0)$ is invertible by Assumption 7 and (2), $J(\beta_0)$ is of full rank by Assumption 9(2).)

Now we will work out the asymptotic distribution of $T^{-1/2} \, \nabla_\theta L_T(h_S(\beta_0))$. A first-order Taylor series expansion about $h(\beta_0)$ gives:

$$T^{-1/2} \, \nabla_\theta L_T(h_S(\beta_0)) = T^{-1/2} \, \nabla_\theta L_T(h(\beta_0)) + \tau^{-1/2}(T^{-1} \, \nabla_\theta^2 L_T(\theta_T^*)) S^{1/2}(h_S(\beta_0) - h(\beta_0)) \tag{30}$$

where θ_T^* lies on the line joining $h(\beta_0)$ and $h_S(\beta_0)$ and $S = \tau T$. By Assumption 6, $T^{-1/2} \, \nabla_\theta L_T(h(\beta_0)) = T^{-1/2} \, \nabla_\theta L_T(\theta_0) \to N(0, B(\theta_0))$. Next, note that since plim $h_S(\beta_0) = h(\beta_0) = \theta_0$ under the null, plim $\theta_T^* = \theta_0$. Hence, by arguments given in Appendix 1, plim $T^{-1}\nabla_\theta^2 L_T(\theta_T^*) = A(\theta_0)$.

Using the result in equation (8), $S^{1/2}(h_S(\beta_0) - h(\beta_0)) \to N(0, A(\theta_0)^{-1} B(\theta_0) A(\theta_0)^{-1})$, since $A_{\beta_0}(h(\beta_0)) = A(\theta_0)$ and $B_{\beta_0}(h(\beta_0)) = B(\theta_0)$ under the null. By Slutsky's theorem, the second term on the right hand side of equation (30) therefore converges in distribution to $N(0, \tau^{-1}B(\theta_0))$. Since $\nabla_\theta L_T(h(\beta_0))$ and $(h_S(\beta_0) - h(\beta_0))$ are independent by construction, the right-hand side of equation (30) converges in distribution to $N(0, (1 + \tau^{-1})B(\theta_0))$. By Slutsky's theorem, the right-hand side of equation (29) therefore converges in distribution to

$$N(0, (1 + \tau^{-1})(J(\beta_0)' A(\theta_0) J(\beta_0))^{-1} J(\beta_0)' B(\theta_0) J(\beta_0) (J(\beta_0)' A(\theta_0) J(\beta_0))^{-1})$$

thereby establishing the result in equation (11).

Proof of Proposition 2

Lemma: Suppose $Y \sim N(0, \Sigma)$, where Y is $n \times 1$ and Σ has rank $n - k$, $0 \leqslant k < n$. Then the distribution of $Y'Y$ is identical to the distribution of $\Sigma_{i=1}^{n-k} \omega_i z_i^2$, where ω_i, $i = 1, \ldots, n - k$ are the non-zero eigenvalues of Σ and $z_i \sim$ i.i.d. $N(0, 1)$, $i = 1, 2, \ldots, n - k$.
Proof of Lemma: Let C be the orthonormal matrix of eigenvectors of Σ. Let Ω be a diagonal matrix whose first $n - k$ diagonal elements are the non-zero eigenvalues of Σ and whose last k diagonal elements are equal to 0; let $\Omega^{1/2}$ and $\Omega^{-1/2}$ be defined in the obvious way. Define $Z \equiv \Omega^{-1/2} C' Y$. Since $C'\Sigma C = \Omega$, $Z \sim N(0, K)$, where K is a diagonal matrix whose first $n - k$ diagonal elements are equal to 1 and whose last k diagonal elements are equal to 0. Letting

z_i denote the ith element of the vector Z, it is clear that $z_i \sim$ i.i.d. $N(0, 1)$, $i = 1, \ldots, n-k$. Note that, since $C'C = I_n$, where I_n is the $n \times n$ identity matrix, $(\Omega^{1/2}Z)'(\Omega^{1/2}Z) = Y'Y$. At the same time $(\Omega^{1/2}Z)'(\Omega^{1/2}Z)$ can be written $\Sigma_{i=1}^{n-k} \omega_i z_i^2$, where the ω_i's are the non-zero eigenvalues of Ω, thereby proving the lemma.

