Queen's Economics Department Working Paper No. 826

Calibration in Macroeconomics

Allan W. Gregory Gregor W. Smith

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

6-1991

CALIBRATION IN MACROECONOMICS

Discussion Paper #826

Allan W. Gregory and Gregor W. Smith

Preliminary Draft
Comments Welcome

June 1991

HANDBOOK OF STATISTICS: VOLUME 10: ECONOMETRICS
G.S. MADDALA, EDITOR
C.R. RAO, SERIES EDITOR

Department of Economics, Queen's University, Kingston Ontario Canada K7L 3N6.
We thank the Social Science and Humanities Research Council of Canada for support of this research, Tiff Macklem for providing data, and Aly Jiwan for research assistance. Helpful comments were provided by colleagues at Queen's.

ABSTRACT

This chapter reviews calibration techniques in macroeconomics. The discussion begins with an outline of the use of calibration in applied work. Next, a simple asset-pricing model is the setting for a demonstration of calibration and for comparison with conventional estimation and testing. Experiments with calibrated models may be formalized as Monte Carlo testing. With the asset-pricing model, we use simulation methods to calculate the exact size of the variance-bounds-type test proposed by Hansen and Jagannathan (1991). Finally, we suggest that calibration is best viewed as an informal guide to model reformulation.

1. Introduction

Empirical questions in macroeconomics often are addressed with dynamic, equilibrium models. Studying such questions involves formulating a model, solving it, assigning parameter values (*i.e.* calibration), and then conducting experiments with the model and evaluating the results. This chapter reviews some methods for parameterization and evaluation. Researchers typically study models by conducting simulations and studying simulated data using statistics (such as moments) which also may be calculated with historical data. The aim may be to deduce some of the model's properties or implications, to compare its properties to those of data, or to test the model formally.[1] In this chapter we take a formal statistical approach to these calibration methods. We refer to illustrative applied studies, but do not provide a comprehensive bibliography.

The remainder of this introductory section describes calibration methods heuristically. Section 2 describes a simple asset-pricing model, which is the setting for an outline of estimation and calibration in macroeconomics. Section 3 discusses testing and evaluating calibrated models and demonstrates several tests using the asset-pricing model. Section 4 surveys some recently proposed tools for evaluation. A brief conclusion follows in section 5.

Calibration exercises begin with the assignment of parameter values. Just as in the calibration of applied general equilibrium models in the study

---

[1] For example, Lucas (1987) and İmrohoroǧlu (1989) use calibrated models to measure the costs of business cycles.

of international trade or public finance (see for example Shoven and Whalley, 1984) parameter values are assigned on the basis of other studies or evidence.[2] For example, dynamic, representative-agent models might be calibrated with reference to averages found in panel data. The rationale is not that identification and estimation are impossible; indeed Singleton (1988) and Smith (1990) note that often standard econometric methods may be applied to estimate parameters. Rather the idea is to strengthen results and discipline modelling by avoiding free parameters. Usually model properties are studied with several different parameter vectors, as a check on sensitivity. Information drawn from other studies usually is informal and does not include standard errors. Canova (1991) and Hoover (1991) discuss this issue. Moreover some method of aggregation is required in order to use parameters estimated from microeconomic panels in representative-agent models (*i.e.* to make sure one is measuring the same thing).[3] If the relationships estimated from microeconomic studies do aggregate, then estimation could also be done in the aggregate data, at least as a check. If those relationships do not aggregate, then using the micro-based estimates may be misleading.

In some cases parameters are set so as to match exactly a statistic generated by the model with one in data. For example, Kydland and Prescott (1982) calibrated the coefficient of relative risk aversion in their business-cycle model by matching the variance of detrended output. This matching constitutes estimation, and is often done by simulation. In section

---

[2] Lau (1984), MacKinnon (1984), and Pagan and Shannon (1985) suggest the use of formal estimation, testing, and sensitivity analysis in applied general equilibrium models.

[3] Heckman (1984) discusses this issue and Rogerson (1988) provides an example.

2 we outline this aspect of these empirical methods.

Once a macroeconomic model is parameterized it can be evaluated and then used to answer quantitative questions. A second statistical method in macroeconomics involves more general comparison of a model's properties with those of data. A typical business-cycle study reports measures of variability and of covariance with output for actual data and for a business-cycle model, for such variables as consumption, investment, and hours; a good example is given in Tables 1.1 and 1.2 of McCallum (1989). This comparison can be viewed as an informal test, which guides reformulation or respecification of the model, particularly when the discrepancies appear to be large. Obviously the inferences drawn from the comparison depend on the variables and moments used. For example, Singleton (1988) and Cogley (1990) show that the detrending method (or spectral bandwidth) considered in calculating moments may itself have a large effect on conclusions.

Studies which use calibration methods in macroeconomics are now too numerous to list, and it is safe to say that the approach is beginning to predominate in the quantitative application of macroeconomic models. One influential example of calibration is the study of business cycles by Kydland and Prescott (1982), in which fluctuations are driven by an unobservable productivity shock. They assigned most parameter values based on microeconomic and trend evidence and studied the model's business-cycle properties by simulation. Further study of equilibrium business-cycle models driven by technology shocks has led to several reformulations. For example, the variability of the labour input (hours worked) in the Kydland-Prescott economy was less than that in detrended historical data for the U.S. This discrepancy was largely resolved in the version of the business-cycle model constructed by Hansen (1985) which includes labour supply variation caused by variation in the number of persons working as well as in hours per worker. Although reformulations like this are suggested partly by theory they also are impelled by comparisons of the model's predictions with data. Sections 3 and 4 discuss several ways to formalize such comparisons.

2. Estimation and Calibration

To keep notation as simple as possible, we shall survey estimation and calibration methods entirely within a simple asset-pricing model, as outlined by Lucas (1978). Singleton (1990) provides a comprehensive outline of asset-pricing theory and evidence. In the example here a representative consumer has preferences represented by the utility functional

$$E\left[\sum_{s=0}^{\infty} \beta^s \, u(c_{t+s}) \,\middle|\, I_t \right], \tag{1}$$

in which E is the expectations operator, I_t is the consumer's information at time t, $\beta \in (0,1)$ is a discount factor, $u(c) = c^{1-\alpha}/(1-\alpha)$ ($\log(c)$ if $\alpha = 1$), and α is a positive, constant coefficient of relative risk aversion. Consumption c_t evolves according to

$$c_{t+1} = x_{t+1} \cdot c_t, \tag{2}$$

so that $\{x_t\}$ is the sequence of consumption growth rates.

