Finite-sample properties of some alternative GMM estimators
Hansen, Lars Peter; Heaton, John; Yaron, Amir
Journal of Business & Economic Statistics; Jul 1996; 14, 3; ABI/INFORM Global
pg. 262

Finite-Sample Properties of Some Alternative GMM Estimators

Lars Peter HANSEN
Department of Economics, University of Chicago, Chicago, IL 60637

John HEATON
Kellogg Graduate School of Management, Northwestern University, Evanston, IL 60208

Amir YARON
G.S.I.A., Carnegie Mellon University, Pittsburgh, PA 15213

We investigate the small-sample properties of three alternative generalized method of moments (GMM) estimators of asset-pricing models. The estimators that we consider include ones in which the weighting matrix is iterated to convergence and ones in which the weighting matrix is changed with each choice of the parameters. Particular attention is devoted to assessing the performance of the asymptotic theory for making inferences based directly on the deterioration of GMM criterion functions.

KEY WORDS: Asset pricing; Generalized method of moments; Monte Carlo.

The purpose of this article is to investigate the small-sample properties of generalized method of moments (GMM) estimators applied to asset-pricing models. Our alternative asymptotically efficient estimators include ones in which the weighting matrix is estimated using an initial (consistent) estimator of the parameter vector, ones in which the weighting matrix is iterated to convergence, and ones in which the weighting matrix is changed for every hypothetical parameter value. The last of these three approaches has not been used very much in the empirical asset-pricing literature, but it has the attraction of being insensitive to how the moment conditions are scaled. In addition, we study the advantages to basing statistical inferences directly on the criterion function rather than on quadratic approximations to it.

We address the following issues:

1. How does the procedure for constructing the weighting matrix affect the small-sample behavior of the GMM criterion function?
2. How do the confidence regions of parameter estimators constructed using the GMM criterion function perform relative to confidence regions based on the usually constructed standard errors?
3. Is the small-sample overrejection often found in studies of GMM estimators reduced when using an estimator in which the weighting matrix is continuously altered?
4. How are the small-sample biases of the GMM estimators affected by the choice of procedure for constructing the weighting matrix?

Because there has been an extensive body of empirical work investigating the consumption-based intertemporal capital asset-pricing model (CAPM) using GMM estimation methods, we use such models as laboratories for our Monte Carlo experiments. As in the work of Tauchen (1986) and Kocherlakota (1990a), all of our experiments come from single-consumer economies with power utility functions. Within the confines of these economies, there is still considerable flexibility in the experimental design. Some of our experimental economies are calibrated to annual time series data presuming a century of data. Other economies are calibrated to monthly postwar data. In the experiments calibrated to annual data and several of the experiments calibrated to monthly data, the moment conditions are nonlinear in at least one of the parameters of interest. Moreover, some of the specifications introduce time nonseparabilities in the consumer preferences that are motivated by either local durability or habit persistence. In the other experiments calibrated to monthly data, the moment conditions are, by design, linear in the parameter of interest. Some of these setups are special cases of the classical simultaneous-equations model. For other setups, the observed data are modeled as being time averaged, introducing a moving average structure in the disturbance terms.

The article is organized as follows. Section 1 describes the alternative estimators we study and the related econometric literature. Section 2 specifies the Monte Carlo environments we use. Section 3 gives an overview of the calculations including a description of how inferences are made based directly on the shape of the criterion functions. Section 4 then presents the results of the Monte Carlo experiments using a lognormal model calibrated to monthly data. Section 5 presents the results calibrated to annual and monthly data using a Markov-chain approximation. Finally, our concluding remarks are in Section 6.

© 1996 American Statistical Association
Journal of Business & Economic Statistics
July 1996, Vol. 14, No. 3

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1. ALTERNATIVE ESTIMATORS AND RELATED LITERATURE

One of the goals of our study is to compare the finite-sample properties of three alternative GMM estimators, each of which uses a given collection of moment conditions in an asymptotically efficient manner. Write the moment conditions as

$$E[\varphi(X_t, \beta)] = 0, \tag{1}$$

where β is the k-dimensional parameter vector of interest. In (1) the function φ has $n \geq k$ coordinates. We assume that $\{1/\sqrt{T} \sum_{t=1}^{T} \varphi(X_t, \beta)\}$ converges in distribution to a normally distributed random vector with mean 0 and covariance matrix $V(\beta)$.

Let $V_T(\beta)$ denote (an infeasible) consistent estimator of this covariance matrix. This latter estimator is typically made operational by substituting a consistent estimator for β, denoted $\{b_T^1\}$. An efficient GMM estimator of the parameter vector β is then constructed by choosing the parameter vector b that minimizes

$$\left[\frac{1}{T}\sum_{t=1}^{T}\varphi(X_t, b)\right]' [V_T(b_T^1)]^{-1} \left[\frac{1}{T}\sum_{t=1}^{T}\varphi(X_t, b)\right]. \tag{2}$$

The first two GMM estimators that we consider differ in the way in which this is accomplished.

Two-Step Estimator. The first estimator, called the *two-step estimator,* uses an identity matrix to weight the moment conditions so that b_T^1 is chosen to minimize

$$\left[\frac{1}{T}\sum_{t=1}^{T}\varphi(X_t, b)\right]' \left[\frac{1}{T}\sum_{t=1}^{T}\varphi(X_t, b)\right]. \tag{3}$$

Let b_T^2 denote the estimator obtained by minimizing (2).

Iterative Estimator. The second estimator continues from the two-step estimator by reestimating the matrix $V(\beta)$ using $V(b_T^{j-1})$ and constructing a new estimator b_T^j. This is repeated until b_T^j converges or until j attains some large value. Let b_T^∞ denote this estimator.

Continuous-Updating Estimator. Instead of taking the weighting matrix as given in each step of the GMM estimation, we also consider an estimator in which the covariance matrix is continuously altered as b is changed in the minimization. Formally let b_T^c be the minimizer of

$$\left[\frac{1}{T}\sum_{t=1}^{T}\varphi(X_t, b)\right]' [V_T(b)]^{-1} \left[\frac{1}{T}\sum_{t=1}^{T}\varphi(X_t, b)\right]. \tag{4}$$

Allowing the weighting matrix to vary with b clearly alters the shape of the criterion function that is minimized. Although the first-order conditions for this minimization problem have an extra term relative to problems with fixed weighting matrices, this term does not distort the limiting distribution for the estimator. [See Pakes and Pollard (1989, pp. 1044–1046) for a more formal discussion and provision of sufficient conditions that justify this conclusion.] An advantage of this estimator relative to the previous two is that it is invariant to how the moment conditions are scaled even when parameter-dependent scale factors are introduced. A simple example of a continuous-updating estimator is a minimum chi-squared estimator used for restricted multinomial models in which the efficient distance matrix is constructed from the probabilities implied by the underlying parameters and hence is parameter dependent.

The three GMM estimators have antecedents in the classical simultaneous-equations literature. Consider estimating a single equation, say $y_t = \beta' x_t + u_t$, where β is the parameter of interest. Let z_t denote the vector of predetermined variables at time t that by definition are orthogonal to u_t. One way to estimate β is to use two-stage least squares, which is our two-step estimator under the additional restrictions that the disturbance term is conditionally homoscedastic and serially uncorrelated. In this case the iterative estimator converges after two steps and hence is the two-step estimator. It is well known that the two-stage least squares estimator is not invariant to normalization. In fact Hillier (1990) criticized the two-stage least squares estimator by arguing that the object that is identified is the direction $[1, -\beta']$ but not its magnitude. Hillier then showed that the conventional two-stage least squares estimator of direction is distorted by its dependence on normalization.

As an alternative, Sargan (1958) suggested an instrumental-variables-type estimator that minimizes

$$\frac{\frac{1}{T}\sum_{t=1}^{T}(y_t - b'x_t)z_t' \left[\frac{1}{T}\sum_{t=1}^{T} z_t z_t'\right]^{-1} \times \frac{1}{T}\sum_{t=1}^{T}(y_t - b'x_t)z_t}{\frac{1}{T}\sum_{t=1}^{T}(y_t - b'x_t)^2} \tag{5}$$

by choice of b. Under the additional restrictions imposed on the disturbance term discussed in the previous paragraph, this is our continuous-updating estimator. Notice that if we ignore the denominator term in (5) and minimize, the solution is the two-stage least squares estimator. By including the denominator term, Sargan showed that, for an appropriate choice of z_t, the solution is the (limited information) quasi-maximum likelihood estimator (using a Gaussian likelihood), which as an estimator of direction is invariant to normalization.

The estimation environments that we study are more complicated than the one just described. Sometimes the moment conditions are not linear in the parameters, and the disturbance terms are often conditionally heteroscedastic and/or serially correlated. As a consequence, the two-step and iterative estimators no longer coincide and the continuous-updating estimator can no longer be interpreted as a quasi-likelihood estimator. The estimation methods remain limited information, however, in that the moment conditions used are typically not sufficient to fully characterize the time series evolution of the endogenous variables.

The second goal of our analysis is to compare the reliability of confidence regions computed using quadratic approximations to criterion functions to ones based directly on the deterioration of the original criterion functions. The former approach is more commonly used in the empirical

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

asset-pricing literature partially because it is easier to implement. The latter approach exploits the chi-squared feature of the appropriately scaled criterion functions. From the vantage point of hypothesis testing, the plausibility of an observed deterioration of the criterion function caused by imposing parameter restrictions can be assessed by using the appropriate chi-squared distribution. Using this same insight, confidence regions can be computed by using the appropriate chi-squared distribution to prespecify some increment in the criterion function and inferring the set of parameter values that imply no more than that increment. Such confidence regions can have unusual shapes and, in fact, may not even be connected. One of the key questions of this investigation is whether or not they lead to more reliable statistical inferences.

One of our reasons for studying the performance of criterion-function-based inference comes from the work of Magdalinos (1994). Within the confines of the classical simultaneous-equations paradigm, Magdalinos studied the performance of alternative tests of instrument admissibility. As a result of his analysis, Magdalinos recommended altering the weighting matrix to embody the restrictions as is done in the continuous-updating method. In addition, he found that test statistics are better behaved using the limited-information maximum likelihood estimator than the two-stage least squares estimator. Recall that the former estimator coincides with our continuous-updating estimator and the latter to our two-step and iterated estimators in the classical simultaneous-equations estimation environment considered by Magdalinos.

Another reason is Nelson and Startz's (1990) criticism of the use of instrumental-variables methods for studying consumption-based asset-pricing models. These authors were concerned about the behavior of instrumental-variables estimators when the instruments are poorly correlated with the endogenous variables. Their arguments were based on analogies to results derived formally for t statistics and overidentifying restrictions tests in the classical simultaneous-equations setting. The question of interest to us is the extent to which criterion-function-based inference and continuous updating can help overcome the concerns of Nelson and Startz. Furthermore, Stock and Wright (1995) provided a theoretical rationale for considering the continuous-updating criterion function instead of other GMM implementations in situations in which the model is poorly identified.

The finite-sample properties of the two-step and iterative GMM estimators in an asset-pricing setting have been studied previously by Tauchen (1986), Kocherlakota (1990a), and Ferson and Foerster (1994). These investigators did not study the properties of the continuous-updating estimator, nor did they study the behavior of criterion-function-based confidence regions. Furthermore, Tauchen (1986) and Kocherlakota (1990a) considered only the case of time-separable preferences for the representative consumer.

2. MONTE CARLO ENVIRONMENT

We consider several Monte Carlo environments to assess the finite-sample properties of the estimators described in Section 1. The data-generating mechanisms are constructed to be consistent with a representative agent consumption-based capital asset-pricing model (CCAPM). Estimators of the parameters of the representative agent's utility function are considered along with tests of the overidentifying conditions implied by the model. We use the CCAPM as the basis of our experiments because GMM has been used extensively in studying this model (e.g., see Dunn and Singleton 1986; Eichenbaum and Hansen 1990; Epstein and Zin 1991; Ferson and Constantinides 1991; Hansen and Singleton 1982). Furthermore, the CCAPM forms the basis for two studies of the finite-sample properties of GMM conducted by Tauchen (1986) and Kocherlakota (1990a).

Preferences and Euler Equations. In the model the representative consumer is assumed to have preferences over consumption given by

$$U_0 = E\left[\sum_{t=0}^{\infty} \delta^t \, \frac{(c_t + \theta c_{t-1})^{1-\gamma} - 1}{1-\gamma}\right], \qquad \gamma > 0, \quad (6)$$

where c_t is consumption at date t. The parameter θ captures some time nonseparability in preferences. Examples of models with time nonseparability in preferences can be found in the work of Abel (1990), Constantinides (1990), Detemple and Zapatero (1991), Dunn and Singleton (1986), Eichenbaum and Hansen (1990), Gallant and Tauchen (1989), Heaton (1993, 1995), Novales (1990), Ryder and Heal (1973), and Sundaresan (1989). If $\theta > 0$, consumption is durable or substitutable over time. If $\theta = 0$, the preferences of the consumer are time-additive. If $\theta < 0$, consumption is complementary over time and the preferences of the representative consumer exhibit habit persistence.

