Rochester Center for Economic Research

An Introduction to the Generalized Method of Moments

Ogaki, Masao

Working Paper No. 314
February 1992

University of Rochester

An Introduction to the Generalized Method of Moments

Masao Ogaki

Rochester Center for Economic Research
Working Paper No. 314

An Introduction to the Generalized Method of Moments

Masao Ogaki

University of Rochester

The Rochester Center for Economic Research
Working Paper No. 314

February 1992

(First Draft, January 1992)

I thank Adrian Pagan for his suggestions, Lars Peter Hansen for clarification, and Changyong Rhee for a conversation that motivated this work. All remaining shortcomings are mine.

An Introduction to the Generalized Method of Moments

Masao Ogaki
Department of Economics
University of Rochester
Rochester, N.Y. 14627-0156

Abstract

The purpose of the present paper is to explain Hansen's (1982) Generalized Method of Moments (GMM) to applied researchers and to give practical guidance as to how GMM estimation should be implemented. The present paper discusses statistical properties of GMM estimators and test statistics. It presents some of the recent developments in the GMM procedure that have been used in applications. In explaining empirical applications, the present paper emphasizes pitfalls that researchers have encountered and how they have avoided them.

Table of Contents

Abstract ii
Table of Contents ii

1. Introduction 1

2. The Generalized Method of Moments 2

    2.1. Moment Restrictions and GMM Estimators
    2.2. Distributions of GMM Estimators
    2.3. Optimal Choice of the Distance Matrix
    2.4. A Chi-Square Test for the Overidentifying restrictions

3. Special Cases 5

    3.1. Ordinary Least Squares
    3.2. Linear Instrumental Variables Regressions
    3.3. Nonlinear Instrumental Variables Estimation

4. Extensions 7

4.1. Sequential Estimation
4.2. GMM with Deterministic Trends
4.3. Cross Sectional Data and Panel Data
4.4. The Minimum Distance Estimation

5. Important Assumptions 12

5.1. Stationarity
5.2. Identification

6. Covariance Matrix Estimation 14

6.1. Serially Uncorrelated Disturbance
6.2. Serially Correlated Disturbance
6.2.1. Unknown Order of Serial Correlation
6.2.2. Known Order of Serial Correlation

7. Hypothesis Testing and Specification Tests 21

8. Empirical Applications 22

8.1. Euler Equation Approach to Models of Consumption
8.1.1. Hansen and Singleton's (1982) Model
8.1.2. Time Aggregation
8.1.3. Habit Formation and Durability
8.1.4. Multi-Good Models
8.1.5. The Cointegration-Euler Equation Approach
8.1.6. Monetary Asset Pricing Model
8.1.7. Seasonality
8.1.8. State-Nonseparable Preferences
8.2. Other Empirical Applications

9. Further Issues 34

9.1. Optimal Choice of Instrumental Variables
9.2. GMM and Semi-Parametric Estimation
9.3. Small Sample Properties

10. Concluding Remarks 38

References 39

1. Introduction

The purpose of the present paper is to explain Hansen's (1982) Generalized Method of Moments (GMM) to applied researchers and to give practical guidance as to how GMM estimation should be implemented.[1] The present paper discusses statistical properties of GMM estimators and test statistics. It presents some of the recent developments in the GMM procedure that have been used in applications. These include sequential (or two step) estimation, GMM with deterministic trends, applications for cross sectional and panel data, some statistics that are often used for hypothesis testing. In explaining empirical applications, the present paper emphasizes pitfalls that researchers have encountered and how they have avoided them.[2]

The rest of the present paper is organized as follows. Section 2 presents the basic GMM framework. Section 3 illustrates how ordinary least squares and liner and nonlinear instrumental variables estimation are embedded in the GMM framework as special cases. Section 4 presents some GMM related statistical procedures that extends the basic GMM in Section 2. These include sequential (or two step) estimation, GMM with deterministic trends, applications of GMM to cross sectional and panel data, and the minimum distance estimation. Section 5 discusses important assumptions for GMM that applied researchers should be aware of. In Section 6, I explain methods for covariance matrix estimation that is necessary to calculate standard errors for GMM estimators and to use the optimal distance matrix for GMM estimation. Section 7 presents Wald, Lagrange

---

[1]Hall (1991) provides a nontechnical introduction to GMM that explains the basic intuition behind GMM.

[2]In the companion paper, Ogaki (1992a), I explain how to use Hansen/Heaton/Ogaki GAUSS GMM package to implement GMM estimation and form test statistics.

multiplier and likelihood ratio type statistics for hypothesis testing and recently developed specification tests. In Section 8, I explain empirical applications. Section 9 discusses optimal choice of instrumental variables and small sample properties of GMM estimators and test statistics. Section 10 contains concluding remarks.

2. The Generalized Method of Moments

This section explains the basic GMM framework.

2.1. Moment Restrictions and GMM Estimators

Let $\{X_t: t=1,2,\ldots\}$ be a collection of random vectors X_t's, β_0 be a p-dimensional vector of the parameters to be estimated, and $f(X_t,\beta)$ a q-dimensional vector of functions. The present paper treats GMM in the context of the time series analysis except for Subsection 4.3, where I discuss applications of GMM to cross sectional data and panel data. Assume that X_t is (strictly) stationary.[3] I refer to $u_t=f(X_t,\beta_0)$ as the disturbance of GMM. Consider (unconditional) moment restrictions

$$E(f(X_t, \beta_0)) = 0. \tag{2.1}$$

Suppose that a law of large numbers can be applied to $f(X_t,\beta)$ for admissible β, so that the sample mean of $f(X_t,\beta)$ converges to its population mean:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} f(X_t, \beta) = E(f(X_t, \beta)) \tag{2.2}$$

with probability one (or in other words, almost surely). The basic idea of GMM is to mimic the moment restrictions (2.1) with the sample mean by minimizing a quadratic form

---

[3]See Section 5 for a definition and a discussion of stationarity.

$$J_T(\beta) = \{\frac{1}{T}\sum_{t=1}^{T} f(X_t,\ \beta)\}' W_T \{\frac{1}{T}\sum_{t=1}^{T} f(X_t,\ \beta)\} \tag{2.3}$$

by choosing β, where W_T is a positive definite matrix, which satisfies

$$\lim_{T\to\infty} W_T = W_0. \tag{2.4}$$

with probability one for a positive definite matrix W_0. We refer W_T and W_0 as the distance matrix or the weighting matrix. Then the GMM estimator, β_T, is the solution of the minimization problem (2.3). Under fairy general regularity conditions, the GMM estimator β_T is a consistent estimator for arbitrary distance matrices.[4] I will discuss how to choose the distance matrix to obtain an (asymptotically) efficient GMM estimator.

2.2. Distributions of GMM Estimators

Suppose that a central limit theorem applies to the disturbance of GMM, $u_t=f(X_t,\beta_0)$, so that $(1/\sqrt{T})\sum_{t=1}^{T}u_t$ has an (asymptotic) normal distribution with mean zero and the covariance matrix Ω in large samples. If u_t is serially uncorrelated, $\Omega=E(u_t^2)$. If u_t is serially correlated,

$$\Omega = \lim_{j\to\infty} \sum_{-j}^{j} E(u_t u_{t-j}). \tag{2.5}$$

Some authors refer to Ω as the long run covariance matrix of u_t. Let $\Gamma=E(\partial f(X_t,\beta)/\partial\beta')$ be the expectation of the $q \times p$ matrix of the derivatives of $f(X_t,\beta)$ with respect to β and assume that Γ has a full column rank. Under a set of regularity conditions, $\sqrt{T}(\beta_T-\beta_0)$ has a (asymptotic) normal distribution with mean zero and the covariance matrix

---

[4]Some regularity conditions that are important for applied researchers are discussed in Section 5.

(2.6) $$\mathrm{Cov}(W_0) = \{\Gamma' W_0 \Gamma\}^{-1}\{\Gamma' W_0 \Omega W_0 \Gamma\}\{\Gamma' W_0 \Gamma\}'^{-1}.$$

in large samples.