Relative prices are calculated by equating them to intertemporal marginal rates of substitution. The price of a one-period, risk-free, discount bond which provides one unit of consumption at time t+1 is given by

$$p_t^f = \beta\ E[u'(c_{t+1})/u'(c_t)|I_t] = \beta\ E[(c_{t+1}/c_t)^{-\alpha}|I_t] = \beta\ E[x_{t+1}{}^{-\alpha}|I_t], \tag{3a}$$

and the price of an equity claim to the consumption stream is given by

$$\begin{aligned} p_t^e &= \beta\ E[(u'(c_{t+1})/u'(c_t))\cdot(c_{t+1} + p_{t+1}^e)|I_t] \\ &= \beta\ E[x_{t+1}{}^{-\alpha}\cdot(c_{t+1} + p_{t+1}^e)|I_t]. \end{aligned} \tag{3b}$$

The structural parameters of this model are $\gamma = (\alpha,\beta)$. Denote the realized history $\{y_t = (p_t\ x_t);\ t = 1,\dots,T+1\}$, where p_t is a vector of asset prices. We next consider estimation under various assumptions about the investigator's knowledge and aims. Estimation and inference are based on the sample moments of the observable variables:

$$W_T = T^{-1}\sum_{t=1}^{T} w(y_t), \tag{4}$$

where T is a sample size, w is a q-dimensional vector of observable continuous functions and W_T is a vector of sample moments of the observable variables. For example, W_T could include the sample means and variances of bond and equity prices.

Next suppose that an economic theory predicts a vector $\tilde{w}(y_t,\gamma)$ where $\gamma \in \Gamma \subset \mathbb{R}^v$ is a v-vector of parameters, and the model determines the endogenous variables from the forcing variable and the parameters. Tildes label the predictions of theory. Denote the true parameter values γ_0 so that unconditionally $E[w(y_t)] = E[\tilde{w}(y_t,\gamma_0)]$. We assume that these conditions are satisfied only at γ_o. The sample analogues to these population conditions may be used for estimation. The researcher is interested in the parameters of the economic model (such as α in the asset-pricing model above) which may be consistently estimated if $q \geq v$ and some identification conditions are satisfied. For the most part we shall assume identification and sufficient regularity. We consider three different methods of obtaining parameter values: first the generalized method of moments (GMM), second the method of simulated moments (MSM), and third calibration.

First, if the econometrician can calculate the theoretical moments then estimation can be based on the generalized method of moments (*i.e.* generalized instrumental variables):

$$\hat{\gamma}_{GMM} = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \; \| W_T - \tilde{W}_T(\gamma) \|, \tag{5}$$

where the asymptotic distribution of the estimator will depend on the norm $\|\cdot\|$. Hansen (1982) shows consistency and asymptotic normality (CAN) using the L^2 norm under regularity and stationarity conditions. For example if the minimization is

$$\hat{\gamma}_{GMM} = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \; [(W_T - \tilde{W}_T(\gamma))^T S_T^{-1} (W_T - \tilde{W}_T(\gamma))], \tag{5'}$$

where S_T equals the sample variance-covariance matrix of the moment condition $W_T - \tilde{W}_T(\gamma)$, then $T^{\frac{1}{2}}(\hat{\gamma}_{GMM} - \gamma_0) \overset{a}{\sim} N(0, (\nabla_{Wo}{}^T S_o^{-1} \nabla_{Wo})^{-1})$, where ∇_{Wo} is the expected Jacobian of the same condition and S_o is the population variance-covariance matrix, both evaluated at γ_o. GMM describes the optimal weighting of moment conditions. Efficiency may be increased if population moments $E[\tilde{w}(y_t,\gamma)]$ can be calculated for the model and used in place of $\tilde{W}_T(\gamma)$.

An example of estimation is suggested by equation (3a). From the tower property of conditional expectations, $Ep_t^f = E\beta x_{t+1}^{-\alpha}$. Thus one could estimate α by GMM with q = v = 1 (presetting β, say) based on the moments: $W_T = T^{-1}\sum_{t=1}^{T} p_t^f$ and $\tilde{W}_T(\gamma) = T^{-1}\sum_{t=1}^{T}\beta\cdot x_{t+1}^{-\alpha}$. In this simple example a constant is the only instrument. Hansen and Singleton (1982) estimate α and β using this method with some additional moment conditions that provide overidentifying restrictions and hence a test of the model.

The advantage of this method is that one can estimate and test the model in (3a) and (3b) without specifying the law of motion for the forcing variable. Data on $\{x_t\}$ are used, and some weak restrictions on its properties are required for asymptotic distribution theory, but there is no parameterization of the $\{x_t\}$ process. The disadvantage of this method is that it cannot be used to predict asset prices since the expectations in (3a) and (3b) are unknown. Solving for asset prices requires further, testable assumptions on the data generation process (DGP). For example, one could restrict prices to be positive or one could test whether the forecasts embodied in asset prices coincide with optimal conditional forecasts. Hansen and Singleton (1983) specify a joint, log-normal distribution for $\{y_t\}$ and hence solve for asset prices and estimate by maximum likelihood. Hansen and Sargent (1980) outline methods for testing cross-equation restrictions in the multivariate $\{y_t\}$ process for linear models.

Macroeconomic models frequently include unobservable or latent variables such as productivity shocks in growth models or, in the case of the asset-pricing example, measurement error in aggregate consumption.[4] Computing the likelihood or even moments can be difficult with a latent variable. In these circumstances a heuristic device is to set the parameters and simulate the model. Then comparing statistics from the simulations to those in data, while varying the parameter settings to seek a good match, amounts to estimation. Formal simulation estimators in economics originated with McFadden (1989), and Pakes and Pollard (1989) (see Hajivassiliou's (1991) review of estimation by simulation in models with limited dependent variables elsewhere in this volume). Simulation estimators sometimes can be constructed without the complete DGP (see McFadden and Ruud, 1990). In macroeconomics Kydland and Prescott (1982) estimated some of their parameters by grid search and simulation, while others were set on the basis of other evidence. Thus their method of parameterization was a hybrid of estimation by simulation and calibration. Several subsequent studies have simply calibrated related models with the Kydland-Prescott parameter settings.

Natural examples of moments which are difficult to calculate analytically in macroeconomics arise from measurement schemes. Typically data are collected by time-averaging, skip sampling, or other schemes the effects of which on moments may be difficult to work out analytically. Simulating the measurement or sampling model along with the underlying economic model provides a very simple estimation method. Other settings for estimation by simulation arise in financial modelling in continuous time.

---

[4]See for example Kydland and Prescott (1982) and Gregory and Wirjanto (1990).

There the conditional likelihood (transition probability density function) often is the solution to a partial differential equation which is very difficult to solve for interesting processes.