We consider estimators of the parameters δ, γ, and θ, as well as tests of the model based on implications of the Euler equations. Let $mus_t \equiv (c_t + \theta c_{t-1})^{-\gamma}$, which can be interpreted as the indirect marginal utility for consumption "services" as measured by $s_t = c_t + \theta c_{t-1}$. Similarly, let $muc_t \equiv (c_t + \theta c_{t-1})^{-\gamma} + \theta\delta E[(c_{t+1} + \theta c_t)^{-\gamma} | \mathcal{F}_t]$, where $\mathcal{F}_t$ gives the information set at time t. The Euler equation for a representative agent's portfolio allocation decision is given by

$$muc_t = E(\delta muc_{t+1} R_{t+1} | \mathcal{F}_t), \quad (7)$$

where R_{t+1} is a gross return on an asset from t to $t+1$. Because aggregate consumption is growing over time, we divided (7) by mus_t to induce stationarity. The (normalized) Euler equation that we consider is then given by

$$\frac{muc_t}{mus_t} = E\left(\delta \, \frac{muc_{t+1}}{mus_t} \, R_{t+1} | \mathcal{F}_t\right). \quad (8)$$

Removing conditional expectations from (8) results in the Euler-equation error

$$\phi_{t+2}(\delta, \gamma, \theta) = \frac{muc_t^*}{mus_t} - \delta \, \frac{muc_{t+1}^*}{mus_t} \, R_{t+1}, \quad (9)$$

where $muc_t^* \equiv (c_t + \theta c_{t-1})^{-\gamma} + \delta\theta(c_{t+1} + \theta c_t)^{-\gamma}$. Note that $E(\phi_{t+2} | \mathcal{F}_t) = 0$. Furthermore, notice that, when θ is not

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 1. Maximum Likelihood Estimator of Monthly Law of Motion

Order of C(L)	Unrestricted log-likelihood	No time avg. Log-likelihood	No time avg. γ	Time averaged Log-likelihood	Time averaged γ
1	3,220.6	—	—	—	—
2	3,237.3	3,214.9	4.55	—	—
3	3,238.8	3,218.1	4.21	3,230.3	4.99

0, o_{t+2} has a first-order moving average [MA(1)] structure. By choosing instruments, z_t, in $\mathcal{F}_t$, unconditional moment conditions are given by

$$E[\varphi_{t+2}(\delta,\gamma,\theta)] = E[z_t\phi_{t+2}(\delta,\gamma,\theta)] = 0. \tag{10}$$

Finally, notice that ϕ_{t+2} can be expressed in terms of consumption ratios and returns, which we take to be stationary processes.

One unpleasant feature of these moment conditions is that they can always be made to be satisfied in a degenerate fashion. Suppose that $\gamma = 0$ and $\delta\theta = -1$; then clearly $muc_t^* = 0$ and the moment conditions are trivially satisfied. Without imposing additional constraints on the parameter vectors, this degeneracy in the moment conditions creates problems for the two-step and iterative estimators. Of course, when time separability is imposed ($\theta = 0$), these problematic parameter values cannot be reached. When θ is permitted to be different from 0, Eichenbaum and Hansen (1990) were led to divide the moment conditions by $1 + \delta\theta$ so that the two-step and iterative estimators not be driven to the degenerate values. We will do likewise. An attractive property of the continuous-updating estimator is its insensitivity to parameter-dependent scale factors and hence to the moment transformation used by Eichenbaum and Hansen (1990). Moreover, the criterion of the continuous-updating estimator does not necessarily tend to 0 in the vicinity $\gamma = 0$ and $\delta\theta = -1$. Although the estimated mean for $\{\varphi_{t+2}\}$ becomes small in the neighborhood of these parameter values, so does the estimated asymptotic covariance matrix, and the criterion function for the continuous-updated estimator plays off this tension.

We build several Monte Carlo environments to simulate returns and consumption growth that are consistent with (8). These are used to assess the finite-sample properties of the estimators of Section 1 based on the moment conditions (10).

2.1 Lognormal Model

Time-Additive Model, No Time Averaging. In our first Monte Carlo environment we model consumption growth and returns directly by assuming that they are jointly lognormally distributed as in the work of Hansen and Singleton (1983). Let $Y(t) = [\log(c_t/c_{t-1})\log R_t^e \log R_t^f]'$, where c_t is aggregate consumption at time t, R_t^e is the gross return on a stock index at time t, and R_t^f is the gross return on a bond at time t. We assume that

$$Y_t = \mu + B(L)\varepsilon_t, \tag{11}$$

where ε_t is a normally distributed three-dimensional random vector that is independent over time and has zero mean and covariance matrix I and where μ is the mean of Y_t. Furthermore, $B(L)$ is a matrix of polynomials in the lag operator. To use this assumption about the dynamics of consumption and returns along with the Euler equation (7), we assume that the preferences of the representative agent are time additive. In this case the Euler-equation error for each return can be written as

$$-\gamma\log(c_{t+1}/c_t) + \log R_{t+1} - \kappa = \eta_{t+1}, \tag{12}$$

where $E(\eta_{t+1}|\mathcal{F}_t) = 0$ and κ is a constant (e.g., see Hansen and Singleton 1983).

The relation (12) implies a set of restrictions on the law of motion (11). To impose these restrictions, we consider several finite-order parameterizations of $B(L)$ and use the methods described by Hansen and Sargent (1991). These parameterizations are of the form

$$B(L) = \frac{C(L)}{\alpha(L)}, \tag{13}$$

where $C(L)$ is a 3×3 matrix of polynomials in the lag operator and $\alpha(L)$ is a 1×1 polynomial in the lag operator. We restrict the polynomial $\alpha(L)$ to be second order and considered several different orders of $C(L)$.

To estimate the constrained law of motion, we used monthly data from 1959.2 to 1992.12. Aggregate consumption is seasonally adjusted real aggregate consumption of nondurables plus services for the United States taken from CITIBASE. These data were converted to a per capita measure by dividing by total U.S. population for each month, obtained from CITIBASE. The equity return is the value-weighted return from the Center for Research in Security Prices (CRSP), and the bond return is the Fama–Bliss risk-free return from CRSP. Each of these return series was converted into a real return using the implicit price deflator for nondurables and services from CITIBASE.

For simplicity we removed the sample mean from the vector Y_t so that the constants in (11) and (12) did not have to be estimated. As a result, the only preference parameter to be estimated is γ. The results of estimating the law of motion (11) using exact maximum likelihood for different orders of the polynomial $C(L)$ are given in Table 1. The column labeled "Unrestricted log-likelihood" reports the log-likelihood in the case of unrestricted estimation of the polynomials in the lag operator. The columns labeled "no time avg." report the log-likelihood and the estimated value of γ under the restrictions implied by (12). In searching for the maximized log-likelihood, larger values of the log-likelihood were found for values of γ larger than 50. The results reported in Table 1 for the constrained models correspond to local maxima. We used the local maximizers for our simulations because they result in plausible values for γ. Notice that there is substantial improvement in the log-likelihood in moving from a first- to a second-order polynomial for $C(L)$ in the unrestricted case. This indicates that more than a first-order polynomial is needed for $C(L)$. There is little improvement in going to a third-order polynomial in the unrestricted case.

For both the second- and third-order polynomial cases, there is great deterioration in the log-likelihood when the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

model restrictions are imposed. This is consistent with the results reported by Hansen and Singleton (1983). Moreover, there is little improvement in the log-likelihood in moving from a second-order to a third-order $C(L)$. For this reason we used the point estimates from the restricted model with a second-order $C(L)$ to conduct our Monte Carlo experiments for the lognormal model with no time averaging.

In assessing the finite-sample properties of GMM estimators in this case, we constructed 500 Monte Carlo samples, each with a sample size of 400. A sample size of 400 approximates the size of available monthly consumption data. We constructed moment conditions based on the Euler-equation errors for both the bond and the stock returns simultaneously. For each Euler-equation error we used one period lagged (log) bond and (log) stock returns and one period lagged (log) consumption growth as instruments. We did not include constants as instruments because the data are simulated under the assumption that they have a zero mean.

Time-Additive Model With Time Averaging. As a further data-generating mechanism, we also consider an example in which the decision interval of the representative agent is much smaller than the interval of the data. Suppose that there are n decision periods within each observation period. For example, if the representative agent's decision interval is a week and data are observed monthly, then n would be approximately 4. The representative agent's utility function at time t is given by

$$U_t = E\left[\sum_{h=0}^{\infty} (\delta_n)^h \frac{(c_{t+h/n})^{1-\gamma} - 1}{1-\gamma} \,\middle|\, \mathcal{F}_t\right]. \tag{14}$$

If we maintain the assumption that consumption and returns are jointly lognormally distributed, the Euler-equation error η_{t+1} is again given by (12). Moreover, this error can be decomposed as

$$\eta_{t+1} = \sum_{h=1}^{n} \zeta_{t+h/n}, \tag{15}$$

where

$$\zeta_{t+h/n} \equiv E(\eta_{t+1}|\mathcal{F}_{t+h/n}) - E(\eta_{t+1}|\mathcal{F}_{t+(h-1)/n}). \tag{16}$$

In this environment, we presume that observed consumption does not correspond to the actual point-in-time consumption of the representative agent but instead is an average of actual consumption over one unit of time. Specifically, suppose that observed consumption, c_t^a, is a geometric average of actual consumption,

$$c_t^a = \left(\prod_{h=1}^{n} c_{t-1+h/n}\right)^{1/n}, \tag{17}$$

and similarly for the observed return, R_t^a. Averaging (12) over time implies that

$$-\gamma \log(c_{t+1}^a/c_t^a) + \log R_{t+1}^a = \kappa + \frac{1}{n} \sum_{\tau=1}^{n} \sum_{h=1}^{n} \zeta_{t-1+(\tau/n)+(h/n)}. \tag{18}$$

Notice that in this case the Euler-equation error

$$\eta_{t+1}^a \equiv \frac{1}{n} \sum_{\tau=1}^{n} \sum_{h=1}^{n} \zeta_{t-1+(\tau/n)+(h/n)} \tag{19}$$

is predictable at time t. However, $E(\eta_{t+1}^a|\mathcal{F}_{t-1}) = 0$ so that instruments can be chosen from the information set at time $t-1$. An instrumental-variables estimator of this model must account for the MA(1) structure of the moment condition.

We estimated the log-linear law of motion (13) for consumption and asset returns under the restriction implied by (18). Because the consumption data are an arithmetic average of consumption expenditures over a period, in applying (18) to actual consumption data we assume that geometric averages and arithmetic averages are approximately the same. This assumption was made by Grossman, Melino, and Shiller (1987), Hall (1988), and Hansen and Singleton (1996). Note also that the returns in (18) are time averaged. For simplicity, in estimating the law of motion (13), we use the monthly CRSP series directly. Furthermore, the model with time averaging imposes a weaker set of restrictions than does the model that takes no account of time averaging because we do not impose the restriction on the first-order autocorrelation of the Euler-equation error of the stock return. In the limit case of continuous decision making, the first-order autocorrelation of the error should be .25, as discussed by Grossman et al. (1987) and Hall (1988).

We consider the case of a third-order polynomial for $C(L)$ and a second-order polynomial for $\alpha(L)$. The results of this estimation are reported in Table 1 in the columns labeled "Time averaged." Notice that the log-likelihood function improves somewhat compared to the case in which time averaging is ignored. The model, however, is still substantially at odds with the data. The estimated value of γ is slightly larger as well.

We used this model to create 500 Monte Carlo draws, each with a sample size of 400. As in the case of no time averaging, we studied estimators based on the Euler equations for both returns. In this case the instruments were (log) stock returns, (log) bond returns, and (log) consumption growth, all lagged two periods.

2.2 Discrete-State Models

In our second set of Monte Carlo environments we follow Tauchen (1986) and Kocherlakota (1990a) and consider a Markov-chain model for aggregate consumption and dividend growth. Aggregate consumption is assumed to represent the endowment of the representative consumer, and dividends represent the cash flow from holding stock. We form one-period stock returns and the returns to holding a one-period (real) discount bond. Each of these returns can be represented as functions of the state of the Markov chain. Construction of these returns in the case of time-additive utility was described by Kocherlakota (1990a) and Tauchen (1986).

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 2. Preference Settings for Discrete State-Space Model

$$U = E\left[\sum_{t=0}^{\infty} \delta^t \frac{(c_t+\theta c_{t-1})^{1-\gamma}-1}{1-\gamma}\right]$$

Preference case	δ	γ	θ
TS1	.97	1.3	0
TS2	1.139	13.7	0
TNS1	.97	1.3	1/3
TNS2	.97	1.3	−1/3
TNS3	.97	1.3	−2/3

Annual Model. To calibrate the first Markov chain, we used the method described by Tauchen and Hussey (1991) to approximate a first-order vector autoregression (VAR) for consumption and dividend growth. The parameters of the VAR are taken from Kocherlakota (1990a) and are given by

$$\begin{bmatrix} \ln \xi_t \\ \ln \lambda_t \end{bmatrix} = \begin{bmatrix} .004 \\ .021 \end{bmatrix} + \begin{bmatrix} .117 & .414 \\ .017 & .161 \end{bmatrix} \begin{bmatrix} \ln \xi_{t-1} \\ \ln \lambda_{t-1} \end{bmatrix} + \varepsilon_t, \tag{20}$$

where ξ_t is the gross growth rate of real annual dividends on the Standard & Poor 500, where λ_t is the gross growth rate of U.S. per capita real annual consumption, and where

$$E(\varepsilon_t \varepsilon_t') = \begin{bmatrix} .01400 & .00177 \\ .00177 & .00120 \end{bmatrix}. \tag{21}$$

Furthermore, ε_t is assumed to be normally distributed and uncorrelated over time. The Markov chain for $[\xi_t; \lambda_t]'$ is chosen to have 16 states.