2.3. Optimal Choice of the Distance Matrix

When the number of moment conditions (q) is equal to the number of parameters to be estimated (p), the system is just identified. In the case of just identified system, the GMM estimator does not depend on the choice of distance matrix. On the other hand, when q>p, there exist overidentifying restrictions and we obtain different GMM estimators for different distance matrices. In this case, one may wish to choose the distance matrix, so that the resulting GMM estimator is asymptotically efficient. Hansen (1982) shows that when the covariance matrix (6) is minimized $W_0 = \Omega^{-1}$.[5] With this choice of the distance matrix, $\sqrt{T}(\beta_T - \beta_0)$ has an (asymptotic) normal distribution with mean zero and the covariance matrix

(2.7) $$\mathrm{Cov}(\Omega^{-1}) = \{\Gamma' \Omega^{-1} \Gamma\}^{-1},$$

in large samples.

Let Ω_T be an consistent estimator of Ω. Then $W_T = \Omega_T^{-1}$ is used to obtain β_T. The resulting estimator is called the optimal or efficient GMM estimator. It should be noted, however, it is optimal given $f(X_t, \beta)$. In the context of instrumental variable estimation, this means that instrumental variables are given. I will discuss optimal choice of instrumental variables in Section 9. Let Γ_T be a consistent estimator for Γ. Then The standard errors of the optimal GMM estimator β_T is calculated

[5] The covariance matrix is minimized in the sense that $\mathrm{Cov}(W_0) - \mathrm{Cov}(\Omega^{-1})$ is a positive semidefinite matrix for any positive definite matrix W_0.

as square roots of the diagonal elements of $T^{-1}(\Gamma_T'\Omega_T^{-1}\Gamma_T)^{-1}$. The appropriate estimation method of Ω depends on the model and this problem is discussed in Section 6. It is usually easier to estimate Γ by $\Gamma_T=(1/T)\sum_{t=1}^{T}(\partial f(X_t,\beta_T)/\partial\beta')$ than to estimate Ω. In linear models or in some simple nonlinear models, analytical derivatives are readily available. In nonlinear models, numerical derivatives are often used.

2.4. A Chi-Square Test for the Overidentifying Restrictions

In the case where there there are overidentifying restrictions (q>p), a chi-square statistic can be used to test the overidentifying restrictions. In the context of Euler equation approach explained in Section 8 of the present paper, this test has been used to test Euler equations that imply the moment conditions for GMM. Hansen (1982) shows that T times the minimized value of the objective function, $TJ_T(\beta_T)$, has an (asymptotic) chi-square distribution with q-p degrees of freedom if $W_0=\Omega^{-1}$ in large samples. This test is sometimes called Hansen's J test.[6]

3. Special Cases

This section shows how linear regressions and nonlinear instrumental variable estimation are embedded in the GMM framework explained in the last section.

3.1. Ordinary Least Squares

Consider a linear model,

$$y_t = x_t'\beta_0 + e_t, \tag{3.1}$$

where y_t and e_t are scalar random variables, x_t is a p-dimensional random

---

[6]See Newey (1985) for an analysis of the asymptotic power properties of this chi-squre test.

vector. OLS estimation can be embedded in the GMM framework by letting $X_t = (y_t, x_t')'$, $f(X_t,\beta)=x_t(y_t-x_t'\beta)$, $u_t=x_te_t$, and $p=q$. Thus the moment conditions (2.1) become the orthogonality conditions:

$$E(x_te_t) = 0. \tag{3.2}$$

Since this is the case of a just identified system, the distance matrix W_0 does not matter. The OLS estimator minimizes $\sum_{t=1}^{T} f(X_t,\beta)^2$. Though this minimization problem is different from the minimization problem (2.3) that the GMM estimator solves, it turns out that the GMM estimator coincides with the OLS estimator in this case. To see this, note that (2.3) can be minimized by setting β_T so that $(1/T)\sum_{t=1}^{T} f(X_t,\beta)=0$ in the case of a just identified system. This implies that $(1/T)\sum_{t=1}^{T} x_ty_t=(1/T)\{\sum_{t=1}^{T} x_tx_t'\}\beta_T$. Thus as long as $\{\sum_{t=1}^{T} x_tx_t'\}$ is invertible, $\beta_T=\{\sum_{t=1}^{T} x_tx_t'\}^{-1}\{\sum_{t=1}^{T} x_ty_t\}$. Hence the GMM estimator β_T coincides with the OLS estimator.

3.2. Linear Instrumental Variables Regressions

Let z_t be a q-dimensional random vector of instrumental variables. Then instrumental variable regressions are embedded in the GMM framework by letting $X_t=(y_t,x_t',z_t')'$, $f(X_t,\beta)=z_t(y_t-x_t'\beta)$, and $u_t=z_te_t$. Thus the moment conditions become the orthogonality conditions

$$E(z_te_t) = 0. \tag{3.3}$$

In the case of a just identified system (q=p), the instrumental variable regression estimator that minimizes $\sum_{t=1}^{T} f(X_t,\beta)^2$ coincides with the GMM estimator.

3.3. Nonlinear Instrumental Variables Estimation

GMM is often used in the context of nonlinear instrumental variable estimation (NLIV). Section 8 explains some examples of applications from Euler Equation Approach. Let $g(x_t,\beta)$ be a k-dimensional vector of functions and $e_t=g(x_t,\beta_0)$. Suppose that there exist conditional moment restrictions, $E[e_t|I_t]=0$, where $E[\cdot|I_t]$ signifies the mathematical expectation conditioned on the information set I_t. Here it is assumed that $I_t \subset I_{t+1}$ for any t. Let z_t be a q×k matrix of random variables that are in the information set I_t.[7] Then by the law of iterative expectations, we obtain unconditional moment restrictions:

$$E[z_t g(x_t,\beta_0)]=0. \tag{3.4}$$

Thus we let $X_t=(x_t',z_t')'$ and $f(X_t,\beta)=z_t g(x_t,\beta)$ in this case.

4. Extensions

This section explains econometric methods that are closely releated with the basic GMM framework in Section 2.

4.1. Sequential Estimation

This subsection discusses sequential estimation (or two step estimation). Consider a system

$$f(X_t,\beta) = \begin{bmatrix} f_1(X_t,\beta^1) \\ f_2(X_t,\beta^1,\beta^2) \end{bmatrix}, \tag{4.1}$$

where $\beta=(\beta^{1\prime},\beta^{2\prime})'$, β^i is a p_i-dimensional vector of parameters, and f_i is a q_i-dimensional vector of functions. Though it is possible to estimate β^1

---

[7] In some applications, z_t is a function of β. This does not cause any problem as long as the resulting $f(X_t,\beta)$ can be written as a function of β and a stationary random vector X_t.

and β^2 simultaneously, it may be computationally convenient to estimate β^1 from $f_1(X_t,\beta^1)$ in the first step estimation and then estimate β^2 from $f_2(X_t,\beta^1,\beta^2)$ in the second step estimation (see, e.g., Barro (1977) and Atkeson and Ogaki (1991) for examples of applications). In general, the asymptotic distribution of the first step estimator affects the asymptotic distribution of the second step estimator (see Newey (1984) and Pagan (1984, 1986)). A GMM computer package with a sequential estimation can be used to calculate the standard errors that take into account of these effects from the first step estimation. If there are overidentifying restrictions in the system, an econometrician may wish to choose the distance matrix in the second step in an efficient way. This problem of the choice of the distance matrix in the second step is analyzed by Hansen, Heaton, and Ogaki (1992).

Suppose that the first step estimator β^1_T minimizes

$$(4.2)\qquad J_{1T}(\beta^1) = \{\frac{1}{T}\sum_{t=1}^{T} f_1(X_t,\ \beta^1)\}' W_{1T}\{\frac{1}{T}\sum_{t=1}^{T} f(X_t,\ \beta^1)\},$$

and that the second step estimator minimizes

$$(4.3)\qquad J_{2T}(\beta^2) = \{\frac{1}{T}\sum_{t=1}^{T} f_1(X_t,\beta^1_N,\beta^2)\}' W_{2T}\{\frac{1}{T}\sum_{t=1}^{T} f(X_t,\beta^1_N,\beta^2)\},$$

where W_{iT} is a positive definite matrix that converges to W_{i0} with probability one. Let Γ_{ij} be the $q_i \times p_j$ matrix $E(\partial f_i/\partial\beta_j')$ for i=1,2 and j=1,2.

