Working Paper Series

Hypothesis Testing and Finite Sample Properties of Generalized Method of Moments Estimators: A Monte Carlo Study

WP 90-12

Ching-Sheng Mao
Federal Reserve Bank of Richmond

THE FEDERAL RESERVE BANK OF RICHMOND
RICHMOND ▪ BALTIMORE ▪ CHARLOTTE

Working Paper 90-12

Hypothesis Testing and Finite Sample Properties of Generalized Method of Moments Estimators: A Monte Carlo Study

Ching-Sheng Mao

Federal Reserve Bank of Richmond

November 1990

1. Introduction

Econometric methods based on the first-order conditions of intertemporal optimization models have gained increasing popularity in recent years. To a large extent, this development stems from the celebrated Lucas critique, which argued forcibly that traditional econometric models are not structural with respect to changes in the economic environment caused by policy regime shifts. The generalized method of moments (GMM) estimation procedure developed by Hansen (1982) is a leading example of a large research program in estimating parameters of taste and technology that are arguably invariant to shifts in policy rules. This estimation procedure has been used by many researchers to estimate nonlinear rational expectations models and has had a major impact on the practice of macroeconometrics.

In this paper I set out to examine the finite sample properties of GMM estimators using conventional Monte Carlo simulation techniques. This study is motivated by the fact that little is known about the performance of the GMM estimation in a small sample setting. Most of the desirable properties of GMM estimators are based on large sample approximations. There does exist some work on similar problems (Tauchen (1986) and Kocherlakota (1988)). The current study differs from these previous studies in several important aspects. First, the model I use to assess the performance of GMM estimators involves not only the saving decision of a representative agent but the leisure decision as well. Previous studies have abstracted from the leisure decision. The second distinct feature of this paper concerns the way that random data are generated. Here, an equilibrium business cycle model was utilized to simulate an artificial economy in which the production technology and the forcing process are explicitly specified. This model has been widely used in the real business cycle literature to calibrate the U.S. economy. Our approach is different from the previous studies in which random data were generated from an endowment (barter) economy.

It is well known that Monte Carlo experiments have serious limitations. In particular, the sampling results they generate can only be applied to the parameter values that were considered by the experiments because the sampling distribution of the estimates may depend in a nontrivial way on the parameter values. With this caution in mind, our experiments were carried out along a number of dimensions in order to make the experimental results as robust as possible. First, the parameter governing intertemporal substitution or relative risk aversion was varied over a relatively wide range of values. This parameter has been very difficult to pinpoint in empirical studies (see Eichenbaum, Hansen, and Singleton (1983), Hall (1989), and Hansen and Singleton (1982, 1983, 1988)). Second, the parameter governing the persistence of the random shock was also varied in order to understand the sensitivity of the sampling distribution of the estimates to shifts in the forcing process. Thirdly, the estimation was performed using various sample sizes, ranging from 50 to 500 observations. The case of 500 observations allows us to see how the asymptotic properties of GMM estimators hold up in this seemingly large sample environment. As it turned out, a sample of this size may not be large enough for making correct inferences. Finally, the estimation was performed using different numbers of lags in forming the instrumental vector. As will be seen, the performance of the GMM estimation and the associated specification test is fairly sensitive to this parameter.

These experiments produce a number of results that are quite different from those of previous studies. Perhaps the most striking finding is that the GMM specification test tends to over-reject the true model even for large samples. This is particularly so when relative risk aversion is high and the lag lengths used to form instruments are relatively large. For moderate sample sizes (say, 300 observations) the rejection rates of the model can be as high as 30% at the significance level of 5%. The poor performance of this specification test is mainly due to the asymptotic sampling bias. In fact, our experiments show that the disparities between the sampling distribution of the test statistic and the asymptotic distribution are fairly substantial. The tails of these sampling distributions are much thicker than those of the asymptotic distribution. In many instances, the objective function which the GMM estimation procedure tries to minimize is also ill behaved. This result explains why the specification test tends to over-reject the model.

Our experiments also indicate that the test results are very sensitive to the lag lengths used to form instruments. Specifically, as the number of lags used to form instruments decreases, the approximations of the objective function and the test statistic are much more accurate, and the performance of the test improves by a significant margin. The rejection rates in these cases are close to the corresponding significance levels, particularly, for large sample sizes. This last result is consistent with that of Tauchen (1986) and strongly suggests that whenever possible shorter lag lengths should be used to form instruments.[1]

The rest of the paper is organized as follows. The next section briefly describes a representative agent model and derives the Euler equations. The GMM estimation procedure is reviewed in section 3. In section 4 I discuss the data generating process and perform a statistical test on the accuracy of the simulated data. The experiment results are discussed in section 5. The final section contains a brief summary and conclusions.

2. The Model

This section lays out a prototype representative agent model which has

---

[1] My results are somewhat different from those of Tauchen who found that the test tends to under-reject the model.

received considerable attention in macroeconomics. This model provides the basic framework in which the GMM estimator will be assessed in this paper.

The economy is assumed to be populated by a large number of identical and infinitely lived consumers. At each date t the representative consumer values service from consumption of a single commodity c_t and leisure l_t. Preferences are assumed to be represented by a constant relative risk aversion (CRRA) utility function:

$$u(c_t, l_t) = \begin{cases} \frac{1}{1-1/\sigma} \left\{ \left[c_t^{\theta} \, l_t^{(1-\theta)} \right]^{1-1/\sigma} - 1 \right\}, & \text{if } \sigma > 0 \text{ and } \sigma \neq 1, \\ \theta \ln c_t + (1-\theta) \ln l_t, & \text{if } \sigma = 1, \end{cases}$$

where $\theta \in (0,1)$. The parameter σ has the interpretation of the intertemporal elasticity of substitution with respect to a "composite good" defined as a Cobb-Douglas function of c_t and l_t.[2]

At each date t the consumer is endowed with a pre-determined capital k_t and earns capital income $r_t k_t$, where r_t is a stochastic one period return (or rental rate) in consumption units. The consumer also receives in period t a wage income $w_t n_t$, where n_t represents hours worked and w_t is the stochastic market-determined wage rate. The total income is divided between consumption and investment so that the budget constraint at time t is $c_t + k_{t+1} - (1-\delta)k_t \leq w_t n_t + r_t k_t$. It is assumed that capital depreciates at the rate $0 \leq \delta \leq 1$ and the agent is endowed with one unit of time each period.

The consumer chooses a sequence of consumption and labor supply, taking prices as given, so as to maximize expected lifetime utility subject to a set of intertemporal budget constraints. Formally, the consumer's problem is

---

[2] The CRRA utility function has been widely used in the real business cycle literature (see, for example, King, Plosser, and Rebelo (1988) and Kydland and Prescott (1983)). Because of its popularity, many empirical studies have attempted to estimate the parameters of this utility function. Note that the inverse of σ is a measure of relative risk aversion.

$$\max_{\{c_t, k_{t+1}, l_t, n_t\}} E_0\left[\sum_{t=0}^{\infty} \beta^t \, u(c_t, l_t) \right], \qquad 0 < \beta < 1,$$

subject to

$$c_t + k_{t+1} - (1-\delta)k_t \le r_t k_t + w_t n_t, \quad \text{for all } t,$$

$$l_t + n_t = 1, \quad \text{for all } t,$$

where β is the time preference discount factor. The information set at time t over which expectations are taken is assumed to contain current and past values of all decision variables and prices. Consumers, however, do not know future wage rates and rental rates.