To illustrate a formal version of this second approach to estimation, suppose that the econometrician measures $\{x_t\}$ with error. In some cases α and β may still be estimatable by GMM but generally estimation (and certainly prediction) requires one to parameterize and simulate the unobserved process. Then the unknown expression for the theoretical moment can be replaced by a simulated moment $\tilde{W}(y_n, \gamma)$, where n indexes simulated observations. For example, suppose that the researcher parameterizes the process for the consumption growth rate as Mehra and Prescott (1985) did. Let the true growth rate, x_n, (observable only by agents) follow a Markov process on a finite, discrete state space $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_J\}$. This process is stationary and ergodic, with transition matrix ϕ,

$$\phi_{ij} = \text{Prob}\ [x_{n+1} = \lambda_j \mid x_n = \lambda_i] \qquad i, j = 1, 2, \ldots, J. \tag{6}$$

The equilibrium or unconditional probabilities are given by

$$\phi_i = \text{Prob}\ [x_n = \lambda_i] \quad \forall\ n. \tag{7}$$

If the current state is (c_n, λ_i), then from equations (3a) and (3b) the prices (relative to one unit of the commodity at time t) of the two assets are

$$p^f(\lambda_i, c_n) = \beta\ E[(x_{n+1}{}^{-\alpha}) | I_n] = \beta \sum_{j=1}^{J} \phi_{ij}\ \lambda_j{}^{-\alpha}, \tag{8a}$$

$$p^e(\lambda_i, c_n) = \beta\, E[(x_{n+1}^{-\alpha})\cdot(c_{n+1} + p^e(\lambda_j, c_{n+1}))\,|\,I_n]$$

$$= \beta \sum_{j=1}^{J} \phi_{ij}\, \lambda_j^{-\alpha}(\lambda_j c_n + p^e(\lambda_j, \lambda_j c_n)). \qquad (8b)$$

The finite-state Markov process allows analytical expressions for the asset prices. In many cases realistic models will not have this feature and a recent and important development in macroeconomics is the use of numerical methods for *solving* models (see Taylor and Uhlig, 1990, and references therein). Burnside (1990) discusses estimation when moments are approximated analytically, rather than by Monte Carlo methods. In this chapter we focus on estimation and testing and hence use an example in which deducing the predictions of the theory is straightforward, once $\{x_n\}$ is simulated.

Next simulations $\{x_n : n = 1,2,3,\ldots,N+1\}$ are drawn from this probability law and functions $\{\tilde{w}(y_n, \gamma)\}$ are calculated using specific values of γ. Then estimation can be based on the method of simulated moments (MSM):

$$\hat{\gamma}_{MSM} = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \parallel W_T - \tilde{W}_N(\gamma) \parallel \qquad (9)$$

where $\tilde{W}_N(\gamma) = N^{-1} \sum_{n=1}^{N} \tilde{w}(y_n, \gamma)$ are the simulated moments. The vector γ now may include parameters of the forcing process. If the rationale for estimation by simulation is that $\{x_t\}$ is measured with error then W_T will include moments of asset prices only. Under ergodicity the two sample moments which are matched in (9) converge, as T and N approach infinity, to two population moments which are equal at γ_0; this equality forms the basis of estimation.

The argument in equation (9) can be rewritten as:

$$W_T - \tilde{W}_N(\gamma) = [W_T - \tilde{W}_T(\gamma)] + [\tilde{W}_T(\gamma) - E\tilde{w}] + [E\tilde{w} - \tilde{W}_N(\gamma)]$$

$$= [W_T - \tilde{W}_T(\gamma)] + s_T + u_N \qquad (10)$$

In equation (10) the first term is the argument of the GMM estimator in (5). The second term, s_T, is a sampling error which arises if GMM or MSM is based on sample moments rather than population moments $E\tilde{w}$. The third term, u_N, is a simulation error. Thus the difference between GMM and MSM estimators depends on the properties of u_N. In many macroeconomic applications no simulation bias arises, but consistency and asymptotic normality results in MSM require some restrictions on the error u_N.

One possibility suggested by (10) is to prove a central limit theorem by applying empirical process methods (see Pollard 1984, 1985) to the simulation residual. Let v_N be the empirical process operator defined as $v_N \equiv N^{\frac{1}{2}}(P - P_N)$, where P is the population probability measure and P_N is the empirical measure, with mass N^{-1} at each observation $\tilde{w}(y_N, \gamma)$. The standardized residual is $N^{\frac{1}{2}}u_N = v_N\tilde{w}(y_n, \gamma)$. An important necessary condition for CAN is that the empirical process $v_N\tilde{w}(y_n, \gamma)$ is stochastically equicontinuous in γ. This smoothness condition requires that Monte Carlo random numbers not be redrawn as γ is varied. It does allow some discontinuities. For example, in finance many applications (*e.g.* with kinked payoffs) involve functions w which are not pointwise differentiable with respect to γ. In these circumstances Taylor's Theorem applied to $\tilde{w}(y_n, \gamma)$ cannot be used to establish asymptotic normality. Pakes and Pollard (1989) describe central limit theorems for such environments; as in (10) these theorems take limiting operations first, then rely on the differentiability of the expectation.

Duffie and Singleton (1990) note two reasons why standard proofs of CAN for MSM in i.i.d. environments may not apply to dynamic models. First, simulations begin from some initial values, yet the unconditional distribution of the forcing variables typically is not known as a function of the parameters. For example, one would simulate the asset-pricing model here by choosing an initial state (c_0 λ_i) then calculating prices as in (8a) and (8b) by drawing consumption growth rates from the conditional probability density function. Geometric ergodicity in $\{x_n\}$ is sufficient for the effects of using arbitrary initial conditions in simulation to die out. Second, in a non-i.i.d. environment the simulated moment depends on the parameters directly through the moment condition (as in GMM) but also indirectly because values of parameters such as ϕ are used in generating the past history of the simulated variables. For example, the simulated mean price of the risk-free asset is

$$\tilde{W}_N(\gamma) = N^{-1} \sum_{n=1}^{N} \beta \cdot x_{n+1}{}^{-\alpha}, \tag{11}$$

which is a function of β and α (just as in GMM) but also depends on the parameters through their use in simulating $\{x_n\}$. Duffie and Singleton provide CAN results for MSM under these conditions.

In fact, in this example one can calculate the unconditional mean of the risk-free price as:

$$E[\tilde{w}(y_t,\gamma)] = E(p_t^f) = \sum_{i=1}^{J} \sum_{j=1}^{J} \phi_i \phi_{ij} \; \lambda_j{}^{-\alpha}. \tag{12}$$

This population moment would provide more efficient estimates than would the simulated moment — both s_T and u_N can be avoided. In general, though, the simulation error u_N will make the MSM estimator less efficient than a comparable GMM estimator. For $P\tilde{w}^2 < \infty$, $v_n\tilde{w}$ is asymptotically distributed as $N(0, P\tilde{w}^2 - (P\tilde{w})^2)$ *i.e.* $N(0, var[\tilde{w}])$. This result is a heuristic version of Lee and Ingram's (1991) finding that if N = T then the asymptotic (as T and $N \to \infty$) variance is twice that of the GMM estimator. The idea is that simulations are independent of the observed data and hence their contribution to the variance is orthogonal to that of the usual GMM component.