In simulating data from this model we chose several values of the preference parameters of the representative consumer. These are presented in Table 2. Preference setting TS1 was used by Tauchen (1986) and preference setting TS2 was used by Kocherlakota (1990a). As shown by Kocherlakota (1990a), these latter parameters, along with the Markov-chain model of endowments, imply first and second moments for asset returns that mimic their sample counterparts. The large value of δ in TS2 is not inconsistent with the existence of an equilibrium in the model because of the large value of γ (Kocherlakota 1990b).

We restrict our attention to "moderate" values of δ and γ in our examination of time-nonseparable preferences, and we consider a range of values of θ. Parameter setting TNS1 introduces a modest degree of durability by letting $\theta = \frac{1}{3}$, and TNS2 introduces habit persistence with $\theta = -\frac{1}{3}$. TNS3 results in a more extreme amount of habit persistence by setting $\theta = -\frac{2}{3}$. The asymmetric (in magnitude) across the specifications of θ is guided in part by the a priori notion that there should only be a limited amount of durability in the goods classified as "nondurable" in National Income and Product Accounts and by empirical evidence for a substantial degree of habit persistence reported by Ferson and Constantinides (1991).

In implementing the estimators, the Euler equations for the stock and bond returns are multiplied by instrumental variables to construct moment conditions. The two instrument sets are listed in Table 3. Monte Carlo results for other moment conditions were given by Hansen, Heaton, and Yaron (1994). GMM estimates and test statistics were computed for 500 replications of a sample size of 100. This sample size corresponds approximately to the length of most annual datasets.

Monthly Model. We repeated some of the experiments using a law of motion calibrated to postwar monthly data. As in the construction of the Markov chain for the annual model, we started with a first-order VAR for consumption and dividend growth. The consumption data used in estimating the VAR were aggregate U.S. expenditures on nondurables and services described in Subsection 2.1. We constructed dividends implied by the monthly CRSP value-weighted portfolio return. These dividends were converted to real dividends using the implicit price deflator for monthly nondurables and services taken from CITIBASE. This dividend series is highly seasonal because of the regular dividend payout policies of most companies. To avoid modeling this seasonality, we let $\xi_t \equiv \log(d_t/d_{t-12})/12$, and we assumed that ξ_t represents the one-period dividend growth of the model. The series $\{\log(d_t/d_{t-12})\}$ appears to be stationary.

The parameters of the VAR estimated using these data are given by

$$\begin{bmatrix} \ln \xi_t \\ \ln \lambda_t \end{bmatrix} = \begin{bmatrix} .0012 \\ .0019 \end{bmatrix} + \begin{bmatrix} -.1768 & .1941 \\ .0267 & -.2150 \end{bmatrix} \begin{bmatrix} \ln \xi_{t-1} \\ \ln \lambda_{t-1} \end{bmatrix} + \varepsilon_t, \tag{22}$$

where

$$E(\varepsilon_t \varepsilon_t') = \begin{bmatrix} .1438 & .0001 \\ .0001 & .0145 \end{bmatrix} \times 10^{-3}. \tag{23}$$

As in the case of the annual Markov-chain model, we approximated the VAR of (22) and (23) with a 16-state Markov chain using the methods of Tauchen and Hussey (1991). The Monte Carlo data consisted of 500 replications of a sample size of 400. We focused exclusively on the more "moderate" preference configuration (TS1), adjusting δ for the shorter sampling interval. (More precisely, we used the twelfth root of .97 in place of .97 for δ.) In addition, we generated Monte Carlo data using nonseparable specification TNS1 and TNS2, again with δ adjusted appropriately.

3. OVERVIEW OF MONTE CARLO RESULTS

3.1 Descriptive Statistics

In describing the results of the various Monte Carlo experiments in Sections 3 and 4, we focus most of our discussion on the following calculations:

Table 3. Moment Conditions

Moment set	Instruments, z_t	Number of moment conditions
M1	R_t^e, R_t^f, λ_t, 1	8
M2	λ_t, 1	4

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

1. We found the minimum value of the criterion function. Call this J_T. The limiting distribution of TJ_T is chi-squared with degrees of freedom equal to the number of moment conditions minus the number of parameters estimated. We used this limiting distribution to test the overidentifying moment conditions.

2. We evaluated the criterion function at the true parameter vector. Call this J_T^{tr}. The limiting distribution of TJ_T^{tr} is chi-squared with degrees of freedom equal to the number of moment conditions. With this limiting distribution, we characterize the family of parameter vectors that look plausible from the standpoint of the moment conditions. Stock and Wright (1995) advocated the use of this type of inference when it is suspected that the parameters are poorly identified. Here we examine whether the limiting distribution of J_T^{tr} provides a reasonable small-sample approximation for conducting this inference.

3. We found the minimum value of the criterion function when γ is constrained to be its true value. Call this J_T^{γ}. Because γ is the only parameter estimated in the lognormal model, in this case J_T^{γ} coincides with J_T^{tr}. The limiting distribution of $T(J_T^{\gamma} - J_T)$ is chi-squared with 1 df. This limiting distribution allows us to construct a confidence region for γ based on the increments of the criterion function from its unconstrained minimum. By evaluating $T(J_T^{\gamma} - J_T)$ we determined whether the true value of γ is in the resulting interval for alternative confidence levels.

4. We constructed the more standard confidence intervals for γ based on a quadratic approximation to the criterion function. In particular, let γ_T be an estimator of γ and σ_T^{γ} be the estimated asymptotic standard error of the estimator γ_T. We study $T(\gamma_T - \gamma)^2/(\sigma_T^{\gamma})^2$, which has an asymptotic chi-squared distribution with 1 df. Notice that this object is just the Wald statistic for the hypothesis that the true value of the parameter is γ. In constructing this statistic for the continuous-updating estimator, the standard errors include a term that reflects the derivative of the GMM weighting matrix with respect to the parameters.

Our Monte Carlo calculations are greatly simplified by our knowledge of the true parameter vector. In empirical work, the corresponding computations would be more complicated. For instance, to construct a confidence interval for γ based on the original criterion function, a researcher would have to characterize numerically the hypothetical values of this parameter that are consistent with a prespecified deterioration in the criterion while concentrating out all of the other parameters. When there are very few remaining components in the parameter vector (in our examples 0, 1, or 2), this concentration is tractable. This approach may become very difficult, however, when the parameter vector is large.

In reporting our Monte Carlo results we use one of the graphical methods advocated by Davidson and McKinnon (1994). For each Monte Carlo setup we computed the empirical distributions of the statistics and compared them to the corresponding chi-squared distributions. The results are plotted on a set of figures constructed as follows. For each probability value (depicted on the x axis), we computed the corresponding chi-squared critical value and the fraction of the actual computed statistics that are above that value (depicted on the y axis). Thus the 45-degree line (depicted as ...) is the appropriate reference for assessing the quality of the limiting distribution. Following Davidson and McKinnon (1994), these plots are referred to as *p-value* plots, and we present the results for the interval [0, .5] because this bounds the region of probability values used in most applications. Although we use probability values as our basis of comparison, confidence intervals at alternative significance levels can be assessed by simply subtracting the probability values from 1.

The figures are organized as follows. For each Monte Carlo setup we first consider a single figure with four graphs titled as follows: "(a) Minimized", "(b) True", "(c) Constrained-Minimized" and "(d) Wald," corresponding to the statistics $TJ_T, TJ_T^{tr}, T(J_T^{\gamma} - J_T)$, and $T(\gamma_T - \gamma)^2/(\sigma_T^{\gamma})^2$, respectively. To provide a formal statistical measure of the distance between the empirical distributions and their theoretical counterparts, on each figure a band around the 45-degree line is plotted using dotted lines. This band is a 90% confidence region based on the Kolmogorov–Smirnov Test. This states that the probability that the maximal difference between the empirical distribution and the theoretical one will lie within those lines is 90%. Maximal differences within these bands are not statistically significant at the 10% significance level. Although we present results for the interval [0, .5], the Kolmogorov–Smirnov confidence region is based on calculating the supremum between the empirical distribution and the 45-degree line over the region [0, 1].

In each graph, the dashed line gives the Monte Carlo results for the two-step estimator, the dot-dash line for the iterated estimator, and the solid line for the continuous-updating estimator. For the minimized criterion function results, there is a necessary ordering between the continuous-updating and the iterative estimator. When the iterative estimator converges, the value of the criterion function can also be obtained by the continuous-updating estimator. Because the continuous-updating estimator minimizes its criterion, this minimized value must be smaller than the criterion for the iterative estimator. As a result, the plot for the minimized criterion of the continuous-updating estimator must lie below the plot for the iterative estimator unless the iterative estimator fails to converge. There is no natural ordering between the results for the two-step estimator and the continuous-updating estimator or between the two-step estimator and the iterative estimator.

To complement our p-value plots, we also provide some results summarizing the performance of the implied parameter estimators. The finite-sample properties of the point estimates are of interest in their own right and in some cases provide additional insights into the behavior of the p-value plots.

3.2 Numerical Search Routines

The two-step and iterative estimators are given by the minimizers of the objective function (2), and the

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

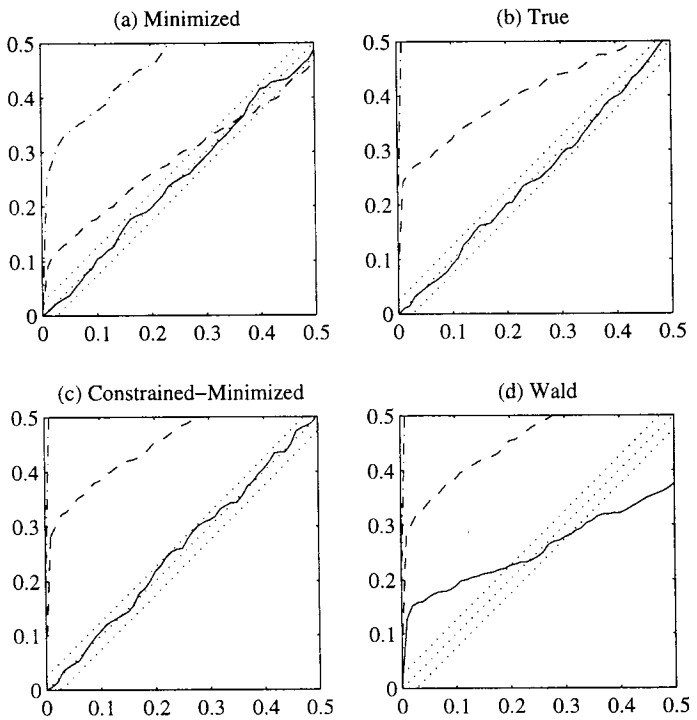

Figure 1. Criterion Functions, Monthly Lognormal Model, No Time Averaging: —·—·—, Iterative; – – –, Two-Step; ———, Continuous-Updating.

continuous-updating estimator is given by the minimizer of the objective function (4). For some of our experiments, the estimators are given by solutions to linear equations. In the other cases we used the numerical optimization routines fminu.m and fmins.m, which are part of the "Optimization Toolbox" for use with MATLAB. The routine fminu.m implements a quasi-Newton method, which is dependent on an initial setting for the parameters. To check whether the results were sensitive to initialization, we considered several different starting values that included the true parameter vector. When this gradient method failed to converge or resulted in unusual estimates, we also used the routine fmins.m, which is a simplex search method. Details on these MATLAB programs can be found in the MATLAB Optimization Toolbox manual. In Section 4 we show that the continuous-updating criterion can make numerical search for the minimizer difficult. As a further check on our numerical results, when we obtained extreme parameter estimates we also examined the continuous-updating criterion over a grid of the parameters. This gave us additional assurance that the estimated parameters were indeed minimizers of the criterion.