The Euler equations that characterize the consumer's equilibrium are

$$u_l(c_t, l_t)/u_c(c_t, l_t) = w_t, \tag{1}$$

$$u_c(c_t, l_t) = \beta \, E_t\left[u_c(c_{t+1}, l_{t+1}) \, (1 - \delta + r_{t+1}) \right], \tag{2}$$

where u_c and u_l represent the marginal utility of consumption and leisure, respectively. These two equations have the usual interpretations. Equation (1) states that the rate of substitution between consumption and leisure in a given period must equal the cost of leisure, which is the real wage rate. Equation (2) implies that in equilibrium the consumer is indifferent between consuming one extra unit of goods today and investing it in the form of capital and consuming tomorrow.

Given the assumed CRRA utility function, (1) and (2) imply the following relation:

$$E_t\left[\beta(c_{t+1}/c_t)^{-1/\sigma} \, (1 - \delta + r_{t+1}) \, (w_{t+1}/w_t)^{(1-\theta)(1/\sigma-1)} - 1\right] = 0. \tag{3}$$

This expression is obtained by solving leisure from (1) and substituting into (2). It states that the expectation of the Euler equation residual ε_{t+1} (i.e., the term defined in the bracket) is zero, conditional on information at time t. That is, any variables contained in the information set I_t should be uncorrelated with ε_{t+1}. These restrictions are commonly referred to by economists as the orthogonality conditions.

The moments restrictions implied by equations like (3) are the building blocks for constructing a large class of instrumental variables estimators of the underlying parameters of the utility function. Many of these estimation procedures, such as the maximum likelihood or two stage least squares (2SLS), require additional distributional assumptions which may not be true.[3] The generalized method of moments procedure proposed by Hansen (1982) does not require such ad hoc restrictions and has gained increasing popularity for estimating nonlinear rational expectations models.

3. The GMM Estimation Procedure

In this section I describe the GMM estimation using the above model as an example. The discussion follows closely that of Hansen and Singleton (1982). The main purpose here is to fix notation, so the discussion will be brief. A rigorous treatment of this subject can be found in Hansen (1982).

Suppose an econometrician who observes time series data $\{\underline{x}_t; t = 1, \cdots, T\}$ on consumption c_t, the rate of return on capital $(1 - \delta + r_t)$, and the real wage rate w_t, wishes to estimate the parameter vector $\underline{\gamma} = [\beta\ \sigma\ \theta]'$.[4] It is assumed that the joint process of $\{\underline{x}_t\}$ is stationary. Let $\varepsilon_{t+1} = h(\underline{x}_{t+1}, \underline{\gamma})$ be the residual defined in (3) and $\underline{z}_t$ a $(q\times 1)$ vector of variables contained in the information set I_t, with q being greater than or equal to the number

---

[3] For example, Hansen and Singleton (1983) estimated an asset pricing model using a maximum likelihood procedure. The same model was estimated by Hall (1989) using a 2SLS procedure. These studies assumed that the logarithm of consumption growth and asset returns follow a joint Gaussian process. As pointed out by Hansen and Singleton (1982), the maximum likelihood estimator is biased and inconsistent if this assumption is false. The finite sample properties of 2SLS estimators were studied by Mao (1989).

[4] For simplicity, both consumers and the econometrician are assumed to observe the gross rate of return on investment so that the depreciation rate δ needs not be estimated.

of parameters to be estimated. As discussed before, the Euler equation (3) implies a set of q population orthogonality conditions:

$$E[g(\underline{x}_{t+1}, \underline{z}_t, \underline{\gamma}_0)] \equiv E[h(\underline{x}_{t+1}, \underline{\gamma}_0)\ \underline{z}_t] = \underline{0}, \tag{4}$$

where E is the unconditional expectation operator and $\underline{\gamma}_0$ is a vector of the true parameter values. The $\underline{z}_t$ used to form the product in equation (4) are the instruments for the estimation. In the context of our example, the vector $\underline{z}_t$ usually includes a constant term and current and lagged values of the rate of growth in consumption, asset returns, and the rate of change in the real wage rate.

Condition (4) is the basis for constructing the GMM estimator of $\underline{\gamma}_0$. It is obtained by choosing the value of $\underline{\gamma}$ that makes the sample counterpart of the population orthogonality conditions "close" to zero. Specifically, let the sample average of the function g be given by

$$\bar{g}_T(\underline{\gamma}) = (1/T) \sum_{t=1}^{T} g(\underline{x}_{t+1}, \underline{z}_t, \underline{\gamma}),$$

where the subscript T indicates that the value of the function depends on the sample size. Note that $\bar{g}_T(\underline{\gamma})$ is a method of moments estimator of $E[g(\underline{x}_{t+1}, \underline{z}_t, \underline{\gamma})]$. Since $\bar{g}_T(\underline{\gamma}) \longrightarrow E[g(\underline{x}_{t+1}, \underline{z}_t, \underline{\gamma})]$ almost surely in $\underline{\gamma}$, and from (4) $E[g(\underline{x}_{t+1}, \underline{z}_t, \underline{\gamma}_0)] = 0$, the value of the function $\bar{g}_T(\underline{\gamma})$, evaluated at $\underline{\gamma} = \underline{\gamma}_0$, should be close to zero for a sufficiently large sample size T. Using this fact, the GMM estimator $\hat{\underline{\gamma}}$ of $\underline{\gamma}_0$ can be obtained by minimizing the quadratic form J_T given by

$$J_T(\underline{\gamma}) = \bar{g}_T(\underline{\gamma})' W_T \bar{g}_T(\underline{\gamma}), \tag{5}$$

where W_T is a $q \times q$ symmetric, positive definite matrix that satisfies $W_T \longrightarrow W$ almost surely, and W is a symmetric and nonsingular constant matrix. The choice of the weighting matrix W_T, which can depend on sample information, defines a metric that makes $\bar{g}_T$ close to zero.

Hansen (1982) showed that under regularity conditions the estimator $\hat{\underline{\gamma}}$ constructed in this way is a consistent estimator of $\underline{\gamma}_0$ with an asymptotic variance-covariance matrix that depends on the choice of the weighting matrix W_T. Hansen also showed that it is possible to select "optimally" a weighting matrix that minimizes (in the matrix sense) the limiting covariance matrix of $\hat{\underline{\gamma}}$. This smallest asymptotic variance-covariance matrix is given by[5]

$$\Sigma^* = \left\{ E\left[\frac{\partial h}{\partial \underline{\gamma}}(\underline{x}_{t+1}, \underline{\gamma}_0)\ \underline{z}_t\right]' W^*\ E\left[\frac{\partial h}{\partial \underline{\gamma}}(\underline{x}_{t+1}, \underline{\gamma}_0)\ \underline{z}_t\right] \right\}^{-1}, \tag{6}$$

where $W^* = (E[g(\underline{x}_{t+1}, \underline{z}_t, \underline{\gamma}_0) g(\underline{x}_{t+1}, \underline{z}_t, \underline{\gamma}_0)'])^{-1}$, which is the inverse of the variance-covariance matrix of the random variable $g(\underline{x}_{t+1}, \underline{z}_t, \underline{\gamma}_0)$. Since both W^* and Σ^* depend on the unknown parameters, they must be estimated.

In order to obtain consistent estimates of the weighting matrix W^* and the covariance matrix Σ^*, Hansen and Singleton (1982) implemented a two-step procedure which will be followed in this paper. Initially, a 2SLS weighting matrix $(\Sigma \underline{z}_t' \underline{z}_t)^{-1}$ is employed to obtain the first step estimates of $\underline{\gamma}_0$. These estimates are used to construct a consistent estimate of W^*,[6] which is then used in the second step to obtain final estimates of $\underline{\gamma}_0$ and Σ^*.