Other simulation methods and variance-reduction techniques can reduce sampling variability. Duffie and Singleton (1990) note that averaging over R independent simulations with N = T yields an asymptotic variance $(1 + R^{-1})$ times the GMM one. By ergodicity the same result holds if there is one long simulation with N = RT. Melino (1991) suggests other improvements: for example, if one is simulating an Itô process, then one need only simulate the predictable component, since the expectation (used in estimation) is unaffected by the martingale component. In many cases one can simply set N very large, calculate population moments, and then apply GMM; in such cases there is little efficiency loss from simulation, which is simply used as a calculation device.

Gregory and Smith (1990) present some Monte Carlo results for GMM and MSM estimators in small samples in this asset-pricing model. They also present nonparametrically estimated densities for examples of these estimators. One drawback in informal MSM is that parameters may be selected even if they are not identifiable. A second general point is that matching properties other than moments may lead to inconsistent estimators.

Nevertheless, estimation and testing sometimes may be based on matching properties other than moments. For example, macroeconomic evidence might be summarized in the coefficients of a vector autoregression or of a linear regression. Given identifiability, one could calculate the population coefficients in the same regressions in a theoretical model and match the two sets in order to estimate parameters. Smith (1989) shows that such matching yields consistent, asymptotically normal estimators provided it is based on regular functions of asymptotically normal variables.

In practice the choice among estimators often hinges on computation. Much remains to be learned about practical estimation in specific applications, particularly with non-differentiabilities. Bossaerts and Hillion (1991) outline applications of MSM in financial models. They describe methods for estimating option-pricing models in which prices are found by simulation (see also Boyle (1977)). CAN results are available, despite non-differentiabilities and dependent errors, provided simulation estimates enter moment conditions linearly.

In some economic models a further approximation arises because moments of a continuous-time process are approximated by moments from a simulated discrete-time process. Bossaerts and Hillion (1991) let the order of discretization grow with the sample size, and interpolate to reduce bias in finite samples. Duffie and Singleton (1990) discuss discretization schemes and asymptotic results. Smith and Spencer (1991) use a simple discretization for MSM estimation in a target-zone model of exchange-rate intervention, in which again theoretical moments cannot be calculated analytically.

So far we have noted that simulation methods may be useful in parameterizing macroeconomic models in which there are unobservable variables or simply analytical intractabilities. But there is a further use for simulation methods even if GMM is feasible: repeated simulation can allow exact (small-sample) estimation. For example, Tauchen (1986) and Gregory and Smith (1990) find numerically that for this asset-pricing DGP and for realistic persistence in consumption growth rates $\hat{\alpha}_{GMM}$ is biased down if α_0 = 2.0 or 4.0 (for N = 100 or 500). Setting N = T and making R large traces out the finite-sample density of $\hat{\alpha}_{GMM}$ and hence can be used to make bias corrections. This is a traditional use of the Monte Carlo method in econometrics.

A third method of parameterization, and an alternative to formal estimation methods such as GMM and MSM, is to assign parameter values with reference to other studies *i.e.* to calibrate. One idea behind this method is simply to reduce uncertainty about a model's predictions, and hence strengthen tests, by assigning parameter values using point estimates from related studies. Gregory and Smith (1990) study mixed estimators in which some parameters are pre-set (calibrated) and others are estimated, as in Kydland and Prescott (1982) or Burnside, Eichenbaum, and Rebelo (1990). Obviously there is a trade-off between efficiency and robustness — generally estimators will be inconsistent if the pre-setting is incorrect but may lead to estimates with lower mean square error if the pre-setting error is not large. The importance of pre-setting a parameter, as opposed to estimating it consistently, can be gauged in sample size: How much larger an historical sample would be required with estimation, as opposed to calibration, to achieve as much precision in some measure? Moreover, in some cases moment conditions can be found which can consistently estimate some parameters even if others are set incorrectly. Gregory and Smith (1990) give an example.

3. Model Evaluation and Testing

Once a model has been formulated, solved, and parameterized its properties can be studied and compared to those in data. Relatively informal comparisons of moments have become very widespread in macroeconomics. These comparisons may illuminate respects in which the model seems inadequate. Of course, an exact match is unduly restrictive since historical moments have sampling variability and so can differ from a model's population moments even if the model is true. Therefore some method for gauging the discrepancy between actual and predicted moments is necessary. Three sources of uncertainty may affect the comparison. First, uncertainty may arise from simulation or from approximation. We shall assume R, the number of replications, is large, so that simulation error can be safely ignored (and we need not distinguish between GMM and MSM, for example). Second, uncertainty arises if parameters are estimated. Third, there is sampling variability in the historical moments themselves.

This section illustrates several techniques of model evaluation using a common data set. The data are annual returns used by Grossman and Shiller (1981) and Mehra and Prescott (1985) for 1889-1979. Consumption is measured as annual per capita real consumption on non-durables and services. The real return on equity is constructed from the corresponding consumption deflator, the annual average of the Standard and Poor Composite Stock Price Index, and annual dividends. The risk-free real return is based on prices of short-term securities (such as Treasury Bills), as in Mehra and Prescott (1985, section

2).

Economic models typically restrict a wide range of properties, and hence one approach to formal testing is based on overidentification. Suppose that v parameters have been estimated using q moment conditions. Hansen (1982) shows that under H_0 T times the minimized value of the criterion in (5') is asymptotically $\chi^2(q-v)$:

$$J = T[(W_T - \tilde{W}(\hat{\gamma}_{GMM}))^T S_T^{-1}(W_T - \tilde{W}(\hat{\gamma}_{GMM}))] \overset{a}{\sim} \chi^2(q-v), \tag{13}$$

where S_T is defined in section 2. This test of overidentifying restrictions allows for sampling variability in moments and for parameter uncertainty. A similar test can be applied if only v moments are first chosen for estimation, and then the remaining q-v are used for testing. An example is given by Burnside, Eichenbaum, and Rebelo (1990).

Example 3.1: Testing Overidentifying Restrictions.