Given an arbitrary W_{10}, the optimal choice of the second step distance matrix is $W_{20}=(\Omega^*)^{-1}$, where

$$(4.4)\qquad \Omega^* = [-\Gamma_{21}(\Gamma_{11}W_{10}\Gamma_{11})^{-1}\Gamma_{11}W_{10},\ I]\ \Omega \begin{bmatrix} -\Gamma_{21}(\Gamma_{11}W_{10}\Gamma_{11})^{-1}\Gamma_{11}W_{10} \\ I \end{bmatrix}.$$

With this choice of W_{20}, $(1/\sqrt{T})\sum_{t=1}^{T}(\beta_T-\beta_0)$ has an asymptotic normal distribution with mean zero and the covariance matrix

$$\{\Gamma_{22}'(\Omega^*)^{-1}\Gamma_{22}]\}^{-1}, \tag{4.5}$$

and $TJ_{2T}(\beta_T^2)$ has an (asymptotic) chi-square distribution with q_2-p_2 degrees of freedom as $T\to\infty$. It should be noted that if $\Gamma_{21}=0$, then the effect of the first step estimation can be ignored because $\Omega^* = \Omega_{22} = E(f_2(X_t,\beta_0)f_2(X_t,\beta_0)')$.

4.2. GMM with Deterministic Trends

This subsection discusses how GMM can be applied to time series with deterministic trends (see Eichenbaum and Hansen (1990) and Ogaki (1988, 1989) for examples of applications). Suppose that X_t is trend stationary rather than stationary. In particular let

$$X_t = d(t,\beta_0^1) + X_t^*, \tag{4.6}$$

where $d(t,\beta_1)$ is a function of deterministic trends such as time polynomials and X_t^* is detrended X_t. Assume that X_t^* is stationary with $E(X_t^*)=0$ and that there are q_2 moment conditions

$$E(f_2(X_t^*, \beta_0^1, \beta_0^2)) = 0. \tag{4.7}$$

Let $\beta=(\beta^{1\prime}, \beta^{2\prime})'$, $f_1(X_t,\beta^1)=X_t-d(t,\beta^1)$ and $f(X_t,\beta)=[f_1(X_t,\beta^1)', f_2(X_t^*,\beta^1,\beta^2)']'$. Then GMM can be applied to $f(X_t,\beta)$ to estimate β^1 and β^2 simultaneously as shown in Hansen, Heaton, and Ogaki (1992).

4.3. Cross-Sectional Data and Panel Data

GMM has been applied to cross-sectional data and panel data. For example, Hotz, Kydland, and Sedlacek (1988), Altug and Miller (1990), Runkle (1991) have applied GMM to panel data and discuss econometric issues in detail. The reader who is interested in econometric issues that are not treated in this subsection is referred to these papers. In this section, I explain a simple method to allow for a general form of serial correlation in each cohort in panel data (see e.g., Atkeson and Ogaki (1991) and Ogaki and Atkeson (1991) for examples of applications.)

Consider a panel data set in which there exist H cohorts (for example, a cohort consists of the individuals in a village). Suppose that cohort h include N_h individuals and that the data set contain T periods of observations. Let $N=\sum_{h=1}^{H} N_h$ be the total number of individuals. It is assumed that N is large compared with T, so that we drive N to infinity with T fixed in considering asymptotic distributions. Assume that individuals $i=1,\ldots,N_1$ are in cohort 1 and $i=N_1+1,\ldots,N_1+N_2$ are in cohort 2, etc. It is assumed that $\lim_{N\to\infty} N_h/N=\delta_h$. Let x_{it} be a random vector of economic variables for an individual i at period t and $f_t(x_{it},\beta)$ be a q_t-dimensional vector of functions. Here $\sum_{t=1}^{T} q_t=q$. Let $X_i=(x_{i1}',\ldots,x_{iT}')'$ and $f(X_i,\beta)=(f_1(x_{i1})',\ldots,f_T(x_{iT})')'$. It is assumed that X_i is identically and independently distributed over the individuals. Assume that there exist q orthogonality conditions:

$$E_N(f(X_i,\beta_0)) = 0, \tag{4.8}$$

where E_N is the unconditional expectation operator over individuals. A subscript N is attached to emphasize that the expectation is taken over individuals.

It is assumed that a law of large number applies to f, so that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=N_{h-1}+1}^{N_h} f(X_i, \beta) = \delta_h E_N(f(X_i, \beta)) \tag{4.9}$$

for each $h=1,\ldots,H$. Let $g_N(X_i,\beta)=(\sum_{i=1}^{N_1} f(X_i,\beta)', \ldots, \sum_{i=N_{H-1}+1}^{N_H} f(X_i,\beta)')'$. Then the GMM estimator β_N minimizes a quadratic form

$$J_N(\beta) = \{\frac{1}{N} g_N(X_i,\ \beta)\}' W_N \{\frac{1}{N} g_N(X_i,\ \beta)\}, \tag{4.10}$$

where W_N is a positive definite matrix, which satisfies

$$\lim_{N\to\infty} W_N = W_0. \tag{4.11}$$

with probability one for a positive definite matrix W_0.

Suppose that a central limit theorem applies to the disturbance of GMM, $u_i=f(X_i,\beta_0)$, so that $(1/\sqrt{N})\sum_{i=N_{h-1}+1}^{N_h} u_i$ has a (asymptotic) normal distribution with mean zero and the covariance matrix Ω_h for large N. Here $\Omega_h=\delta_h E_N(u_i u_i')$ for any individual i in cohort h. Let Ω be a matrix that has Ω_h in the h-th diagonal blocks for $h=1,\ldots,H$ and zeros elsewhere. With these modifications, the GMM framework explained in Section 2 can be applied to this problem with all limits taken as $N\to\infty$ instead of $T\to\infty$. For example, $W_0=\Omega^{-1}$ leads to an efficient GMM estimator and $NJ(\beta_N)$ has an asymptotic chi-square distribution with this choice of the distance matrix.

4.4. *The Minimum Distance Estimation*

The Minimum Distance Estimation (MDE) provides a convenient way to obtain an efficient estimator that imposes nonlinear restrictions (see, e.g, Chiang (1956), Ferguson (1958), and Chamberlain (1982, 1984)) and a test statistic for the restrictions. The MDE is closely related to GMM and a GMM program can be used to implement the MDE (see, e.g., Ogaki (1992a)). Suppose that θ_T is an unrestricted estimator for a *p+s* vector of parameters θ_0, and that $\sqrt{T}(\theta_T-\theta_0)$ converges in distribution to a normal random vector with the covariance matrix Ω. Consider nonlinear restrictions such that

$$(4.12) \qquad \phi(\beta_0) = \theta_0,$$

where β_0 is a p-dimensional vector of parameters. The MDE estimator, β_T, minimizes a quadratic form

$$(4.13) \qquad J_T(\beta) = \{\phi(\beta)-\theta_T\}' W_T \{\phi(\beta)-\theta_T\},$$

for a positive definite matrix W_T that converges to a positive definite matrix W_0 with probablity one. As with the GMM estimation, $W_0=\Omega^{-1}$ is the optimal choice of the distance matrix and $TJ_T(\beta_T)$ has an asymptotic chi-square distribution with s degrees of freedom. The null hypothesis that (4.12) holds is rejected when this statistic is lager than critical values obtained from chi-square distributions.

5. Important Assumptions

In this section, I discuss two assumptions under which large sample properties of GMM estimators are derived. These two assumptions are important in the sense that applied researchers have encountered cases where these assumptions are obviously violated unless special cares are taken.