The GMM estimation provides a convenient test of the overidentifying restrictions implied by the model. In particular, Hansen (1982) showed that

---

[5]The expression (6) is the smallest asymptotic variance-covariance matrix of the GMM estimators among all possible choices of weighting matrices, holding constant the sequence of instruments. This expression, therefore, depends on the choice of instruments. Hansen (1985) developed a method to calculate the greatest lower bound of this covariance matrix as the instruments vary over an admissible set. This method can be used to select instruments that yield the (asymptotically) smallest covariance matrix. Tauchen (1986) studied the properties of the GMM estimator using this optimal procedure. This technique is not considered in this paper.

[6]The weighting matrix can be estimated by the inverse of the sample variance covariance matrix of the random variable g evaluated at the initial estimates of the parameter vector. This estimate may be modified to account for the autocorrelation of the disturbance term (see Hansen and Singleton (1982)). My estimation procedure sets the order of the moving average term to zero.

the statistic $TJ_T(\hat{\underline{\gamma}})$, which is the sample size times the minimized value of the objective function, is distributed asymptotically as a chi-squared random variable with degrees of freedom equal to the dimension of $g(\underline{x}_{t+1},\underline{z}_t,\underline{\gamma})$ less the number of parameters being estimated. This statistic has been used in many studies to test the overall specification of the underlying economic model. One of the objectives of this paper is to understand the behavior of this statistic and to evaluate the consequences of this specification test in a finite sample environment.

4. The Data Generating Process

The model I used to generate artificial data is a standard real business cycle model which has received considerable attention in the literature. In this section I briefly outline this model and describe a numerical method to obtain equilibrium solutions.

Consider the following optimization problem of a central planner:

$$\max_{\{c_t,k_{t+1},l_t,n_t\}} E_0\left[\sum_{t=0}^{\infty} \beta^t u(c_t,l_t)\right], \qquad 0 < \beta < 1,$$

subject to

$$c_t + k_{t+1} - (1-\delta)k_t \le \lambda_t F(k_t,n_t), \quad \text{for all } t,$$

$$l_t + n_t = 1, \quad \text{for all } t,$$

where $F(k_t,n_t)$ is a constant returns to scale technology and λ_t is a positive random shock. This model is identical to the consumer's problem except that income is generated from an endogenous production process. Solutions of this optimization problem will be used below to generate time series data for the sampling experiments. Note that in equilibrium the rental rate and the wage rate equal the marginal product of capital and labor, which can be used to generate data on prices once the model is solved.

As in King, Plosser, and Rebelo (1988), I assume that the technology is given by a Cobb-Douglas production function, i.e., $\lambda_t F(k_t, n_t) = \lambda_t k_t^{\alpha} n_t^{1-\alpha}$, where $\alpha \in (0, 1)$. Also, the technology shock is assumed to follow a discrete stationary Markov process with a transition matrix that is structured in such a way to yield a first-order autoregressive representation of the process, i.e., $\ln \lambda_{t+1} = \rho \ln \lambda_t + u_{t+1}$, where $\rho \in (0, 1)$ and u_{t+1} is an i.i.d. random disturbance. Using a technique proposed by Rebelo and Rouwenhorst (1989), I employed a five-state Markov chain to approximate this AR(1) process.[7]

A discrete dynamic programming algorithm will be used to solve the above maximization problem. This method has recently been applied to solve a large class of dynamic models where equilibria are optimal (e.g., Christiano (1989) and Rebelo and Rouwenhorst (1989)). Details of this method can be found in Bertsekas (1976) and will not be presented here. Basically, this numerical method approximates the policy functions for capital and labor on a finite number of grid points over the state space. Starting from an initial guess, the numerical procedure iterates on the value function of the problem using the conventional successive approximation algorithm (see Bertsekas (1976)). The equilibrium solutions for capital and labor are obtained when the value function converges to a fixed point. Once the capital stock and labor hours are solved, other quantities and prices can be derived. These solutions can then be used to generate pseudodata for the sampling experiments.

The values of parameters used in solving the model are $\alpha = 0.3$, $\delta = 0.1$, $\beta = 0.96$ and $\theta = 0.3$. These parameters will be held constant throughout the experiment. Two parameters that are of special interest are ρ and σ, which will be varied in solving the model. The benchmark value for ρ is set to 0.9

---

[7] It should be pointed out that the number of states used for the shock is not important for the results of this paper. I assume that the technology shock lies on five distinct points over a bounded interval. The mean and variance of the log of the shock are 0 and 0.001, respectively. The benchmark value for ρ is set to 0.9. These figures are well within the values used in the real business cycle literature.

and later changed to 0.0. The parameter σ will take four different values, i.e., σ = 0.1, 0.5, 1.0, and 2.5.

It is essential that the artificial data constructed in this fashion are reasonably accurate. To ensure this accuracy, I adopted 2500 grid points for the capital stock. These grids were defined over the ergodic set of capital in order to improve accuracy.[8] In addition, a statistical test suggested by Den Haan and Marcet (1989) was performed to test whether the simulated data satisfy the orthogonality conditions implied by the Euler equation (3).[9] For each case under consideration a sample path containing 3000 observations were generated from the model and the statistic calculated. The results of this experiment are given in Table 1 where the probability value (i.e., the tail area) of the test statistic is indicated in the parenthesis. It is clear form this table that the values of the chi-squared statistic are small and statistically insignificant, indicating that the orthogonality conditions are satisfied by the solutions. These results justify use of the simulated data in the GMM estimation, which is based on the same set of restrictions tested by the above statistical procedure.

---

[8] Intuitively, the ergodic set of capital is a set of numbers $\{k \mid \underline{k} \le k \le \bar{k}\}$ such that its complement has probability zero. This means that once capital falls into this set, it stays there forever and never moves out. Restricting capital over this smaller set avoids wasting grid points and thereby improves accuracy. In my numerical procedure the ergodic set is approximated by first solving the problem over the feasible set using coarse grids. The implied ergodic set is then used in the second run to define a new range for capital. The process continues until the number of grids contained in the ergodic set exceeds 90 percent of grids being used. For more discussion on the concept of ergodicity, see Brock and Mirman (1972) and Sargent (1980).

[9] The Den Haan-Marcet statistic for testing the orthogonality condition (4) is $m = \tilde{\underline{\gamma}}'[\Sigma \underline{z}_t'\underline{z}_t][\Sigma \underline{z}_t'\underline{z}_t \varepsilon_{t+1}^2]^{-1}[\Sigma \underline{z}_t \underline{z}_t']\tilde{\underline{\gamma}}$, where $\tilde{\underline{\gamma}}$ is a vector of OLS estimates in a regression of ε_{t+1} on the instruments $\underline{z}_t$. This statistic has an asymptotic chi-squared distribution with degrees of freedom equal to the dimension of $\underline{z}_t$. The value of this statistic should be "small" if the orthogonality conditions are satisfied.

5. Simulation Results

The pseudodata generated from the artificial economy are used in this section to estimate the parameter $\underline{\gamma}$ of the utility function, using the GMM procedure described in section 3.[10] These experiments were carried out along several dimensions in order to assess the robustness of the sampling results. The following are some pertinent features of the experimental design.

The experiments were organized along two types of perturbations. The first perturbation concerns variations of those parameters that are important for generating pseudodata that have different stochastic properties. These parameters include ρ, which controls the persistence of the random shock, and σ, which controls relative risk aversion or intertemporal substitution of consumption and leisure. For these two parameters the following setups were considered: $\rho = 0, 0.9$, and $\sigma = 0.1, 0.5, 1.0$, and 2.5.