4. MONTE CARLO RESULTS, LOGNORMAL MODEL

For the case of time-averaged data, $\{\varphi_t\}$ has an MA(1) structure, as we discussed in Section 2.1. To account for this, the estimator of $V_T(b)$ was computed by first considering an estimator of the form

$$V_T(b) = \frac{1}{T}\left\{\sum_{t=1}^{T}\varphi_T(X_t,b)\varphi_T(X_t,b)' + \sum_{t=2}^{T}[\varphi_T(X_{t-1},b)\varphi_T(X_t,b)' + \varphi_T(X_t,b)\varphi_T(X_{t-1},b)']\right\}, \quad (24)$$

where $\varphi_T(X_t,b) = \varphi(X_t,b) - 1/T\sum_{t=1}^{T}\varphi(X_t,b)$. When this estimator was not positive definite, we used an estimator proposed by Durbin (1960) (see also Eichenbaum, Hansen, and Singleton 1988). Durbin's estimator is obtained by first approximating the MA(1) model with a finite-order autoregression. The residuals from this autoregression are used to approximate the innovations. Then the parameters of the MA(1) model are estimated by running a regression of the original time series onto a one-period lag of the "approximate" innovations. Finally, an estimate of $V_T(b)$ is formed using the estimated MA coefficients and sample covariance matrix for the residuals. This procedure has the advantage that the finite-order MA structure of $\{\varphi_t\}$ is imposed and the estimator is positive semidefinite by construction. It does, however, rely on the choice of a finite-order autoregression to use in the approximation. In implementing the estimator we ran a 12th-order autoregression in the initial stage. Although this covariance matrix estimator was used in searching for the parameter estimates, it was not needed at any of the converged parameter values.

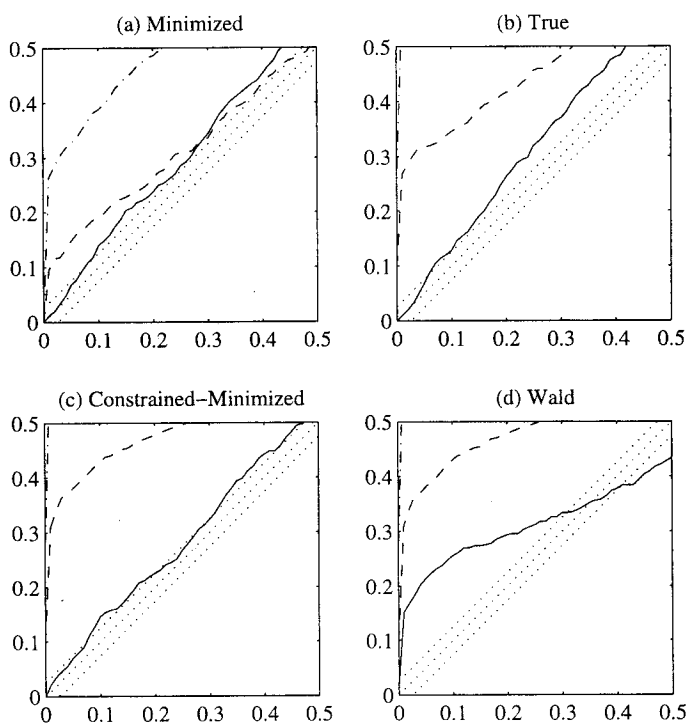

Figure 2. Criterion Functions, Monthly Lognormal Model, Time Averaging: —·—·—, Iterative; – – –, Two-Step; ———, Continuous-Updating.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Table 4. Properties of Estimators of γ, Lognormal Model

Property	Continuous-updating	Iterative	Two-step
A. No time averaging, true $\gamma = 4.55$			
Median	3.72	1.64	1.73
Mean	7,171.17	1.81	2.06
Truncated mean*	4.47	1.81	2.06
10% quantile	−3.67	−.23	−.20
90% quantile	18.75	4.00	4.33
B. Time averaging, true $\gamma = 4.99$			
Median	3.48	1.75	1.74
Mean	−518.14	1.88	2.04
Truncated mean*	3.18	1.88	2.04
10% quantile	−10.48	−.29	−.66
90% quantile	11.52	4.22	4.85

* Estimates with absolute values greater than 100 were excluded from the computation of the truncated means.

Criterion Functions. Figures 1 and 2 report the properties of the criterion functions for the two Monte Carlo experiments. Figure 1 is for the case of no time averaging of the data, and Figure 2 is for the case of time averaging. Figures 1(d) and 2(d) report the results using the Wald (approximate quadratic) criteria. The results for the Wald and constrained-minimized criteria are identical for the iterative and two-step estimators. This occurs because the model is linear in the parameters and the weighting matrix is fixed in constructing $(J_T^{tr} - J_T)$. For the continuous-updating estimator, the results for the Wald and the constrained-minimized criteria are different due to the dependence of the weighting matrix on the hypothetical parameter values.

Notice that the small-sample distributions of the minimized criterion functions for the iterative and two-step estimators are greatly distorted. The small-sample tests of the overidentifying restrictions based on the minimized criterion values are too large, leading to overrejections of the model when using these estimators. The minimized criterion function for the continuous-updating estimator is much better behaved, and the small-sample distribution is very close to being χ^2 for both Monte Carlo experiments. Tests of the overidentifying restrictions of the models using the minimized value of the criterion function of the continuous-updating estimator have the correct size for the model without time averaging and similarly for the model with time averaging for probability values less than about .1. Even for probability values greater than .1, the distribution of the minimized criterion for the continuous-updating estimator is not greatly distorted.

The finite-sample coverage probabilities of the three ways of constructing confidence regions for γ are depicted in subplots (b), (c), and (d) in Figures 1 and 2. Recall that the constrained-minimized and Wald criteria coincide when they are based on the two-step and iterative estimators but differ when the continuous-updating estimator is used. The small-sample coverage probabilities are greatly distorted for the intervals constructed with the iterative and two-step estimators in all cases. In particular, they do not contain the true parameter values as often as is to be expected from the limiting distribution. In the case of the continuous-updating estimator, for low p values coverage rates are too small for confidence intervals built from the Wald criteria, but the distortion is substantially smaller than with the other two estimators. Finally, the coverage rates of the confidence regions implied by the true and constrained-minimized criteria for the continuous-updating estimator accord well with the asymptotic distribution and are clearly better than the coverage rates for the other estimators.

These Monte Carlo results for the lognormal model support the following remedy for the concerns raised by Nelson and Startz (1990). From the standpoint of hypothesis testing and confidence-interval construction, use of the continuous-updating criterion is much more reliable than the other methods we study. The tests of the overidentifying restrictions based on the continuous-updating estimator do not reject too often and in fact are quite well approximated by the limiting distribution. Although confidence intervals based on the Wald criteria can be badly distorted, particularly for the two-step and iterative estimators, confidence regions constructed from the continuous-updating criteria have coverage probabilities that are close to the ones implied by the asymptotic theory. This occurs for confidence sets based on both the constrained-minimized criterion in panel (c) and the true criterion in panel (b). The latter result supports the recommendation of Stock and Wright (1995) to base confidence intervals on the level of the continuous-updating criterion.

Parameter Estimates. Table 4 reports summaries of measures of central tendency for the three estimators of γ along with 10% and 90% quantiles. The medians for the two-step and iterative estimators are considerably lower than the true value of γ, whereas the median bias for the continuous-updating estimator is much smaller. The distribution for the continuous-updating estimator, however, is also more dispersed, as evidenced by the larger increment between the 10% and 90% quantiles. Moreover, the Monte Carlo sample means for the continuous-updating estimator are much more severely distorted than they are for the other two estimators. The enormous sample means for the continuous-updating estimator occur because in the case of no time averaging and of time averaging there were 23 and 31 samples, respectively, in which the estimates are, in absolute value, larger than 100. When these are removed from the Monte Carlo samples, the sample means of the continuous-updating estimator are closer to the true values than are the means for the other two estimators.

Recall that the analog to the two-step and iterative estimator in the classical simultaneous-equations model is two-stage least squares and that the analog to the continuous-updating estimator is limited-information (quasi) maximum likelihood. It is known from the literature that there are settings in which the two-stage least squares estimator has finite moments, but the limited-information maximum likelihood estimator does not (e.g., see Mariano and Sawa 1972; Sawa 1969). In light of these theoretical results and our Monte Carlo findings, the continuous-updating estimator is not an attractive alternative to the other estimators we consider if our bases of comparison are the (untruncated) moment properties [or even relative squared errors as in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

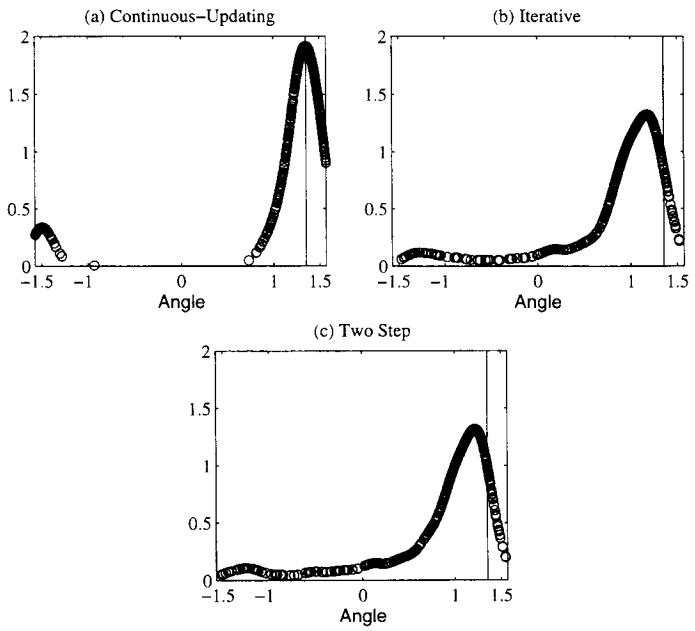

Figure 3. Smoothed Distribution of Estimated Angle Implied by (1, γ_T), Monthly Lognormal Model, No Time Averaging.

the work of Zellner (1978)]. On the other hand, Anderson, Kunitomo, and Sawa (1982) advocated use of the limited information estimator over the two-stage least squares estimator because, among other things, the median bias of the former estimator is smaller. We also find less distortion in the medians for the continuous-updating estimator in our experiments.

As we noted previously, one attractive attribute of the continuous-updating estimator is its invariance to ad hoc (parameter dependent) transformations of the moment conditions. For example suppose that we reparameterize (12) as

$$\gamma_0 \log(c_{t+1}/c_t) + \gamma_1 \log R_{t+1} = \eta_{t+1}, \tag{25}$$

assuming that $\kappa = 0$. In (25) the parameters γ_0 and γ_1 are not uniquely identified by the moment conditions, whereas their ratio is. The parameterization in (12) allows identification of γ_0 by setting $\gamma_1 = 1$. The continuous-updating estimator of γ in (12) is invariant to how this identification is achieved via restrictions on (25). Notice however that the two-step and iterative estimators are sensitive to the chosen normalization. Hillier (1990) achieved identification in a different manner by making the direction from the origin defined by (γ_0, γ_1) in two-dimensional space the object of interest. Although this angle is identified, the length of the ray from the origin along which the true value of (γ_0, γ_1) is not. Hillier's defense of limited information maximum likelihood over conventional two-stage least squares is that the former is a better estimator of direction.

To see whether such a conclusion might well extend to comparisons between the continuous-updating estimator and the other two estimators we consider, we report smoothed distributions of the estimated "direction" in Figures 3 and 4. We measure direction by the angle (as measured in radians) between the horizontal axis and the point $(1, \gamma)$. Because there is still a sign normalization that must be imposed for identification, we restrict attention to the interval $[-\pi/2, \pi/2]$. In smoothing the histogram, we used Gaussian kernel with a bandwidth of .1. The value of the density estimate is plotted at each of the sample points using a circle. The shape of the smoothed distribution along with the mass of the plotted circles provides evidence about the small-sample distribution of the parameter estimators. Notice that the primary modes of the continuous-updating angle estimator are very close to the true parameter values, but the modes of the other two angle estimators are distorted. Moreover, the density estimates for the modal angle are larger for the continuous-updating method. The Monte Carlo distributions for the continuous-updating angle estimator also, however, have secondary modes near $-\pi/2$, corresponding to large-in-magnitude estimates of γ with the wrong sign.

