Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained From Financial Market Data

George Tauchen
Department of Economics, Duke University, Durham, NC 27706

The article examines the properties of generalized method of moments GMM estimators of utility function parameters. The research strategy is to apply the GMM procedure to generated data on asset returns from stochastic exchange economies; discrete methods and Markov chain models are used to approximate the solutions to the integral equations for the asset prices. The findings are as follows: (a) There is variance/bias trade-off regarding the number of lags used to form instruments; with short lags, the estimates of utility function parameters are nearly asymptotically optimal, but with longer lags the estimates concentrate around biased values and confidence intervals become misleading. (b) The test of the overidentifying restrictions performs well in small samples; if anything, the test is biased toward acceptance of the null hypothesis.

KEY WORDS: Monte Carlo simulation; Integral equations; Markov chain.

1. INTRODUCTION

The so-called Lucas critique is now widely understood. The argument that the parameters of traditional econometric models are not invariant with respect to shifts in policy rules, although not universally accepted (e.g., Sims 1982), is found by many to be compelling enough to have had a major impact on macroeconometrics. There is now much ongoing work aimed at redirecting econometrics toward estimation of underlying structural "taste and technology" parameters that are arguably invariant to a wider class of interventions than are more traditional parameter estimates. This work can be viewed as one step in a larger research program wherein estimates of the taste and technology parameters will ultimately be used to guide the calibration of stochastic general equilibrium models, which will then be used to simulate the economy's responses to a variety of interventions. This research program is nowhere near complete, although it seems likely that completion of the program will absorb much research effort in macroeconomics over the next few years.

There are basically two (not necessarily conflicting) strategies for obtaining econometric estimates of structural parameters of rational expectations models. One approach, an example of which is Sargent's (1978) work on dynamic labor demand schedules, is to estimate by parametric maximum likelihood a reduced-form var model subject to the restrictions imposed on the var by the underlying optimizing model. This approach is entirely analogous to full information maximum likelihood for the simultaneous-equations model. Although maximum likelihood estimation has several optimum properties, its implementation in a rational-expectations context is difficult. For instance, it requires the econometrician to make distributional assumptions about random errors, and it requires a complete and credible description of the stochastic environment in which economic agents operate. At present there is little guidance from theory for making these specifications. In addition, this approach is not well suited to handling nonlinearities beyond what can be reasonably modeled in a quadratic-linear setup.

The second approach is the generalized method of moments (GMM) instrumental variables procedure set out by Hansen (1982) and Hansen and Singleton (1982, 1983). The procedure has been used in several financial market applications (Brown and Gibbons 1985; Dunn and Singleton 1984; Rotenberg 1984) and in labor market applications as well (Biddle 1984). The basic idea behind the approach is to apply the GMM estimator directly to orthogonality conditions implied by the first-order conditions of agents' intertemporal optimization problems. Specifically, residuals are formed by using realized values when conditional expectations appear in the first-order conditions and the instruments are comprised of variables known at the time the expectation is formed. The procedure is a limited-information method analogous to two-stage least squares. Among its attractive features are that it does not require strong distributional assumptions nor a complete description of the agents' environment. As noted by Garber and King (1983), however, the procedure does entail the implicit assumption that the first-order conditions are "exact,"

in the sense that the functional form of the agent's objective function is known to the econometrician and not subject to unobserved random fluctuations. Moreover, the GMM approach does not incorporate nearly as wide a class of specification tests as those developed by Newey (1985) and Tauchen (1985a) for parametric maximum likelihood models. Nonetheless, it still appears that because of the attractive characteristics just noted, the GMM procedure will be one of the main econometric methods used in the longer-term research program outlined previously.

GMM parameter estimates are complicated nonlinear functions of the data. The nonlinearities arise directly through those introduced in the objective functions of the agents and indirectly through the estimation of the "optimal" weight matrix for the second-step minimization. All of the econometric theory justifying inference from GMM estimates is asymptotic, and often the sample sizes are not that large in relationship to the number of orthogonality conditions used in the estimation. In addition, there is little guidance from econometric theory for choosing the appropriate lag lengths for forming the instruments. Results such as those of Hayashi and Sims (1983) are suggestive but not that helpful for practice with a given fixed-sized data set, because these results are based on an iterated-limit argument wherein the limit is first taken with respect to sample size, after which a second limit is taken with respect to the dimensionality of the instrument set.

The purpose of this article is to investigate the small sample properties of the GMM procedure applied to financial market data. Among the questions to be addressed are the properties of the estimates and the quality of the asymptotic approximations for small versus large instrument sets. In addition, the validity of the test of the overidentifying restrictions will be examined. Information on the small-sample properties of this procedure is important for the research program discussed before. There should be a clear awareness of the conditions under which the procedure can give biased and misleading point and interval estimates. It almost goes without saying that a large-scale policy simulation based on invalid and misleading parameter estimates could be worse than useless.

The main idea of the article is first to build small-scale artificial economies patterned in a general way on the ideas in Lucas (1978). Then pseudodata consisting of realizations of the dividends and general equilibrium asset prices are generated randomly, and the GMM estimation technique is applied to these data. For each artificial economy, the entire process is replicated a large number of times, thus providing a large random sample of estimates of the underlying preference parameters that is then subjected to further statistical analysis. Although the research strategy is simple enough to describe, there are a number of interesting methodological problems encountered in implementing it that are discussed further later.

Monte Carlo methods are not the only ones used previously by econometricians to analyze instrumental variables estimators. For example, Phillips (1982) and others have made extensive use of various analytical approximations to study instrumental variables estimators like two-stage least squares. These approximations, however, exploit certain features of the estimation context that are not present in the case of the GMM procedure. Specifically, the approximations rest on the fact that the underlying structural model is linear and that the estimator is therefore a known and relatively mild nonlinear function of a few sufficient statistics from the sample, and, due to a normality assumption, the joint distribution of the sufficient statistics is known. None of these restrictive conditions are present in the case of the GMM estimator, which is why Monte Carlo sampling experiments are employed.

The principal findings of this study are as follows: First, the GMM estimate of the curvature parameter of the constant relative risk aversion utility function can be biased, in certain not unrealistic circumstances, with the magnitude of the bias being as large as the asymptotic standard error. The direction of the bias depends on the covariance structure of dividends and consumption. Generally, a strong positive correlation between dividend growth rates and consumption growth rates leads to an upward bias in the estimate of the curvature parameter, whereas a lack of correlation leads to a downward bias. Second, the test of the overidentifying restrictions performs reasonably well in moderate-sized samples. If anything, the test is slightly biased in favor of accepting the null hypothesis, which is consistent with some analytical results of Shapiro (1986) for a different but related model. A caveat that deserves mention, however, is that the sampling experiments reported in this article pertain to the one-asset case. As noted by Singleton (1985), the observed rejections of the specification with this test that are often encountered in practice appear to be due to the inability of the consumption-beta model to explain adequately the cross-asset structure of returns; subsequent work will explore more fully the properties of this test for the multi-asset case. The third finding pertains to the widely divergent parameter estimates that are sometimes encountered in practice when different lag lengths are used in forming instrument sets. In general, with short lag lengths nearly asymptotically optimal estimates of the curvature parameter can be expected, but as lag length increases, the sampling distribution of the estimates becomes increasingly concentrated around severely biased parameter values. Thus, contrary to what is sometimes proposed, it is probably inadvisable to use the range of estimates obtained with different instrument sets as an indicator of the reliability of the procedure or to perform Hausman (1978)-type specification tests on pairwise differences between the estimates, which should be zero asymptotically. Among a large set of estimates produced with various instrument sets, the most credence should

probably be placed on those estimates obtained with the smallest instrument sets, because the confidence intervals will thereby be the most reliable.

The remainder of this article is organized as follows: Section 2 sets forth much of the notation, reviews the main ideas behind the GMM technique, and discusses some work by Hansen (1985a) on asymptotic lower bounds for the asymptotic variance–covariance matrix of GMM estimates. Section 3 gives the details of the construction of the data-generating process; Section 4 presents the findings from the sampling experiments; and Section 5 contains the concluding remarks.

2. THE GMM PROCEDURE AND APPLICATIONS TO FINANCIAL MARKETS

2.1 Orthogonality Conditions and the Estimation Method

Consider a representative agent who has real wealth available for immediate consumption c_t or for investment into M different real assets. Let p_{it} denote the ex-dividend real price of the ith asset and let d_{it} denote the real dividend. The agent's intertemporal budget constraint is

$$c_t + \sum_{i=1}^{M} q_{i,t}p_{it} = \sum_{i=1}^{M} q_{i,t-1}(p_{it} + d_{it}),$$

where q_{it} is the amount of the ith asset that the agent carries into period $t + 1$. The right side of the budget constraint consists of the total real wealth available to the agent at time t, and the left side consists of the various expenditures on consumption and assets that the agent makes at time t. The agent's intertemporal utility function is assumed to be additively separable,

$$\text{lifetime utility} = E_t\left[\sum_{k=0}^{\infty} \beta^k u(c_{t+k})\right],$$

where $E_t[\ ]$ denotes the conditional-expectations operator given all information available to the agent at time t, the parameter $0 < \beta < 1$ is the subjective rate of time preference, and the per-period utility function $u(\cdot)$ satisfies $u' > 0$ and $u'' < 0$. Under regularity conditions that, among other things, restrict the agent from continuously rolling over debt, there will exist a unique solution to the agent's constrained optimization problem. The first order conditions for maximizing lifetime utility subject to the budget constraint are

$$p_{i,t}u'(c_t) = \beta E_t[u'(c_{t+1})(p_{i,t+1} + d_{i,t+1})],$$

$$i = 1, 2, \ldots, M. \quad (2.1)$$

These first-order conditions, which can be derived using a variety of different analytical methods (see, e.g., Lucas 1978; Grossman and Shiller 1982), are fundamental to modern asset pricing models. They form the basis for the GMM estimation technique that will be described presently, and they will also form the basis for the data-generating mechanism for the Monte Carlo work in Sections 3 and 4.

Hansen and Singleton (1982, 1983) developed methods for estimating the underlying parameters of the representative agent's utility function by applying the GMM estimation method to orthogonality conditions implied by the first-order conditions (2.1). Their methods do not require additive separability of the intertemporal utility function, but because of computational constraints only the additively separable case is considered in this article. The basic method proceeds as follows. Suppose it is reasonable to assume that the per-period utility function is a member of a parametric family, $u(c) = u(c; \gamma_0)$, where γ_0 denotes the true but unknown value of the parameter and the parameter space for $\gamma_0 \in \Gamma$ is some finite dimensional space. The parameter vector γ_0 determines the characteristics—for example, the curvature properties—of the per-period utility function. The subjective discount factor for the intertemporal utility function is also an unknown parameter denoted by β_0. (A subscript 0 will be used to denote "true" values when the identification is important.) The complete parameter vector to be estimated will be denoted by $\theta_0 = (\gamma_0, \beta_0)$, and its dimension is $p \times 1$. Define the error functions

$$e_{it}(\theta) = \frac{\beta u'(c_{t+1}; \gamma)(p_{i,t+1} + d_{i,t+1})}{u'(c_t; \gamma)p_{it}} - 1,$$

$$i = 1, 2, \ldots, M, \quad (2.2)$$

where $\theta = (\gamma, \beta)$. Then the first-order conditions of the agent's optimization problem imply that each of the error functions satisfies

$$E_t[e_{it}(\theta_0)] = 0; \quad (2.3)$$

that is, the conditional expectation of each of the error functions evaluated at the true parameter value is zero. Equation (2.3) states that the errors have mean zero conditional on the information set at time t, where the term "error" is used to denote the corresponding error function evaluated at the true parameter value. (The term "information set" means a sigma field of events.) Throughout this article I only consider the case in which lagged errors $e_{i,t-k}(\theta_0)$ for $k \geq 1$ are in the information set, so by (2.3) the errors are serially uncorrelated. This case covers a wide class of estimation problems, including most estimation using equity market data. There are, however, some interesting estimation problems with forward-market data (e.g., Hansen and Hodrick 1983) in which this is not the case, but investigation of them is deferred to later work. Equation (2.3) implies that the errors must be uncorrelated with variables in the information set I_t, and thus if we let z_t denote a finite dimensional vector of random variables that are I_t measurable, then by the law of iterated expectations,

$$E[e_{it}(\theta_0)z_t] = 0, \quad i = 1, 2, \ldots, M. \quad (2.4)$$

The z_t used to form the products in (2.4) are the instruments for the estimation. In typical applications to financial market data, the set of instruments usually

consists of current and lagged values of consumption (or its rate of growth) and asset returns. A more compact way to write (2.4) is

$$E[e_t(\theta_0) \otimes z_t] = 0$$

or

$$E[g_t(\theta_0)] = 0, \tag{2.5}$$

where $e_t(\theta_0)$ is an $M \times 1$ column vector comprised of the $e_{it}(\theta_0)$ and $g_t(\theta) = e_t(\theta) \otimes z_t$.