The second perturbation concerns two aspects of estimation that directly affect the finite sample properties of GMM estimators. These parameters are the sample size, T, and the number of lags used to form instruments, NLAG. Throughout the experiment the instrumental vector $\underline{z}_t$ is selected to include a constant and the lagged values of consumption growth c_{t+1}/c_t, the return on investment $(1-\delta+r_{t+1})$, and the rate of change in the real wage rate w_{t+1}/w_t. The following cases were considered: NLAG = 1 and 2, and T = 50, 150, 300, and 500. For each of these experiments, 400 repetitions were carried out and sampling statistics calculated. Within each experiment the same set of data were used for different sample sizes and different lag lengths in order to reduce variability among experiments. This variance reduction technique was frequently used in Monte Carlo study to control intra-experiment variations

---

[10] The numerical routine I used to carry out the estimation is a GAUSS program written by David Runkel and Gregory Leonard. In order to verify its accuracy I checked this program against a GMM subroutine provided in the RATS package. These two numerical routines yield virtually identical results.

(see Hendry (1984)).

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

Table 2 displays the estimated mean, the standard deviation (SD), and the median of the GMM estimates of β, σ, and θ. As previously mentioned, the values for the two parameters β and θ are fixed throughout the experiments, which are 0.96 and 0.3, respectively. The four control parameters are ρ, σ, T, and NLAG.

In panel A I report the results of the first set of experiments which involve a highly persistent shock (ρ = 0.9). The number of lags used to form instruments in these experiments was 2 so that there are 7 variables included in the instrumental vector $\underline{z}_t$ (i.e., a constant plus two lagged values of consumption growth, asset returns, and wage growth). It is clear from this table that the estimate $\hat{\beta}$ performs extremely well regardless of the σ values or the sample sizes used. The estimated mean and median are almost identical to the pseudo value of β. The standard deviation is very small, indicating that the sampling distribution of $\hat{\beta}$ is tightly concentrated around the true value. This result, which is also true for the other experiments displayed in panels B and C, indicates that the parameter β can be reliably estimated using the GMM procedure.

The performance of $\hat{\sigma}$ and $\hat{\theta}$ is less clear. Except for cases where σ is small (i.e., σ = 0.1), the GMM procedure tends to underestimate σ. Both the mean and median of $\hat{\sigma}$ are below the pseudo value that was used to generate the data. However, as the sample size increases these central measures converge to the true value, which is to be expected. For most cases the true value of σ is within one standard deviation of $\hat{\sigma}$ about its mean. This result suggests that the magnitudes of the bias might not be quantitatively important. This, however, is not true for the estimate of θ. As the Table shows, the estimate

$\hat{\theta}$ is severely biased, particularly, for cases where σ is close to one. This is not surprising because when $\sigma = 1$ the parameter θ is not identifiable and, therefore, cannot be estimated with any precision.[11] The Table shows that the dispersion of $\hat{\theta}$ when $\sigma = 1$ is very large, and both mean and median are skewed toward negative values, which is meaningless. To some extent, this result applies also to the case of $\sigma = 0.5$. As the Table shows, the estimate $\hat{\theta}$ perform relatively better as σ moves away from one. In fact, when σ is small (i.e., $\sigma = 0.1$) the estimate $\hat{\theta}$ converges from below to the true value as the sample size increases, and when σ is large (i.e., $\sigma = 2.5$) it converges from above to the true value as the sample size increases. Note that for small samples $\hat{\theta}$ is still severely biased regardless of the σ values.

Panel B displays the sampling results for cases where the random shocks are purely temporary ($\rho = 0$). The number of lags used to form instruments is the same as before. For these experiments, only the cases of $\sigma = 0.5$ and 2.5 were considered. As can be seen, the chief difference here is that the estimate of the curvature parameter σ is upward biased, which is in contrast to the results displayed in panel A. Although it seems apparent that this difference is due to the forcing process that was used to generate the data, it is not easy to identify the specific sources that cause these biases. For example, the estimated correlation coefficient between consumption growth and asset returns when $\sigma = 0.5$ is higher in the case of $\rho = 0$ than in the case of $\rho = 0.9$. However, when $\sigma = 2.5$ this correlation becomes smaller for $\rho = 0$. There does not appear to exist a clear pattern in the correlation structure of the simulated data that helps explain or identify the bias. It is interesting to note that the sampling distribution of $\hat{\sigma}$ (and $\hat{\theta}$) gets tighter

---

[11] Note that the utility function becomes additively separable when $\sigma = 1$. In this case, the marginal utility of consumption does not depend on the leisure decision, which implies that the wage term will not appear in equation (3). As a result, the parameter θ cannot be identified.

as the shock becomes less persistent.

Panel C gives the sampling results where the number of lags used to form instruments was decreased from 2 to 1. The value of the parameter ρ is still 0.9. Examination of the results indicate that both $\hat{\sigma}$ and $\hat{\theta}$ appear to perform better than the first set of experiments in terms of the central measures. The mean as well as the median of these two estimates are closer to the true values. Although the dispersion (i.e., standard deviation) of the sampling distribution tends to rise as NLAG decreases from 2 to 1, the magnitudes do not seem unusually large. This last result, which will be made more clear below, is somewhat different from that of Tauchen (1986) who found that there is a strong bias/variance trade-off as NLAG increases.

Bias and Root Mean Squared Error (RMSE) of $\hat{\sigma}$ and $\hat{\theta}$

Table 3 contains some specific statistics regarding the performance of $\hat{\sigma}$ and $\hat{\theta}$. Two conventional measures were computed. The first measure, bias, is the sample average of the estimates less the true parameter value, and the second measure, RMSE, is the root mean squared error about the true parameter value. For the purpose of comparison, these two measures were divided by the estimated standard deviation of the estimates, and the results were given in the brackets in the Table.[12] Notice that the standardized RMSE should have a value that is close to but greater than one.

Several conclusions regarding the accuracy of $\hat{\sigma}$ seem apparent from Table 3. First, the estimate $\hat{\sigma}$ is in general biased, but the magnitudes are not very large. This result is consistent with the material presented in Table 2. Except for a few cases the bias is about half of the estimated standard deviation. The normalized RMSE is more or less around the anticipated value

---

[12] Because the estimated standard deviation is itself a random variable, this division introduces some noises in the standardized measures.

of one. Note that for shorter lag lengths (see panel C) the bias as well as the normalized RMSE are smaller than those displayed in panel A. This result indicates that the slightly higher standard deviation that is associated with NLAG = 1 is dominated by the improvement in performance in terms of the bias. The above findings, which are also true for the θ estimates, suggest that the smaller standard deviation that might be obtained by using longer lag lengths may be outweighed by the larger bias. Later we will see that statistical inferences based on longer lag lengths usually yield misleading conclusions.

The results of Table 3 also show that the performance of $\hat{\sigma}$ and $\hat{\theta}$ depends on the values of σ and ρ in a nontrivial way (comparing panels A and B). For example, when σ is small (i.e., $\sigma = 0.5$), the accuracy of these two estimates deteriorates as ρ gets smaller. But when σ is large (i.e., $\sigma = 2.5$), their performance improves as ρ becomes smaller. This conclusion, however, is not unambiguous because, as mentioned before, the standard deviations of $\hat{\sigma}$ and $\hat{\theta}$ decrease with the value of ρ. Thus, in terms of the normalized measures of bias and RMSE, the performance of these two estimates (especially, for smaller sample sizes) becomes worse as the value of ρ gets smaller. Because of the inherent nonlinearity of the model and the GMM estimation procedure, it is difficult to identify the sources that cause these disparities.

Testing the Overidentifying Restrictions

As discussed in section 3, the GMM estimation provides a general test of the specification of the model. In particular, the restrictions that the Euler equation residual should be uncorrelated with variables contained in the information set constitute a set of overidentifying restrictions that can be tested. The statistic used to perform this test is the sample size times the minimized value of the objective function. This statistic is distributed asymptotically as a chi-squared random variable with degrees of freedom equal to the number of overidentifying restrictions.[13] This subsection summarizes the results of this specification test.