Continuous-Updating Criterion Function. The criterion function for the continuous-updating estimator can sometimes lead to extreme outliers for the minimizing value of γ. This occurs in the two Monte Carlo experiments for some of the trials, as we discussed previously. To see why this can occur, suppose for simplicity that there is a single return under consideration, no time averaging, and several instruments. The moment conditions are constructed using

$$\psi(X_t, g) \equiv [\log R_{t+1} - g \log(c_{t+1}/c_t)] z_t, \tag{26}$$

where z_t is a vector of instruments and $E[\psi(X_t, \gamma)] = 0$ [see Eq. (12)]. Because the moment conditions are linear in g, the criteria for the iterated and two-step estimators are quadratic in g. In contrast, the criterion for the continuous-updating estimator converges as g gets large. To see this, observe that for a large value of g the sample average of $\psi(X_t, g)$ is approximately g times the sample mean of $\log(c_{t+1}/c_t) z_t$ and the sample covariance is approximately

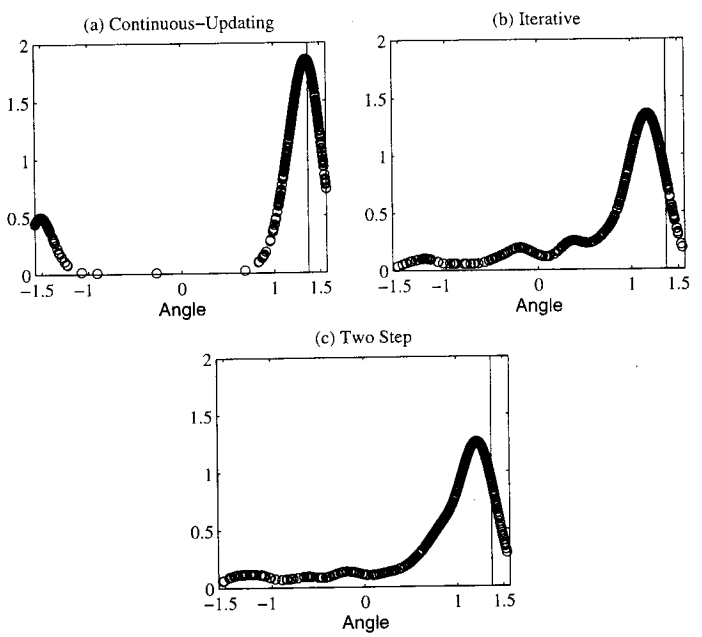

Figure 4. Smoothed Distribution of Estimated Angle Implied by (1, γ_T), Monthly Lognormal Model, Time Averaging.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

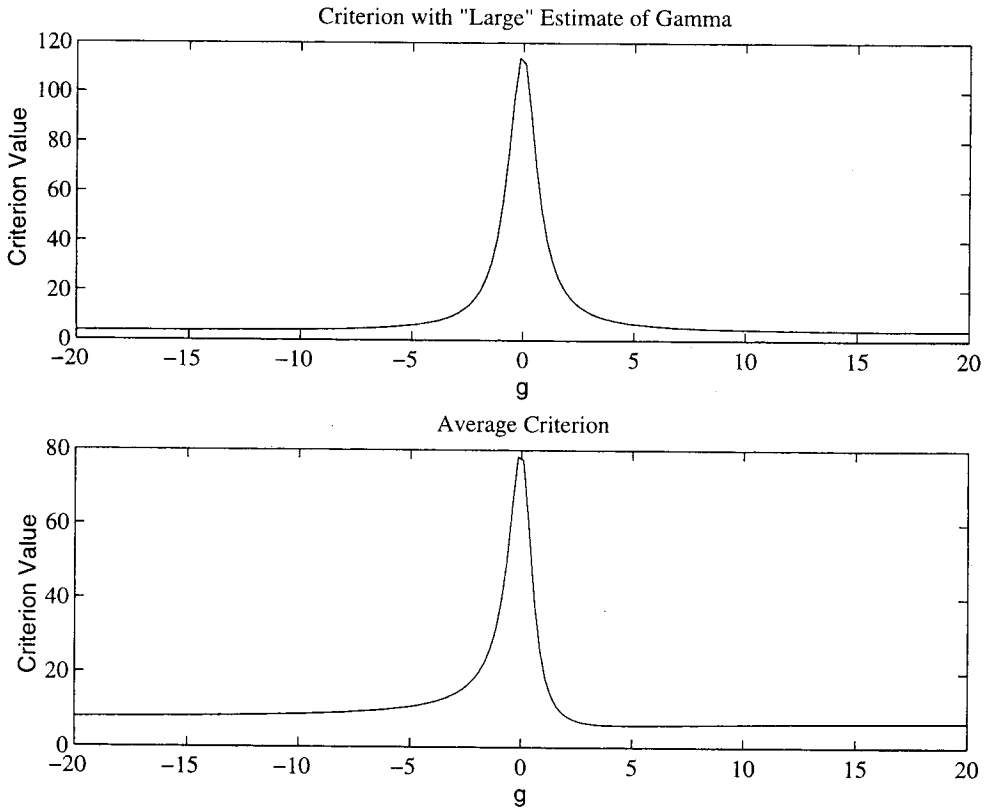

Figure 5. Criterion Function for Continuous-Updating Estimator, Monthly Lognormal Model, No Time Averaging.

g^2 times the sample covariance of $\log(c_{t+1}/c_t)z_t$. Therefore, for large g, the criterion function is approximately a quadratic form that is $(1/T)$ times the chi-squared test statistic for the null hypothesis that

$$E[\log(c_{t+1}/c_t)z_t] = 0. \tag{27}$$

As a result it is possible for the minimized criterion for the continuous-updating criterion to be minimized at a very large value of g.

To further illustrate this potential problem, the upper plot in Figure 5 is of the criterion function for the continuous-updating estimator for a Monte Carlo draw in which the value of g that minimizes the criterion function is 673,750.4. The lower plot in Figure 5 gives the average value of the criterion function over the Monte Carlo experiments. These results are for the case of no time averaging. Notice that in the upper plot the criterion function approaches its lowest value as g becomes large in absolute value. Even in the lower plot the criterion function asymptotes to a local minimum for large negative values of g. The numerical search used to implement the estimator could be complicated by the flat sections of the criterion function and the search routine could end up spuriously searching in the direction of very large values of g. This is particularly true for a gradient-based method (such as fminu.m in MATLAB) in which the routine attempts to set the gradient of the criterion function to 0. When the parameter vector is of low dimension, this problem can easily be assessed by gridding the parameter vector and evaluating the criterion function at the grid points. This is what we did to make sure that large estimates of g were not due to numerical problems. When the parameter vector is of large dimension, however, implementing the continuous-updating estimator may sometimes be difficult.

Estimation of γ and δ. The Monte Carlo environments of the lognormal and Markov-chain models differ in the assumed law of motion for consumption and returns and in the number of estimated parameters. To allow a more direct comparison with the Markov-chain results, we now consider a Monte Carlo experiment with the log-linear law of motion, where both γ and δ are estimated. This second parameter allows us to check whether the favorable performance of the continuous-updating estimator reported in Figures 1 and 2 is unique to the single-parameter setup.

To implement this case, we assume that $\delta = .97^{1/12}$ and that the mean of the (log of) consumption growth is equal to its sample mean. These parameter values, along with the model for $B(L)$ in (13), imply a set of restrictions on the unconditional mean of Y_t given in (11). Monte Carlo samples of Y_t were drawn under these restrictions. To estimate δ and γ we used moment conditions M1 of Table 3.

The results of this Monte Carlo experiment are displayed in Figure 6. As in Figures 1 and 2, rejection rates of the overidentifying restrictions in the case of the continuous-updating estimator are similar to those predicted by the asymptotic distributions, at least for probability values of less than .2. Moreover, as before, the small-sample distributions of the minimized criteria of the two-step and iterative estimators are greatly distorted. In this case, however, a test based on the minimized criterion of the two-step estimator would underreject the restrictions of the model. Coverage

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

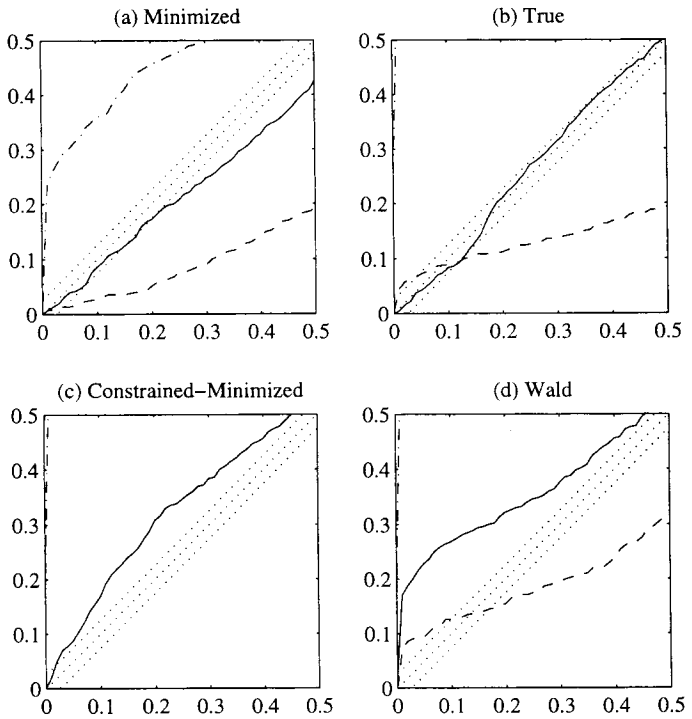

Figure 6. Criterion Functions, Time Additive Model, δ and γ Estimated: —·—·—, Iterative; – – –, Two-Step; ———, Continuous-Updating.

probabilities of the true values of both δ and γ based on the "true" criterion accord well with the asymptotic distribution when using the continuous-updating estimator. The coverage probabilities for the other two estimators, however, are substantially at odds with the theory.

Referring now to panel (c), the coverage rates for the true-constrained criterion are somewhat too small for the continuous-updating estimator. The coverage probabilities are much more seriously distorted, however, for the two-step and iterative estimators just as in Figures 1 and 2. Moreover, confidence intervals based on the true-constrained criterion for the continuous-updating estimator are a little better behaved than those based on the Wald statistic for all three estimators. To summarize, the results displayed in Figure 6 are largely consistent with the results in which γ is the only parameter estimated.

5. MONTE CARLO RESULTS, MARKOV–CHAIN MODELS

5.1 Time-Separable Preferences, Annual Data

First we consider the results for time-separable preferences ($\theta = 0$) using the Markov-chain model calibrated to annual data. For these runs $V_T(b)$ is computed as a simple covariance matrix estimator. The resulting p-value plots are depicted in Figures 7–9 for different combinations of the preference settings and moment sets given in Tables 2 and 3. Features of the Monte Carlo distributions for the point estimates of γ are reported in Table 5.

Results for $\gamma = 1.3$ and $\delta = .97$. We start by discussing the results obtained using the more "moderate" values of the preference parameters TS1 that were used by Tauchen (1986). Figure 7 displays the results for the first set of eight moment conditions M1. Note from panel (b) that the GMM criterion functions evaluated at the true parameter values are larger than predicted by the asymp-

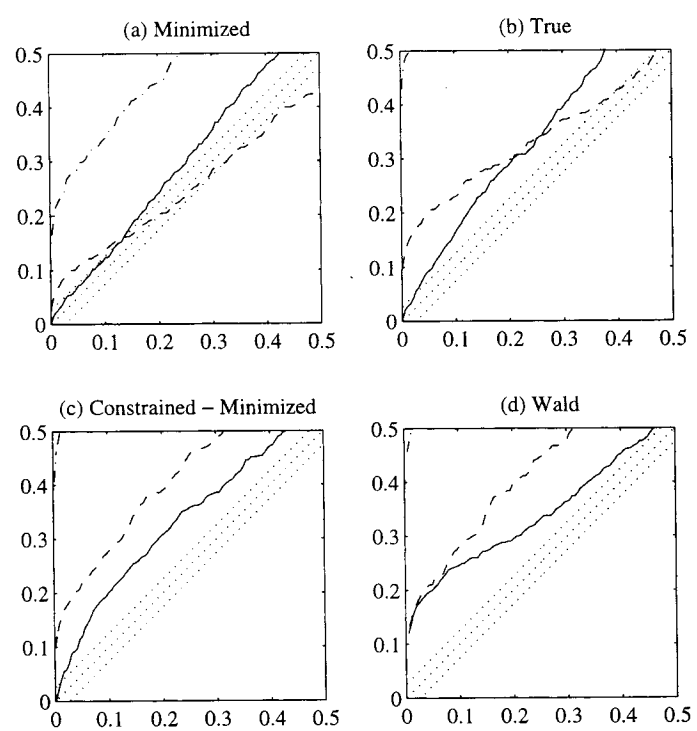

Figure 7. Criterion Functions, Annual Markov-Chain Model, $\delta = .97$, $\gamma = 1.3$, $\theta = 0$, Moment Conditions M1: —·—·—, Iterative; – – –, Two-Step; ———, Continuous-Updating.