5.1. Stationarity

In Hansen (1982), X_t is assumed to be (strictly) stationary.[8] A time series $\{X_t:-\infty<t<\infty\}$ is stationary if the joint distribution of $\{X_t, \ldots, X_{t+\tau}\}$ are identical to those of $\{X_{t+s},\ldots,X_{t+s+\tau}\}$ for any τ and s. Among other things, this implies that unconditional moments $E(X_t)$ and $E(X_tX_{t+\tau}')$ cannot

---

[8] In the context of cross sectional data discussed in Subsection 4.3, this means identical distributions for X_i.

depend on t for any τ when these moments exist. Thus this assumption rules out deterministic trends, autoregressive unit roots, and unconditional heteroskedasticity.[9] On the other hand, conditional moments $E(X_{t+\tau}|I_t)$ and $E(X_{t+\tau}X_{t+\tau+s}'|I_t)$ can depend on I_t. Thus the stationarity assumption does *not* rule out the possibility that X_t has conditional heteroskedasticity. It should be noted that it is not enough for $u_t=f(X_t,\beta_0)$ to be stationary. It is required that X_t is stationary, so that $f(X_t,\beta)$ is stationary for admissible β that is not necessarily equal to β_0 (see Subsection 8.1.4 for an example in which $f(X_t,\beta_0)$ is stationary but $f(X_t,\beta)$ is not for other values of β).

Since many macroeconomic variables exhibits nonstationarity, this assumption can be easily violated in applications unless a researcher is careful. As explained in Subsection 4.2, nonstationarity in the form of the trend stationarity can be treated with ease. In order to teat another popular form of nonstationarity, unit-root nonstationarity, researchers have used transformations such as first differences or growth rates of variables (see Section 8 for examples).

5.2. Identification

Another important assumption of Hansen (1982) is related to

---

[9] Gallant (1986) and Gallant and White (1988) show that the strict stationarity assumption can be relaxed for GMM. They allow for unconditional heteroskedasticity that the stationarity assumption rules out. This does *not* mean that X_t can exhibit nonstationarity by having deterministic trends, autoregressive unit roots, or an integrated GARCH representation. Some of their regularity conditions are violated by these popular forms of nonstationarity and X_t needs to be detrended if it is trend staionary. For this reason, I emphasize the strict stationarity assumption in the context of time series applications rather than the fact this assumption can be relaxed.

identification. Let

(5.1) $$J_0(\beta) = \{E[f(X_t, \beta)]\}' W_0 \{E[f(X_t, \beta)]\}.$$

Then the identification assumption is that β_0 is a unique minimizer of $J_0(\beta)$.[10] Since $J_0 \geq 0$, β_0 is obviously a minimizer. This assumption requires $J_0(\beta)$ to be strictly positive for any other β. A case where this assumption is obviously violated is that $f(X_t, \beta) \equiv 0$ for some β which does not have any economic meanings (see Section 8 for examples).

6. Covariance Matrix Estimation

An estimates of Ω is necessary to calculate asymptotic standard errors of the GMM estimator from (2.6) and to utilize the optimal distance matrix Ω^{-1}. This section discusses estimation methods for Ω. In the following, I assume that a consistent estimator β_T for β_0 is available to form an estimate of u_t by $f(X_t, \beta_T)$. In most applications, the first stage GMM estimator is obtained by setting $W_T = I$, and then Ω_T is estimated from the first stage GMM estimate β_T. The second stage GMM estimator is formed by setting $W_T = \Omega_T^{-1}$. This procedure can be iterated by using the second stage GMM estimate to form the distance matrix for the third stage GMM estimator, and so on. Kocherlakota's (1990) and Ferson and Foerster's (1991) Monte Carlo simulations suggest that the GMM estimator and test statistics have better small sample properties when this procedure is iterated. It is preferable to iterate this procedure until a convergence is obtained. In some nonlinear models, this may be costly in terms of time. In such cases,

---

[10] Hansen, Heaton, and Ogaki (1992) show that a sequence of the sets of the minimizers for (2.3) converges to the set of the minimizers with probability one when all regularity conditions except for this identification assumption hold.

it is recommended that the third stage GMM to be used because the gains from further iterations may be small.

6.1. Serially Uncorrelated Disturbance

This subsection treats the case where $E(u_i u_j)=0$ for $i\neq j$. In the contest of time series data, this means that there is no serial correlation. In the context of the cross sectional and panel data model in Subsection 4.3, this means that the disturbance is uncorrelated across households, even though it can be serially correlated.

In this case, Ω can be estimated by $(1/T)\sum_{t=1}^{T}f(X_t,\beta_T)f(X_t,\beta_T)'$. In the models considered in Section 3, this is White's (1980) heteroskedasticity consistent estimator. For example, consider the NLIV model explained in Section 3. In this model, $u_t=z_t g(X_t,\beta_0)$ and $(1/T)\sum_{t=1}^{T}f(X_t,\beta_T)f(X_t,\beta_T)'=(1/T)\sum_{t=1}^{T}z_t g(X_t,\beta_T)g(X_t,\beta_T)'z_t'$. Note that u_t is serially uncorrelated if $e_t=g(X_t,\beta_0)$ is in the information set I_{t+1} because $E(u_t u'_{t+j})=E(E(u_t u'_{t+j}|I_{t+1}))=E(u_t E(u'_{t+j})|I_{t+1}))=0$ for $j\geq 1$. In some cases, conditional homoskedasticity is assumed and an econometrician may wish to impose this on his estimate for Ω. In this case $(1/T)\sum_{t=1}^{T}z_t\{(1/T)\sum_{t=1}^{T}g(X_t,\beta_T)g(X_t,\beta_T)'\}z_t'$ is used to estimate Ω.

6.2. Serially Correlated Disturbance

This subsection treats the case where the disturbance is serially correlated in the context of time series analysis.

6.2.1. Unknown Order of Serial Correlation

In many applications, the order of serial correlation is unknown. Let $\Phi(\tau)=E(u_t u'_{t-\tau})$ and

$$(6.1) \qquad \Phi_T(\tau) = \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} f(X_t,\beta_T)f(X_{t-j},\beta_T)' & \text{for } j\geq 0 \\ \frac{1}{T}\sum_{t=j+1}^{T} f(X_{t+j},\beta_T)f(X_t,\beta_T)' & \text{for } j<0. \end{cases}$$

Many estimators for Ω in the literature have the form

$$(6.2) \qquad \Omega_T = \frac{T}{T-p}\sum_{\tau=-T+1}^{T-1} k(\frac{\tau}{S_T})\ \Phi_T(\tau),$$

where $k(\cdot)$ is a real-valued kernel, and S_T is a band-width parameter. The factor T/(T-p) is a small sample degrees of freedom adjustment. See Andrews (1991) for examples of kernels. Hansen (1982) and White's (1984, p.152) estimator corresponds with the truncated kernel; Newey and West's (1987a) estimator, with the Bartlett kernel; and Gallant's (1987, p.533), with the Parzen kernel. The estimators corresponding with these kernels place zero weights for $\Phi_T(\tau)\geq S_T$, so that S_T-1 is called the lag truncation number. Andrews (1991) advocates an estimator which uses the Quadratic Spectral (QS) kernel, which does not place zero weights for any $\Phi(\tau)$ for $|\tau|\leq T-1$.[11]

One important problem is how to choose the bandwidth parameter S_T. Andrews (1991) provides formulas for optimal choice of the bandwidth parameter for a variety of kernels. Unfortunately, these formulas include unknown parameters. Andrews proposes automatic bandwidth estimators in which these unknown parameters are estimated from the data. His method involves two steps. The first step is to parameterize the law of motion of the disturbance u_t and to estimate the parameterized law of motion. The second step is to estimate the parameters for the optimal bandwidth

---

[11]Hansen (1990b) relaxes an assumption made by these authors to show consistency of the kernel estimators are relaxed in Hansen (1990b).

the estimated law of motion. In his Monte Carlo simulations, Andrew uses a AR(1) parameterization for each term of the disturbance. This seems to work well in the models he considers.