Table 4 reports the proportion of time that the model was rejected out of the 400 repetitions. The rejection rates were calculated at the nominal significance levels of 5% and 10%. The degrees of freedom of these tests are also indicated in the Table. Since the model is correctly specified in all experiments, the rejection rates (i.e., type I errors) should be close to the corresponding significance level, in particular, for large sample sizes. As the Table clearly demonstrates, the model restrictions were rejected much more frequently than expected, particularly, for cases where $\rho = 0.9$ and NLAG = 2. Looking at panel A, it is striking that even for a relatively large sample (i.e., T = 500), the rejection rates are more than 10% in most cases and can be as high as 29% for $\sigma = 0.1$.[14] As should be expected, the rejection rates increase as the sample size decreases. For a small sample such as T = 50, these rejection rates are in the range of 30% to 55%, depending on the values of σ. Note that the model is rejected more frequently as σ becomes smaller. These results are sharply different from those of Tauchen (1986) who found that, in a somewhat different context, the rejection rates of the model were more or less in line with the significance levels.

Table 4 also shows that the specification test is very sensitive to the number of lags used to form instruments. As shown in panel C (NLAG = 1), the rejections rates are much lower than those associated with NLAG = 2. As the sample size increases, these rejection rates converge to the nominal rates.

---

[13] In our example, the number of overidentifying restrictions is equal to the dimension of the instrumental vector less the number of parameters being estimated (i.e., q-3).

[14] An experiment using 1000 observations was performed for the case of $\sigma = 0.1$. The performance of the test did not improve very much. The rejection rate dropped from 29% to 26% at 5% significance level. This finding suggests that the sampling bias is quite substantial for small σ values. This point will be addressed in more detail later on.

These findings clearly suggest that the risk of making incorrect inferences increases with the lag lengths. To check this conclusion more carefully, I increased NLAG to 4 for some experiments and found that the performance of the test worsen dramatically and the rejection rates appear to depend more on NLAG than on the sample size.[15]

In addition to the results reported here, some further experiments were conducted in order to see the sensitivity of the test with respect to certain aspects of the estimation procedure.[16] Specifically, instead of using the two-step procedure suggested by Hansen and Singleton (1982), the number of iterations was increased until the minimized values of the objective function differ by less than 1 percent over consecutive iterations. These experiments indicate that the estimation usually converges in 4 to 6 steps and the final value of the objective function is not very different from that using the two steps procedure. Consequently, the test results are virtually identical to those reported in Table 4.[17]

Behavior of the Test Statistic and the Objective Function

The results reported so far strongly suggest that rejection of the model is likely due to the asymptotic sampling bias. This subsection provides some pertinent information concerning the behavior of the test statistic as well as the objective function.

A useful way to pose our problem is as follows: In order to make correct inferences (say, to reject the model about 5% of time at the corresponding nominal rate), what is the correct critical value that should be used for the

---

[15] For example, the rejection rates under $\rho = 0.9$ and $\sigma = 0.5$ were 46% and 40% for T = 50 and 500, respectively. These results are not reported and can be obtained from the author upon request.

[16] The following experiments were suggested to me by Martin Eichenbaum.

[17] It should be cautioned that for these experiments only a limited number of cases were considered.

test statistic? The last two columns of Table 4 contain relevant statistics to answer this question. The first column lists the values of a chi-squared random variable at 5% significance level. These numbers, which were used to perform the test, represent the theoretical values that should be used if the sampling distribution of the test statistic is "close" to a true chi-squared distribution. The last column displays the critical values that should have been used for making correct inferences. These figures were calculated from the approximate distribution of the test statistic. It is clear that the sampling values are much higher than the theoretical values, particularly, for longer lag lengths. The two values become closer for NLAG = 1 although some minor disparities do exist. This simple exercise suggests that the tails of the sampling distribution of the test statistic might be thicker than the theoretical distribution.

Figures 1a and 1b plot the "inverted" sampling distribution of the test statistic for $\sigma = 0.5$ and 2.5, respectively. For comparisons, these figures were plotted against a theoretical chi-squared distribution. The heights of these inverted distributions represent the marginal probabilities of the test statistic at the corresponding values shown on the horizontal axis. A dotted line is used to mark the 5% significance level. These figures show that the asymptotic sampling biases are quite substantial for NLAG = 2. As suggested above, the rear ends of the sampling distribution are thicker than those of the true distribution. This is the reason why the specification test tends to over-reject the model. The sampling distribution appears to converge to the true distribution as the sample size increases. However, there still exist some significant disparities for T = 500. On the other hand, the case of NLAG = 1 performs much better. The two distributions are almost identical for T = 500, a result which is consistent with the results of Table 4.

The reliability of the GMM estimator and the associated test is directly related to the behavior of the objective function. As Hansen (1982) pointed out, the values of the objective function evaluated at the true parameter values are distributed asymptotically as chi-square statistics with degrees of freedom equal to the number of orthogonality conditions. If this minimum chi-squared property is violated, one expects a poor performance of the GMM estimator and the associated specification test. Figure 2 plots the distribution of the objective function for σ = 0.1 and 2.5 using NLAG = 2. For comparisons, the true chi-squared distribution is also plotted. As can be seen, these figures are very similar to those in figure 1a and 1b. In particular, the sampling distribution of the objective function is far away from the asymptotic distribution when σ is small. These results explain why GMM estimators perform poorly for small σ values. Again, the sampling distribution converges to the true distribution as the sample size increases.

6. Conclusions

In this paper I have examined the performance of the GMM estimation in a simple neoclassical model which has received considerable scrutiny in recent years. Most empirical studies have found that this model or a similar one did not seem consistent with the actual data. The findings of this article suggest that rejection of the model might have been caused by an inadequacy of the asymptotic approximations. For cases where the curvature parameter of the utility function is small or risk aversion is high, these sampling biases are fairly substantial and could lead to inappropriately high rejections of the true model. In order to make correct inferences, it may require a large number of observations which are simply not feasible in practice. One way to alleviate this problem is to use shorter lag lengths in forming instruments. To some extent, the better performance with shorter lag lengths is due to a reduction in the number of restrictions that are being tested. But, more important, it is because the sampling errors are smaller and the estimators perform better.

Another implication suggested by our experiments is that the curvature parameter of the utility function can be reasonably estimated using the GMM procedure. This result is contrary to the existing belief that it might be very difficult to pin down this parameter. This perception is perhaps due to the wide range of estimates that exist in the literature. Our findings suggest that, if the model is correctly specified, failure to uncover this curvature parameter accurately may result from measurement errors that plague the data. Such errors are particularly characteristic of consumption data.

To keep the cost of simulation to its minimum, several important issues have not been addressed in this paper. In particular, problems concerning the power of the specification test were ignored. Consequently, very little can be said about the power of the test against false specifications. Recent studies by Singleton (1985) have shown that it is feasible to discriminate among competing economic models within the GMM environment. Such endeavors will require much more careful and finer calibrations of the model and are left for future studies.

References

Bertsekas, D. P. *Dynamic Programming and Stochastic Control*. New York: Academic Press, 1976.

Brock, W. A., and L. Mirman. "Optimal Economic Growth and Uncertainty: The Discounted Case," *Journal of Economic Theory*, 4(1972), 479-513.

Christiano, L. J. "Solving a Particular Growth Model by Linear Quadratic Approximation and by Value Function Iteration," Discussion Paper #9, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis, January 1989.

Den Haan, W. J., and A. Marcet. "Accuracy in Simulations," Manuscript, Carnegie-Mellon University, 1989.

Eichenbaum, M., L. P. Hansen, and K. J. Singleton. "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choices Under Uncertainty," Manuscript, Carnegie-Mellon University, 1983.