To illustrate the test in equation (13) we first calculate GMM estimates of α and β by transforming equations (3a) and (3b) to give the following moment conditions:

$$E[(\beta x_{t+1}^{-\alpha} \cdot r_{t+1}^{f} - 1) \cdot z_t] = 0, \tag{14a}$$

$$E[(\beta x_{t+1}^{-\alpha} \cdot r_{t+1}^{e} - 1) \cdot z_t] = 0, \tag{14b}$$

where $r_{t+1}^f = 1/p_t^f$ and $r_{t+1}^e = (p_{t+1}^e + c_{t+1})/p_t^e$ are the real, annual, *gross* returns on bonds and equity and $z_t \subset I_t$ is a vector of instruments in agents' information set. In this example $z_t = (1\ x_t\ x_{t-1}\ r^f_t\ r^f_{t-1}\ r^e_t\ r^e_{t-1})$. With seven instruments and two equations $q = 14$ and $v = 2$. The parameter estimates are $\hat{\beta}_{GMM} = 1.030$ (0.0681) and $\hat{\alpha}_{GMM} = 9.747$ (3.383), with standard errors given in parentheses. Note that the estimate of β is outside the range to which it would usually be restricted in calibration. The value of the J-statistic is 29.689 and comparison to its asymptotic distribution $\chi^2(12)$ gives a prob-value of 0.0031 so that the model is rejected at the usual significance levels.

Many tests of calibrated models can be viewed as more detailed studies of the dimensions in which a model might fail empirically, because complete calibration of a model is not required to test it (as example 3.1 illustrates). The aim of calibrating a model economy is to conduct experiments in which its properties are derived and compared to those of an actual economy. One way to describe this comparison is to use the language of classical statistical inference: if the actual properties could not have been generated by the economic model except with very low probability then the model is rejected and otherwise it is not. But nothing hinges on viewing the comparison as a classical test. For example, one could treat the comparison simply as a measurement exercise in which one gauges the proportion of some observed variance, say, which the theoretical model can reproduce.

In contrast to the test in example 3.1, most tests of calibrated models evaluate models while ignoring parameter uncertainty. For example, Cecchetti, Lam, and Mark (1990) test a fully-parameterized version of the consumption-based asset-pricing model by comparing the unconditional moments of its predicted asset prices with those in historical data. Their asymptotic test allows for sampling variability in moments and in parameters of the $\{x_t\}$ process, and shows the effect of fixing α and β as is done in calibration.

However, an advantage in testing calibrated models is that exact procedures are available because the complete DGP is specified. Thus an alternative to asymptotic tests in a fully calibrated model is to use the sampling variability of the *simulated* moment to gauge the closeness of the historical moment to the model's population moment. To conduct this test, one simulates repeatedly with same sample size as is available in historical data and then forms the empirical density of the simulated moments $\{\tilde{W}_N(\gamma)_r: r = 1,\ldots,R\}$. With this density one can calculate critical values, or treat the historical moment as a critical value. For example, the proportion of the sequence $\{\tilde{W}_{Nr}\}$ that exceeds W_T gives the size of the one-sided test implicit in the comparison of the historical and population moments. Since N = T the inference is exact as R becomes large. The same principle can underlie joint tests, which are valid provided the matrix of moments is of full rank.

Monte Carlo testing can be traced back to Barnard (1963) and an economic application is given by Theil and Shonkwiler (1986). Gregory and Smith (1991) use this method to calculate the sizes of tests of various parameterizations of the simple asset-pricing model used in this chapter. We next provide two examples.

Example 3.2: The Size of the Mehra-Prescott Test.

This example calculates the size (probability of falsely rejecting a true model) of the test Mehra and Prescott (1985) conducted with a calibrated version of the asset-pricing model outlined above. They parameterized the model as follows: $\alpha = 1.5$, $\beta = 0.99$, $\lambda_1 = 0.982$, $\lambda_2 = 1.054$, and

$$\phi = \begin{bmatrix} 0.43 & 0.57 \\ 0.57 & 0.43 \end{bmatrix}.$$

The rationale for this calibration is that the population moments of this consumption growth-rate process match those of the U.S. annual sample for 1889-1979. Based on various applied studies which estimate α Mehra and Prescott concluded that α is probably between 1 and 2. They examined the sensitivity of findings to values of α between 0 and 10 as well as values for β between 0 and 1. As a particular example, consider $\alpha = 1.5$ and $\beta = 0.99$. This choice of α lies outside the 95% confidence interval around $\hat{\alpha}_{GMM}$ estimated from the annual returns data in example 3.1.

Table 1 lists moments implied by the model and found in data, as is typically done in calibration studies. The first column presents the first two population moments of consumption growth, the risk-free return, and the equity premium for the model. The second column of the table gives the same moments for U.S. historical data. The population equity premium rate $(Er_t^e - Er_t^f)$ is 0.20%.

Table 1: Population and Sample Moments
(Mehra-Prescott Model)

Moment	Model	1889-1979	95% Confidence Interval
$\mu(x_t)$	1.8	1.83	(1.2, 2.4)
$std(x_t)$	3.60	3.57	(3.6, 3.6)
$\mu(r_t^f-1)$	3.51	0.80	(3.4, 3.6)
$std(r_t^f-1)$	0.8	5.67	(0.8, 0.8)
$\mu(r_t^e-r_t^f)$	0.20	6.18	(-0.6, 1.0)
$std(r_t^e-r_t^f)$	3.9	16.67	(3.7, 4.0)

Note: Values are given in percent terms. μ denotes a mean, std denotes a standard deviation, r_t^f is the gross, real return on the risk-free asset (T-bill), r_t^e is the gross, real return on equity. Returns are measured as percentages. Confidence intervals are based on R = 1000 replications.

Mehra and Prescott tested the model by examining whether it could (with α and β in the restricted range) generate both a population value for the risk-free rate of return less than 4% and an equity premium of at least 6.2%, which is the value for the historical sample for the U.S. from 1889-1979. Formalizing this test using Monte Carlo methods requires no auxiliary assumptions, since the model is completely parameterized. The proportion of simulations (with N = T) which satisfy the criterion given above is the size of the test. Gregory and Smith (1991) find this proportion to be zero, so that the model is very unlikely to have generated the data. Some other parameterizations lead to positive prob-values. The same method can be used to test the model or formalize comparisons using other moments shown in Table 1 and to construct confidence intervals. For example, the third column in Table 1 shows the interquantile ranges from the empirical density function of the simulated moments.

Example 3.3: The Size of the Hansen-Jagannathan Bounds Test.