The condition (2.5) is the basis for the GMM technique. Suppose the econometrician has data on asset prices, dividends, and the instruments. Define the function

$$\bar{g}_T(\theta) = \left(\frac{1}{T}\right) \sum_{t=1}^{T} g_t(\theta),$$

which is the sample average of the $g_t(\theta)$'s, where T denotes sample size. Then, under regularity conditions $\bar{g}_T(\theta) \rightarrow E[g_t(\theta)]$ almost surely uniformly in θ as $T \rightarrow \infty$. But since $E[g_t(\theta_0)] = 0$, then choosing the estimator $\hat{\theta}$ in such a way as to minimize the magnitude of $\bar{g}_T(\hat{\theta})$ will give a consistent estimator of θ_0. In particular, consider the estimator $\hat{\theta}$ obtained by minimizing the quadratic form

$$Q_T(\theta) = \bar{g}_T(\theta)' W_T \bar{g}_T(\theta), \tag{2.6}$$

where W_T is a symmetric nonsingular weighting matrix that satisfies $W_T \rightarrow W$ almost surely, where W is symmetric and nonsingular. Note that the dimension of the right side of (2.6) is MR, where R is the number of components of the instrument vector. An order condition for identification is $MR \geq p$. If MR is equal to the number of parameters p to be estimated, then in general it is possible to choose $\hat{\theta}$ so that $g_T(\hat{\theta})$ equals zero exactly and the value of the objective function (2.6) will be zero at the estimate. On the other hand, if MR exceeds p, then $\bar{g}_T(\theta)$ cannot in general equal zero for any θ, and only asymptotically will the minimized value of the objective function vanish. Hansen (1982) showed that under regularity conditions the estimator $\hat{\theta}$ that minimizes (2.6) is a consistent estimator of θ_0 with an asymptotic variance–covariance matrix Σ_W that depends on the limiting weighting matrix W. Hansen also showed that, among all possible W's, the one that minimizes (in the matrix sense) the limiting variance–covariance matrix is

$$W_0 = (E[g_t(\theta_0)g_t(\theta_0)'])^{-1}.$$

The matrix W_0 is just the inverse of the variance–covariance matrix of the random variable $g_t(\theta_0)$. If the sequence of weighting matrices satisfies $W_T \rightarrow W_0$ almost surely, then the limiting variance–covariance matrix of the GMM estimator is

$$\Sigma^0 = [(E[\partial g_t(\theta_0)'/\partial\theta])W_0(E[\partial g_t(\theta_0)/\partial\theta'])]^{-1}. \tag{2.7}$$

In addition, the statistic $TQ_T(\hat{\theta})$, which is the sample size times the minimized value of the objective function (2.6), is asymptotically distributed as a chi-squared ran-

dom variable with degrees of freedom equal to the dimension of $g_t(\theta)$ less the number of estimated parameters. This statistic provides a test of the overidentifying restrictions inherent in (2.5).

2.2 The Lower Bound for the Variance–Covariance Matrix of the GMM Estimator

The expression (2.7) is the smallest asymptotic variance–covariance matrix of the GMM estimator taken over all possible choices of sequences of weighting matrices $W_T \rightarrow W$, holding constant the sequence of instruments $\{z_t\}$. To emphasize the dependence of the asymptotic distribution on the instruments, I will, in this subsection, write the asymptotic variance–covariance matrix as Σ_z^0. Hansen (1985a) characterized lower bounds for Σ_z^0 as z_t varies over an admissible set of instruments z_t measurable with respect to the information set I_t. Specifically, Hansen gave an expression for

$$\Sigma^{00} = \inf\{\Sigma_z^0 : z_t \text{ is } I_t \text{ measurable}\},$$

where the infimum is to be interpreted in the matrix sense. Hansen actually derived the lower bound for a more general situation in which the errors are possibly serially correlated, although here I shall make use only of his results as they apply to the serially uncorrelated case. The lower bound Σ^{00} is entirely analogous to the Rao–Cramer lower bound for M estimators, and it plays a role in GMM theory similar to that of the Rao–Cramer bound in maximum likelihood theory. Since the lower bound Σ^{00} and its characterization is important for the subsequent sections, a brief overview of its derivation follows.

In the next paragraph it is shown that any GMM estimator $\hat{\theta}$ based on a sequence weighting matrices W_T such that $W_T \rightarrow W$ almost surely is asymptotically equivalent to another GMM estimator $\tilde{\theta}$ defined to be the minimizer of the quadratic form

$$\bar{e}_{HT}(\theta)'\bar{e}_{HT}(\theta), \tag{2.8}$$

where $\bar{e}_{HT}(\theta)$ is given by

$$\bar{e}_{HT}(\theta) = \left(\frac{1}{T}\right) \sum_{t=1}^{T} H_t e_t(\theta), \tag{2.9}$$

and where H_t is a $p \times M$ matrix whose elements are measurable with respect to the information set I_t. Note that since H_t has exactly as many rows as there are parameters to estimate, then in general $\bar{e}_{HT}(\tilde{\theta})$ will equal zero and the minimized value of the objective function (2.8) will vanish at $\tilde{\theta}$. Thus, the equivalence between the estimators will mean that we can always view the GMM procedure as being asymptotically equivalent to another procedure in which the estimator is obtained by multiplying the error function $e_t(\theta)$ by a $p \times M$ instrument matrix H_t, averaging the products over the sample, and then solving the set of implicit equations

$$\bar{e}_{HT}(\theta) = 0, \tag{2.10}$$

for the estimator. In this equivalent formulation the expression for the asymptotic variance–covariance matrix of the estimates is

$$(E[H_t \partial e_t/\partial \theta'])^{-1}(E[H_t e_t e_t' H_t'])(E[H_t \partial e_t/\partial \theta'])^{-1'},$$
(2.11)

where e_t and its partial derivative are evaluated at θ_0, which is suppressed for legibility.

To establish the asymptotic equivalence of these two estimators, note that $\hat\theta$, which is the solution to the optimization problem

$$\min_\theta \{\bar{g}_T(\theta)' W_T \bar{g}_T(\theta)\},$$

where $W_T \to W$ almost surely, is asymptotically equivalent to the solution to the optimization problem

$$\min_\theta \{\bar{g}_T(\theta)' W \bar{g}_T(\theta)\}.$$

The first-order condition for this last optimization problem is

$$(\partial \bar{g}_T'(\theta)/\partial \theta) W \bar{g}_T(\theta) = 0.$$

The asymptotic distribution of the solution to this system of equations will be the same as the asymptotic distribution to the system of equations given by $V \bar{g}_T(\theta) = 0$, where the $p \times M$ matrix $V = (\partial g(\theta_0)'/\partial \theta) W$ and $g(\theta) = E[g_t(\theta)]$. Note that although the matrix V depends on the underlying true parameter value θ_0, it does not depend on the argument θ, and thus it is just a matrix of constants. Furthermore, given the definitions of the g's, the left side of the last equation is

$$\left(\frac{1}{T}\right) \sum_{t=1}^{T} V g_t(\theta) = \left(\frac{1}{T}\right) \sum_{t=1}^{T} V[z_t \otimes e_t(\theta)]$$

$$= \left(\frac{1}{T}\right) \sum_{t=1}^{T} H_t e_t(\theta)$$

$$= \bar{e}_{HT}(\theta),$$

where the matrix H_t is just linear combinations of the elements of V and z_t. This last equality establishes the asymptotic equivalence of $\bar\theta$ and $\hat\theta$. Thus, as previously noted, we can always view any GMM estimator as being equivalent to one obtained by multiplying the error function $e_t(\theta)$ by a $p \times M$ instrument matrix H_t and then solving the set of implicit equations given by (2.10).

Hansen's characterization of the lower bound matrix is as follows: Define

$$D_t = \partial e_t(\theta)/\partial \theta', \qquad D_t^* = E_t[D_t].$$

The rows of the $M \times p$ matrix D_t are the gradients of the error functions with respect to θ, and the matrix D_t^* contains the conditional expectation of these gradients. Now put

$$\Phi_t = E_t[e_t(\theta_0)e_t(\theta_0)'],$$

which is the conditional variance–covariance matrix of the errors. Then the lower bound for the variance–covariance matrix of the parameter estimates is

$$\Sigma^{00} = (E[D_t^{*'} \Phi_t^{-1} D_t^*])^{-1}.$$
(2.12)

The lower bound matrix Σ^{00} has the interpretation of being the asymptotic variance–covariance matrix of a GMM estimator based on an "optimal" choice of the instrument matrix H_t in (2.9). Specifically, if we define the instrument matrix

$$H_t^0 = D_t^{*'} \Phi_t^{-1},$$
(2.13)

then Σ^{00} is precisely the variance–covariance matrix of the GMM estimator using H_t^0 for the instrument matrix, as can be seen by substituting (2.13) into (2.11). In general, so long as the function $H_t^0 e_t(\theta)$ satisfies the regularity conditions given in Hansen (1985a), then this represents the most efficient choice of the instruments possible. There is a very deep connection between GMM estimation using the instrument matrix H_t^0 and maximum likelihood estimation. Chamberlain (1983) demonstrated that in an iid setting, GMM estimation using this instrument matrix is equivalent to nonparametric maximum likelihood, where the probability distribution of the data is estimated nonparametrically along with the parameters of interest. These instruments cannot be used in practice, though, because the matrix H_t^0 depends on the underlying parameter vector θ_0, which is unknown. It is natural to conjecture, however, that if the dimensionality of the instrument vector z_t used in (2.4) grows with sample size in a suitable manner so that the space spanned by z_t is ultimately the full information set I_t, then the limiting variance–covariance matrix of the GMM estimator will be Σ^{00}. This conjecture is motivated by the results of Hayashi and Sims (1983) for linear models. (As noted in the Introduction, the Hayashi–Sims results are not quite the appropriate results, even for the linear case, because of the two-step limiting process in which sample size first tends to infinity and then the dimensionality of the instrument vector tends to infinity.)

To my knowledge, there is no proof yet in the literature for the conjecture just cited. In the sampling experiments, however, it will be possible to gain some insight into the small-sample properties of the asymptotically optimal quasi-estimator using the instrument matrix H_t^0, since the data-generating process and the "true" parameter vector are available and the required conditional expectations for calculating H_t^0 can be computed. (The term quasi-estimator is used because the matrix H_t^0 depends on the true parameter value.) In addition, the lower bound matrix Σ^{00} can also be computed so that the sampling variability of $\hat\theta$ in infinite samples can be compared with its theoretical asymptotic lower bound.

3. THE DATA-GENERATION PROCESS

The design of the data-generation process follows in a general way the setup presented by Lucas (1978). I

use a one-agent exchange economy with externally given stochastic laws of motion for the consumption endowment and asset dividends. A set of integral equations that are to be solved for the implied asset prices is comprised of the first-order linear conditions for the agent's intertemporal optimization problems. Following Mehra and Prescott (1985) the integral equations are reexpressed in terms of growth rates to allow for (borderline) nonstationarities in the laws of motion for consumption and dividends. For reasons discussed more fully later, a discrete state space approach is adopted for solving the integral equations and generating pseudodata for the sampling experiments.

The use of Monte Carlo methods in a rational expectation context does raise a point noted by Sims in oral remarks at the 1984 Summer Meeting of the Econometric Society. Because the random number generator is deterministic, so is the artificial economy, but the economic agents, who are presumably "rational," are never aware of this determinacy. A response to this point is to note that an estimator's sampling distribution is a complex multivariate integral on a very high dimensional space, and the Monte Carlo method is used only to get good numerical approximations (which are otherwise unavailable) to this integral.

3.1 The Asset Pricing Equations

Write the first-order conditions (2.1) from the agent's intertemporal optimization problem as

$$p_{it} = \beta E_t \left[\frac{u'(c_{t+1})}{u'(c_t)} (p_{i,t+1} + d_{i,t+1}) \right],$$
$$i = 1, 2, \ldots, M. \quad (3.1)$$

As previously noted, we will assume that this is an exchange economy in which the laws of motion for the consumption endowment and the dividend streams from the M assets are externally given, and we will use the first-order conditions to solve for the implied asset prices. There is a slight difference between this setup and Lucas's in that consumption and the M dividends will be treated as having a nondegenerate $(M + 1)$-dimensional probability distribution, whereas Lucas (1978) defined the consumption endowment as the sum of the dividends from the assets, implying a singular joint distribution for dividends and consumption. The setup here, however, is really no different nor any more or less general than Lucas's, because we can always think that there is an $(M + 1)$st asset that has been deleted from the list and that consumption is the sum of the dividends from all $M + 1$ assets.