Hall, R. E. "Intertemporal Substitution in Consumption," *Journal of Political Economy*, 96(1989), 339-357.

Hansen, L. P. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(1982), 1029-1054.

———————— "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators," *Journal of Econometrics*, 30(1985), 293-238.

————————, and K. J. Singleton. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50(1982), 1269-1286.

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

————————, and K. J. Singleton. "Efficient Estimation of Linear Asset Pricing Models with Moving Average Errors," Manuscript, University of Chicago, April 1988.

Hendry, D. F. "Monte Carlo Experimentation in Econometrics," Chapter 16, *Handbook of Econometrics*, Vol. II, A. Griliches and M. D. Intriligator, eds. Amsterdam: North-Holland, 1984.

King, R. G., C. Plosser, and S. Rebelo. "Production, Growth and Business Cycles I. The Basic Neoclassical Model," *Journal of Monetary Economics*, May 1988.

Kocherlakota, N. R. "In Defense of the Time and State Separable Utility-Based Asset Pricing Model," Working paper 63, Department of Finance, Northwestern University, October 1988.

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

Mao, C. S. "Estimating Intertemporal Elasticity of Substitution: The Case of Log-Linear Restrictions," *Economic Review*, Federal Reserve Bank of Richmond, 75(1989), 3-14.

Rebelo, S., and G. Rouwenhorst. "Linear Quadratic Approximations versus Discrete State Space Methods: A Numerical Evaluation," Manuscript, University of Rochester, March 1989.

Sargent, T. J. "Tobin's q and the Rate of Investment in General Equilibrium. *Carnegie-Rochester Conference Series on Public Policy*, 12, K. Brunner and A. H. Meltzer, eds. Amsterdam: North- Holland, 1980.

Singleton, K. J. "Testing Specifications of Economic Agents' Intertemporal Optimum Problems in the Presence of Alternative Models," *Journal of Econometrics*, 30(1985), 391-413.

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

Table 1
The Den Haan-Marcet Statistic for Testing the Orthogonality Conditions
(Fixed parameters: $\alpha = 0.3$, $\delta = 0.1$, $\beta = 0.96$, $\theta = 0.3$, $\rho = 0.9$)

σ	NLAG = 1	NLAG = 2
0.1	1.20 (0.88)	4.91 (0.67)
0.5	2.14 (0.71)	2.73 (0.91)
1.0	2.49 (0.65)	3.51 (0.83)
2.5	2.50 (0.65)	8.44 (0.30)

Note:
1. The statistic is computed based on a sample of 3000 observations.
2. The instruments are chosen to include a constant and the lagged values of consumption growth, asset returns and wage growth.
3. NLAG = number of lags used to form instruments.
4. Test results are similar for smaller sample size and larger degrees of freedom, which are not reported.

Table 2: Moments of the Sampling Distributions of $\hat{\beta}$, $\hat{\sigma}$ and $\hat{\theta}$

Parameters		$\hat{\beta}$ (β_0 = 0.96)			$\hat{\sigma}$			$\hat{\theta}$ (θ_0 = 0.3)		
σ_0	T	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
				Panel A: ρ = 0.9, NLAG = 2						
0.1	50	0.960	0.003	0.960	0.125	0.060	0.110	0.084	0.078	0.086
	150	0.960	0.002	0.960	0.120	0.048	0.110	0.170	0.113	0.150
	300	0.960	0.001	0.960	0.124	0.058	0.109	0.241	0.172	0.204
	500	0.960	0.001	0.960	0.117	0.038	0.110	0.251	0.097	0.228
0.5	50	0.960	0.001	0.960	0.406	0.089	0.387	-0.150	0.420	-0.107
	150	0.960	0.001	0.960	0.449	0.119	0.422	-0.006	0.176	-0.024
	300	0.960	0.000	0.960	0.474	0.118	0.449	0.094	0.237	0.076
	500	0.960	0.000	0.960	0.478	0.092	0.461	0.121	0.745	0.133
1.0	50	0.960	0.001	0.960	0.719	0.212	0.686	-0.887	2.003	-0.692
	150	0.960	0.001	0.960	0.850	0.355	0.773	-0.416	4.125	-0.562
	300	0.960	0.000	0.960	0.937	0.443	0.841	-0.737	4.212	-0.438
	500	0.960	0.000	0.960	0.934	0.216	0.887	-0.523	5.567	-0.337
2.5	50	0.960	0.000	0.960	1.433	0.518	1.323	1.996	1.892	1.757
	150	0.960	0.000	0.960	1.789	0.736	1.614	0.985	0.529	0.685
	300	0.960	0.000	0.960	2.129	0.913	1.858	0.617	0.297	0.574
	500	0.960	0.000	0.960	2.233	0.788	2.028	0.502	0.210	0.494
				Panel B: ρ = 0.0, NLAG = 2						
0.5	50	0.960	0.001	0.960	0.724	0.082	0.732	-0.321	0.442	-0.246
	150	0.960	0.000	0.960	0.601	0.073	0.602	0.116	0.154	0.138
	300	0.960	0.000	0.960	0.554	0.053	0.554	0.213	0.085	0.224
	500	0.960	0.000	0.960	0.536	0.042	0.535	0.245	0.060	0.253
2.5	50	0.960	0.001	0.960	3.317	0.469	3.327	0.468	0.062	0.483
	150	0.960	0.000	0.960	2.837	0.330	2.837	0.382	0.074	0.392
	300	0.960	0.000	0.960	2.682	0.231	2.688	0.348	0.061	0.356
	500	0.960	0.000	0.960	2.604	0.190	2.602	0.329	0.049	0.335
				Panel C: ρ = 0.9, NLAG = 1						
0.1	50	0.960	0.004	0.960	0.093	0.052	0.086	0.104	0.168	0.091
	150	0.960	0.002	0.960	0.098	0.061	0.083	0.197	0.233	0.152
	300	0.960	0.001	0.960	0.099	0.066	0.084	0.283	0.788	0.192
	500	0.960	0.001	0.960	0.098	0.057	0.084	0.261	0.235	0.207
0.5	50	0.960	0.002	0.960	0.413	0.121	0.385	-0.063	0.386	-0.072
	150	0.960	0.001	0.960	0.468	0.183	0.423	0.079	0.730	0.061
	300	0.960	0.000	0.960	0.498	0.211	0.451	0.242	0.666	0.148
	500	0.960	0.000	0.960	0.480	0.092	0.456	0.242	0.228	0.190
1.0	50	0.960	0.001	0.960	0.783	0.337	0.702	-0.654	2.479	-0.574
	150	0.960	0.001	0.960	0.906	0.368	0.806	-0.640	5.088	-0.301
	300	0.960	0.001	0.960	0.977	0.342	0.881	-0.446	5.587	-0.268
	500	0.960	0.000	0.960	0.979	0.273	0.920	-0.536	5.404	-0.184
2.5	50	0.960	0.001	0.960	1.607	0.673	1.439	1.381	2.842	1.164
	150	0.960	0.000	0.960	2.028	0.970	1.747	0.709	0.512	0.655
	300	0.960	0.000	0.960	2.353	0.977	2.003	0.468	0.285	0.432
	500	0.960	0.000	0.960	2.375	0.843	2.140	0.422	0.208	0.407

Note: Statistics are computed based on 400 iterations of each experiment.