Models also may be evaluated using more general diagnostic procedures. Examples are variance bounds tests (see Shiller (1987)). The idea behind such tests stems from writing asset prices as products of expected payoffs and an intertemporal marginal rate of substitution. For example, in equation (3a)

$$p_t^f = E[\beta \, x_{t+1}{}^{-\alpha} | I_t], \tag{15}$$

where p_t^f is the price of an asset which pays 1 in all states and $\beta \, x_{t+1}{}^{-\alpha}$ is the intertemporal marginal rate of substitution (IMRS) between time t and time t+1. Denote this IMRS as m, a scalar and let std denote a standard deviation. Then $p_t^f = E[m_{t+1} | I_t]$, and $std(p_t^f) = std(E[m_{t+1} | I_t]) < std(m_{t+1})$; because the price in the economic model is the conditional expected value of a random variable it should be less variable than the actual random variable. Thus the variance of observed, risk-free asset prices can provide a lower bound on the variance of the IMRS. In this sense the variability of the risk-free price can allow a test of a specific economic model of m.

In discussing calibration we have described attempts to match moments from theory with those in data. Bounds tests use much less information in evaluating the model but can still serve as useful diagnostics and show why certain models might be rejected in statistical tests. They are weak tests because they do not require a complete parametric model of m, as Hansen and Jagannathan (1991) show. We next describe their general diagnostic, for the special case in which there is a unit payoff (a risk-free asset).

Let r be a vector of returns on assets with payoffs one period hence and p be the corresponding price vector. Standard asset pricing models give rise to the following unconditional pricing relation:

$$E[mr] = \iota; \tag{16}$$

where ι is a vector of ones of the same dimension as r. Examples include equations (3a) and (3b), where $m_{t+1} = \beta x_{t+1}^{-\alpha}$ and $r_{t+1} = (1/p_t^f\ (c_{t+1} + p_{t+1}^e)/p_t^e)^T$.[5] Following Hansen and Jagannathan assume that $E|m|^2 < \infty$, $E|r|^2 < \infty$, $E|rr^T|$ is nonsingular, and $E|p| < \infty$. Although m cannot be calculated directly from observations of returns, imagine constructing its population least-squares projection onto r, denoted $m^* = r^T\delta$. Suppose in constructing m* we require that it satisfy the restriction from theory given in (16). Then $E[rr^T\delta] = \iota$ and $\delta = E[rr^T]^{-1}\iota$. Notice that we can construct m* using the unconditional second moment of r. Importantly since the portfolio contains a unit payoff then $Em^* = Em = Ep_t^f$,

---

[5] While the bounds may be developed using a conditioning set (say information available to the agents at time t, as in Gallant, Hansen and Tauchen, 1990) we choose the simpler unconditional relation.

the expected price of a unit payoff.[6] Moreover, std(m) ≥ std(m*), because m* is based on a projection.

In applications the population moments are estimated by the sample moments; denote the estimator $\hat{m}_T$. One can find the standard deviation bound of a candidate IMRS, std(m*), and compare it to std($\hat{m}_T$). The idea is similar to that in example 3.2 in that it evaluates the credibility of a candidate model. We next use simulation methods to determine the sampling variability involved in this bounds test.

We first determine the population standard deviation of the intertemporal marginal rate of substitution from the same, fully parameterized economy used in example 3.1. We denote this std[m]; this is the object which will be bounded from below by std[m*]. We also calculate the population bound std[m*]. These values are given in the first two columns of Table 2. The first row uses only the risk-free asset while the second row uses the risk-free asset and the equity. The latter case gives a much tighter bound, since it uses more payoffs.

Next we simulate the model R = 1000 times and calculate an estimated std[$\hat{m}_N$]$_r$ r = 1,...R based on simulated samples of N = 90 observations.The third column in Table 2 gives the (.05,.95) interquantile range for these bounds. There is considerably more sampling variability in the bound when simulated equity payoffs are included, in part due to near singularities in

---

[6] Hansen and Jagannathan also consider more realistic cases in which there is no risk-free payoff. They also discuss how to exploit the restriction that m is positive.

the sample mean of rr^T.

The interquantile range can be thought of as the non-rejection region for the test with size 10%. Testing may be done using actual data by determining whether the observed sample estimate std[$\hat{m}_T$] lies in this region. This estimate is shown in the fourth column of the Table. The Table uses data for 1889-1979 so that T = N.

Table 2: Size of the Bounds Test
(Standard Deviation of the Intertemporal Marginal Rate of Substitution)

Prices	std[m]	std[m*]	(.05,.95) std[$\hat{m}_N$]$_r$	std[$\hat{m}_T$]	p-value
p^f_t	0.051	0.00717	(.00709, .00721)	0.059 (0.009)	.000
p^f_t, p^e_t	0.051	0.049	(.00739, .229)	0.372	.000

First consider the properties of the bound which uses only the risk-free returns. From the first row of the table, std[$\hat{m}_T$] is slightly larger than std[m]. The asymptotic standard deviation of std[$\hat{m}_T$] is given beneath it in parentheses and shows that the sampling variability in std[$\hat{m}_T$] could rationalize std[$\hat{m}_T$] < std[m]. In this sense the asset-pricing model is able to generate realistic volatility in the IMRS to be consistent with that implied by risk-free returns alone.

However, the exact 95% interquantile range for std[$\hat{m}_N$] for the one-asset case does not contain std[$\hat{m}_T$] and the probability of observing std[$\hat{m}_T$] in a sample of 90 observations under this asset-pricing model is zero. Thus the exact test finds that the model cannot generate sufficient volatility in small samples to account for the volatility in the data.

The same two tests can be conducted with two asset return sequences, as in the second row of Table 2. In this case we were not able to calculate the asymptotic standard error of $std[\hat{m}_T]$ because of a numerical singularity. With two assets $std[\hat{m}_T]$ rises to 0.372 and this lower bound seems be much larger than the standard deviation of the model's IMRS, 0.051. Under the null hypothesis that the model is true we find that it does not generate a bound of 0.372 in 1000 replications so that the one-sided prob-value is 0. Even though the sampling variability of the estimated lower bound is much larger with two assets it is not large enough to account for the large lower bound in the data.

We have examined whether there is enough sampling variability so that the model could generate the observed, historical bound $std[\hat{m}_T]$ *i.e.* it compares $std[m^*]$ to $std[\hat{m}_T]$. A separate though related question is: what is the probability of finding $std[\hat{m}_T] > std[m]$ when the model is true? Recall that $std[m^*] < std[m]$. For this test one again uses the variability of the $std[\hat{m}_N]_r$ and compares $std[m^*]$ to $std[m]$. The one-sided prob-value for this test with one asset is 0.00 and with two assets it is 0.68. Thus in the artificial economy with bounds based on two returns 68% of the lower bounds exceed the actual standard deviation of the IMRS.

These prob-values and confidence regions are themselves estimates, but increasing the number of replications or local smoothing of the empirical density of the $std[\hat{m}_N]_r$ had no effect to several decimal places. When calculating E[m] and std[m] for a candidate model of the IMRS it seems simple to use simulation to gauge the exact sampling variability of the Hansen-Jagannathan bound under the null.