I now rewrite the first-order conditions (3.1) in such a way as to allow for nonstationarities in the laws of motion for consumption and dividends. The motivation for this step is the common empirical finding that, like many of the macroaggregates, per-capita consumption and measures of aggregate dividends are best described

as stochastic processes that are stationary in growth rates instead of levels [see Hansen and Singleton (1982) and the empirical results in the next section]. Mehra and Prescott (1985) presented an approach for doing this that is generalized here in the obvious way. Define the dividend growth variables $x_{it} = d_{it}/d_{i,t-1}$ for $i = 1, 2, \ldots, M$, and define the consumption growth variable $w_t = c_t/c_{t-1}$. It is assumed that the per-period utility function is of the constant relative risk aversion family,

$$u(c) = [1/(1-\gamma)]c^{1-\gamma}. \quad (3.2)$$

Then, letting $v_{it} = p_{it}/d_{it}$ denote the price dividend ratio for the ith asset, Equation (3.1) can be written in terms of these new variables as

$$v_{it} = \beta E_t[(1 + v_{i,t+1})(w_{t+1})^{-\gamma} x_{i,t+1}]. \quad (3.3)$$

It is further assumed that the random variables $x_t = (x_{1t}, x_{2t}, \ldots, x_{Mt})$ and w_t are jointly stationary first-order Markov processes with conditional cumulative probability distribution function $F(w', x' \mid w, x) = \Pr[w_{t+1} \leq w', x_{t+1} \leq x' \mid w_t = w, x_t = x]$, where the vector inequality involving x_{t+1} is interpreted in the obvious elementwise sense and the primes denote values for "one period hence," not the transpose operator, with the distinction obvious from the context in subsequent expressions. Thus, given this Markovian structure, the values w and x when the event $\{w_t = w$ and $x_t = x\}$ occurs completely characterize the state of the system, so the equilibrium price dividend ratios will be functions $v_i(w, x)$ of w and x. These functions are the solutions to the following set of integral equations:

$$v_i(w, x) = \beta \int [1 + v_i(w', x')]$$
$$\times (w')^{-\gamma} x_i' \, dF(w', x' \mid w, x),$$
$$i = 1, 2, \ldots, M. \quad (3.4)$$

Under regularity conditions analogous to those given by Lucas (1978), these "asset pricing" equations admit unique positive solutions for the price dividend ratios.

Clearly, this device of reexpressing the asset pricing equations in terms of the price dividend ratios and consumption and dividend growth relies in an essential way on the presumed homotheticity of preferences for consumption across periods that is implicit in the use of the constant relative risk aversion utility function. The homotheticity assumption, however, is crucial to avoid degeneracies when introducing nonstationarities in a representative agent framework.

3.2 The Discrete State-Space Model

Generation of the pseudodata for the sampling experiments requires solution of the integral equations (3.4) for the price dividend ratios. If the state variables of the economy are continuous-valued random variables, however, the integral equations admit explicit solutions (or solutions that are feasible to approximate

in Monte Carlo work) only under very special circumstances, and these special circumstances constitute a far too narrow set of possibilities for conducting interesting sampling experiments. Hence, a discrete state-space approach is adopted. This approach is motivated in part by Mehra and Prescott (1985), who used a very coarse two-state model in their study of the "equity premium." Specifically, the laws of motion for consumption growth and dividend growth are taken to be a finite-space Markov chain. With a finite-state Markov chain, as with the data-generation process, the solutions of the integral equations are simply the solutions to a large set of linear simultaneous equations. The parameters of the discrete Markov chain process are determined in such a way that the statistical properties of realizations of the Markov chain approximate closely the properties of an underlying continuous-valued autoregression for the state variables. Perhaps the primary advantage of using an underlying autoregression to calibrate the Markov chain is that it then becomes straightforward to use empirical estimates from fitted autoregressions and other information to assign interesting and meaningful values to the laws of motion of the state variables. Theorem 3.1 in Section 3.3 justifies the methodology by verifying that with a sufficiently fine discrete state space the results of sampling experiments become arbitrarily close to what would be obtained with a continuous state space.

The notation for the discrete state space model is as follows. Let there be N states of nature indexed by the variable s, and let $\bar{w}(s)$ and $\bar{x}_i(s)$ denote, respectively, the values of consumption growth and dividend growth for the ith asset in state s. Let $\bar{x}(s)$ denote the $M \times 1$ vector of dividend growths in state s. The transition probabilities will be denoted by

$$\pi(s, s') = \Pr[w_{t+1} = \bar{w}(s'),$$

$$x_{t+1} = \bar{x}(s') \mid w_t = w(s), x_t = \bar{x}(s)].$$

Thus in this notation the $(M + 1) \times 1$ vector (w_t, x_t) of consumption and dividend growths follows a discrete Markov chain with transition probability matrix $\Pi = [\pi(s, s')]$. Furthermore, the integral equations (3.4) for the price dividend ratios are written as

$$\bar{v}_i(s) = \beta \sum_{s'=1}^{N} \pi(s, s')[1 + \bar{v}_i(s')]w(s')^{-\gamma}x_i(s'), \quad (3.5)$$

where $\bar{v}_i(s)$ denotes the value of the ith price dividend ratio in state of nature s. These equations are simply a system of linear equations for the $\bar{v}(s)$ that can be solved fairly quickly in practice even when there is a rather large number of states, as is the case in multivariate situations.

I now describe the method used to generate the Markov chain for consumption and dividend growth. The basic idea is to start with a continuous-valued autoregression and then take a discrete approximation to it.

Specifically, consider the autoregression

$$\log(x_{it}) = b_i + \sum_{l=1}^{M} a_{il}\log(x_{l,t-1}) + a_{i,S}\log(w_{t-1}) + \varepsilon_{i,t},$$

and $\qquad\qquad\qquad i = 1, 2, \ldots, M,$

$$\log(w_t) = b_S + \sum_{l=1}^{M} a_{Sl}\log(x_{l,t-1}) + a_{S,S}\log(w_{t-1}) + \varepsilon_{S,t},$$

where the b_i are constants, the a_{ij} are autoregressive coefficients, $S = M + 1$, and the ε_{it} constitute an $S \times S$ vector white noise process with variance–covariance matrix denoted by Ω. The method of taking a discrete approximation can actually handle autoregressions with lag lengths of any finite order, although throughout this article I will use only first-order autoregressions. A more compact way to write this last system is

$$y_t = b + Ay_{t-1} + \varepsilon_t, E_{t-1}[\varepsilon_t] = 0, \text{var}(\varepsilon_t) = \Omega, \quad (3.6)$$

where y_t is an $S \times 1$ vector comprised of the $\log(x_{it})$ and $\log(w_t)$. Here we think of the ε_{it}, and hence the y_{it}, as having a continuous distribution over R^S or some subset of it. One should keep in mind that the term *continuous-valued* here pertains to the characteristics of the range space of the random variables ε_{it} and y_{it}, not to the time index. That is, time is taken as discrete with a fixed interval length; the range space of the ε_{it} and y_{it} is continuous, and the range space of the approximating Markov chain is discrete.

Tauchen (1985b) developed a method for finding a Markov chain that approximates well the statistical properties of the autoregressive (AR) model (3.6). An overview of the method follows. The first step is to transform linearly the system (3.6) to one with no intercept and a diagonal variance–covariance matrix for the innovations. The transformation is

$$y_t \longrightarrow B[y_t - (I - A)^{-1}b], \quad (3.7a)$$

$$A \longrightarrow F = BAB^{-1}, \quad (3.7b)$$

where B is a lower triangular matrix such that $\Sigma = B\Omega B^{-1}$ is a diagonal matrix. This new system is written as

$$y_t = Fy_{t-1} + \eta_t, \quad \text{var}(\eta_t) = \Sigma(\text{diagonal}), \quad (3.8)$$

where y_t now denotes the variables after the transformation (3.7). It is assumed that the components of η_t are jointly normally distributed and thus are mutually independent. The assumption of normality is for convenience only and is not essential to the method. It is also assumed that the system (3.8) is stationary and thus the matrix equation $\Sigma^* = F\Sigma^*F' + \Sigma$ has a unique solution for the variance–covariance matrix Σ^* in the stationary distribution of y_t. The square root of the ith diagonal element of Σ^* will be denoted by σ_i^*.

A useful notational convention will be to employ a tilde to indicate the discrete-valued Markov process $\tilde{y}_t = (\tilde{y}_{1t}, y_{2t}, \ldots, \tilde{y}_{St})'$ that is the approximating process

for y_t in (3.8) and to employ an overbar to indicate the values $\bar{y}_i(k)$ that $\bar{y}_{it}$ assumes. The values $\bar{y}_i(k)$ are taken as equispaced over an interval $[-P\sigma_i^*, P\sigma_i^*]$ that extends for P standard deviations σ_i^* in each direction from zero, where P is a small number. (The value $P = 3$ has been found to work well in practice.)

Thus if K_i values are taken for the range of $\bar{y}_{it}$, then the kth is

$$\bar{y}_i(k) = \Delta_i[k - (K_i + 1)/2],$$

where $\Delta_i = 2P\sigma_i^*/(K_i - 1)$. Any possible state of the approximating process is an event of the form

$$[\bar{y}_t = (\bar{y}_1(k_1), \bar{y}_2(k_2), \ldots, \bar{y}_S(k_S))],$$

and there are $N = K_1 \cdot K_2 \cdots K_S$ such events in total. Let the integer s be a label for these events so that each number s corresponds to an arrangement $(k_1, k_2, \ldots, k_S)$ of integers with $1 \le k_i \le K_i$ for $i = 1, 2, \ldots, S$.

The transition probabilities $\Pi(s, s') = \Pr[\bar{y}_{t+1}$ in state $s' \mid \bar{y}_t$ in state $s]$ are assigned as follows. Let $(k_1, k_2, \ldots, k_S)$ and $(k_1', k_2', \ldots, k_S')$ be the arrangements corresponding to s and s', respectively, and put

$$\underset{S \times 1}{\mu} = A \begin{bmatrix} \bar{y}_1(k_1) \\ \cdot \\ \cdot \\ \cdot \\ \bar{y}_S(k_S) \end{bmatrix} \tag{3.9}$$

The components μ_i of μ are the expected values of the continuous-valued process $y_{i,t+1}$ conditional on $y_t = (\bar{y}_1(k_1), \ldots, \bar{y}_S(k_S))'$. The idea is to center the density of the innovation $\eta_{i,t+1}$ over μ and then assign the probability of getting $\bar{y}_{i,t+1} = \bar{y}_i(k_i')$ to be the probability mass that is within $\pm\Delta_i/2$ of $\bar{y}_i(k_i')$, with a suitable "fixup" at the endpoints. Specifically, put

$$h(s, k_i') = \Pr\left[\bar{y}_i(k_i') - \frac{\Delta_i}{2} < \mu_i + \eta_i \le \bar{y}_i(k_i') + \frac{\Delta_i}{2}\right]$$

$$\text{if } 2 \le k_i' \le K_i - 1$$

$$= \Pr\left[\bar{y}_i(K_i) - \frac{\Delta_i}{2} < \mu_i + \eta_i\right] \quad \text{if } k_i' = K_i$$

$$= \Pr\left[\mu_i + \eta_i \le \bar{y}_i(1) + \frac{\Delta_i}{2}\right] \quad \text{if } k_i' = 1,$$

where η_i is a random variable with the same distribution as the innovation $\eta_{i,t+1}$, and $h(s, k_i')$ depends on s through the dependence of μ_i on s in (3.9). Then $h(s, k_i')$ is taken to be the probability that $\bar{y}_{i,t+1} = \bar{y}_i(k_i')$ given $\bar{y}_t$ in state s. By the conditional independence that is achieved via the transformation (3.7), the transition probability can be taken as

$$\pi(s, s') = \prod_{i=1}^{S} h(s, k_i').$$

The final step of the method is to reverse the transformation (3.7) so that the discrete system corresponds

to (3.6) with intercepts and a nondiagonal covariance matrix. This only affects the values assumed by the approximating process with no effect on the transition probabilities.

The algorithm used to generate realizations $\{\bar{y}_1, \bar{y}_2, \ldots, \bar{y}_T\}$ of the Markov process is as follows. Put $\bar{y}(s) = (\bar{y}_1(k_1), \ldots, \bar{y}_S(k_S))'$ for $s = 1, 2, \ldots, N$, where as before the arrangement of integers $(k_1, \ldots, k_S)$ corresponds to the state labeled s. The recursive scheme is

1. Let s = previous state; that is, $\bar{y}_{t-1} = \bar{y}(s)$.
2. Draw u uniform from $[0, 1]$.
3. Let the next state s' be the smallest number such that $\pi(s, 1) + \pi(s, 2) + \cdots + \pi(s, s') \ge u$.
4. Set $s' \rightarrow s$ and $\bar{y}_t = \bar{y}(s')$. Return to step 1 if t is less than T; otherwise, terminate.

To initialize the algorithm, the initial state is drawn from the stationary distribution of the Markov chain in a manner analogous to steps 2 and 3.