Now we will prove the result in Proposition 2. A first-order Taylor series expansion about β_0 yields:

$$T^{1/2}(\hat{\theta}_T - h_S(\hat{\beta}_T)) = T^{1/2}(\hat{\theta}_T - h_S(\beta_0)) - J_S(\beta_T^*)T^{1/2}(\hat{\beta}_T - \beta_0) \tag{31}$$

where β_T^*, which varies from row to row of $J_S(\cdot)$, lies on the line joining $\hat{\beta}_T$ and β_0. Since plim $\hat{\beta}_T = \beta_0$, plim $\beta_T^* = \beta_0$. Thus, by Assumption 10(iv), plim $J_S(\beta_T^*) = J(\beta_0)$. Note that

$$T^{1/2}(\hat{\theta}_T - h_S(\beta_0)) = T^{1/2}(\hat{\theta}_T - \theta_0) - \tau^{-1/2}S^{1/2}(h_S(\beta_0) - \theta_0) \tag{32}$$

By the result in equation (9), the first term on the right-hand side of equation (32) converges in distribution to $N(0, A^{-1}BA^{-1})$, where A denotes $A(\theta_0)$ and B denotes $B(\theta_0)$. By the result in equation (8), the second term on the right-hand side of equation (32) converges in distribution to $N(0, \tau^{-1}A^{-1}BA^{-1})$. Since the two terms on the right-hand side of equation (32) are independent, it follows that $T^{1/2}(\hat{\theta}_T - h_S(\beta_0)) \to N(0, (1+\tau^{-1})A^{-1}BA^{-1})$.

Let J denote $J(\beta_0)$. Since

$$T^{1/2}(\hat{\beta}_T - \beta_0) \to N(0, (1+\tau^{-1})(J'AJ)^{-1}J'BJ(J'AJ)^{-1})$$

it follows that $T^{1/2}(\hat{\beta}_T - \beta_0)$ and $(J'AJ)^{-1}J'AT^{1/2}(\hat{\theta}_T - h_S(\beta_0))$ have the same limiting distribution. Returning to equation (31), these results show that $T^{1/2}$ $(\hat{\theta}_T - h_S(\hat{\beta}_T))$ has the same limiting distribution as $(I_n - J(J'AJ)^{-1}J'A)T^{1/2}(\hat{\theta}_T - h_S(\beta_0))$, where I_n is the $n \times n$ identity matrix. This random variable in turn converges in distribution to

$$N(0, (1+\tau^{-1})A^{-1}V\Lambda V'A^{-1})$$

where $B = VV'$ and Λ is defined in the statement of Proposition 2.

Define $Z_T \equiv (1+\tau^{-1})^{-1/2}(\hat{A}_T^{-1}\hat{V}_T)^{-1}T^{1/2}(\hat{\theta}_T - h_S(\hat{\beta}_T))$, where $\hat{A}_T$, $\hat{V}_T$, and $\hat{B}_T \equiv \hat{V}_T\hat{V}_T'$ are consistent estimates of, respectively, A, V, and B. Note that $Z_T'Z_T = Q_T$, where Q_T is the test statistic defined by equation (12). Moreover, using the result at the end of the preceding paragraph, $Z_T \to N(0, \Lambda)$. The limiting distribution of Q_T is therefore identical to the distribution of $Y'Y$, where $Y \sim N(0, \Lambda)$. Thus, by the Lemma stated above, the limiting distribution of Q_T is identical to the distribution of $\Sigma_{i=1}^{n-k} \lambda_i c_i^2$, where the λ_i's are the $n-k$ non-zero eigenvalues of Λ and $c_i \sim$ i.i.d. $N(0, 1)$.