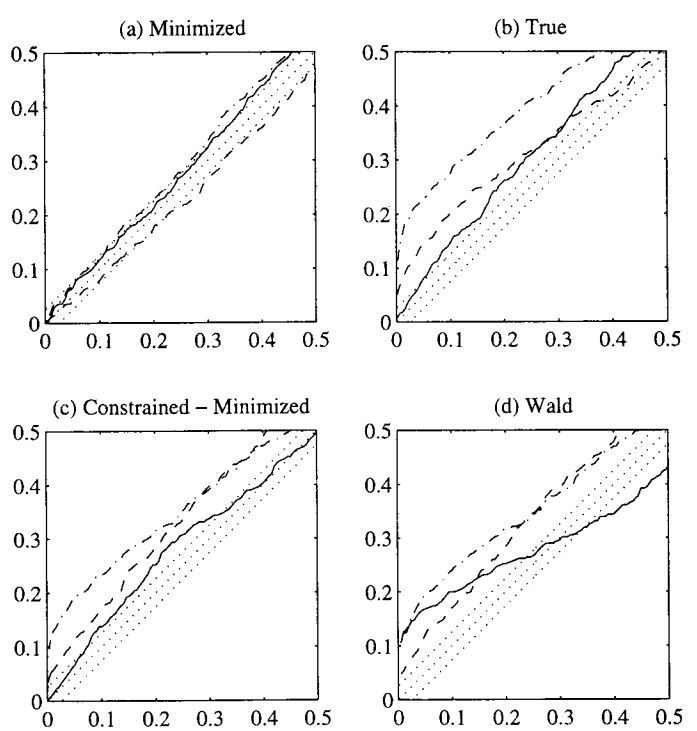

Figure 8. Criterion Functions, Annual Markov-Chain Model, $\delta = .97$, $\gamma = 1.3$, $\theta = 0$, Moment Conditions M2: —·—·—, Iterative; – – –, Two-Step; ———, Continuous-Updating.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

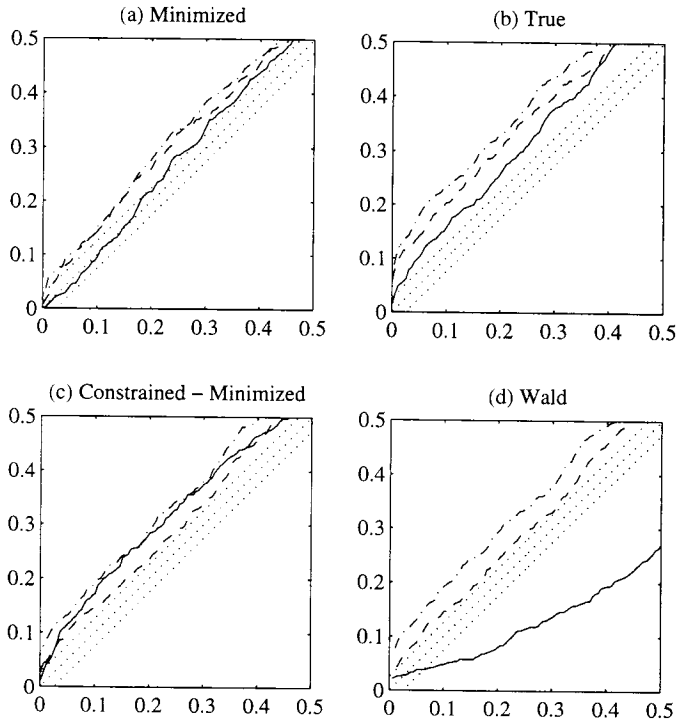

Figure 9. Criterion Functions, Annual Markov-Chain Model, δ = 1.139, γ = 13.7, θ = 0, Moment Conditions M2: —·—·, Iterative; – – –, Two-Step; ———, Continuous-Updating.

totic theory. Although the asymptotic approximations are better with the continuous-updating estimator, at least for smaller probability values, the criterion evaluated at the true parameters is still too large. On the other hand, the minimized criteria functions are much better behaved, especially for the continuous-updating estimator with probability values less than .15 [see Fig. 7(a)]. Confidence intervals based on criteria-function behavior performed poorly in this setting, although they performed better for the continuous-updating estimator than for the other two estimators. Moreover, the criterion-function-based confidence intervals for the continuous-updating estimator proved to be more reliable than the confidence intervals based on the Wald statistic.

Figure 8 reports plots for the set of four moment conditions M2. The reduction in moment conditions (relative to M1) leads to an improvement in the underlying central limit approximations depicted in panel (b). This is especially true for the continuous-updating criterion function except at large probability values. The minimized criterion-function values used to test the overidentifying restrictions behave as predicted by the asymptotic theory for all three estimators. The criterion-function-based confidence intervals work quite well for the continuous-updating estimator but not for confidence intervals based on the Wald statistic [compare panels (c) and (d)].

In summary, asymptotic theory gives a poor guide for statistical inference in the case of moment conditions M1. Presumably this occurs because of the many moment conditions relative to the sample size. For the smaller moment condition set M2, the asymptotic approximations work considerably better. For both moment conditions the asymptotic distributions for the overidentifying restrictions tests and criterion-function-based inference are more reliable when based on the continuous-updating estimator.

Recall that the minimized value of the continuous-updating criterion is always less than or equal to that of the limiting criterion of the iterative estimator. Tauchen (1986) reported cases in which the minimized criterion function for the iterative estimator led to underrejection of the overidentifying restrictions. In these cases the underrejection of the overidentifying restrictions was more pronounced when using the continuous-updating criterion function. See Hansen et al. (1994) for an analysis of one of these cases.

Of course, assessing the reliability of the asymptotic theory as applied to the different parameter estimators is a different question than assessing the performance of the parameter estimators themselves. In regard to this latter question, the results in the first portion of Table 5 show that the continuous-updating estimator has considerably less bias in the median than the other two estimators in the case of M1. The iterative estimator, however, has much less dispersion in this case as measured by the width between the .10 and .90 quantiles. In the case of M2, there is little median bias for all three estimators. On the other hand, the dispersion in the continuous-updating estimator is smaller than that of the other two estimators.

Table 5. Properties of Estimators of γ, Markov-Chain Model, Annual Data

Property	Continuous-updating	Iterative	Two-step
A. TS1-M1, true γ = 1.3, δ = .97			
Mean	1.90	.85	1.08
Truncated mean*	1.90	.85	1.08
Median	1.15	.70	.78
10% quantile	0.61	.13	−1.25
90% quantile	5.76	1.60	4.09
B. TS1-M2, true γ = 1.3, δ = .97			
Mean	1.36	1.35	1.35
Truncated mean*	1.53	1.35	1.35
Median	1.17	1.10	1.25
10% quantile	.61	−1.35	−2.71
90% quantile	4.10	3.51	5.59
C. TS2-M2, true γ = 13.7, δ = 1.139			
Mean	6.29	13.68	14.10
Truncated mean*	14.80	13.68	14.10
Median	14.14	13.10	13.43
10% quantile	8.96	9.01	9.85
90% quantile	22.68	19.40	19.54

* Estimates with absolute values greater than 100 were excluded from the computation of the truncated means.

Results for γ = 13.7 and δ = 1.139. Next we consider results using the preference specification considered by Kocherlakota (1990a) (TS2 in Table 1). We only consider the performance of GMM estimators obtained using moment conditions M2. Our results are displayed in Figure 9 and Table 5. With this change in parameter configuration, the results for the M2 moment conditions are similar to those in Figure 8. In regard to the parameter estimates,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

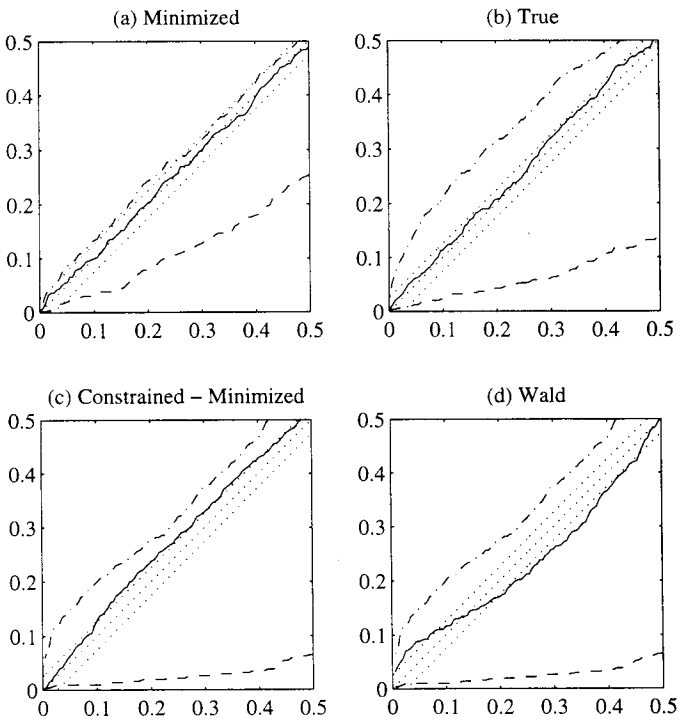

Figure 10. Criterion Functions, Monthly Markov-Chain Model, $\delta = .97^{1/12}$, $\gamma = 1.3$, $\theta = 0$, Moment Conditions M1: —·—·, Iterative; – – –, Two-Step; ———, Continuous-Updating.

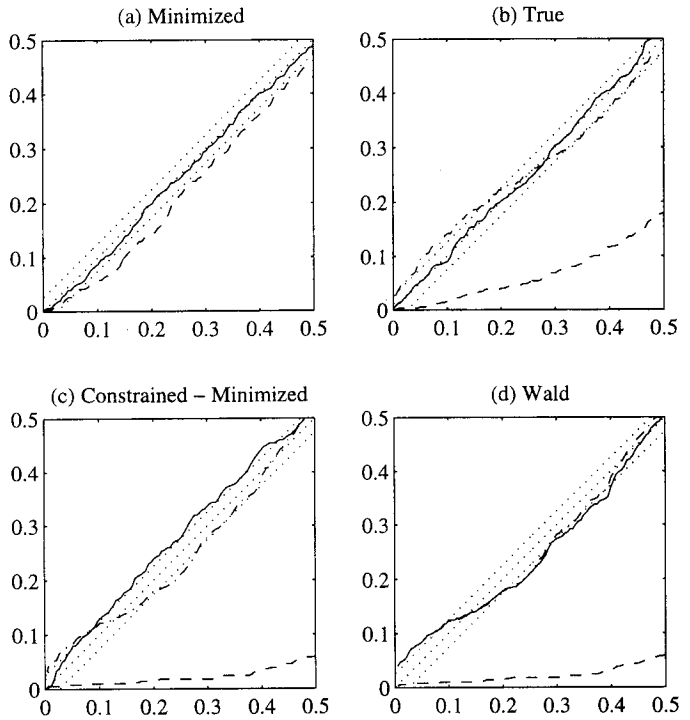

Figure 11. Criterion Functions, Monthly Markov-Chain Model, $\delta = .97^{1/12}$, $\gamma = 1.3$, $\theta = 0$, Moment Conditions M2: —·—·, Iterative; – – –, Two-Step; ———, Continuous-Updating.

again the median bias is small relative to the dispersion for all three estimators. The continuous-updating estimator displays somewhat more dispersion than the other two estimators.

5.2 Time-Separable Preferences, Monthly Data

To explore the extent to which the limiting distribution provides a better guide for inference for larger sample sizes (with less extreme data points), we redid some of our calculations using simulations calibrated to monthly data as described in Subsection 2.2. We focused exclusively on the more "moderate" preference configuration adjusting δ accordingly. In this case we only looked at estimators constructed using moment conditions M1 and M2. We are particularly interested in moment set M1 because of its common use in practice when analyzing postwar data. Our results are reported in Figures 10 and 11 and Table 6. Notice that all of the asymptotic approximations are consistently reliable for the continuous-updating estimation method. In sharp contrast, large-sample inferences for the two-step estimator are of particularly poor quality with the exception of the overidentifying restrictions test using M2. Moreover, it is of note that the iterative estimates and the continuous-updating estimates are very close to one another when M2 is used. This is reflected in quantiles reported in Table 6 as well as in the "Minimized" and "Constrained-Minimized" graphs. Presumably, the reason for this is that the weighting matrix tends to be a relatively "flat" function of the parameters.

In regard to the parameter estimates of γ, both the continuous-updating estimator and the iterative estimators have distributions that are much more concentrated around the true parameter value than the distributions for the two-step estimator (again see Table 6). The performance of the two-step estimator could potentially be improved by using a different weighting matrix in the first step. For example, the residuals from nonlinear two-stage least squares applied to each Euler equation could be used to estimate the asymptotic covariance matrix of the moment conditions. Another possibility is to use the covariance matrix of the prices of the "synthetic" securities implicit in the use of instrumental variables. See Hansen and Jagannathan (1993) for a discussion of this weighting matrix.

Table 6. Properties of Estimators of γ, Markov-Chain Model, Monthly Data: True $\gamma = 1.3$, $\delta = .97^{1/12}$

Property	*Continuous-updating*	*Iterative*	*Two-step*
	A. TS1-M1		
Mean	1.37	1.25	1.57
Truncated mean*	1.37	1.25	1.57
Median	1.28	1.18	1.17
10% quantile	1.02	.96	−2.12
90% quantile	1.85	1.61	4.63
	B. TS1-M2		
Mean	1.38	1.38	1.84
Truncated mean*	1.38	1.38	1.84
Median	1.27	1.27	1.27
10% quantile	1.00	1.00	−2.95
90% quantile	1.89	1.89	6.50

* Estimates with absolute values greater than 100 were excluded from the computation of the truncated means.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

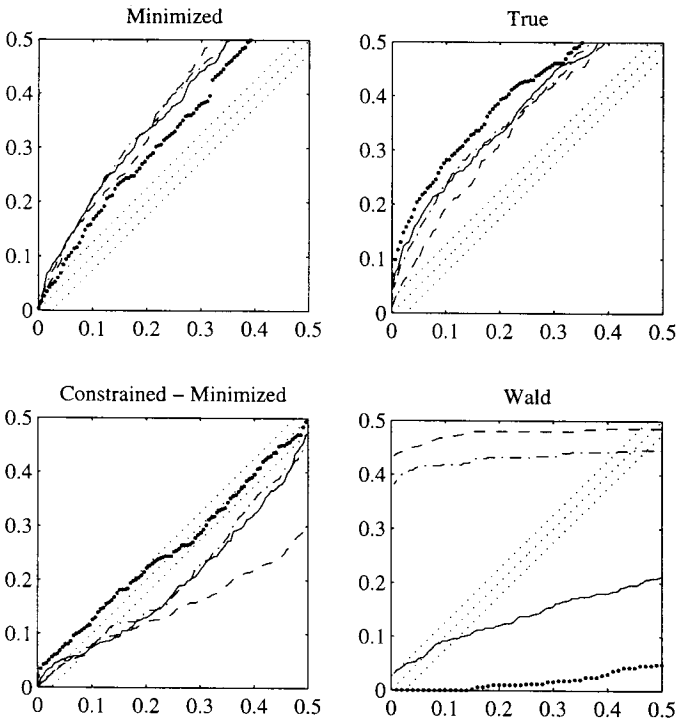

Figure 12. Criterion Functions, Annual Markov-Chain Model, $\delta = .97^{1/12}$, $\gamma = 1.3$, Time Nonseparable, Moment Conditions M2: —·—·, $\theta = \frac{1}{3}$; - - -, $\theta = 0$; ———, $\theta = -\frac{1}{3}$; · · ·, $\theta = -\frac{2}{3}$.