3.3 Theoretical Justification

Variables such as asset prices, dividends, consumption, and so forth are naturally modeled as continuous-valued random variables. This is obviously the case despite the fact that, because of the structure of the monetary system and the manner in which asset prices are quoted, the measurements of these variables are actually discrete-valued. The range spaces of the measurements are so fine that knowledge of their discreteness is not operational information; one obtains a better approximation by thinking of them as being continuous. In this article I therefore do not view the discrete state-space approach as an effort to model the extremely fine state space of the measured data, because this is evidently impossible. Instead, the discrete state-space approach is used to get good numerical approximations in the sampling experiments compared with what would be obtained if the continuous case could be simulated.

The general approach of using discrete methods to obtain approximations to the solutions of integral equations was developed by Kantorovich, and a good overview was given by Wouk (1979, chaps. 7 and 8). The specific method used here is the method of solution by numerical quadrature. The method is based on the one given by Wouk (1979, pp. 149–156), but with a different but analogous integration rule. The conditional distribution of the Markov process is the *approximate kernel*, and the space R^M containing the solutions to the system of linear equations (3.5) is the *numerical space*. An attractive feature of this particular method and its implementation here is that it not only gives approximate solutions to the integral equations, it also provides a probability model in the form of the Markov process suitable for generating pseudodata for the sampling experiments. Other methods, such as collocation, gener-

ally do not also provide a probability model along with the approximate solution.

It is apparent from the construction of the approximating Markov process that the joint distribution of realizations of the Markov process converges to that of the underlying autoregression. Specifically, let $\tilde{Y}_{KT} = (\tilde{y}_{K1}, \tilde{y}_{K2}, \ldots, \tilde{y}_{KT})$ denote generically a realization of length T of the Markov process when a "grid size" of K is used. A suitable index of the grid size is the smallest of the number of discrete points used for the components of the $\tilde{y}$'s; that is, $K = \min\{K_i\}$. Then by construction $\text{law}(\tilde{y}_{KT}) \rightarrow \text{law}_{\text{AR}}$ as $K \rightarrow \infty$, where law_{AR} denotes the joint distribution of a realization of length T from the underlying AR process.

By virtue of the Skorohod construction (Serfling 1980, p. 23; Breiman 1968, p. 283), we can always identify weak convergence with almost sure convergence. Skorohod's result states that given the weak convergence of $\tilde{Y}_{KT}$ to Y_T, where $\text{law}(Y_T) = \text{law}_{\text{AR}}$, then in general there exist random variables $\tilde{Y}_{KT}^*$ and Y_T^* defined on a common probability space with $\text{law}(\tilde{Y}_{KT}^*) = \text{law}(\tilde{Y}_{KT})$ and law $(Y_T^*) = \text{law}_{\text{AR}}$ such that $\tilde{Y}_{KT}^* \rightarrow Y_T^*$ almost surely. This has some useful implications. Inspection of the proof of Skorohod's result (Breiman 1968, pp. 283–284) reveals that the method used in the proof to define the $\tilde{Y}_{KT}^*$ is the same as that used here to generate the Monte Carlo realizations of the Markov process. Thus I can take $\tilde{Y}_{KT}^* = \tilde{Y}_{KT}$; that is, the Monte Carlo realizations are just the random variables of the Skorohod construction and hence $Y_T^* = Y_T$. Therefore, any continuous function of the data, say $\Psi(\tilde{Y}_{KT})$, converges almost surely to $\Psi(Y_T)$. Moreover, the Monte Carlo realizations of the estimator $\hat{\theta}$ are also almost surely convergent.

Theorem 3.1. Let $\hat{\theta}_K = \hat{\theta}(\tilde{Y}_{KT})$ denote generically a realization of the Monte Carlo estimator of the quasi-true value of the parameter when a grid of size K is used. Then $\hat{\theta}_K \rightarrow \hat{\theta}(Y_T)$ almost surely as $K \rightarrow \infty$.

In cases in which the objective function has multiple minimums, the estimator $\hat{\theta}_K$ is interpreted to be a measurable selection from the set minimizers and the almost sure convergence is interpreted as meaning that lim sup of the set of minimizers is contained in the limiting set of minimizers. To sketch a proof, observe that $\hat{\theta}_K$ is a minimizer of the GMM objective function when the realization of the data is $\tilde{Y}_{KT}$. But this objective function converges almost surely as $K \rightarrow \infty$ to the GMM objective function when the realization is Y_T, uniformly in θ on compact sets, and the conclusion follows by standard arguments. It is to be emphasized that the randomness here pertains to the random numbers generated in the sampling experiments and the limit is taken with respect to the grid size K, not with respect to T.

To interpret the result, suppose that the limiting objective function has a unique minimizer, which in general is true when T is large. Then the result implies that for almost every realization of T drawings from the uni-

form random number generator used in the data-generating algorithm (see the end of Sec. 3.2), there is a grid size K^* such that if L_1, $L_2 \geq K^*$, then $|\hat{\theta}_{L_1} - \hat{\theta}_{L_2}| \leq \varepsilon$, where ε is arbitrary. Here $\hat{\theta}_{L_1}$ and $\hat{\theta}_{L_2}$ refer to estimates based on pseudodata obtained with grid sizes L_1 and L_2 but generated from the same T uniform random variables in the algorithm. In other words, except on a null set, for each group of T uniform u's we have to find that after K becomes large enough, then experimenting with finer grid sizes does not matter much.

4. RESULTS

To be in compliance with the standards set forth by Hoaglin and Andrews (1975) for documenting Monte Carlo work, Section 4.1 gives a detailed description of the experimental design and numerical methods. Section 4.2 contains some additional documentation and an accounting of the main findings from the experiments.

4.1 Experimental Design and Methods

The experiments reported here all pertain to a case in which the econometrician has data on returns from holding one asset and on per-capita consumption. The underlying autoregressive model for the period-to-period proportional growth in the dividend, $x_t = d_t/d_{t-1}$, and the growth in consumption, $w_t = c_t/c_{t-1}$, is

$$\log(x_t) = b_1 + a_{11}\log(x_{t-1}) + a_{12}\log(w_{t-1}) + \varepsilon_{1t}$$

$$\log(w_t) = b_2 + a_{21}\log(x_{t-1}) + a_{22}\log(w_{t-1}) + \varepsilon_{2t},$$

$$(4.1)$$

where ε_{1t} and ε_{2t} are jointly normally distributed white noise random variables with second moments ω_1^2, ω_2^2, and ω_{12}.

Using the method described in the previous section, a discrete-valued Markov chain $\{\tilde{x}_t, \tilde{w}_t\}$ was taken as an approximation to the system (4.1). In the approximation each of $\tilde{x}_t$ and $\tilde{w}_t$ could assume one of eight values giving 64 states of nature in total. For each of the experimental settings of the pseudotrue values of the discount factor β and the curvature parameter γ, the asset's price dividend ratio $v(s)$ is computed as a function of the state of nature s by solving the system of linear equations (3.5). Realizations of the discrete process $\{\tilde{x}_t, \tilde{w}_t\}$ were generated using the method described in Section 3.2, and the implied price dividend ratio $\tilde{v}_t$ was calculated concurrently. Extensive trial testing with the discrete approximation to (4.1) showed that over the ranges of the a's, b's, and ω's that are used here, the discrete approximation is exceptionally good—nearly exact. In what follows, think of the actual data-generation process as being (4.1), with little reference to the approximating discrete model, although to avoid possible confusion tildes will appear over the relevant quantities in formulas involving the pseudodata.

The notation for the description of the estimation method is as follows: Define the error function

$$e_t(\theta) = \beta R_{t+1} \bar{w}_{t+1}^{-\gamma} - 1, \tag{4.2}$$

where $\theta = (\beta, \gamma)$ and where

$$R_{t+1} = [(1 + \bar{v}_{t+1})/\bar{v}_t] \bar{x}_{t+1} \tag{4.3}$$

is the asset's gross return from t to $t + 1$ per dollar invested in it. As can be easily checked, the expression (4.2) for the error function is equivalent to the expression (2.2). Now define the instrument vector

$$z_t = (1, \bar{w}_t, \ldots, \bar{w}_{t-L+1}, R_t, \ldots, R_{t-L+1})', \tag{4.4}$$

where $L = \mathrm{NLAG} \geq 1$ denotes the number of lags used to form the instrument vector. The use of lags of consumption growth and asset returns as instruments is consistent with current practice in applied work (see Hansen and Singleton 1982). Put

$$g_t(\theta) = e_t(\theta) z_t \tag{4.5}$$

and

$$\bar{g}_T(\theta) = \left(\frac{1}{T}\right) \sum_{t=1}^{T} g_t(\theta). \tag{4.6}$$

In all cases sufficient data were generated to "prime" the lags appropriately so that reported econometric sample sizes of $T = 50$ and $T = 75$ correspond to exactly that many terms in the sum in (4.6).

The estimation method to be investigated is the Han-

sen (1982) two-step procedure. In the first step a preliminary estimate $\hat{\theta}^{(1)}$ is obtained by minimizing the quadratic form

$$Q_T^1(\theta) = \bar{g}_T(\theta)' W_T^1 \bar{g}_T(\theta), \tag{4.7}$$

where W_T^1 is a matrix of preliminary weights comprised of the inverses of the diagonal elements of the matrix $(1/T) \sum_{t=1}^{T} g_t(\theta) g_t(\theta)'$ evaluated at the starting values for the nonlinear minimization of $Q_T^1(\theta)$. The purpose of using the preliminary weighting matrix W_T^1 instead of an identity matrix is to keep all calculations across experiments on roughly the same numerical scale during the nonlinear optimization. Given the estimate $\hat{\theta}^1$ from the first-step estimation, the weighting matrix

$$W_T^2 = \left[\left(\frac{1}{T}\right) \sum_{t=1}^{T} g_t(\hat{\theta}^1) g_t(\hat{\theta}^1)' \right]^{-1} \tag{4.8}$$

is formed and the function

$$Q_T^2(\theta) = \bar{g}_T(\theta)' W_T^2 \bar{g}_T(\theta) \tag{4.9}$$

is minimized to obtain the final estimate $\hat{\theta}$. Estimates of the variance–covariance matrix of $\hat{\theta}$ were calculated as

$$T^{-1}[(\partial \bar{g}(\hat{\theta})'/\partial \theta) \, W_T^2 \, (\partial \bar{g}(\hat{\theta})/\partial \theta')]^{-1}. \tag{4.10}$$

This expression takes account of conditional heteroscedasticity.

The primary computer used for the calculations was an IBM 3081 mainframe. The nonlinear optimization

Table 1. Pseudotrue Parameter Values and Other Parameter Settings

Experi-ment	Gamma	Beta	Autoregressive matrix		Error covariance		Intercepts
1A	.30	.97	−.500	.000	.01	.00	.000
			.000	−.500	.00	.01	.000
2A	.30	.97	−.500	.000	.0001	.00	.000
			.000	−.500	.00	.01	.000
3A	.30	.97	−.010	.000	.01	.00	.000
			.000	−.500	.00	.01	.000
4A	.30	.97	−.010	.000	.0001	.00	.000
			.000	−.500	.00	.01	.000
5A	.30	.97	−.500	.000	.01	.00	.000
			.000	−.100	.00	.01	.000
6A	.30	.97	−.500	.000	.0001	.00	.000
			.000	−.100	.00	.01	.000
7A	.30	.97	−.100	.000	.01	.00	.000
			.000	−.100	.00	.01	.000
8A	.30	.97	−.100	.000	.0001	.00	.000
			.000	−.100	.00	.01	.000
9A	.30	.97	.073	.620	.013	$r = .49$	.0034
			.015	−.122	$r = .49$	.002	.0220
10A	.30	.97	.070	.000	.01	.00	.000
			.000	−.120	.00	.002	.000
11A	.30	.97	.070	.600	.01	.00	.000
			.000	−.120	.00	.002	.000
12A	.30	.97	.070	.000	.01	$r = .50$	.000
			.000	−.120	$r = .50$	.002	.000
13A	.30	.97	.070	.000	.01	.00	.000
			.000	−.120	.00	.002	.020

NOTE: Experiments 1B–13B are the same as 1A–13A except gamma = 1.30.

was carried out in double precision FORTRAN using the DFP algorithm in the Goldfeld–Quandt package. The coding of the FORTRAN routines was checked by replicating the estimation on test data sets using PROC SYSNLIN in the Statistical Analysis System (SAS) package. In addition, extensive checking was carried to ensure convergence of the optimization routine. The random number routine for generating realizations of the Markov chain was the GGUBFS in the International Mathematical and Statistical Libraries (IMSL) package. The initial calculations to determine the parameters of the approximating Markov chain and the lower bounds for the variance–covariance matrix of the estimates were done in PROC MATRIX in SAS, and most of the postprocessing of data from the sampling experiments was done in SAS.