Proof of Proposition 3

First we will show the consistency of $\tilde{\beta}_T$. Let $Z_T(\beta) = (\hat{\theta}_T - h_S(\beta))'W_T(\hat{\theta}_T - h_S(\beta))$. Note first that plim $Z_T(\beta) = (\theta_0 - h(\beta))'W(\theta_0 - h(\beta))$. Moreover, Assumption 10(4) guarantees that this convergence is uniform in β. By Assumption 9(1), $Z_T(\beta)$ is a continuously differentiable function of β in an open neighbourhood $N(\beta_0)$ of β_0. Define $\beta^* \equiv \arg\min_\beta Z(\beta)$. From the first-order conditions to this problem, β^* must satisfy $\theta_0 = h(\beta^*)$, or $\beta^* = h^{-1}(\theta_0)$. (Assumption 9(2) guarantees that, in a neighbourhood of β_0, h^{-1} is well defined.) Under the null hypothesis, $\theta_0 = h(\beta_0)$, so that $\beta^* = \beta_0$. Since β_0 is an interior point of C by assumption, there exists an open subset E of C such $\beta_0 \in E$. The conditions of Theorem 4.1.2 in Amemiya (1985) are therefore satisfied, implying that there exists a consistent root of the equation $\nabla Z_T(\beta) = 0$.

Now we will show asymptotic normality. By definition, the EMSM estimator $\tilde{\beta}_T$ satisfies:

$J_S(\tilde{\beta}_T)' W_T(\hat{\theta}_T - h_S(\tilde{\beta}_T)) = 0$. A first-order Taylor series expansion about β_0 yields:

$$J_S(\beta_0)' W_T(\hat{\theta}_T - h_S(\beta_0)) + D_T(\beta_T^*)(\tilde{\beta}_T - \beta_0) = 0 \tag{33}$$

where $D_T(\beta_T^*) = -J_S(\beta_T^*)' W_T J_S(\beta_T^*) + C_T(\beta_T^*)$, $C_T(\beta)$ is a $k \times k$ matrix with i,jth element $c_{ij}(\beta) = (\hat{\theta}_T - h_S(\beta))' W_T \partial^2 h_S(\beta)/\partial\beta_i \partial\beta_j$, and β_T^*, which varies from row to row of $D_T(\cdot)$, lies on the line joining $\tilde{\beta}_T$ and β_0. Since plim $\beta_T^* = \beta_0$, $\text{plim}(\hat{\theta}_T - h_S(\beta_T^*)) = 0$ under the null hypothesis. Using arguments analogous to those used in the proof of Proposition 1, it is clear that plim $c_{ij}(\beta_T^*) = 0$, so that plim $C_T(\beta_T^*) = 0$. It follows that plim $D_T(\beta_T^*) = J(\beta_0)' WJ(\beta_0)$ (plim $J_S(\beta_T^*) = J(\beta_0)$ by Assumption 10(4) and plim $W_T = W$ by construction).

Rearranging equation (33) yields:

$$T^{1/2}(\tilde{\beta}_T - \beta_0) = D_T(\beta_T^*)^{-1} J_S(\beta_0)' W_T T^{1/2}(\hat{\theta}_T - h_S(\beta_0)) \tag{34}$$

As shown in the proof of Proposition 2,

$$T^{1/2}(\hat{\theta}_T - h_S(\beta_0)) \to N(0, (1 + \tau^{-1})A(\theta_0)^{-1}B(\theta_0)A(\theta_0)^{-1})$$

Since plim $D_T(\beta_T^*)^{-1} J_S(\beta_0)' W_T = (J(\beta_0)' WJ(\beta_0))^{-1} J(\beta_0)' W$, applying Slutsky's theorem to the right-hand side of equation (34) yields the result in equation (15).