Another problem is to choose the kernel. One serious problem with the truncated kernel is that the corresponding estimator is not guaranteed to be positive semidefinite. Andrews (1991) discusses that the QS kernel is an optimal kernel in the sense that it minimizes asymptotic MSE among the estimators of the form (6.3) that are guaranteed to be positive semidefinite. His Monte Carlo simulations show that the QS kernel and the Parzen kernel work better than Bartlett kernel in most of the models he consider. He also finds that even the estimators based on the QS kernel and the Parzen kernel are not satisfactory in the sense that the standard errors calculated from these estimators are not accurate in small samples when the amount of autocorrelation is large.

Because the estimators of the form (6.3) do not seem satisfactory, Andrews and Monahan (1990) propose an estimator based on a VAR prewhitening. The intuition behind this is that the estimators of the form (6.2) only takes care of MA components of u_t and cannot handle the AR components well in small samples. The first step in the VAR prewhitening method is to run a VAR of the form

$$(6.3) \qquad u_t = A_1 u_{t-1} + A_2 u_{t-2} + \dots + A_n u_{t-n} + v_t.$$

Note that the model (6.3) need not be a true model in any sense. Then the estimated VAR is used to form an estimate v_t and an estimator of the form (6.2) is applied to the estimated v_t to estimate the long-run variance of v_t, Ω^*. The estimator based on the QS kernel with the automatic bandwidth parameter can be used to v_t for example. Then the sample counterpart of the formula

$$(6.4) \qquad \Omega = [I - \sum_{\tau=1}^{n} A_\tau]^{-1}\Omega^*[I - \sum_{\tau=1}^{n} A'_\tau]^{-1}$$

is used to form an estimate of Ω. Andrews and Monahan uses the VAR of order one in their Monte Carlo simulations. Their results suggest that the prewhitened kernel estimator performs better than the nonprewhitened kernel estimators for the purpose of calculating standard errors of estimators.[12]

In sum, Monte Carlo evidence suggests that the VAR prewhitened QS or Parzen kernel estimator with Andrews's (1991) automatic bandwidth parameter is to be recommended. Though the QS kernel estimator may be preferred to the Parzen kernel estiamator because of its asymptotic optimality, it takes more time to calculate the QS kernel estimators than the Parzen kernel estimators. This difference may be important when estimation is repeated many times.

6.2.2. Known Order of Serial Correlation

In some applications, the order of serial correlation is known. Assume that the order of serial correlation is known to be s. For example, consider the NLIV model explained in Section 3. Suppose that e_t is in the information set I_{t+s+1}. In multi-period forecasting models, s is greater than one (see Hansen and Hodrick (1980, 1983) and Section 8 of the present paper for examples). Then $E(u_t u'_{t+j}) = E(E(u_t u'_{t+j} | I_{t+s+1})) = E(u_t E(u'_{t+j}) | I_{t+s+1})) = 0$ for $j \geq s+1$. Thus the order of serial correlation of u_t is s and u_t has an MA(s) structure in this case.

---

[12] Park and Ogaki's (1991b) Monte Carlo simulations suggest that the VAR prewhitening improves estimators of Ω in the context of cointegrating regressions.

In this case, there exit restrictions that $\Phi(\tau)=0$ for $|\tau|>s$, which I shall call the zero restrictions on autocovariances. It is likely that imposing these zero restrictions on the estimator of Ω leads to a more efficient estimator.[13] Since $\Omega=\sum_{\tau=-s}^{s}\Phi(\tau)$ in this case, a natural estimator is

$$(6.5) \qquad \Omega_T = \frac{T}{T-p}\sum_{\tau=-s}^{s}\Phi_T(\tau),$$

which corresponds with the truncated kernel estimator. Hansen and Hodrick (1980) study a multi-period forecasting model that leads to $s\geq1$. They use (6.5) with conditional homoskedasticity imposed as discussed at the end of Subsection 6.1 in the present paper. Their method of calculating the standard errors for linear regressions is known as Hansen-Hodrick correction.

A possible problem with the estimator (6.1) is that Ω_T is not guaranteed to be positive semidefinite if $s\geq1$. In applications, researchers often encounter cases where Ω_T is invertible but is not positive semidefinite. If this happens, $W_T=\Omega_T^{-1}$ should not be used to form the optimal GMM estimator (see, e.g., Newey and West (1987a)). There exist at least two ways to handle this problem. One way is to use Eichenbaum, Hansen, and Singleton's (1988) modified Durbin's method. The first step of this method is to estimate the VAR (6.3) for a large n by solving the Yule Walker equations. The second step is to estimate an MA(s) representation

---

[13] In some applications, the order of serial correlation may be different for different terms of u_t. The econometrician may wish to impose these restrictions.

(6.6) $$u_t = B_1 v_{t-1} + \ldots + B_s v_{t-s} + e_t,$$

by running estimated u_t on estimated lagged v_t's. Then the sample counterpart of

(6.7) $$\Omega = (I + B_1 + \ldots + B_s)E(e_t e_t')(I + B_1 + \ldots + B_s)'$$

is used to form an estimate of Ω that imposes the zero restrictions. One problem with this method is that this is not reliable when the number of elements in u(t) is large compared with the sample size because two many parameters in (6.3) need to be estimated. The number elements in u(t) need to be kept as small as possible when this method is to be used.

Another method is to use one of the kernel estimators of the form (6.2) (or VAR prewhitened kernel estimators if s is large) that are guaranteed to be positive semidefinite. When this method is used, the zero restrictions should *not* be imposed even though $\Phi(\tau)$ is known to be zero for $|\tau|>s$. In order to illustrate this in an simple example, consider the case where s=1 and Newey-West's (1987) Bartlett kernel estimator is used. Then

(6.6) $$\Omega_T = \frac{T}{T-p} \sum_{\tau=-\ell}^{\ell} \frac{T-|\tau|}{T} \Phi_T(\tau),$$

where $\ell = S_T - 1$ is the lag truncation number. If $\ell=1$ is used to impose the zero restrictions, then Ω_T converges to $\Phi(0)+(1/2)\Phi(1)+(1/2)\Phi(-1)$, which is not equal to Ω. Thus ℓ needs to be increased as T is increased to obtain a consistent estimator. On the other hand, if $\ell>1$ is used and the zero restrictions are imposed by setting $\Phi_T(\tau)$ in (6.6) to zero for $|\tau|>1$, then the resulting estimator is no longer guaranteed to be positive semidefinite.

7. Hypothesis Testing and Specification Tests

In this section, I discuss specification tests and Wald, Lagrange Multiplier (LM), and likelihood ratio type statistics for hypothesis testing. Gallant (1987), Newey and West (1987b), and Gallant and White (1988) have considered these three test statistics, and the Eichenbaum, Hansen and Singleton (1988) considered the third test for hypothesis testing for GMM (or more general estimation method that includes GMM as a special case).

Consider s nonlinear restrictions

$$H_0: R(\beta_0)=r \tag{7.1}$$

where R is a $s\times 1$ vector of functions. The null hypothesis H_0 is tested against the alterative that $R(\beta_0)\neq r$. Let $\Lambda=\partial R/\partial\beta'$ and Λ_T be an consistent estimator for Λ. It is assumed that Λ is of rank s. If the restrictions are linear, then $R(\beta_0)=\Lambda\beta_0$ and Λ is known. Let β_T^u be an unrestricted GMM estimator and β_T^r be a GMM estimator that is restricted by (9.1). It is assumed that $W_0=\Omega^{-1}$ is used for both estimators.