There are 26 sampling experiments labeled 1A–13A and 1B–13B. Table 1 displays the pseudotrue values of the parameters of the representative agent's utility function for the experiments. In Experiments 1A–13A the curvature parameter γ for per-period utility function

$$u(c) = (1 - \gamma)^{-1} c^{1-\gamma} \tag{4.11}$$

is .30, whereas in Experiments 1B–13B it is 1.30. Mehra and Prescott (1985) reviewed the literature on estimates of this parameter and concluded that "sensible" values for γ should be less than 2.0; this conclusion motivates the choices here. For all experiments the subjective discount factor β is set to .97, which corresponds to a subjective discount rate of about .03. Although the experimental results are not invariant to the choice of β, extensive preliminary calculations indicated that the lower bound for the asymptotic standard deviations of the estimated parameters are virtually unaffected by the choice of β over a very wide range of values (from .40 to .99), and hence this parameter is left fixed across all experiments.

Table 1 also displays the experimental settings for the parameters of the data-generation process (4.1) for dividend and consumption growth. Since the parameter settings are the same for the A and B groups, the letter suffix will be omitted in the subsequent discussion whenever the distinction between the two groups of experiments is irrelevant. Experiments 1–8 are basically test cases in which the autoregressive matrix is diagonal, as is the covariance matrix of the innovations. Although the asset's dividend evolves independently of consumption in these cases, the rate of return on this asset is strongly correlated with consumption, as is clear from inspection of formula (3.1) for the asset's price. In these eight experiments the variance of the innovation ω_2^2 in the consumption growth equation is set to .01, whereas the variance of the innovation to dividend growth is set to .01 or .0001. The experimental results depend only on the ratio of these variances and not on their individual levels. Both consumption growth and dividend growth are negatively autocorrelated in these experiments, so

the levels of each variable tend to regress toward zero after an innovation. In Experiments 1 and 2 the degree of autocorrelation is unrealistically strong (AR coefficients of $-.50$). This choice of the AR coefficients generates very small standard asymptotic standard errors for the coefficient estimates, because the magnitude of the standard errors is inversely related to the magnitude of the AR coefficients. In these two experiments the parameters can thereby be expected to be estimated precisely and the asymptotic approximations should be fairly good, thus providing a check on the validity of the overall calculations, including the coding of the routines to evaluate the function and its derivatives, the nonlinear optimization package, and the random number generator. In Experiments 3–8 the degree of autocorrelation in the two series becomes progressively weaker, with the values $-.10$ for the AR coefficients in Experiments 7 and 8 being much more realistic.

The parameter settings for Experiment 9 were determined on the basis of the following autoregression fitted with actual U.S. data on consumption and dividend growth:

$$\log(x_t) = \begin{array}{l} .003 \\ (.013) \end{array} + \begin{array}{l} .073 \\ (.115) \end{array} \log(x_{t-1})$$
$$+ \begin{array}{l} .620 \\ (3.44) \end{array} \log(w_{t-1}), \quad \hat{\omega}_1^2 = .0130$$

$$\log(w_t) = \begin{array}{l} .022 \\ (.005) \end{array} + \begin{array}{l} .015 \\ (.040) \end{array} \log(x_{t-1})$$
$$- \begin{array}{l} .122 \\ (.119) \end{array} \log(w_{t-1}), \quad \hat{\omega}_2^2 = .0016. \tag{4.12}$$

The residual correlation is .492. When the autoregression was estimated in levels instead of first differences, the diagonals elements of the matrix of AR coefficients were very close to unity, indicating that a model with stationary first differences is an appropriate specification. The data set for the autoregression is similar to that used by Mehra and Prescott (1985), although the data were independently collected. The dividend series is from the Standard and Poors index. The consumption deflator was used to convert to constant dollars the dividends and per-capita consumption series. The data are annual with the period of fit being 1980–1982. Annual data are used for model calibration here, because they make more plausible the assumption of a first-order autoregression for the state variables, and a first-order autoregression is needed to keep the number of states in the approximating Markov chain small enough to be computationally feasible. It is recognized that much of the work with the GMM procedure applied to financial market data uses quarterly or monthly data, so the conclusions here need to be qualified accordingly. Nevertheless, much can be learned about the properties of the procedure by studying its properties as applied to data generated by the aforementioned autoregression.

Although the parameter settings for the AR model

in Experiment 9 are based on actual data and thus are arguably more "realistic" than those of Experiments 1–8, the results for this experiment are a confounding of the effects of three characteristics of Experiment 9 that are not present in Experiments 1–8. These characteristics are (a) the feedback from consumption growth to dividend growth, (b) the contemporaneous correlation between the innovations, and (c) the nonzero intercepts. Experiments 10–13 are designed to disentangle the separate effects of these characteristics. Experiment 10 is the "base" case for this effort in which the autoregressive coefficients and the innovation variances are set to the diagonal elements of those for Experiment 9 after some inconsequential rounding. In Experiments 11–13 each of these effects is introduced in turn. Experiment 11 has feedback from consumption growth to dividend growth, Experiment 12 has contemporaneous correlation between the innovations, and Experiment 13 has a nonzero intercept in the consumption growth equation (note the very small and insignificant intercept in the dividend growth equation that is ignored in what follows).

4.2   Findings From the Sampling Experiments

The following are some pertinent features of the experiments. There are 100 replications within each experiment using econometric sample sizes of 50 and 75. (The qualitative conclusions from the more important experiments 1, 7, and 9–13 were validated by repeating them with 500 and 1,000 replications, and no essential differences were uncovered.) For each sample size the parameter estimates and associated statistics were obtained using NLAG = 1 through NLAG = 4 to form the instrument vector (4.4) and using the "optimal" instruments discussed in Section 2.2. These last instruments are based on knowledge of the underlying structure and hence are not operational for applied work, although the results based on them do provide a benchmark. In many instances, however, the objective function was very ill-behaved when these instruments were used in conjunction with sample sizes of 50 or 75, and valid results were often not available (see Sec. 4.2.1). Finally, the experiments are "blocked" in the sense that for each econometric sample size within an experiment the estimation using the different lag lengths and the "optimal" instruments was over exactly the same data. This variance reduction technique induces a strong correlation across results with different lag lengths, so the effects of varying the lag lengths are estimated relatively precisely. Because of the extreme nonlinearity of the structural model generating the data and because of the nonlinearity of the estimation method, however, there is a lack of strong prior information, precluding use of other standard variance reduction techniques. Such information is essential to achieve variance reduction (see Rubinstein 1981).

The GMM method uses lagged endogenous variables as instruments, and there are strong nonlinearities inherent in the structure and the estimation method. In addition, the restriction that the estimator must lie in a compact set, though common in theoretical work, is rarely operational in practice and is not imposed in the sampling experiments. It seems likely then that the Hansen two-step estimator does not possess moments. Thus, in the reporting of the results, care will be taken to report statistics like medians and interquartile ranges that always exist whether or not the estimator has moments. In addition, the natural moment-based statistics are also reported and, as it turns out, "infinite moment" problems generally are not apparent, which is consistent with Sargan (1982).

4.2.1   General Characteristics of the Sampling Distribution of $\hat{\gamma}$.

Table 2 shows the lower bound on the asymptotic standard error and summary statistics on the sampling distribution of the estimate of the curvature parameter γ. The lower bound was computed as the square root of $(1/T)$ times the appropriate diagonal element of the lower bound on the asymptotic variance-covariance matrix given in (2.12). The lower bound and summary statistics are reported only for Experiments 1, 7, and 9. The reason for selecting only 1 and 7 from the first eight for reporting in detail is that for the odd-numbered experiments in this group the experimental conditions are somewhat more stringent for the estimator, and thus more interesting, because of the larger innovation variance for dividend growth and because the general characteristics of the sampling distributions show a gradual progression between the extremes of 1–2 and 7–8 as the series become less autocorrelated. Experiments 10–13 will be discussed subsequently.

Examination in Table 2 of the sampling distribution of $\hat{\gamma}$ in Experiments 1A and 1B indicates that, under the rather favorable conditions of these experiments, the estimator performs reasonably well. The medians are quite close to the true values and the other quantiles are more or less symmetrically placed about the true parameter values. The lower bound for the asymptotic standard deviations for the estimator range between .173 and .271, depending on the number of observations and the parameter settings of the experiments. Generally the interquartile ranges are right around their anticipated values of 1.35 times this standard error. The right two columns of Table 2 contain the means and standard deviations of the estimates computed over the 100 replications. The means are very close to the medians and the standard errors are close to the theoretical lower bound, indicating that full asymptotic efficiency is achieved using very short lag lengths under the conditions of these experiments. Although these conditions are perhaps too favorable to make it possible to draw firm conclusions about the estimator's performance in actual practice, the results do suggest that there are no fundamental deficiencies in the GMM procedure itself nor in the methods used here to study its properties.

Table 2. Characteristics of the Sampling Distribution of Estimated Gamma

	Parameters			Quantiles					Other Statistics		
Exp.	Nobs	NLAG	MSTD	$Q_{.01}$	$Q_{.05}$	Med.	$Q_{.95}$	$Q_{.99}$	IQR	Mean	SD
1A	50	1	.173	−.22	−.05	.26	.66	1.01	.232	.275	.215
	50	2	.173	−.14	−.01	.25	.57	.97	.212	.259	.179
	50	3	.173	−.15	−.02	.24	.50	.89	.211	.242	.170
	50	4	.173	−.14	−.03	.22	.51	.86	.180	.226	.161
	50	*	.173	−.21	−.04	.28	.73	1.15	.203	.287	.221
	75	1	.141	−.11	.06	.29	.62	.86	.218	.309	.168
	75	2	.141	−.14	.05	.26	.56	.75	.209	.283	.161
	75	3	.141	−.14	.03	.25	.54	.76	.212	.267	.156
	75	4	.141	−.08	.02	.24	.55	.78	.211	.258	.157
	75	*	.141	−.07	.03	.30	.65	.82	.217	.319	.168
1B	50	1	.271	.70	.93	1.25	1.87	2.33	.339	1.306	.289
	50	2	.271	.45	.80	1.17	1.71	2.04	.301	1.192	.259
	50	3	.271	.62	.75	1.09	1.52	1.76	.247	1.106	.219
	50	4	.271	.41	.76	1.06	1.37	1.75	.282	1.050	.203
	50	*	.271	.94	.99	1.32	1.99	6.72	.431	1.426	.628
	75	1	.221	.92	1.00	1.26	1.81	3.21	.308	1.342	.310
	75	2	.221	.83	.94	1.22	1.77	2.35	.259	1.262	.259
	75	3	.221	.86	.90	1.17	1.53	2.09	.222	1.192	.196
	75	4	.221	.71	.87	1.13	1.44	1.76	.210	1.140	.176
	75	*	.221	.91	1.04	1.31	1.97	4.53	.321	1.393	.408
7A	50	1	1.134	−3.25	−1.70	.15	2.59	4.69	1.004	.238	1.187
	50	2	1.134	−1.52	−1.12	.06	1.15	1.89	.618	.106	.578
	50	3	1.134	−1.54	−.71	.07	.89	1.87	.578	.090	.502
	50	4	1.134	−1.63	−.65	.10	.66	1.00	.363	.049	.419
	50	*	1.134	NA							
	75	1	1.098	−4.13	−1.80	.09	1.38	2.38	.672	.056	.870
	75	2	1.098	−1.79	−.63	.10	1.13	2.50	.518	.101	.547
	75	3	1.098	−1.45	−.63	.06	.87	1.09	.447	.063	.418
	75	4	1.098	−.92	−.51	.08	.73	1.14	.432	.077	.369
	75	*	1.098	NA							
7B	50	1	2.119	−3.07	−1.42	.60	3.47	6.76	1.141	.662	1.382
	50	2	2.119	−1.85	−.83	.32	1.79	3.14	.785	.332	.758
	50	3	2.119	−2.45	−.79	.28	1.28	2.11	.651	.231	.624
	50	4	2.119	−1.02	−.61	.26	.93	1.19	.426	.234	.420
	50	*	2.119	NA							
	75	1	1.730	−2.34	−.79	.63	2.27	4.82	.858	.642	.911
	75	2	1.730	−1.23	−.42	.36	1.17	3.76	.685	.428	.596
	75	3	1.730	−.74	−.43	.35	1.19	1.83	.559	.361	.467
	75	4	1.730	−.87	−.25	.31	1.02	1.26	.465	.319	.405
	75	*	1.730	NA							
9A	50	1	4.257	−12.73	−2.09	1.62	5.97	27.71	3.258	1.748	3.917
	50	2	4.257	−3.97	−.09	2.04	4.53	6.69	2.219	1.932	1.683
	50	3	4.257	−2.06	−.88	2.23	4.44	5.36	1.707	2.092	1.419
	50	4	4.257	−1.63	−.15	2.03	4.12	4.47	1.566	2.015	1.201
	50	*	4.257	NA							
	75	1	3.476	−14.97	−8.20	1.37	5.91	10.19	2.997	.833	3.181
	75	2	3.476	−5.25	−1.47	1.72	4.40	10.78	2.354	1.684	2.049
	75	3	3.476	−2.11	−.67	1.88	4.34	5.14	1.696	1.921	1.393
	75	4	3.476	−1.41	−.38	1.83	3.88	6.10	1.356	1.884	1.208
	75	*	3.476	NA							
9B	50	1	3.647	−24.82	−2.78	1.94	9.05	31.66	2.715	2.082	5.308
	50	2	3.647	−8.18	−.52	2.08	5.19	9.42	1.830	2.005	1.961
	50	3	3.647	−1.28	.21	2.01	4.03	8.73	1.599	2.069	1.356
	50	4	3.647	−.12	.14	2.07	3.88	7.62	1.326	2.129	1.255
	50	*	3.647	NA							
	75	1	2.978	−9.63	−3.08	1.91	5.41	17.28	2.457	1.778	2.967
	75	2	2.978	−4.16	−.52	1.85	4.47	6.56	1.617	1.891	1.545
	75	3	2.978	−2.48	.05	2.19	3.90	6.11	1.565	1.990	1.222
	75	4	2.978	−.94	.37	1.97	3.64	5.65	1.294	2.023	1.034
	75	*	2.978	NA							

NOTE: Exp = experiment; Nobs = econometric sample size; NLAG = number of lags used to form instruments; MSTD = theoretical lower bound for asymptotic standard error; IQR = interquartile range; Mean = average of estimates over 100 replications; SD = standard deviation over 100 replications; NA = not available due to convergence problems.
* Asymptotically optimal instruments.