Table 3: Bias and Root Mean Squared Error (RMSE) of $\hat{\sigma}$ and $\hat{\theta}$

Parameters		$\hat{\sigma}$				$\hat{\theta}$			
σ_0	T	Bias	Norm. Bias	RMSE	Norm. RMSE	Bias	Norm. Bias	RMSE	Norm. RMSE
		Panel A: ρ = 0.9, NLAG = 2							
0.1	50	0.025	[0.42]	0.065	[1.08]	-0.216	[-2.78]	0.230	[2.96]
	150	0.020	[0.42]	0.052	[1.08]	-0.130	[-1.15]	0.172	[1.52]
	300	0.024	[0.42]	0.062	[1.08]	-0.059	[-0.34]	0.181	[2.31]
	500	0.017	[0.46]	0.041	[1.09]	-0.049	[-0.50]	0.109	[1.12]
0.5	50	-0.094	[-1.06]	0.129	[1.46]	-0.450	[-1.07]	0.615	[1.46]
	150	-0.051	[-0.43]	0.129	[1.09]	-0.306	[-1.73]	0.353	[2.00]
	300	-0.026	[-0.22]	0.121	[1.02]	-0.206	[-0.87]	0.314	[1.32]
	500	-0.022	[-0.23]	0.094	[1.03]	-0.179	[-0.24]	0.765	[1.03]
1.0	50	-0.280	[-1.33]	0.351	[1.66]	-1.187	[-0.59]	2.326	[1.16]
	150	-0.150	[-0.42]	0.385	[1.09]	-0.716	[-0.17]	4.182	[1.01]
	300	-0.063	[-0.14]	0.447	[1.01]	-1.037	[-0.24]	4.368	[1.03]
	500	-0.066	[-0.30]	0.225	[1.04]	-0.823	[-0.15]	5.621	[1.01]
2.5	50	-1.067	[-2.06]	1.185	[2.29]	1.696	[0.90]	2.539	[1.34]
	150	-0.711	[-0.97]	1.022	[1.39]	0.685	[1.30]	0.865	[1.64]
	300	-0.371	[-0.41]	0.984	[1.08]	0.317	[1.07]	0.434	[1.46]
	500	-0.267	[-0.34]	0.831	[1.05]	0.202	[0.96]	0.291	[1.39]
		Panel B: ρ = 0.0, NLAG = 2							
0.5	50	0.224	[2.74]	0.238	[1.91]	-0.621	[-1.41]	0.762	[1.72]
	150	0.101	[1.38]	0.124	[1.70]	-0.184	[-1.20]	0.240	[1.56]
	300	0.054	[1.02]	0.076	[1.43]	-0.087	[-1.02]	0.122	[1.43]
	500	0.036	[0.87]	0.055	[1.32]	-0.055	[-0.91]	0.081	[1.35]
2.5	50	0.817	[1.74]	0.942	[2.01]	0.168	[2.70]	0.179	[2.88]
	150	0.337	[1.02]	0.472	[1.43]	0.081	[1.09]	0.110	[1.48]
	300	0.182	[0.79]	0.293	[1.27]	0.048	[0.78]	0.077	[1.27]
	500	0.104	[0.55]	0.216	[1.14]	0.029	[0.57]	0.057	[1.15]
		Panel C: ρ = 0.9, NLAG = 1							
0.1	50	-0.007	[-0.14]	0.053	[1.01]	-0.196	[-1.17]	0.258	[1.54]
	150	-0.002	[-0.04]	0.061	[1.00]	-0.103	[-0.44]	0.254	[1.09]
	300	-0.001	[-0.02]	0.066	[1.00]	-0.017	[-0.02]	0.787	[1.00]
	500	-0.002	[-0.04]	0.057	[1.00]	-0.039	[-0.17]	0.238	[1.01]
0.5	50	-0.087	[-0.72]	0.149	[1.23]	-0.363	[-0.94]	0.529	[1.37]
	150	-0.032	[-0.18]	0.186	[1.01]	-0.221	[-0.30]	0.762	[1.04]
	300	-0.002	[-0.01]	0.211	[1.00]	-0.058	[-0.09]	0.668	[1.00]
	500	-0.021	[-0.22]	0.094	[1.02]	-0.058	[-0.25]	0.234	[1.03]
1.0	50	-0.217	[-0.64]	0.400	[1.19]	-0.954	[-0.38]	2.653	[1.07]
	150	-0.094	[-0.26]	0.389	[1.03]	-0.940	[-0.18]	5.168	[1.02]
	300	-0.023	[-0.07]	0.342	[1.00]	-0.746	[-0.13]	5.629	[1.01]
	500	-0.021	[-0.08]	0.273	[1.00]	-0.836	[-0.15]	5.462	[1.01]
2.5	50	-0.893	[-1.33]	1.117	[1.66]	1.081	[0.38]	3.038	[1.07]
	150	-0.472	[-0.49]	1.077	[1.11]	0.409	[0.80]	0.655	[1.28]
	300	-0.147	[-0.15]	0.987	[1.01]	0.168	[1.36]	0.331	[1.16]
	500	-0.125	[-0.15]	0.851	[1.01]	0.122	[0.59]	0.241	[1.16]

Note: 1. Statisitics are computed based on 400 iterations of each experiment.
2. Figures reported in the brackets are normalized Bias and RMSE [e.g., BIAS / (estimated standard deviation)].

Table 4: Rejection Rate (%) of the Overidentifying Restrictions

Parameters σ_0	T	Degrees of Freedom	Rejection Rate at Significance Level of 5 %	10 %	Critical Value at 5% Significance Level True	Sampling
			Panel A: ρ = 0.9, NLAG = 2			
0.1	50	4	55	65	9.49	21.09
	150	4	43	50	9.49	31.42
	300	4	31	39	9.49	28.84
	500	4	29	38	9.49	23.68
0.5	50	4	40	52	9.49	20.39
	150	4	31	40	9.49	24.04
	300	4	20	28	9.49	18.83
	500	4	12	19	9.49	15.18
1.0	50	4	36	45	9.49	18.99
	150	4	24	32	9.49	21.23
	300	4	18	24	9.49	17.87
	500	4	12	18	9.49	13.94
2.5	50	4	29	41	9.49	20.15
	150	4	25	33	9.49	18.96
	300	4	19	21	9.49	16.58
	500	4	13	26	9.49	14.20
			Panel B: ρ = 0.0, NLAG = 2			
0.5	50	4	25	33	9.49	15.41
	150	4	17	25	9.49	14.16
	300	4	13	17	9.49	13.85
	500	4	11	16	9.49	12.91
2.5	50	4	17	26	9.49	13.45
	150	4	13	17	9.49	13.13
	300	4	9	13	9.49	10.56
	500	4	7	13	9.49	10.46
			Panel C: ρ = 0.9, NLAG = 1			
0.1	50	1	12	18	3.84	6.55
	150	1	12	19	3.84	6.26
	300	1	7	13	3.84	4.52
	500	1	9	15	3.84	5.10
0.5	50	1	16	22	3.84	9.72
	150	1	8	12	3.84	5.18
	300	1	6	11	3.84	4.47
	500	1	6	11	3.84	3.99
1.0	50	1	14	20	3.84	8.47
	150	1	7	12	3.84	4.78
	300	1	6	12	3.84	4.47
	500	1	6	11	3.84	3.86
2.5	50	1	12	18	3.84	7.36
	150	1	7	12	3.84	4.66
	300	1	7	12	3.84	4.36
	500	1	5	11	3.84	3.76

Note: Statistics are computed based on 400 iterations of each experiment.