Proof of Proposition 4

To simplify notation let A denote $A(\theta_0)$, let B denote $B(\theta_0)$, and let J denote $J(\beta_0)$. We will show, first, that $D \equiv (J'AJ)^{-1}J'BJ(J'AJ)^{-1} - (J'AB^{-1}AJ)^{-1}$ is positive semi-definite, and, second, that the rank of $D = \min(k, n-k)$. Since D is $k \times k$, this implies that D is positive definite if $n - k \geqslant k$. D is positive semi-definite if and only if $E \equiv J'AB^{-1}AJ - J'AJ(J'BJ)^{-1}J'AJ$ is positive semi-definite. Write E as follows:

$$E = J'A(P^{-1})'(I_n - Q(Q'Q)^{-1}Q')P^{-1}AJ$$

where $B = PP'$, $Q = P'J$, and I_n is the $n \times n$ identity matrix. The symmetric idempotent matrix $(I_n - Q(Q'Q)^{-1}Q')$ is positive semi-definite with rank $n - k$. E is therefore positive semi-definite. The matrices D and E have the same rank. Since E is $k \times k$, E has rank k if $k \leqslant n - k$ and has rank $n - k$ otherwise.

Proof of Proposition 5

Recall that $\tilde{\beta}_T^*$ is the optimal EMSM estimator, i.e. the EMSM estimator that results when the asymptotic weighting matrix $W = W^* \equiv (A^{-1}BA^{-1})^{-1}$ (A denotes $A(\theta_0)$ and B denotes $B(\theta_0)$). A first-order Taylor series expansion about β_0 yields:

$$T^{1/2}(\hat{\theta}_T - h_S(\tilde{\beta}_T^*)) = T^{1/2}(\hat{\theta}_T - h_S(\beta_0)) - J_S(\beta_T^*)T^{1/2}(\tilde{\beta}_T - \beta_0) \tag{35}$$

where β_T^*, which varies from row to row of $J_S(\cdot)$, lies on the line joining $\tilde{\beta}_T^*$ and β_0. Since plim $\tilde{\beta}_T^* = \beta_0$, plim $\beta_T^* = \beta_0$. Thus, by Assumption 10(4), plim $J_S(\beta_T^*) = J(\beta_0)$.

As shown in the proof of Proposition 2, $T^{1/2}(\hat{\theta}_T - h_S(\beta_0)) \to N(0, (1 + \tau^{-1})(W^*)^{-1})$. Letting J denote $J(\beta_0)$, it is clear from Proposition 3 that $T^{1/2}(\tilde{\beta}_T^* - \beta_0)$ has the same limiting distribution as $(J'W^*J)^{-1}J'W^*T^{1/2}(\hat{\theta}_T - h_S(\beta_0))$. Returning to equation (35), these results show that $T^{1/2}(\hat{\theta}_T - h_S(\tilde{\beta}_T^*))$ has the same limiting distribution as

$$(I_n - J(J'W^*J)^{-1}J'W^*)T^{1/2}(\hat{\theta}_T - h_S(\beta_0))$$

This random variable in turn converges in distribution to $N(0, (1+\tau^{-1})(V')^{-1}\Omega V^{-1})$, where $W^* = VV'$ and $\Omega = I_n - V'J(J'W^*J)^{-1}J'V$.

Define $G_T = (1+\tau^{-1})^{-1/2}V'T^{1/2}(\hat{\theta}_T - h_S(\tilde{\beta}_T^*))$ and note that $G_T'G_T = Z_T$, where Z_T is the test statistic defined in Proposition 5. Using the result at the end of the preceding paragraph, $G_T \to N(0,\Omega)$. The limiting distribution of G_T is therefore identical to the distribution of $Y'Y$, where $Y \sim N(0,\Omega)$. Thus, by the Lemma stated above, the limiting distribution of G_T is identical to the distribution of $\Sigma_{i=1}^{n-k}\omega_i c_i^2$, where the ω_i's are the $n-k$ non-zero eigenvalues of Ω and $c_i \sim$ i.i.d. $N(0,1)$. Since Ω is an idempotent matrix of rank $n-k$, its non-zero eigenvalues are all equal to 1. Hence $G_T \to \chi^2(n-k)$.