These particular Monte Carlo experiments are ones in which the unconditional moment restrictions provide much more identifying information about the power parameter γ than any of the other experiments we report on. Not only are estimates more accurate than the estimates obtained from the Monte Carlo experiments calibrated to annual data, but also the estimates from the lognormal Monte Carlo experiments reported in Section 3 that have the same sample size. Presumably the main reason for this latter disparity is that the Markov-chain models are calibrated to dividend behavior rather than return behavior. As is well known from the empirical asset-pricing literature, the dividend calibrations imply returns that are less volatile than historical time series because of some fundamental model misspecification.

5.3 Time-Nonseparable Preferences, Annual Data

As we discussed in Subsection 5.1, the continuous-updating estimator generally provides more reliable inference in the case of time-separable preferences when data are generated from the annual Markov-chain model. Even for that estimator, however, it is only when moment conditions M2 are used that the distributions of the criteria TJ_T (minimized), TJ_T^{tr} (true) and $T(J_T^{\gamma} - J_T)$ (constrained-minimized) accord well with the corresponding chi-squared distributions. For these reasons we consider results with time-nonseparable preferences using only moment conditions M2 and only for the continuous-updating estimator. To construct our Monte Carlo datasets we used the three time-nonseparable settings of the parameters listed in Table 2 as TNS1, TNS2, and TNS3 ($\theta = \frac{1}{3}, \theta = -\frac{1}{3}$, $\theta = -\frac{2}{3}$). We also used data generated with $\theta = 0$ but still estimated with the parameter θ. Unlike the case of time-separable preferences when the restriction $\theta = 0$ is imposed in the estimation, the estimator $V_T(b)$ of the asymptotic covariance matrix accommodates an MA(1) structure in the Euler-equation errors. As in Section 3, we used the $V_T(b)$ estimator given in (24) except when it was not positive definite, in which case we shifted to Durbin's (1990) estimator with a fourth-order autoregression.

Figure 12 presents the p-value plots for the criterion functions for the different settings for θ. The p-value plots of Figures 1, 2, 6, and 7–11 considered results for fixed preference parameters and the three different estimators. The same Monte Carlo data were used for the experiments for each estimator in these plots.

In contrast to the time-separable case (the solid line in Fig. 8), the distribution of the minimized criteria imply small-sample overrejection of the moment conditions for each of the settings for θ. Furthermore, even when evaluated at the true parameters, the criterion functions are not distributed as a chi-square. This occurs for all four settings of θ including $\theta = 0$ (time-separable preferences). Evidently the estimator of the asymptotic covariance matrix of the moment conditions, which assumes an MA(1) structure for the errors, causes small-sample distortion of the GMM criterion function. Notice that the distributions of the Wald statistics are very far from being chi-squared. Consistent with the results reported in Section 4 and Subsections 5.1

Table 7. Properties of Estimators of γ and θ, Markov-Chain Model, Annual Data: Continuous-Updating Estimator, True $\gamma = 1.3$, $\delta = .97$

Property	γ	θ
A. TNS1-M2, true $\theta = \frac{1}{3}$		
Mean	4.56	5.99×10^5
Truncated mean*	4.56	.83
Median	1.71	.52
10% quantile	.64	.10
90% quantile	10.83	1.25
B. TS1-M2, true $\theta = 0$		
Mean	5.88	−.34
Truncated mean*	3.67	−.34
Median	.73	−.12
10% quantile	.00	−.92
90% quantile	10.92	.18
C. TNS2-M2, true $\theta = -\frac{1}{3}$		
Mean	3.05	−.46
Truncated mean*	2.69	−.46
Median	.97	−.42
10% quantile	.01	−.93
90% quantile	8.28	−.01
D. TNS3-M2, true $\theta = -\frac{2}{3}$		
Mean	.41	.07
Truncated mean*	2.16	−.64
Median	1.22	−.69
10% quantile	.04	−.94
90% quantile	5.79	−.40

* Estimates with absolute values greater than 100 were excluded from the computation of the truncated means.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

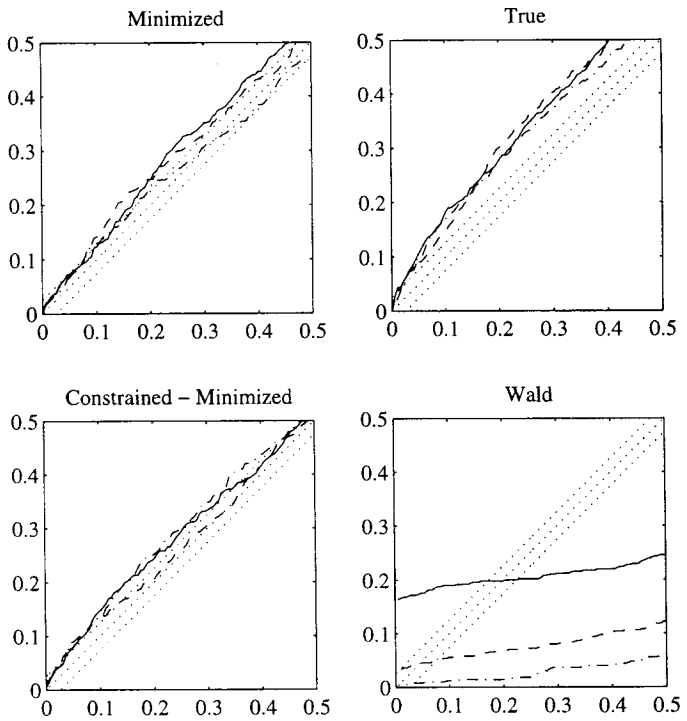

Figure 13. Criterion Functions, Monthly Markov-Chain Model, $\delta = .97^{1/12}$, $\gamma = 1.3$, Time Nonseparable, Moment Conditions M1: —·—·, $\theta = \frac{1}{3}$; – – –, $\theta = 0$; ———, $\theta = -\frac{1}{3}$.

and 5.2, confidence intervals for γ constructed using the criterion function perform much better than those based on the Wald statistic.

Table 7 reports statistics summarizing the properties of the estimators of γ and θ. The estimator of γ does not in general perform as well as in the time-separable preference case (see Table 5 for the comparison). The median for γ_T is substantially below the true value of γ in the case of $\theta = 0$ and $\theta = -\frac{1}{3}$ and above for $\theta = \frac{1}{3}$. Furthermore, the dispersion of the estimators of γ is considerably larger than when time separability is correctly imposed, at least in part due to having to estimate an additional parameter and to accommodate MA(1) terms in the estimator of the asymptotic covariance matrix $V(b)$. Regarding the estimators of θ, there is substantial dispersion for each case as evidenced by the 10% and 90% quantiles. Notice further that there is some median bias in the estimators of θ for the cases of $\theta = \frac{1}{3}, 0$, and $-\frac{1}{3}$ (panels A, B, and C of Table 7). In summary, the annual data do not permit simultaneous estimation of θ and γ with any reasonable precision, at least for moment condition M2.

5.4 Time-Nonseparable Preferences, Monthly Data

Using the monthly Markov-chain model, we also examined the time-nonseparable model for $\theta = \frac{1}{3}, 0$, and $-\frac{1}{3}$ and for moment conditions M2. We further considered moment conditions M1 because these conditions are often used in practice and because the continuous-updating estimators demonstrated reasonable small-sample properties under time-separable preferences. We did not consider the case of $\theta = -\frac{2}{3}$ with the monthly model because the Markov-chain model implies that the covariance matrix of the Euler-equation errors is close to being singular at this value of θ. Once again we used the estimator of $V(b)$ given by (24). When Durbin's (1960) estimator was necessary we used a twelfth-order autoregression.

Figure 13 reports the results for criterion functions using moment conditions M1. Under M1, the minimized criterion performs reasonably well for all three settings of θ. Tests of the overidentifying restrictions of the model have the correct small-sample size in this case. Notice further that the criterion $T(J_T^{\gamma} - J_T)$ (constrained-minimized) is close to being chi-squared distributed but that the distribution of the Wald statistic is very far from chi-squared. Hence, we continue to find that inferences based directly on criterion functions are much more reliable than those based on quadratic approximations to the criterion func-

Table 8. Properties of Estimators of γ and θ, Markov-Chain Model, Monthly Data: Continuous-Updating Estimator, True $\gamma = 1.3$, $\delta = .97^{1/12}$

Property	γ	θ
A. TNS1-M1, true $\theta = \frac{1}{3}$		
Mean	13.56	.35
Truncated mean*	2.90	.35
Median	1.27	.34
10% quantile	.79	.22
90% quantile	7.35	.48
B. TS1-M1, true $\theta = 0$		
Mean	16.68	−.01
Truncated mean*	4.18	−.01
Median	1.29	.01
10% quantile	.38	−.23
90% quantile	13.46	.19
C. TNS2-M1, true $\theta = -\frac{1}{3}$		
Mean	6.03	−.42
Truncated mean*	5.82	−.42
Median	1.24	−.35
10% quantile	.00	−.98
90% quantile	17.95	.09
D. TNS1-M2, true $\theta = \frac{2}{3}$		
Mean	153.96	.66
Truncated mean*	11.09	.66
Median	1.45	.37
10% quantile	.48	.16
90% quantile	410.29	1.16
E. TNS1-M2, true $\theta = 0$		
Mean	62.51	−.14
Truncated mean*	17.89	−.14
Median	1.40	.04
10% quantile	.03	−.66
90% quantile	142.79	.26
F. TNS2-M2, true $\theta = -\frac{1}{3}$		
Mean	12.98	−.40
Truncated mean*	12.78	−.40
Median	.35	−.58
10% quantile	.01	−.91
90% quantile	40.99	.17

* Estimates with absolute values greater than 100 were excluded from the computation of the truncated means.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

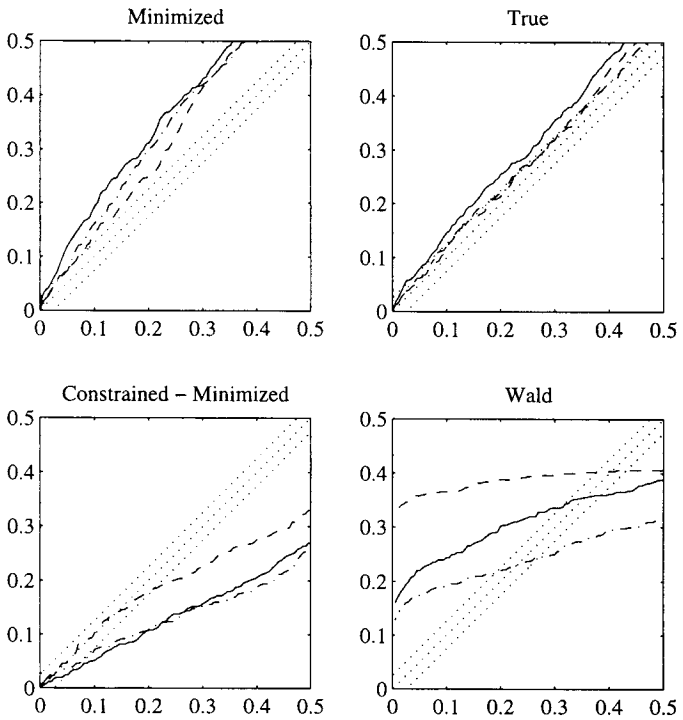

Figure 14. Criterion Functions, Monthly Markov-Chain Model, $\delta = .97^{1/12}$, $\gamma = 1.3$, Time Nonseparable, Moment Conditions M2: —·—·, $\theta = \frac{1}{3}$; – – –, $\theta = 0$; ———, $\theta = -\frac{1}{3}$.

tions. Finally, we report summary statistics of the central tendency of the estimators of γ and θ in Table 8, panels A, B, and C. Lack of prior knowledge of the parameter θ again causes the estimator of γ to be much less precise as measured by the distance between the 90% and 10% quantiles. (For instance, compare the first column of Table 6 to panel B of Table 8.)

Figure 14 presents the p-value plots for moment conditions M2. In this case the distribution of the minimized criterion functions and the constrained-minimized criteria are not chi-squared. The model's overidentifying conditions are rejected too often for all of the parameter settings, and confidence intervals for the parameter γ have the wrong coverage probabilities. With moment conditions M2, allowing for time nonseparability results in a substantial number of very large estimates of γ as reflected by the size of the 90% quantiles (see Table 8, panels C, D, and E). It appears that the addition of returns as instruments (the difference between moment conditions M1 and M2) improves the performance of the estimator and the quality of the central limit approximations.