The Wald test statistic is

$$T(R(\beta_T^u)-r)'[\Lambda_T(\Gamma_T'\Omega_T^{-1}\Gamma_T)^{-1}\Lambda_T']^{-1}(R(\beta_T^u)-r), \tag{7.2}$$

where Ω_T, Γ_T, and Λ_T are estimated from β_T^u. The Lagrange multiplier test statistic is

$$\begin{aligned} LM_T = \frac{1}{T}\sum_{t=1}^{T}&f(X_t,\beta_T^r)'\Omega_T^{-1}\Gamma_T\Lambda_T'(\Lambda_T\Lambda_T')^{-1} \\ &[\Lambda_T(\Gamma_T'\Omega_T^{-1}\Gamma_T)^{-1}\Lambda_T']^{-1}(\Lambda_T\Lambda_T')^{-1}\Lambda_T\Gamma_T'\Omega_T^{-1}\sum_{t=1}^{T}f(X_t,\beta_T^r), \end{aligned} \tag{7.3}$$

where Ω_T, Γ_T, and Λ_T are estimated from β_T^r. Note that in linear models LM_T is equal to (9.2), where Ω_T, Γ_T, and Λ_T are estimated from β_T^r

rather than β^{u}_{T}. The likelihood ratio type test statistic is

$$C_T = T(J_T(\beta^r_T) - J_T(\beta^u_T)), \tag{7.4}$$

which is T times the difference between minimized value the objective function when the parameters are restricted and the minimized value of the objective function when the parameters are unrestricted. It is important that the same estimator for Ω is used for both unrestricted and restricted estimation for the C_T test statistic. Under a set of regularity conditions, all three test statistics have asymptotic chi-square distributions with s degrees of freedom. The null hypothesis is rejected when these statistics are larger than critical values obtained from chi-square distributions.

Existing Monte Carlo evidence suggests that the small sample distributions of the Lagrange multiplier test and the likelihood ratio type test are better approximated by their asymptotic distributions than those of the Wald test (see Gallant (1987)). Another disadvantage of the Wald test is that the test result for nonlinear restrictions depends on the parameterization in general (see, e.g., Gregory and Veall (1985) and Phillips and Park (1988)).

Though the chi-square test for the overidentifying restrictions discussed in Section 2 has been frequently used as a specification test in applications of GMM, other specification tests applicable to GMM are available. These include tests developed by Singleton (1985), Andrews and Fair (1988), Hoffman and Pagan (1988), Andrews (1990), Ghysels and Hall (1990a, 1990b, 1990c), Hansen (1990a), Dufour, Ghysels, and Hall (1991). Some of these tests are discussed by Hall (1991).

8. Empirical Applications

The GMM estimation has been frequently applied to rational expectations models. This section discusses examples of these applications. The purpose is not to provide a survey of the literature but to illustrate applications. I will discuss problems that researchers have encountered in applying GMM and how they have solved them. In this section, I use the notations for the NLIV model in Section 3 in explaining the econometric formulations.

8.1. Euler Equation Approach to Models of Consumption

8.1.1. Hansen and Singleton's (1982) Model

Hansen and Singleton (1982) show how to apply GMM toa Consumption-Based Capital Asset Pricing Model (C-CAPM). Consider an economy in which a representative agent maximizes

$$(8.1) \qquad \sum_{t=1}^{\infty} \delta^t E(U(t)|I_0)$$

subject to a budget constraint. Hansen and Singleton (1982) uses an isoelastic intraperiod utility function

$$(8.2) \qquad U(t) = \frac{1}{1-\alpha}(C_t^{1-\alpha} -1),$$

where C_t is real consumption at t and $\alpha>0$ is the reciprocal of the intertemporal elasticity of substitution (α is also the relative risk aversion coefficient for consumption in this model). The standard Euler equation for the optimization problem is

$$(8.3) \qquad \frac{E[\delta C_{t+1}^{-\alpha} R_{t+1}|I_t]}{C_t^{-\alpha}} = 1,$$

where R_{t+1} is the (gross) real return of any asset. The observed C_t they use is obviously nonstationary, though it is not clear what form of nonstationarity that it takes. Hansen and Singleton uses C_{t+1}/C_t in their econometric formulation, which is assumed to be stationary.[14] Then let $\beta=(\delta,\alpha)$, $X_t=(C_{t+1}/C_t, R_{t+1})'$, and $g(X_t,\beta)=\delta(C_{t+1}/C_t)^{-\alpha}R_{t+1}-1$ in the notations for the NLIV model in Section 2 of the present paper.[15] Stationary variables in I_t, such as the lagged values of X_t, are used for instrumental variables z_t. In this case, u_t is in I_{t+1}, and hence u_t is serially uncorrelated. Hansen and Singleton (1984) find that the chi-square test for the overidentifying restrictions reject their model especially when nominal risk free bond returns and stock returns are used simultaneously. Their finding is consistent with Mehra and Prescott's (1985) equity premium puzzle. When the model is rejected, the chi-square test statistic does not provide much guidance as to what causes the rejection. Hansen and Jagannathan (1991) develop a diagnosis that could provide such guidance.

8.1.2. Time Aggregation

The use of consumption data for the C-CAPM is subject to a time aggregation problem because consumers can make decisions at intervals much finer than the observed frequency of the data and because the observed data consist of average consumption over a period of time.

It is not possible for GMM to take into account of the effect of the time aggregation in nonlinear models in general. For example, Heaton (1991) uses the method of simulated moments (MSM) for his nonlinear asset pricing

---

[14] In the following, assumptions about trend properties of consumption are maid. These assumptions need to be satisfied by equilibrium consumption. The simplest example of economies that satisfy these assumptions is an endowment economy without production in which endowments of consumption satisfy them.

[15] When multiple asset returns are used, $g(X_t,\beta)$ becomes a vector of functions.

model with time-nonseparable preferences in taking the time aggregation into account. Bossaerts (1989), Duffie and Singleton (1989), MacFadden (1989), Pakes and Pollard (1989), Lee and Ingram (1991), and Pearson (1991) have studied asymptotic properties of MSM. It is easier to take into account of the effect of time aggregation in linear models. In linear models, the time aggregation means that the disturbance has an MA(1) structure and the instrumental variables need to be lagged one period more compared with the econometric model that does not take into account of the time aggregation (see, e.g, Grossman, Melino, and Shiller (1987), Hall (1988), and Hansen and Singleton (1988) for applications to C-CAPM and Heaton (1990) and Christiano, Eichenbaum, and Marshall (1991) for applications to Hall's (1978) type models of permanent income hypothesis.

8.1.3. Habit Formation and Durability

Many researchers have considered effects of time-nonseparability in preferences on asset pricing. Let us replace (8.2) by

$$(8.4) \qquad U(t) = \frac{1}{1-\alpha}(S_t^{1-\alpha} - 1),$$

where S_t is service flow from consumption purchases. Purchases of consumption and service flows are related by

$$(8.5) \qquad S_t = a_0 C_t + a_1 C_{t-1} + a_2 C_{t-2} + \dots .$$

This type of specification for time-nonseparability has been used by Mankiw (1982), Hayashi (1985), Dunn and Singleton (1986), Eichenbaum, Hansen, and Singleton (1988), Ogaki (1988, 1989), Eichenbaum and Hansen (1990), Heaton (1990, 1991), Cooley and Ogaki (1991), Ferson and Constantinides (1991), Ferson and Harvey (1991), and Ogaki and Park (1992) among others.[16] Depending on values of a_τ's, the model (8.4) leads to a model with habit formation and/or durability. Constantinides (1990) argues that habit formation could help solving the equity premium puzzle and shows how the intertemporal elasticity of substitution and the relative risk aversion coefficient depend on a_τ's and α in a habit formation model.

In this subsection, I will discuss applications by Ferson and Constantinides (1991), Cooley and Ogaki (1991), and Ogaki and Park (1992) to illustrate econometric formulations.[17] In their models, it is assumed that $a_\tau=0$ for $\tau\geq2$. Let us normalize a_0 to be one, so that $\beta=(\delta,\alpha,a_1)$. The asset pricing equation takes the form

$$(8.6) \qquad \frac{E[\delta\{S_{t+1}^{-\alpha}+\delta a_1 S_{t+2}^{-\alpha}\}R_{t+1}|I_t]}{E[S_t^{-\alpha}+\delta a_1 S_{t+1}^{-\alpha}|I_t]} = 1.$$

Then let $e_t^0=\delta(S_{t+1}^{-\alpha}+\delta a_1 S_{t+2}^{-\alpha})R_{t+1}-(S_t^{-\alpha}+\delta a_1 S_{t+1}^{-\alpha})$. Though Euler equation (8.5) implies that $E(e_t^0|I_t)=0$, this cannot be used as the disturbance for GMM because both of the two regularity assumptions discussed in Section 5 of the present paper are violated. These violations are caused by nonstationarity of C_t and three sets of trivial solutions, $\alpha=0$ and $1+\delta a_1=0$; $\delta=0$ and $\alpha=\infty$; and $\delta=0$ and $a_1=\infty$ with a positive α. In order to solve these problems, Ferson and Constantinides (1991) defines $e_t=e_t^0/S_t^{-\alpha}$, which avoids the trivial solutions. Since $(1+a_1)C_t^{-\alpha}$ is in I_t, $E(e_t|I_t)=0$. The disturbance is a function of $S_{t+\tau}/S_t$ $(\tau=1,2)$ and R_{t+1}. When C_{t+1}/C_t and R_t are assumed to be

---

[16] Some of these authors allow for a possibility of a deterministic technological progress in the transformation technology (8.4).