The sampling distributions for Experiments 7 and 9 in Table 2 are much more diffuse, which is to be expected given the larger lower bounds on the estimator's standard deviations. The larger lower bounds are due primarily to the reduction in the degree of autocorrelation in the consumption growth process [see Tauchen (1985c) for a more detailed discussion of the determinants of the magnitude of the asymptotic lower bound]. Because of the spreading of the distributions, the downward bias of the estimate that is present throughout Experiments 1–8 is more apparent in the results for 7A and 7B than in those for 1A and 1B. Interestingly, in Experiment 9, which is the one calibrated on the basis of actual data, the estimator is upward biased. The primary difference between Experiments 9 and 1–8 is that in the former dividend and consumption growth are positively correlated whereas in the latter they are independent, suggesting that positive association between consumption and dividend growth tends to counteract the downward bias and may produce upward bias. This conjecture will be explored more fully in Section 4.2.2.

Because of convergence problems, Table 2 does not contain information on the sampling distribution of the "optimal" instrument quasi-estimator in Experiments 7 and 9. In fact, for each of the experiments numbered 5 or higher the estimation algorithm either failed to reach a minimum or reached a putative minimum with implausibly large estimates of the variance–covariance matrix due to a nearly rank-deficient derivative matrix in about half of the 100 replications. As a check on the experimental methods, some of the experiments were repeated using these instruments and econometric sample sizes as large as 1,000 observations. With so many observations, the convergence problems were absent and the asymptotic approximations were very accurate. This suggests that the "optimal" instrument quasi-estimator might have very undesirable finite sample properties, except under special conditions or for unrealistically large sample sizes. It is interesting to note, however, that under the special conditions of Experiment 1, where reliable results are available for reporting in Table 2, the use of the optimal instruments tends to remove the bias and center the sampling distribution on the pseudotrue value.

4.2.2 The Variance/Bias Trade-Off.

All of the sampling distributions summarized in Table 2 indicate the presence of a strong variance/bias trade-off as the number of lags used for forming the instruments is increased from NLAG = 1 to NLAG = 4. As NLAG increases, the measures of dispersion decrease, and in many instances they decrease to values well below their asymptotic theoretical lower bounds. On the other hand, the magnitude of the bias almost uniformly increases with lag length.

Table 3 contains the relevant statistics for assessing the characteristics of the variance/bias trade-off and for identifying the determinants of the sign and magnitude

of the bias. The table shows two standardized measures of bias: $bias_1^0$ is the difference between the arithmetic average of the estimates and the true parameter value divided by the lower bound for the asymptotic standard deviation; $bias_2^0$ is the same except that the median is used in place of the average. The table also shows $RMSE^0$, which is the root mean squared error about the true value divided by the lower bound on the standard deviation, and MAD^0, which is the median absolute deviation about the true value divided by the same lower bound. One should note that the lower bound used for the standardization is not random, and thus the division does not introduce any additional noise into the Monte Carlo estimates of bias and estimation accuracy. Because of the standardization, the anticipated value for the $RMSE^0$ is unity or slightly above that and the anticipated value for the MAD^0 is $(2/\pi)^{1/2}$ or just above that.

Several conclusions about the variance/bias trade-off seem apparent from Table 3. For long lag lengths, either measure of bias can be up to 50% of the asymptotic standard error, and in a few instances it can be as large as 90% of the asymptotic standard error. On the other hand, for short lag lengths the bias is usually a much smaller fraction of the asymptotic standard error. This is the case in Experiments 1A and 1B, which are the "check" experiments with very favorable parameter settings, and it is also the case in the other experiments, where the conditions are much less favorable and more realistic. Generally speaking, under a strict mean squared error criterion the increase in the magnitude of the bias with lag length is associated with an improvement of the overall performance of the estimator as indicated by a reduction in $RMSE^0$, although this is not the case with the MAD^0. But nearly always the reduction in $RMSE^0$ beyond that achieved in moving from NLAG = 1 to NLAG = 2 is quite small. Of course the conclusions one would draw from these results depend on one's loss function. Nevertheless, it seems apparent that under any loss function that penalizes to some extent possible misleading inference due to a bias that is large in magnitude relative to the standard error, the general recommendation for applied work should be to keep the lag lengths short. The improvement in performance that is achieved from using longer lags would most likely be outweighed by the increased risk of incorrect inference.

The results of Table 3 also help identify the determinants of the direction of the estimator's bias. As previously noted, the estimator is downward biased in Experiments 1–8 whereas it is upward biased in Experiment 9. Experiments 10–13 help isolate the various effects that are confounded in Experiment 9. Comparing the A experiments with the B experiments for 9–13 shows that, other things equal, a larger true γ tends to move the bias downward, in the sense of making the negative biases more negative and reducing the positive biases. Comparing Experiment 10 with 11 and 12 indicates that the introduction of positive feedback from consumption

Table 3. Standardized Measures of Bias and Performance

Exp.	Nobs	NLAG	$Bias_1^0$	$RMSE^0$	$Bias_2^0$	MAD^0	Exp.	Nobs	NLAG	$Bias_1^0$	$RMSE^0$	$Bias_2^0$	MAD^0
1A	50	1	−.14	1.24	−.23	.69	10B	50	1	−.10	.84	−.15	.41
	50	2	−.23	1.06	−.29	.57		50	2	−.21	.54	−.18	.31
	50	3	−.34	1.04	−.35	.73		50	3	−.28	.49	−.30	.34
	50	4	−.43	1.02	−.46	.62		50	4	−.29	.40	−.27	.30
	50	*	−.08	1.27	−.11	.60							
	75	1	.06	1.24	−.07	.78		75	1	−.39	1.28	−.28	.44
	75	2	−.12	1.06	−.28	.79		75	2	−.35	.71	−.44	.54
	75	3	−.23	1.03	−.35	.79		75	3	−.40	.55	−.42	.43
	75	4	−.30	1.02	−.42	.85		75	4	−.40	.53	−.41	.41
	75	*	.13	1.27	.00	.70							
1B	50	1	.02	1.06	−.19	.66	11A	50	1	.20	.81	.11	.38
	50	2	−.40	1.03	−.48	.72		50	2	.14	.57	.06	.27
	50	3	−.71	1.08	−.77	.83		50	3	.10	.41	.11	.28
	50	4	−.92	1.19	−.89	.90		50	4	.09	.34	.07	.23
	50	*	.46	2.35	.07	.67							
	75	1	.19	1.40	−.18	.70		75	1	.09	.95	.16	.54
	75	2	−.17	1.18	−.36	.69		75	2	.10	.63	.11	.38
	75	3	−.49	1.01	−.59	.72		75	3	.17	.50	.15	.25
	75	4	−.72	1.07	−.77	.83		75	4	.12	.43	.11	.24
	75	*	.42	1.88	−.77	.64							
7A	50	1	−.05	.88	−.11	.41	11B	50	1	−.21	.76	−.22	.36
	50	2	−.14	.45	−.18	.24		50	2	−.24	.48	−.25	.34
	50	3	−.16	.40	−.17	.22		50	3	−.24	.40	−.23	.31
	50	4	−.19	.36	−.15	.20		50	4	−.23	.39	−.22	.26
	75	1	−.22	.82	−.19	.36		75	1	−.03	.80	−.08	.40
	75	2	−.18	.53	−.18	.28		75	2	−.22	.67	−.21	.38
	75	3	−.22	.44	−.22	.29		75	3	−.20	.49	−.22	.31
	75	4	−.20	.39	−.20	.27		75	4	−.18	.47	−.20	.28
7B	50	1	−.30	.72	−.33	.45	12A	50	1	.32	.95	.34	.51
	50	2	−.46	.58	−.46	.47		50	2	.29	.57	.30	.40
	50	3	−.50	.58	−.48	.48		50	3	.31	.50	.28	.37
	50	4	−.50	.58	−.49	.49		50	4	.28	.44	.30	.32
	75	1	−.38	.65	−.33	.46		75	1	.31	1.30	.37	.54
	75	2	−.50	.61	−.46	.54		75	2	.39	.68	.43	.49
	75	3	−.54	.60	−.48	.55		75	3	.35	.61	.36	.42
	75	4	−.56	.62	−.49	.56		75	4	.34	.52	.38	.44
9A	50	1	.34	.98	.31	.42	12B	50	1	−.12	.90	.04	.44
	50	2	.38	.55	.41	.42		50	2	−.04	.59	.02	.33
	50	3	.42	.54	.46	.46		50	3	−.03	.55	.03	.24
	50	4	.40	.49	.41	.41		50	4	.01	.36	.05	.21
	75	1	.15	1.10	.31	.51		75	1	.08	1.31	−.04	.50
	75	2	.40	.70	.41	.46		75	2	−.02	.62	−.01	.35
	75	3	.47	.61	.45	.48		75	3	−.03	.49	.04	.29
	75	4	.46	.58	.44	.45		75	4	−.02	.40	−.03	.27
9B	50	1	.21	1.46	.18	.34	13A	50	1	.16	1.20	−.01	.39
	50	2	.19	.57	.21	.31		50	2	−.04	.52	−.02	.31
	50	3	.21	.43	.19	.25		50	3	−.03	.42	−.02	.22
	50	4	.23	.41	.21	.25		50	4	−.04	.33	−.08	.23
	75	1	.16	1.00	.20	.48		75	1	−.09	1.17	−.09	.34
	75	2	.20	.55	.18	.30		75	2	−.08	.61	−.08	.34
	75	3	.23	.47	.30	.37		75	3	−.07	.44	−.13	.29
	75	4	.24	.42	.22	.30		75	4	−.09	.40	−.16	.27
10A	50	1	.08	.97	−.08	.44	13B	50	1	−.16	.83	−.26	.44
	50	2	.02	.58	−.09	.24		50	2	−.31	.56	−.40	.43
	50	3	−.02	.41	−.06	.20		50	3	−.25	.50	−.32	.36
	50	4	−.07	.33	−.10	.20		50	4	−.27	.39	−.31	.35
	75	1	−.01	1.15	−.02	.49		75	1	−.11	1.10	−.18	.45
	75	2	−.01	.69	−.09	.31		75	2	−.31	.61	−.23	.38
	75	3	−.02	.50	−.04	.25		75	3	−.35	.56	−.35	.36
	75	4	−.04	.38	−.03	.21		75	4	−.42	.55	−.39	.41

NOTE: Nobs = econometric sample size; NLAG = number of lags used to form instruments; $Bias_1^0$ = (average-true value)/(minimum asymptotic standard deviation); RMSE = (root MSE about true value)/(minimum asymptotic standard deviation); $Bias_2^0$ = (median-true value)/(minimum asymptotic standard deviation); MAD = (median absolute deviation about true value)/(minimum asymptotic standard deviation).
* Asymptotically optimal instruments.

growth to dividend growth or the introduction of positive correlation between the variables' innovations tends to raise the bias. On the other hand, comparing Experiment 10 with Experiment 13 shows that the nonzero intercept in the consumption growth equation has virtually no effect on the bias. Thus it is clearly the combined effects of the two sources of positive association between the dividend and consumption growth variables in the autoregression that produce the upward bias in Experiment 9, and the bias is less severe in 9B because of the countervailing influence of the larger underlying γ_0.