The three tests outlined in the examples — based on orthogonality restrictions, on matching the mean equity premium, and on a bound on the variance of the IMRS — all reject the model. In this case they yield similar results, although they use different information. For example, the rejection in example 3.1 does not require a parametric model of the consumption growth rate x_t. It thus suggests that reformulating the asset-pricing functional in (3a) and (3b) (as opposed to the forcing process in (6) and (7) only) is necessary. Numerous reformulations of this and other asset-pricing models have sought to generate features such as a larger mean equity premium and a more variable IMRS than those generated by the DGP here.

4. Further Topics in Model Evaluation

Models also may be evaluated according to their ability to reproduce features of standard 'windows' applied to historical data. For example, one could study a linear regression or vector autoregression which has been fitted to data, and calculate the population regression coefficients in the same statistical window implied by a fully calibrated model. Again sampling variability can be taken into account to gauge the distance between the two. Such comparisons can highlight particular directions towards which reformulated models might aim. Backus, Gregory, and Zin (1989) study the term structure of interest rates along these lines. Kwan (1990) matches impulse response functions in a business-cycle model.

So far we have discussed evaluating calibrated models by comparing moments in data with those predicted by the theoretical model. A stronger test is to seek to match entire densities, even if moments do exist.[7] In the asset-pricing example one could seek to match the density of the real interest rate. Such methods are used informally in international finance, where target-zone models of exchange rates generate exchange-rate densities which are compared to those in data. A difficulty with this approach is that standard Kolmogorov-Smirnov tests cannot be used because neither density is known analytically.

Several proposals for model evaluation have been made which return to the issue of parameter uncertainty. One possibility is simply to examine the sensitivity of findings to parameter settings by reporting results for various sets of parameters. An alternative suggested by Kwan (1990) is to formulate a prior density and use formal Bayesian methods. Denote this prior by $\pi(\gamma)$, which is conditional on information in other studies, for example. In most calibration studies π is sharp and has point mass at specific values. The joint, predictive density of simulated, endogenous variables is

$$f(\{\tilde{p}_n: n=1,\ldots,N\}) = \int f(\{\tilde{p}_n: n=1,\ldots,N\} | \gamma) \cdot \pi(\gamma) d\gamma.$$

Model evaluation is based on moments or other statistics calculated in the simulated data. Kwan proposes a formal measure of distance between simulated

[7] The same matching could be applied to densities of sample moments. For example, one could compare the distribution of means (calculated for each decade) of the risk-free interest rate with the simulated density of means each based on ten observations. Or one could compare the distribution of mean returns for different countries with the same simulated density. We thank Doug Willson for this suggestion.

and actual moments, given by the probability that $\tilde{W}_N$ lies in some neighbourhood around the historical sample moment W_T. This is the same idea as above — the randomness in the simulated (rather than actual) moment is used to measure distance.

As in the classical case inference is essentially exact (given a large number of simulations), as it would be in a standard Bayesian analysis based on the likelihood function of the data $f(\{p_t\}|\gamma)$. While $f(\{\tilde{p}_n\}|\gamma)$ is not needed explicitly, since one can simply draw from the model (resampling from $\pi(\gamma)$ with replacement) the density of the sample moment or the moments of the sample moment's density also are unavailable analytically. These can be calculated by Monte Carlo integration, perhaps with importance sampling. Canova (1991) formalizes mixed calibration/estimation exercises as Bayesian, along these lines.

In some cases calibrated models are not complete probability models and hence are not intended to mimic the complete random properties of the series under study. These models cannot be evaluated statistically, or fairly treated as null hypotheses, unless they are augmented by some random variables, perhaps interpreted as measurement error. For example, Hansen and Sargent (1980) observe that stochastic singularities often arise in dynamic economic models. An example can be given in the asset-pricing model. Recall that $p_t^f = E[\beta x_{t+1}^{-\alpha}|I_t]$, and suppose that to test the model an investigator (having studied statistical properties of the IMRS) proposes that the expectation be modelled as $E[\beta x_{t+1}^{-\alpha}|x_t] = g(x_t)$. In that case the predicted asset price is a deterministic function of x_t. Since such deterministic relationships are rarely detected in data, this model would be rejected. It can be made testable if there is some component of the information set used in forecasts which is not observed by the investigator so that $p_t^f = g(x_t) + \varepsilon_t$, where $\varepsilon_t = E[\beta x_{t+1}{}^{-\alpha} | I_t] - E[\beta x_{t+1}{}^{-\alpha} | x_t]$. A further example is given by Altuğ (1989) who begins with a one-shock business cycle model and then augments variables with idiosyncratic error. Watson (1990) proposes identifying the error process by the requirement that its variance be as small as possible. He also proposes goodness-of-fit measures which evaluate the contribution of the original model to accounting for movements in the endogenous variables, by measuring the variance of error necessary to match theoretical and data properties.

5. Conclusion

Although we have attempted to give a formal statistical interpretation to some aspects of calibration in macroeconomics, it perhaps is best viewed as an informal guide to reformulating a theoretical model. Setting parameter values (*i.e.* calibrating), simulating a model, and comparing properties of simulations to those of data often suggests fruitful modifications of the model. Precisely this method has led to numerous modifications of the simple asset-pricing model used an as expository device in this chapter. Further statistical formalization and refinement of the methods used to evaluate calibrated models will help improve economic models.

References

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

Backus, D.K., Gregory, A.W., and Zin, S.E. (1989) Risk Premiums in the Term Structure: Evidence from Artificial Economies. *Journal of Monetary Economics* 24, 371-399.

Barnard, G.A. (1963), Discussion of 'The Spectral Analysis of Point Processes' by M.S. Bartlett, *Journal of the Royal Statistical Society Series B*, 25, 294.

Bossaerts, P. and Hillion, P. (1991) Arbitrage Restrictions Across Financial Markets: Theory, Methodology, and Tests. California Institute of Technology, Social Science Working Paper 751.

Boyle, P.P. (1977) Options: A Monte Carlo Approach. *Journal of Financial Economics* 4, 323-338.

Burnside, C. (1990) Asymptotic Properties of Method of Moments Estimators Based on Numerical Solutions to an Economic Model. mimeo, Department of Economics, Queen's University.

Burnside, C., Eichenbaum, M., and Rebelo, S. (1990) Labor Hoarding and the Business Cycle. Working Paper No. 3556, National Bureau of Economic Research.

Canova, F. (1991) Simulating Dynamic General Equilibrium Models Using Bayesian Techniques. mimeo, Department of Economics, Brown University.