6. CONCLUDING REMARKS

In this article we examined the finite-sample properties of three alternative GMM estimators that differ in the way in which the moment conditions are weighted. Particular attention was paid to both the performance of tests of overidentifying restrictions and to comparing alternative ways of constructing confidence sets. In documenting finite-sample properties, we used several different specifications of the CCAPM. The experiments differed substantially in the amount of sample information that there is about the parameters of interest. Although the experiments do not uniformly support the conclusion that one estimator dominates the others, some interesting patterns emerged.

1. Continuous updating in conjunction with criterion-function-based inference often performed better than other methods for annual data; however, the large-sample approximations are still not very reliable.
2. In monthly data the central limit approximations for the continuous-updating estimation method applied in conjunction with the criterion-function-based method of inference performed well in most of our experiments, including ones in which the parameters are estimated very accurately and ones in which there is a substantial amount of dispersion in the estimates.
3. Confidence intervals constructed using quadratic approximations to the criterion function performed very poorly in many of our experiments.
4. The continuous-updating estimator typically had less median bias than the other estimators, but the Monte Carlo sample distributions for this estimator sometimes had much fatter tails.
5. The tests for overidentifying restrictions are, by construction, more conservative when the weighting matrix is continuously updated, and in many cases this led to a more reliable test statistic.

Our reason for exploring criterion-function-based inferences and continuous updating is to assess some simple ways of making GMM inferences more reliable. Moreover, when continuous updating is used in conjunction with criterion-function-based inferences, the large-sample approximations become invariant to parameter-dependent transformations of the moment conditions. In this article we have made no attempt to explore the ramifications for power of the resulting statistical tests. Moreover, from the standpoint of obtaining point estimates, we see no particular advantage to using continuous updating when minimizing GMM criterion functions. For example, continuous updating can indeed fatten the tails of the distributions of the estimators. In this sense continuous updating sometimes inherits the defects of maximum likelihood estimators relative to two-stage least squares estimators in the classical simultaneous-equations environment.

Our Monte Carlo experiments for monthly data were sufficiently successful to convince us to reexamine some of the empirical evidence for the CCAPM. In most tests of the CCAPM, the model's overidentifying conditions are rejected (e.g., see Hansen and Singleton 1982). Because the two-step or iterative estimator is typically used in practice, one potential explanation for these rejections could be the tendency of these estimators to result in overrejection of the model in small samples. To assess this possibility we estimated the time-separable and time-nonseparable models using the continuous-updating estimator. We used the consumption and return data described in Subsection 2.1 along with moment conditions M1 given in Table 3.

Estimation of the time-separable model resulted in point estimates of δ and γ of .25 and 720.65, respectively. This

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

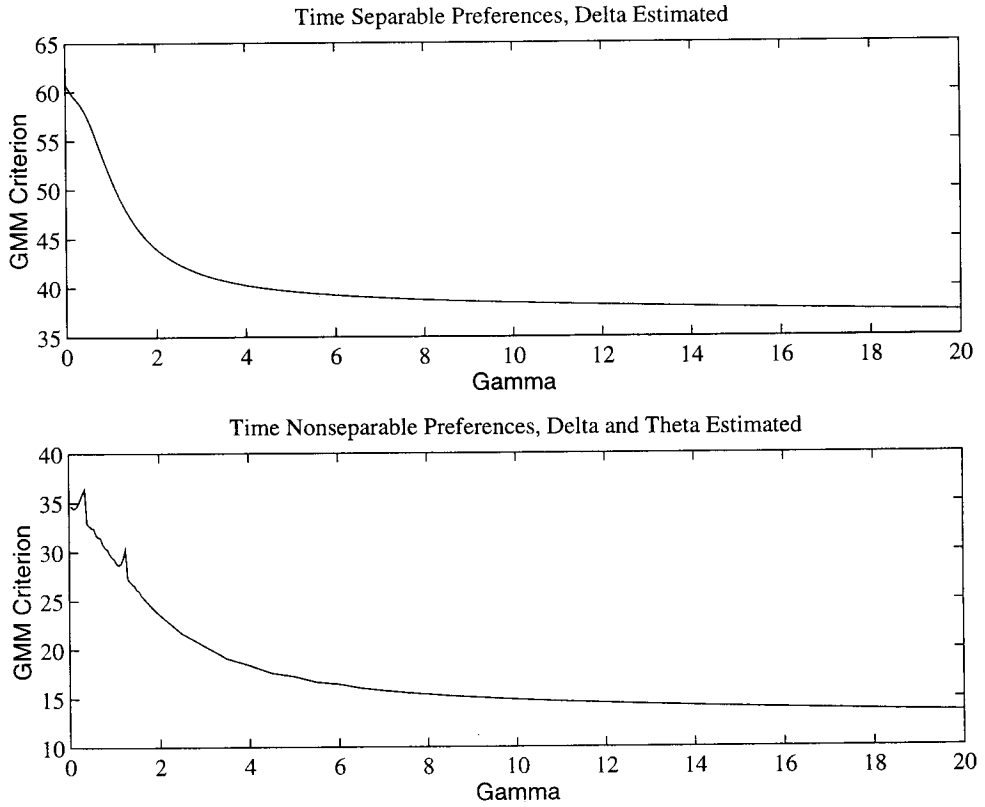

Figure 15. GMM Criterion Functions, Monthly Data, γ Constrained, Moment Conditions M1.

is an example in which the tail behavior of the criterion results in large estimated value of γ. The minimized GMM criterion was 5.94 with an implied p value of .43; hence, it appears that the continuous-updating estimator implies that the model is not at odds with the data. The estimate parameters, however, are very far from being economically plausible. As we found in several of our Monte Carlo experiments, with the continuous-updating estimator, extreme point estimates of the parameters are possible. In those cases, however, there typically was little deterioration in the criterion function when evaluated near the true parameter values, so in practice it is important to evaluate the criterion function at plausible values of the parameters. In this case we restricted γ to the range [0, 20] and estimated δ for each hypothetical value of γ. The resulting minimized criterion as a function of γ is plotted in the top panel of Figure 15. Notice that for this range of γ the minimized criterion function is well above 30, where the implied p value is essentially 0. As a result, once a plausible set of parameters is considered, the model is still rejected when using the continuous-updating estimator.

In estimation of the time-nonseparable model, the point estimates of the parameters were also quite implausible with estimates of δ, γ, and θ of 1.20, 267.96, and .32, respectively. The bottom panel of Figure 15 presents the criterion function for the continuous-updating estimator with γ restricted to the range [0, 20]. At $\gamma = 20$ the criterion reaches a minimum of 13.55 with an implied p value of .035. As a result, even at this extreme value for γ the model is still substantially at odds with the data.

In summary, although the continuous-updating estimator does not save the CCAPM, the experiments that we have presented provide evidence that it should be of use in many GMM estimation environments.

ACKNOWLEDGMENTS

We thank Renee Adams, Wen Fang Liu, Alexander Monge, Marc Roston, George Tauchen, and an anonymous referee for helpful comments and Marc Roston for assisting with the computations. Financial support from the National Science Foundation (Hansen and Heaton) and the Sloan Foundation (Heaton and Yaron) is gratefully acknowledged. Additional computer resources were provided by the Social Sciences and Public Policy Computing Center. Part of this research was conducted while Heaton was a National Fellow at the Hoover Institution.

[Received January 1995. Revised July 1995.]

REFERENCES

Abel, A. (1990), "Asset Prices Under Habit Formation and Catching Up With the Joneses," *A.E.R. Papers and Proceedings*, 80, 38–42.

Anderson, T. W., Kunitomo, N., and Sawa, T. (1982), "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator," *Econometrica*, 50, 1009–1028.

Constantinides, G. (1990), "Habit Formation: A Resolution of the Equity Premium Puzzle," *Journal of Political Economy*, 98, 519–543.

Davidson, R., and McKinnon, J. G. (1994), "Graphical Methods for Investigating the Size and Power of Hypothesis Tests," Discussion Paper 903, Queens University, Dept. of Economics.

Detemple, J., and Zapatero, F. (1991), "Asset Prices in an Exchange Economy With Habit Formation," *Econometrica*, 59, 1633–1657.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Dunn, K., and Singleton, K. (1986), "Modeling the Term Structure of Interest Rates Under Nonseparable Utility and Durability of Goods," *Journal of Financial Economics*, 17, 27–55.

Durbin, J. (1960), "The Fitting of Time Series Models," *Review of the International Statistical Institute*, 28, 233–244.

Eichenbaum, M., and Hansen, L. (1990), "Estimating Models With Intertemporal Substitution Using Aggregate Time Series Data," *Journal of Business & Economic Statistics*, 8, 53–69.

Eichenbaum, M., Hansen, L., and Singleton, K. (1988), "A Time Series Analysis of a Representative Agent Model of Consumption and Leisure Choice Under Uncertainty," *Quarterly Journal of Economics*, 103, 51–78.

Epstein, L., and Zin, S. (1991), "Substitution, Risk Aversion, and Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *Journal of Political Economy*, 199, 263–286.

Ferson, W., and Constantinides, G. (1991), "Habit Formation and Durability in Aggregate Consumption: Empirical Tests," *Journal of Financial Economics*, 29, 199–240.

Ferson, W., and Foerster, S. (1994), "Finite Sample Properties of the Generalized Method of Moments in Tests of Conditional Asset Pricing Models," *Journal of Financial Economics*, 36, 29–55.

Gallant, A., and Tauchen, G. (1989), "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes," *Econometrica*, 57, 1091–1120.

Grossman, S. J., Melino, A., and Shiller, R. (1987), "Estimating the Continuous-Time Consumption-Based Asset-Pricing Model," *Journal of Business & Economic Statistics*, 5, 315–327.

Hall, R. (1988), "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 96, 339–357.

Hansen, L., Heaton, J., and Yaron, A. (1994), "Finite Sample Properties of Some Alternative GMM Estimators," Working Paper 3728-94-EFA, Massachusetts Institute of Technology, Sloan School of Management.

Hansen, L., and Jagannathan, R. (1993), "Assessing Specification Errors in Stochastic Discount Factor Models," unpublished manuscript.

Hansen, L., and Sargent, T. (1991), "Exact Linear Rational Expectations Models," in *Rational Expectations Econometrics*, eds. L. Hansen and T. Sargent, Boulder: Westview Press, pp. 45–75.

Hansen, L., and Singleton, K. (1982), "Generalized Instrumental Variable Estimation 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.

——— (1996), "Efficient Estimation of Linear Asset Pricing Models With Moving-Average Errors," *Journal of Business & Economic Statistics*, 14, 53–68.

Heaton, J. (1993), "The Interaction Between Time-Nonseparable Preferences and Time Aggregation," *Econometrica*, 61, 353–385.

——— (1995), "An Empirical Investigation of Asset Pricing With Time Nonseparable Preferences," *Econometrica*, 63, 681–717.

Hillier, G. (1990), "On the Normalization of Structural Equations: Properties of Direction Estimators," *Econometrica*, 58, 1181–1194.

Kocherlakota, N. R. (1990a), "On Tests of Representative Consumer Asset Pricing Models," *Journal of Monetary Economics*, 26, 285–304.

——— (1990b), "A Note on the 'Discount' Factor in Growth Economies," *Journal of Monetary Economics*, 25, 43–48.

Magdalinos, A. M. (1994), "Testing Instrument Admissibility: Some Refined Asymptotic Results," *Econometrica*, 62, 373–405.

Mariano, R. S., and Sawa, T. (1972), "The Exact Finite-Sample Distribution of the Limited-Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables," *Journal of the American Statistical Association*, 67, 159–163.

Nelson, C. R., and Startz, R. (1990), "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument Is a Poor One," *Journal of Business*, 63, 125–164.

Novales, A. (1990), "Solving Nonlinear Rational Expectations Models: A Stochastic Equilibrium Model of Interest Rates," *Econometrica*, 58, 93–111.

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

Ryder, H. E., Jr., and Heal, G. M. (1973), "Optimal Growth With Intertemporally Dependent Preferences," *Review of Economic Studies*, 40, 1–31.

Sargan, J. (1958), "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415.

Sawa, T. (1969), "The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimator," *Journal of the American Statistical Association*, 64, 923–937.

Stock, J., and Wright, J. (1995), "Asymptotics for GMM Estimators With Weak Instruments," unpublished manuscript, Harvard University, Kennedy School.

Sundaresan, S. (1989), "Intertemporally Dependent Preferences and the Volatility of Consumption and Wealth," *Review of Financial Studies*, 2, 73–89.

Tauchen, G. (1986), "Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained From Financial Market Data," *Journal of Business & Economic Statistics*, 4, 397–425.

Tauchen, G., and Hussey, R. (1991), "Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models," *Econometrica*, 59, 371–396.

Zellner, A. (1978), "Estimation of Functions of Population Means and Regression Coefficients Including Structural Coefficients; A Minimum Expected Loss (MELO) Approach," *Journal of Econometrics*, 8, 127–158.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.