[17] Eichenbaum, Hansen, and Singleton (1988) and Eichenbaum and Hansen (1990) consider similar models with nonseparability across goods in preferences.

stationary, $S_{t+\tau}/S_t$ and the disturbance can be written as a function of stationary variables.

One problem that researchers have encountered in these applications is that $C_{t+1}+a_1C_t$ becomes a negative number when a_1 is close to minus one. A GMM computer program stalls when it tries such non-admissible values of a_1 that makes $C_{t+1}+a_1C_t$ negative for any t. This may happen in a nonlinear search for β_T or in calculating numerical derivatives for example. Atkeson and Ogaki (1991) have encountered similar problems in estimating fixed subsistence levels from panel data. One way to avoid this problem is to program the function $f(X_t,\beta)$, so that the program returns very large numbers as the values of $f(X_t,\beta)$ when non-admissible values are used. It is necessary to avoid calculating derivatives using these large values of $f(X_t,\beta)$ if numerical derivatives are used. This can be done by modifying a program to calculate numerical derivatives.[18]

8.1.4. Multiple-Good Models

Mankiw, Rotemberg, and Summers (1985), Dunn and Singleton (1986), Eichenbaum, Hansen, and Singleton (1988), Eichenbaum and Hansen (1990), among others have estimated versions of multiple-good C-CAPM. I illustrate basic economic formulations in these multiple-good models in the context of a simple model with a nondurable good and a durable good.

Let us replace (8.2) by Houthakker's (1960) addilog utility function that Miron (1986) and Ogaki (1988,1989) have estimated among others:

---

[18] Ogaki (1992a) explains these modifications for Hansen/Heaton/Ogaki GMM package.

$$U(t) = \frac{1}{1-\alpha}(C_t^{1-\alpha} - 1) + \frac{\theta}{1-\eta}(K_t^{1-\eta} - 1), \tag{8.7}$$

where C_t is nondurable consumption and K_t is household capital stock from purchases of durable consumption good D_t.[19] The stock of durables is assumed to depreciate at a constant rate 1-a, where $0 \le a < 1$:

$$K_t = aK_t + D_t \; . \tag{8.8}$$

Alternatively, K_t can be considered as service flow in (8.5) with $a_\tau = a^\tau$. When $\alpha \neq \eta$, preferences are not quasi homothetic. In applications, the data for K_t is constructed from data for initial stock K_0 and data for D_t for $t=1,\ldots,T$. Let P_t be the intratemporal relative price of durable consumption and nondurable consumption. Then the intraperiod first order condition that equates the relative price with the marginal rate of substitution is

$$P_t = \frac{\theta E(\sum_{\tau=1}^{\infty} \delta^\tau a^\tau K_{t+\tau}^{-\eta} \mid I_t)}{C_t^{-\alpha}} \; . \tag{8.9}$$

Assume that D_{t+1}/D_t is stationary. Then $K_{t+\tau}/D_t$ is stationary for any τ because $K_{t+\tau}/D_t = \sum_{\tau=0}^{\infty} a^\tau D_{t+\tau}/D_t$. From (8.8),

$$\frac{P_t C_t^{-\alpha}}{D_t^{-\eta}} = \theta E[\sum_{\tau=1}^{\infty} \delta^\tau a^\tau \left(\frac{K_{t+\tau}}{D_t}\right)^{-\eta} \mid I_t]. \tag{8.10}$$

---

[19]Since the addilog utility function is not quasi-homothetic in general, the distribution of initial wealth affects the utility function of the representative consumer. The existence of a representative consumer under complete markets is discussed by Ogaki (1990) for general concave utility functions and by Atkeson and Ogaki (1991) for extended addilog utility functions.

Assume that the variables in I_t are stationary.[20] Then (8.10) implies that the $P_tC_t^{-\alpha}/D_t^{-\eta}$ is stationary because the right hand side of (8.10) is stationary. Taking natural logs, we conclude that $\ln(P_t)-\alpha\ln(C_t)+\eta\ln(D_t)$ is stationary. This restriction is called the stationarity restriction.

From (8.9), define

$$(8.11) \qquad e_t^0 = P_tC_t^{-\alpha} - (1-\delta aF)^{-1}\theta K_t^{-\eta},$$

where F is the forward operator. The first order condition (8.9) implies that $E(e_t^0|I_t)=0$. One problem is that e_t^0 involves $K_{t+\tau}$ for τ from 0 to infinity, so that e_t^0 cannot be used as the disturbance for GMM. To solve this problem, define $e_t=(1-\delta aF)e_t^0$. Note that e_t involves only C_t, C_{t+1}, and K_t and that $E[e_t|I_t]=0$. Hence e_t forms the basis of GMM. The only remaining problem is to attain stationarity. One might think it is enough to divide e_t^0 by $K_t^{-\eta}$, so that the resulting e_t is stationary as implied by the stationarity restriction. It should be noted that it is *not* enough for $e_t=g(X_t,\beta_0)$ to be stationary but it is necessary for $g(X_t,\beta)$ to be stationary for $\beta\neq\beta_0$. Hence if α and η are unknown and C_t or D_t is difference stationary, GMM cannot be applied to the first order condition (8.9).[21] Ogaki (1988, 1989) assume that C_t and D_t are trend stationary and apply the method explained in Section 4 to utilize the detrended version of e_t. In these applications, the restrictions on the trend coefficients and the curvature parameters α and η implied by the stationarity restriction are

---

[20] If I_t include nonstationary variables, assume that the right hand side of (8.9) is the same as the expectation conditioned on the stationary variables in I_t.

[21] Cointegrating regressions can be used for this case as explained below.

imposed on the GMM estimators. Imposing the stationarity restrictions helps obtaining more reasonable point estimates for α and η in these applications.

Eichenbaum, Hansen, and Singleton (1988) and Eichenbaum and Hansen (1990) use the Cobb-Douglas utility function, so that α and η are known to be one. They allow preferences to be nonseparable across goods and time-nonseparable, but the stationarity restriction is shown to hold. In this case, the stationarity restriction implies that $P_t C_t^{-1}/K_t^{-1}$ is stationary. This transromation does not involve any unknown parameters to be estimated. Hence this transformation is used to apply GMM to their intraperiod first order conditions.

8.1.5. The Cointegration-Euler Equation Approach

When at least one of C_t and D_t is difference stationary, the stationarity restriction implies cointegration that Engle and Granger (1987) defines. Ogaki (1988) and Ogaki and Park (1992) propose to estimate the curvature parameters α and η of the addilog utility function, using a cointegrating regression.[22] Cooley and Ogaki (1991) combine this cointegration approach with Euler Equation approach based on GMM in a two-step procedure. In the first step, curvature parameters are estimated from a cointegrating regression. In the second step, we use this estimated value of α in the asset pricing equation (8.3) and estimate only δ.[23] This two step procedure does not alter the asymptotic distributions of GMM

---

[22] Ogaki and Park (1992) use Park's (1990) canonical cointegrating regressions and Park and Ogaki's (1991) seemingly unrelated canonical regressions (also see Ogaki (1992b,1992c)).

[23] In applications of Cooley and Ogaki (1991) and Ogaki and Park (1992), time-nonseparability in preferences are allowed for C_t and the asset pricing equation (8.6) is used to estimate δ and a_1.

estimators and test statistics because the cointegrating regression estimator for α is super consistent and converges at a faster rate than $T^{1/2}$.