Cecchetti, S.G., Lam, P. and Mark, N.C. (1989) The Equity Premium and the Risk Free Rate: Matching the Moments. mimeo, Ohio State University

Cogley, T. (1990) Spurious Business Cycle Phenomena in HP Filtered Time Series. mimeo, Department of Economics, University of Washington.

Duffie, D. and Singleton, K.J. (1990) Simulated Moments Estimation of Markov Models of Asset Prices. Technical Working Paper No. 87, National Bureau of Economic Research.

Gallant, A.R., Hansen, L.P., and Tauchen, G. (1990) Using Conditional Moments of Asset Payoffs to Infer the Volatility of Intertemporal Marginal Rates of Substitution. *Journal of Econometrics* 45, 141-179.

Gregory, A.W. and Smith, G.W. (1990) Calibration as Estimation. *Econometric Reviews* 9, 57-89.

Gregory, A.W. and Smith, G.W. (1991) Calibration as Testing: Inference in Simulated Macroeconomic Models. forthcoming in the *Journal of Business and Economic Statistics*

Gregory, A.W. and Wirjanto, T. (1990) The Effect of Sampling Error on the Time Series Behavior of Consumption Data: A Comment." forthcoming in the *Journal of Econometrics*.

Grossman, S.J. and Shiller, R.J. (1981) The Determinants of the Variability of Stock Market Prices. *American Economic Review*

Hajivassiliou, V.A. (1991) Simulation Estimation Methods for Limited Dependent Variable Models. chapter __, this volume.

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, 1029-1054.

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

Hansen, L.P. and Singleton, K. (1982) Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models. *Econometrica* 50, 1269-1286.

Hansen, L.P. and Singleton, K. (1983) Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns. *Journal of Political Economy* 91, 249-265.

Hansen, L.P. and Jagannathan, R. (1991) Implications of Security Market Data for Models of Dynamic Economies. *Journal of Political Economy* 99, 225-262.

Heckman, J. (1984) Comments on the Ashenfelter and Kydland papers. In *Essays On Macroeconomic Implications of Financial and Labor Markets and Political Processes*, ed. K. Brunner and A. Meltzer. Carnegie-Rochester Conference Series on Public Policy 21, 209-224. Amsterdam: North-Holland.

Hoover, K.D. (1991) Calibration versus Estimation: Standards of Empirical Assessment in the New Classical Macroeconomics. Working Paper 72, University of California, Davis.

İmrohoroǧlu, A. (1989) Cost of Business Cycles with Indivisibilities and Liquidity Constraints. *Journal of Political Economy* 97, 1364-1383.

Kwan, Y.-k. (1990), Bayesian Model Calibration with An Application to a Non-Linear Rational Expectations Two-Country Model, mimeo, Graduate School of Business, University of Chicago

Kydland, F.E. and Prescott, E.C. (1982) Time to Build and Aggregate Fluctuations. *Econometrica* 50, 1345-1370.

Lau, L.J. (1984) Comments. In H.E. Scarf and J.B. Shoven, eds. *Applied General Equilibrium Analysis*. Cambridge University Press.

Lee, B.-S., and Ingram, B.F. (1991), Simulation Estimation of Time Series Models, *Journal of Econometrics*, 47, 197-205.

Lucas, R.E., Jr. (1978) Asset Prices in an Exchange Economy. *Econometrica* 46, 1426-1445.

Lucas, R.E., Jr. (1987) *Models of Business Cycles.* Yrjö Jahnsson Lectures. Oxford: Basil Blackwell.

MacKinnon, J. (1984) Comments. In H.E. Scarf and J.B. Shoven, eds. *Applied General Equilibrium Analysis.* Cambridge University Press.

McCallum, B.T. (1989) Real Business Cycle Models. chapter 1 in R.J. Barro, ed. *Modern Business Cycle Theory.* Harvard University Press.

McFadden, D. (1989) A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration. *Econometrica* 57, 995-1026.

McFadden, D. and Ruud, P.A. (1990) Estimation by Simulation. mimeo, Department of Economics, University of California, Berkeley.

Mehra, R and Prescott, E.C. (1985) The Equity Premium: A Puzzle. *Journal of Monetary Economics* 15, 145-161.

Melino, A. (1991) Estimation of Continuous-Time Models in Finance. Working Paper 9115, Department of Economics, University of Toronto.

Pagan, A. and Shannon, J (1985) Sensitivity Analysis for Linearized Computable General Equilibrium Models. In J. Piggott and J. Whalley, eds. *New Developments in Applied General Equilibrium Analysis.* Cambridge University Press.

Pakes, A. and Pollard, D. (1989) Simulation and the Asymptotics of Optimization Estimators. *Econometrica* 57, 1027-1057.

Pollard, D. (1984) *Convergence of Stochastic Processes.* New York: Springer-Verlag.

__________ (1985) New Ways to Prove Central Limit Theorems. *Econometric Theory* 1, 295-314.

Rogerson, R. (1988) Indivisible Labor, Lotteries, and Equilibrium. *Journal of Monetary Economics* 21, 3-16.

Shiller, R.J. (1987) The Volatility of Stock Market Prices. *Science* 235, 33-37.

Shoven, J.B. and Whalley, J. (1984) Applied General-Equilibrium Models of Taxation and International Trade. *Journal of Economic Literature* 22, 1007-1051.

Singleton, K. (1988) Econometric Issues in the Analysis of Equilibrium Business Cycle Models. *Journal of Monetary Economics* 21, 361-386.

Singleton, K. (1990) Specification and Estimation of Intertemporal Asset-Pricing Models. In B. Friedman and F. Hahn, eds. *Handbook of Monetary Economics.* Amsterdam: North-Holland.

Smith, A.A. (1989) Estimation of Dynamic Economic Models by Simulation: A

Comparison of Two Approaches. mimeo, Duke University

Smith, A.A. (1990) Econometric Evaluation of a Real Business Cycle Model Using Simulation Methods. chapter 3, Ph.D. dissertation, Duke University.

Smith, G.W. and Spencer, M.G. (1991) Estimation and Testing in Models of Exchange-Rate Target Zones and Process Switching. In P. Krugman and M. Miller, eds. *Exchange Rate Targets and Currency Bands*, Cambridge University Press.

Tauchen, G. (1986) Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data. *Journal of Business and Economic Statistics* 4, 397-416.

Taylor, J.B. and Uhlig, H. (1990) Solving Nonlinear Stochastic Growth Models: A Comparison of Alternative Solution Methods. *Journal of Business and Economic Statistics* 8, 1-17.

Theil, H., and Shonkwiler, J.S. (1986), Monte Carlo Tests of Autocorrelation, *Economics Letters*, 20, 157-160.

Watson, M.W. (1990), Measures of Fit for Calibrated Models, mimeo, Department of Economics, Northwestern University.