Figure 1a: Test Statistic
$1-\int_0^x dF(y)$: Asymptotic vs. Sampling
(parameters: $\sigma = 0.5$, $\rho = 0.9$)

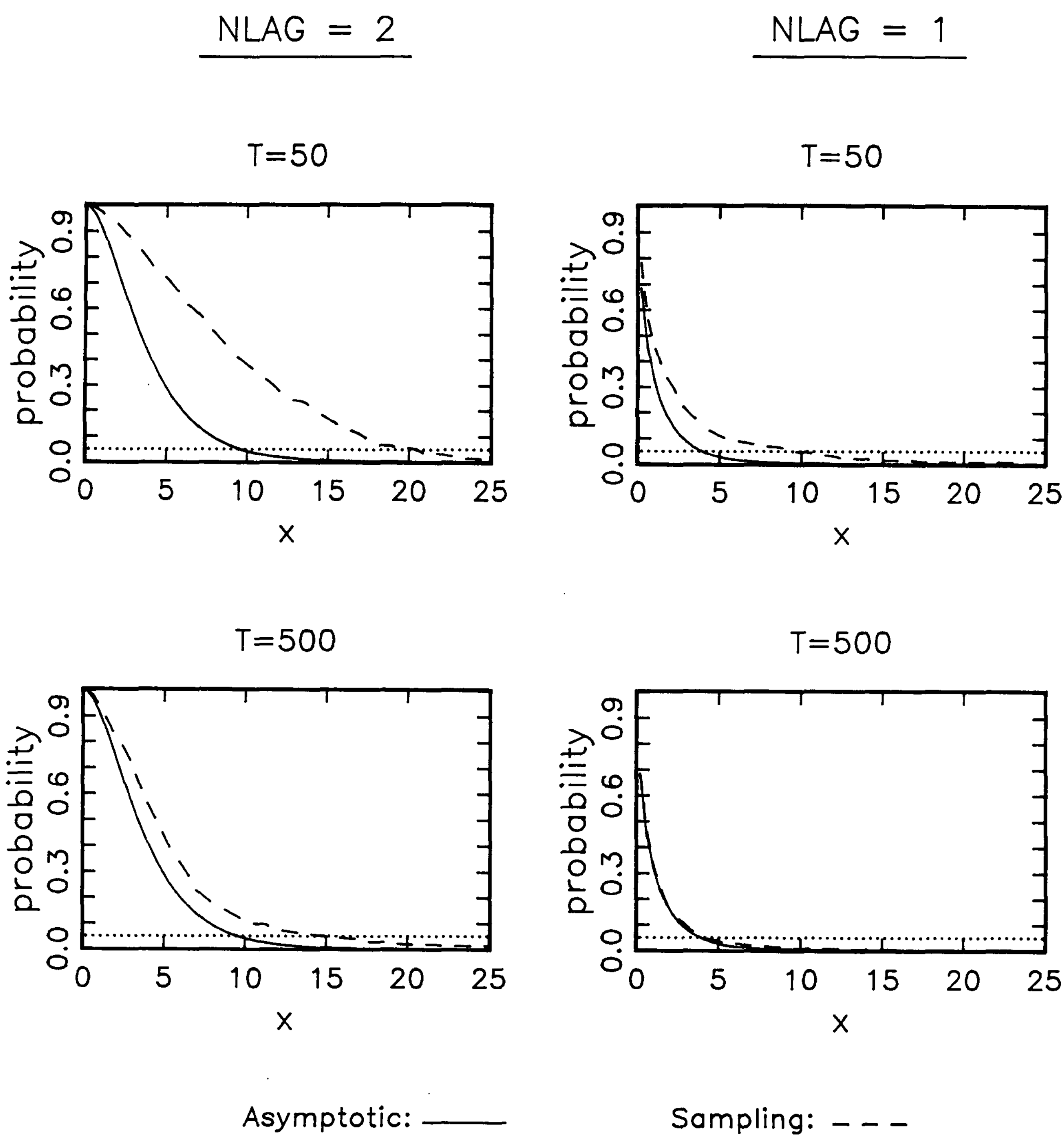
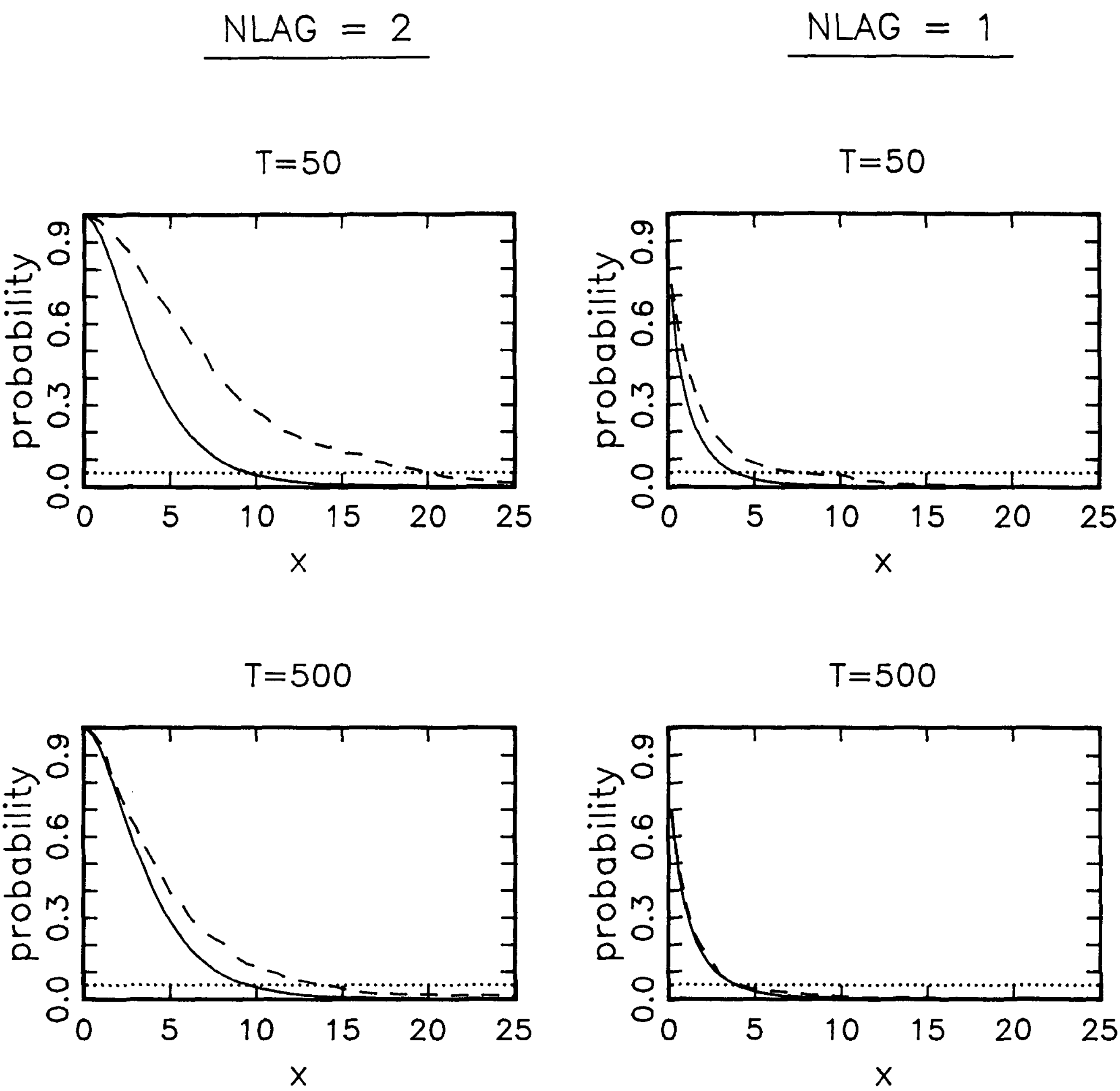


Figure 1b: Test Statistic
$1-\int_0^x dF(y)$: Asymptotic vs. Sampling
(parameters: σ = 2.5, ρ = 0.9)



Asymptotic: ——— Sampling: – – –

Figure 2: Objective Function
$1-\int_0^x dF(y)$: Asymptotic vs. Sampling
(parameters: $\rho = 0.9$, NLAG = 2)