ACKNOWLEDGEMENTS

This paper is based on Chapters 2 and 3 of the author's PhD dissertation at Duke University. He would like to thank the members of his thesis committee, John Geweke, Bob Marshall, Jean-François Richard, and George Tauchen, for their guidance. The author would also like to thank for their comments seminar participants at the University of Pennsylvania and Queen's University and participants at the conference on 'Econometric Inference Using Simulation Techniques' held in June 1992 at Erasmus University in Rotterdam, the Netherlands. Finally he would like to acknowledge financial support from the Social Sciences and Humanities Research Council of Canada.

REFERENCES

Altug, S. (1989), 'Time-to-build and aggregate fluctuations: some new evidence', *International Economic Review*, **30**, 889–920.

Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press, Cambridge, MA.

Andrew, D. W. K. (1991), 'Heteroskedasticity and autocorrelation consistent covariance matrix estimation', *Econometrica*, **59**, 817–858.

Bartle, R. G. (1966), *The Elements of Integration*, Wiley, New York.

Christiano, L. J. (1988), 'Why does inventory investment fluctuate so much?', *Journal of Monetary Economics*, **21**, 247–280.

Christiano, L. J. (1990), 'Linear-quadratic approximation and value-function iteration', *Journal of Business and Economic Statistics*, **8**, 99–113.

Domowitz, I., and H. White (1982), 'Misspecified models with dependent observations', *Journal of Econometrics*, **20**, 35–58.

Duffie, D., and K. J. Singleton (1988), 'Simulated moments estimation of Markov models of asset prices', manuscript, Graduate School of Business, Stanford University.

Gourieroux, C., and A. Monfort (1991), 'Simulation-based inference: a survey with special reference to panel data models', Working Paper No. 9106, INSEE.

Gourieroux, C., A. Monfort, and E. Renault (1992), 'Indirect inference', manuscript, CREST-CEPREMAP, Département de la Recherche-INSEE, and GREMAQ-IDEI.

Greenwood, J., Z. Hercowitz, and G. W. Huffman (1988), 'Investment, capacity utilization, and the real business cycle', *American Economic Review*, **78**, 402–417.

Hansen, G. D. (1985), 'Indivisible labor and the business cycle', *Journal of Monetary Economics*, **16**, 402–417.

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica*, **50**, 1029–1054.

Kydland, F., and E. C. Prescott (1982), 'Time to build and aggregate fluctuations', *Econometrica*, **50**, 1345–1370.

Lee, B., and B. F. Ingram (1991), 'Simulation estimation of time series models', *Journal of Econometrics*, **47**, 197–205.

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

McFadden, D., and P. Ruud (1990), 'Estimation by simulation', manuscript, MIT and University of California, Berkelely.

McGrattan, E. R. (1990), 'Solving the stochastic growth model by linear-quadratic approximation', *Journal of Business and Economic Statistics*, **8**, 41–44.

Newey, W. K., and K. D. West (1987), 'A simple, positive definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica*, **55**, 703–708.

Ogaki, M. (1992), 'An introduction to the generalized method of moments', Working Paper No. 314, The Rochester Center for Economic Research.

Pakes, A., and D. Pollard (1989), 'Simulation and the asymptotics of optimization estimators, *Econometrica*, **57**, 1027–1058.

Perron, P. (1989), 'The Geat Crash, the oil price shock, and the unit root hypothesis', *Econometrica*, **57**, 1361–1402.

Smith, Jr, A. A. (1990), *Three Essays on the Solution and Estimation of Dynamic Macroeconomic Models*, PhD dissertation, Duke University.

Tauchen, G. (1985), 'Diagnostic testing and evaluation of maximum likelihood models', *Journal of Econometrics*, **30**, 415–443.