A useful way to summarize the findings on the variance/bias trade-off is to follow the recommendations of Hendry (1984) to report response surfaces fitted to the experimental results. The following display shows the relevant subsets of the coefficients from regressions of the standardized bias and the logarithm of the standardized RMSE on the design variables of the experiments:

$$\text{bias}_1^0 = -.096 - .173D_2 - .279D_3 - .361D_4$$
$$\quad\quad (.078) \quad (.054) \quad (.055) \quad (.068)$$

$$\quad\quad .150 + .372D_2\,\rho + .588D_3\,\rho + .729D_4\,\rho$$
$$\quad\quad (.125) \quad (.122) \quad (.124) \quad (.139) \quad ,$$

$$\log(\text{RMSE}^0) = -.683 - .242D_2 - -315D_3 - .338D_4$$
$$\quad\quad (.143) \quad (.061) \quad (.065) \quad (.077)$$

$$\quad\quad .655 - .648D_2\,\rho - .769D_3\,\rho - .990D_4\,\rho$$
$$\quad\quad (.199) \quad (.111) \quad (.125) \quad (.165) \quad ,$$

where the D_j are dummy variables for NLAG = 2, 3, or 4, and ρ is the correlation between the consumption growth and dividend growth innovations. Each regression is based on 208 observations and included as explanatory variables the other parameters of the underlying AR model and the econometric sample size T, the coefficients of which are not reported. The standard errors in parenthesis are "generalized" standard errors (White 1984, chap. 6) that take into account heteroscedasticity and error correlation due to the experimental blocking. As is clear from these regressions, when ρ = 0 the bias becomes larger in magnitude (more negative) at longer lag lengths and the variance decreases, with a net reduction in the RMSE. The interaction effect with ρ is also evident, since with positive ρ the bias increases with lag length and the reduction in the RMSE is exaggerated.

4.2.3 Implications for Inference.

One standard on which any estimation method should be evaluated is the reliability of the method's specification tests for the overall adequacy of the fitted model. For the GMM procedure, the restriction that the errors are uncorrelated with all variables that are measurable with respect to the information set I_t and in particular uncorrelated with the instruments used in the estimation is an overidentifying restriction that provides a means for performing an omnibus test of the specification of the model. The test statistic is the (suitably scaled) minimized value of the

objective function. Specifically, the statistic is $TQ_T^2(\hat{\theta})$, where the notation is given in Equation (4.9). The test rejects the specification when the statistic exceeds the critical chi-squared value with degrees of freedom equal to the number of orthogonality conditions used in the estimation—that is, the dimension of the vector $g_t(\theta)$ minus the number of estimated parameters.

Table 4 displays the rejection rates for this specification test for some of the more interesting sampling experiments. The table shows the rejection rates at nominal significance levels of 10% and 5%. Since the model is correctly specified in all experiments, the anticipated value for each rejection rate is the corresponding significance level. The experimental results clearly indicate that under the conditions here the test performs quite well, as there are very few statistically significant deviations from the anticipated values of the rejection rates. It is interesting to note that in most instances the rejection rates are below the nominal rates. This is consistent with the analytical results of Shapiro (1986), who showed that in a different though related estimation situation the exact marginal significance level approaches the nominal level from below as the sample size grows. (Shapiro showed specifically that the exact marginal significance level stays above the nominal level only finitely often as $T \to \infty$.) The experimental evidence presented here in conjunction with Shapiro's results indicates that, if anything, the test is slightly biased toward accepting the specification of the model. This suggests that the preponderance of rejections with this test that are observed in actual practice (see Rotenberg 1984) may not be due to an inadequacy of the asymptotic approximations, but rather may be construed as evidence against the model's specification.

Another standard for evaluating an estimation method is the reliability of its interval estimates of the parameters. Table 4 also contains statistics relevant for determining the accuracy of the confidence intervals constructed with the GMM procedure. The table shows specifically the proportion of times in the 100 replications that individual (one-dimensional) 95% confidence intervals for β and γ cover the underlying pseudotrue value. The confidence intervals were constructed using the standard errors implied by the estimate (4.10) for the estimator's asymptotic variance–covariance matrix. For short lag lengths, the intervals are evidently very reliable as the coverage rates are rarely statistically significantly different from the anticipated value of .95. For the longer lag lengths of NLAG = 3 or NLAG = 4, however, the intervals are clearly inaccurate, because they include the pseudotrue parameter value far too infrequently, and this is much more the case for the intervals for the curvature parameter γ.

Figures 1–3 provide a graphical summary of the accuracy of the interval estimates of γ and also further evidence on the characteristics of the variance/bias trade-off with lag length. The vertical height in each panel is the proportion of the replications that the 95% confi-

Table 4. Coverage Rates for 95% Confidence Intervals and Rejection Rates for the Test of the Overidentifying Restrictions

Exp.	Nobs	NLAG	Coverage Rates		Rejection Rates		Exp.	Nobs	NLAG	Coverage Rates		Rejection Rates	
			Beta	Gamma	10%	5%				Beta	Gamma	10%	5%
1A	50	1	.92	.93	.10	.02	10B	50	1	.96	.91	.08	.04
	50	2	.94	.92	.08	.03		50	2	.95	.85*	.05	.03
	50	3	.90	.93	.09	.03		50	3	.94	.74*	.07	.04
	50	4	.87	.92	.14	.01		50	4	.91	.76*	.07	.03
	75	1	.98	.90*	.10	.03		75	1	.97	.94	.05	.04
	75	2	.95	.90*	.12	.05		75	2	.95	.86*	.04	.02
	75	3	.91	.88*	.12	.05		75	3	.92	.78*	.07	.04
	75	4	.94	.86*	.09	.03		75	4	.91	.66*	.07	.03
1B	50	1	.95	.91	.11	.04	11A	50	1	.97	.99	.03	.01
	50	2	.93	.85*	.13	.07		50	2	.92	.99	.04	.02
	50	3	.87*	.72*	.10	.06		50	3	.91	.98	.06	.03
	50	4	.85*	.61*	.10	.01		50	4	.86	.93	.10	.01
	75	1	.96	.94	.11	.06		75	1	.98	.99	.08	.03
	75	2	.95	.87*	.12	.07		75	2	.98	.95	.08	.02
	75	3	.91	.82*	.11	.04		75	3	.99	.91	.07	.03
	75	4	.92	.76*	.10	.04		75	4	.94	.88*	.07	.04
7A	50	1	.95	.97	.14	.02	11B	50	1	.99	.97	.15	.09
	50	2	.90	.96	.08	.04		50	2	.96	.93	.09	.04
	50	3	.84*	.93	.04	.00*		50	3	.95	.86*	.08	.02
	50	4	.82*	.85*	.03	.02		50	4	.95	.83*	.03	.01
	75	1	.99	1.00*	.07	.04		75	1	.97	.99	.07	.03
	75	2	.99	.95	.08	.04		75	2	.96	.93	.06	.03
	75	3	.98	.89*	.04	.01		75	3	.95	.92	.06	.03
	75	4	.96	.86*	.05	.03		75	4	.95	.88*	.08	.03
7B	50	1	.97	.77*	.08	.04	12A	50	1	.97	.96	.06	.01
	50	2	.92	.49*	.09	.03		50	2	.94	.84*	.07	.03
	50	3	.92	.31*	.03	.02		50	3	.90*	.71*	.06	.03
	50	4	.92	.18*	.08	.02		50	4	.88*	.69*	.04	.02
	75	1	.94	.76*	.06	.01		75	1	.97	.97	.03	.00*
	75	2	.92	.50*	.09	.04		75	2	.96	.91	.02	.01
	75	3	.92	.35*	.07	.04		75	3	.95	.84*	.03	.01
	75	4	.93	.24*	.06	.04		75	4	.92	.76*	.07	.04
9A	50	1	.86*	.88*	.06	.03	12B	50	1	.96	.98	.04	.02
	50	2	.74*	.71*	.07	.02		50	2	.94	.97	.04	.00*
	50	3	.63*	.54*	.03	.02		50	3	.93	.99	.06	.00*
	50	4	.57*	.48*	.04	.00*		50	4	.92	.94	.05	.02
	75	1	.94	.92	.07	.04		75	1	1.00*	.99	.08	.02
	75	2	.77*	.80*	.07	.00*		75	2	.96	.98	.07	.02
	75	3	.69*	.63*	.06	.04		75	3	.96	.98	.03	.01
	75	4	.63*	.59*	.06	.00*		75	4	.94	.95	.04	.00*
9B	50	1	.92	.99	.05	.02	13A	50	1	.98	1.00*	.07	.02
	50	2	.88*	.96	.06	.03		50	2	.95	.98	.01	.01
	50	3	.80*	.87*	.06	.01		50	3	.95	.96	.04	.00*
	50	4	.77*	.82*	.07	.02		50	4	.95	.91	.04	.00*
	75	1	.99	.97	.11	.04		75	1	.99	.97	.04	.02
	75	2	.93	.96	.08	.04		75	2	.99	.97	.06	.03
	75	3	.87*	.92	.07	.00*		75	3	.98	.97	.05	.02
	75	4	.84*	.82*	.04	.00*		75	4	.96	.96	.08	.01
10A	50	1	.98	1.00*	.04	.03	13B	50	1	.92	.95	.07	.03
	50	2	.98	.99	.05	.01		50	2	.85*	.86*	.04	.02
	50	3	.96	.94	.05	.00*		50	3	.85*	.80*	.07	.01
	50	4	.92	.97	.06	.01		50	4	.82*	.71*	.04	.02
	75	1	.93	1.00*	.05	.04		75	1	.97	.95	.06	.03
	75	2	.93	1.00*	.04	.04		75	2	.91	.88*	.05	.03
	75	3	.93	1.00*	.05	.02		75	3	.86*	.82*	.06	.03
	75	4	.92	.99	.06	.01		75	4	.76*	.72*	.09	.05

NOTE: Exp = experiment; Nobs = econometric sample size; NLAG = number of lags used to form instruments.
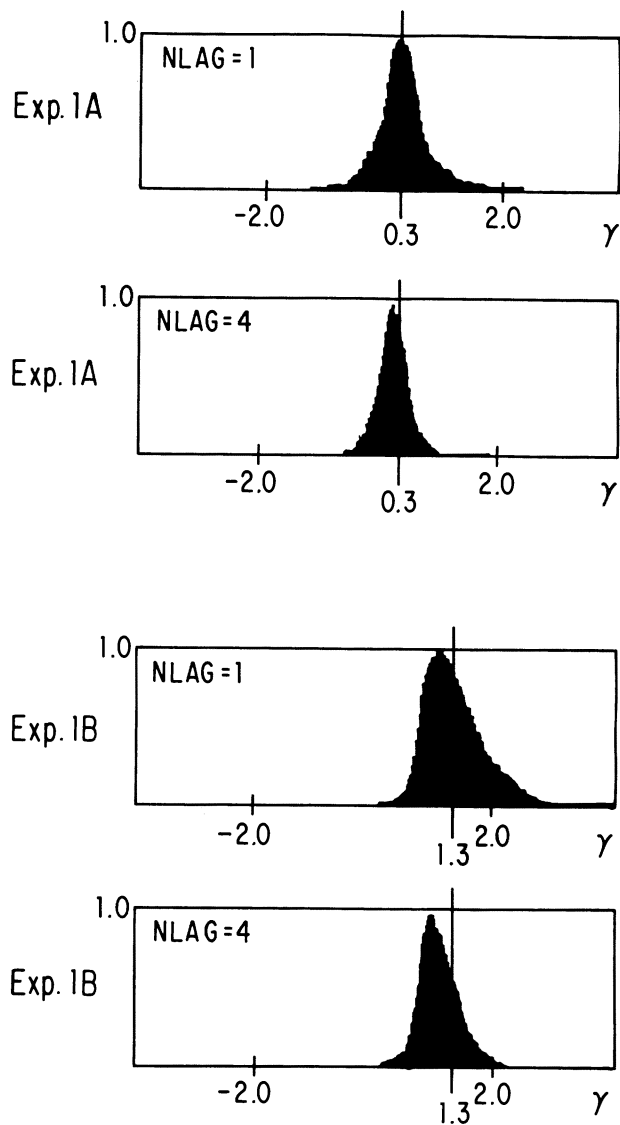*Significantly different from expected value at the 95% level.

Figure 1. Coverage Rates of 95% Confidence Intervals for Lag Lengths of 1 and 4 in Experiments 1A and 1B. The height indicates the proportion of the replications that a 95% confidence interval covered the value of γ along the horizontal axis. The scales of the horizontal axis are not the same for Experiments 1A and 1B. The solid vertical line extending outside the rectangle marks the pseudotrue value of the parameter γ. The vertical height should be .95 at the pseudotrue value if the confidence interval is unbiased.

dence interval about $\hat\gamma$ covered the corresponding value of γ on the horizontal axis. Another interpretation of the figures is that each panel is a plot of the function $h(\gamma) = (1/\text{NREP})\Sigma I[\gamma \in k\text{th confidence interval}]$, where $I[-]$ denotes the 0–1 indicator function, NREP is the number of replications equal to 100 in this case, and the index k for the summation runs over the replications. The function h should attain the value .95 at the pseudotrue values that are indicated in the figures.

The downward bias in $\hat\gamma$ in Experiments 1 and 7 is evident in Figures 1 and 2, and the bias becomes more severe at larger γ_0 and at the small values for magnitudes of the diagonal elements of the AR matrix for the data generation process. The upward bias in Experiment 9

is also evident. Interestingly, the estimator is close to being unbiased in Experiment 9B with NLAG = 1, but this is really due to the approximate cancellation of two biases—one of which is the upward bias induced by the positive correlation between consumption and dividends, and the other is the larger pseudotrue γ_0.

Concerning the variance/bias trade-off, note that with the longer lag length of NLAG = 4 the confidence intervals clearly tend to cluster more tightly around biased values. This is especially evident in Experiments 1B, 7B, 9A, and 9B. If we consider in particular Experiment 9A in Figure 3, then with NLAG = 1 the intervals are quite wide on average but the coverage rate at the pseudotrue value is close to .95; on the other hand, with
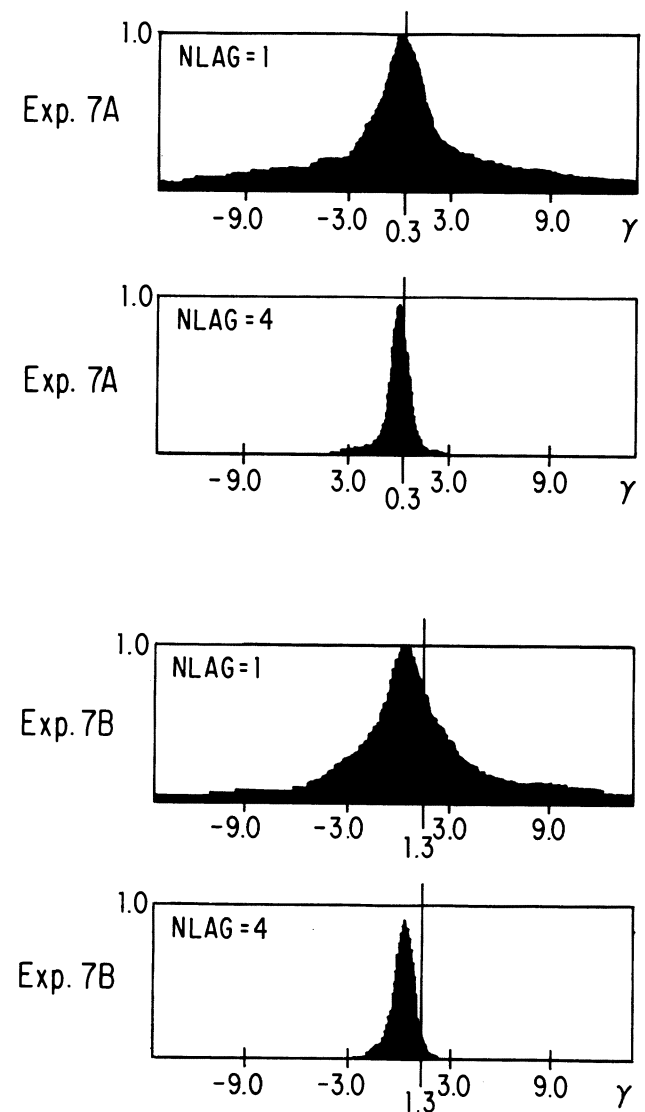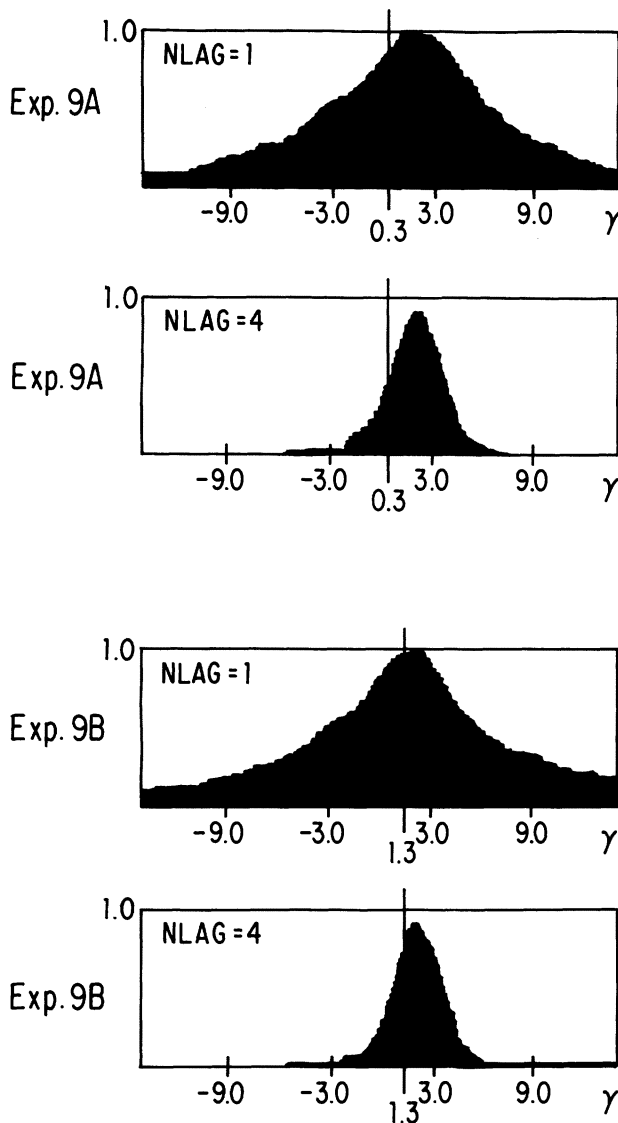

Figure 2. Coverage Rates of 95% Confidence Intervals for Lag Lengths of 1 and 4 in Experiments 7A and 7B. The height indicates the proportion of the replications that a 95% confidence interval covered the value of γ along the horizontal axis. The scales of the horizontal axis are not the same for experiments 7A and 7B. The solid vertical line extending outside the rectangle marks the pseudotrue value of the parameter γ. The vertical height should be .95 at the pseudotrue value if the confidence interval is unbiased.

Figure 3. Coverage Rates of 95% Confidence Intervals for Lag Lengths of 1 and 4 in Experiments 9A and 9B. The height indicates the proportion of the replications that a 95% confidence interval covered the value of γ along the horizontal axis. The scales of the horizontal axis are not the same for Experiments 9A and 9B. The solid vertical line extending outside the rectangle marks the pseudotrue value of the parameter γ. The vertical height should be .95 at the pseudotrue value if the confidence interval is unbiased.

NLAG = 4 the intervals are much too optimistic and are centered on much too large a value of γ. As previously noted, increasing NLAG beyond 1 or 2 may bring about marginal reductions in the mean squared error (see Table 3). Nevertheless, it seems clear from this experiment and from the others depicted in the figures that this reduction is associated with a sharp increase in the likelihood of making misleading inferences about the curvature parameter γ. The magnitude of this parameter is an important determinate of the size of risk premia in stochastic equilibrium models, and thus it is important to have a reliable interval estimate of it. Therefore, it seems inescapable to conclude that short

lag lengths are preferred to longer ones in applied work with the GMM procedure.

5. CONCLUSIONS

The methodology and findings of this article suggest an interesting topic for future research—namely, bootstrapping GMM estimates. Specifically, it is practicable to use the methods like those set out in Sections 3 and 4 to calibrate small-scale artificial economies using the estimated values of the structural parameters from a GMM estimation. By applying the GMM estimator to realizations of pseudodata from the model economy, one could get additional insight into the characteristics of the bias and variability of the GMM estimates that were obtained using the actual data. The bootstrap would require a modest amount of additional VAR estimation of the laws of motion of the predetermined variables, and one would have to recognize that these laws of motion are really low-dimensional approximations to the actual laws of motion. That is, the proposed bootstrap for the GMM estimator should be termed a "limited information" or "approximate" bootstrap where the data-generating process (DGP) is acknowledged to be an approximation of the actual DGP. Nonetheless, it is reasonable to conjecture that such a bootstrap could be very informative, although this needs to be explored further in future work.

ACKNOWLEDGMENTS

REFERENCES

Biddle, J. (1984), "The Intertemporal Elasticity of Supply of Labor: Another Look," unpublished manuscript, Duke University, Dept. of Economics.

Breiman, L. (1968), *Probability*, Reading, MA: Addison–Wesley.

Brown, D. P., and Gibbons, M. R. (1985), "A Simple Econometric Approach for Utility-Based Asset Pricing Models," *The Journal of Finance*, 60, 359–381.

Chamberlain, G. (1983), "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," Working Paper 8307, University of Wisconsin, Social Systems Research Institute.

Dunn, K. B., and Singleton, K. J. (1984), "Modeling the Term Structure of Interest Rates Under Nonseparable Utility and Durability of Goods," unpublished manuscript, Carnegie-Mellon University, Dept. of Economics.

Garber, P. M., and King, R. (1983), "Deep Structural Excavation? A Critique of Euler Equation Methods," Technical Paper 31, National Bureau of Economic Research.

Grossman, S. J., and Shiller, R. (1982), "Consumption Correlated-

ness and Risk Measurement in Economies With Non-Traded Assets and Heterogeneous Information," *Journal of Financial Economics,* 10, 195–210.

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica,* 50, 1029–1055.

——— (1985a), "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators," *Journal of Econometrics,* 30, 203–238.

Hansen, L. P., and Hodrick, R. J. (1983), "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," in *Exchange Rates and International Macroeconomics,* ed. J. A. Frenkel, Chicago: University of Chicago Press, pp. 113–152.

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

——— (1983), "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy,* 91, 249–265.

Hausman, J. (1978), "Specification Testing in Econometrics," *Econometrica,* 46, 1251–1271.

Hayashi, F., and Sims, C. A. (1983), "Nearly Efficient Estimation of Time Series Models With Predetermined, but Not Strictly Exogenous, Instruments," *Econometrica,* 51, 783–798.

Hendry, D. F. (1984), "Monte Carlo Experimentation in Econometrics," in *The Handbook of Econometrics* (Vol. 2), eds. Z. Griliches and M. Intriligator, pp. 939–975.

Hoaglin, D. C., and Andrews, D. F. (1975), "The Reporting of Computation-Based Results in Statistics," *The American Statistician,* 29, 122–126.

Lucas, R. E., Jr. (1978), "Asset Prices in an Exchange Economy," *Econometrica,* 46, 1429–1445.

Mehra, R., and Prescott, E. C. (1985), "The Equity Premium: A Puzzle," *Journal of Monetary Economics,* 15, 145–162.

Newey, W. (1985), Maximum Likelihood Specification Testing and Instrumented Score Tests," *Econometrica,* 53, 1047–1071.

Phillips, P. C. B. (1982), "Best Uniform and Modified Padé Approximants to Probability Densities in Econometrics," in *Advances in Econometrics,* ed. W. Hildebrand, Cambridge, U.K.: Cambridge University Press.

Rotenberg, J. J. (1984), "Interpreting the Statistical Failures of Some Rational Expectations Macroeconomic Models," *American Economic Review,* 74, 188–193.

Rubinstein, R. Y. (1981), *Simulation and the Method of Monte Carlo,* New York: John Wiley.

Sargan, J. D. (1982), "On Monte Carlo Estimates of Moments That Are Infinite," *Advances in Econometrics,* 1, 267–299.

Sargent, T. J. (1978), "Estimation of Dynamic Labor Demand Schedules Under Rational Expectations," *Journal of Political Economy,* 86, 1009–1044.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics,* New York: John Wiley.

Shapiro, A. (1986), "Asymptotic Theory of Overparameterized Structural Models," *Journal of the American Statistical Association,* 81, 142–149.

Sims, C. A. (1982), "Policy Analysis With Econometric Models," *Brookings Papers on Economic Activity,* 1, 107–152.

Singleton, K. J. (1985), "Testing Specifications of Economic Agent's Intertemporal Optimum Problems Against Non-Nested Alternatives," *Journal of Econometrics,* 30, 391–413.

Tauchen, G. (1985a), "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics,* 30, 415–443.

——— (1985b), "Finite-State Markov Approximations to Univariate and Vector Autoregressions," *Economics Letters,* 20, 177–181.

——— (1985c), "A Note on the Asymptotic Lower Bound for the Covariance Matrix of the GMM Estimator of the Parameters of Agents' Utility Functions," *Economics Letters,* 20, 151–155.

White, H. (1984), *Asymptotic Theory for Econometricians,* Orlando, FL: Academic Press.

Wouk, A. (1979), *A First Course in Applied Functional Analysis,* New York: John Wiley.