Cooley and Ogaki (1991) proposes a specification test *a la* Hausman (1978) based on the likelihood ratio type statistic (discussed in Section 7 of the present paper) that tests the cross equation restriction for the cointegrating regression and the GMM disturbance on α. This test has powers against the factors that make the two estimates different, such as nonseparability in preferences across goods, measurement errors, and liquidity constraints.

8.1.6. Monetary Models

Monetary models have been estimated by applying GMM to Euler equations and/or intratemporal first order conditions. Singleton (1985), Ogaki (1988), Finn, Hoffman, and Schlagenhauf (1990), and Bohn (1991) estimate cash-in-advance models, and Poterba and Rotemberg (1987), Eckstein and Leiderman (1989), and Finn, Hoffman, and Schlagenhauf (1990), Imrohoroglu (1991) estimate money-in-the-utility-function (MIUF) models. It turns out that in cash-in-advance models, basically the same asset pricing equation as (8.3) holds as long as the cash-in-advance constraints are binding and C_t is a cash good (in the terminology of Lucas and Stokey (1987)). However, nominal prices of consumption, nominal consumption, nominal asset returns are aligned over time in a different way in monetary models than they are in Hansen and Singleton's (1982) model. Information available to agents at time t is also considered in a different way. As a result, instrumental variables are lagged one period more compared with Hansen-Singleton's model and u_t has an MA(1) structure in nonlinear models (the time aggregation has the same effects in linear models as discussed above). There is some tendency for chi-square test statistics for the overidentifying restrictions to be more favorable for the timing conventions suggested by cash-in-advance models (see Finn, Hoffman, and Schlagenhauf (1990) and Ogaki (1988)). Ogaki (1988) focuses on monetary distortions in relative prices for a cash good and a credit good and do not find monetary distortions in the U.S. data he examines.

8.1.7. Seasonality

Miron (1987) augment Hansen and Singleton's (1982) model by including seasonal taste shifters and argues that the empirical rejection of C-CAPM by Hansen and Singleton (1982) and others might be attributable to the use of seasonally adjusted data. Although this is theoretically possible, English, Miron, and Wilcox (1989) find that seasonally unadjusted quarterly data reject asset pricing equations at least as strongly as seasonally adjusted data. Ogaki (1988) also finds similar empirical results for seasonally unadjusted and adjusted data in the system that involves both asset pricing equations and intraperiod first order conditions. These studies have used seasonally unadjusted quarterly data. Ferson and Harvey (1991) construct seasonally unadjusted monthly data and estimate a C-CAPM with time nonseparable preferences. They find that seasonal habit persistence is empirically relevant. Braun and Evans (1991) estimate a model of seasonal fluctuations that includes a Christmas taste shift in the econometric system including an Euler equation for consumption.

8.1.8. State-Nonseparable Preferences

Epstein and Zin (1991) estimate a model with state-nonseparable preference specification in which the life time utility level V_t at period t is defined recursively by

$$(8.12)\qquad V_t = \{C_t^{1-\alpha} + \delta\{E[V_{t+1}^{1-\alpha}|I_t]\}^{(1-\rho)/(1-\alpha)},$$

where $\alpha>0$ and $\rho>0$. The asset pricing equation for this model is

$$(8.13)\qquad E[\delta^*(R^e_{t+1})^{\eta}(C_{t+1}/C_t)^{\theta}R_{t+1}] = 1,$$

for any asset return R_{t+1}, where $\delta^*=\delta^{(1-\alpha)/(1-\rho)}$, $\eta=(\rho-\alpha)/(1-\rho)$, $\theta=-\rho(1-\alpha)/(1-\rho)$, and R^e_{t+1} is the (gross) return of the optimal portfolio (R^e_{t+1} is the return from period t to t+1 of a security that pays C_t every period forever). Epstein and Zin use the value-weighted return of shares traded on the New York Stock Exchange as R^e_{t+1}. Thus Roll's (1977) critique of CAPM is relevant here as Epstein and Zin discusses.

Even though (8.13) holds for $R_{t+1}=R^e_{t+1}$, the identification assumption discussed in Section 5 is violated for this choice of R_{t+1} because there exists a trivial solution, $(\delta^*,\eta,\theta)=(1,\ 1,\ 0)$, for $g(X_t,\beta)=0$. When multiple returns that include R^e_{t+1} are used simultaneously, then the whole system can satisfy the identification assumption but the GMM estimators for this partially unidentified system are likely to have bad small sample properties. A similar problem arises when R_{t+1} does not include R^e_{t+1} but include multiple equity returns whose linear combination is close to R^e_{t+1}. It should be noted that Epstein and Zin avoid these problems by carefully choosing returns to be included as R_{t+1} in their system.

8.2. Other Empirical Applications

Hansen and Sargent (1982) develop a method to apply GMM to Hansen and Sargent's (1980, 1981a) linear rational expectations models, imposing nonlinear restrictions implied by Wiener-Kolmogorov prediction formulas (see, e.g., Hansen and Sargent (1981b) on a VAR representation. West (1989) extends Hansen and Sargent's (1982) formulas to deterministic terms. Maximum likelihood estimation has been used more frequently for Hansen and Sargent's type linear rational expectations models than GMM (see, e.g, Sargent (1979, 1981a, 1981b), Eichenbaum (1984), and Finn (1989), Giovannini and Rotemberg (1989)), though Hansen and Sargent's (1982) method can be applied to these models. West (1987, 1988a) and Eichenbaum (1990) apply GMM to linear Euler equations in Hansen and Sargent's type model. West (1987, 1988a) uses West's (1988b) results when difference stationary variables are involved.[24]

Singleton (1988) discusses the use of GMM in estimating real business cycle models. Christiano and Eichenbaum (1990) develop a method to estimate real business models, using GMM and apply their method to U.S. data. Braun (1990), Burnside, Eichenbaum, and Rebelo (1990), Braun and Evans (1991) have estimated real business cycle models, among others.

9. Further Issues

9.1 Optimal Choice of Instrumental Variables

In the NLIV model discussed in Section 3 of the present paper, there are infinitely many possible instrumental variables because any variable in I_t can be used as an instrument. Hansen (1985) characterizes a greatest lower bound for the asymptotic covariance matrices of the alternative GMM estimators, an efficiency bound, and optimal instruments that attain the bound. Since it can be time consuming to obtain optimal instruments, an

---

[24] It should be noted that West (1989b) treats the special case of one difference stationary regressor with nonzero drift, which is relevant for his applications cited here. His results do not extend to multiple regressors (see, e.g, Park and Phillips (1988)).

econometrician may wish to compute an estimate of the efficiency bound to assess efficiency losses from using ad hoc instruments. Hansen (1985) provides a method for calculating this bound and optimal instruments for models with conditionally homoskedastic disturbance terms with an invertible MA representation. Hansen, Heaton, and Ogaki (1988) extend this method to models with conditionally heteroskedastic disturbances and models with an MA representation that is not invertible.[25] Hansen and Singleton (1988) calculate these bounds and optimal instruments for a continuous time financial economic model.

9.2. GMM and Semi-Parametric Estimation

In many empirical applications, the density of the random variables is unknown. Chamberlain (1987), Newey (1988), and Hansen (1989) among others have studied the relationship between GMM estimators and efficient semi-parametric estimators in this environment. Technically, Hansen (1989) shows that the GMM efficiency bound coincides with the semi-parametric efficiency bound for finite parameter maximum likelihood estimators for dependent processes. Chamberlain (1987) shows similar results for independently and identically distributed processes.

In order to give an intuitive explanation for the relationship between GMM and semi-parametric estimation, let us consider a simple model that is a special case of the models that Newey (1988) studies:[26]

---

[25]Heaton and Ogaki (1991) provide an algorithm to calculate efficiency bounds for a continuous time financial economic model based on Hansen, Heaton, and Ogaki's (1988) method.

[26]The materials that follows in this subsection was suggested by Adrian Pagan.

(9.1) $$y_t = x_t'\beta_0 + e_t,$$

where the disturbance e_t is a scalar i.i.d. random variable with unknown symmetric density $\phi(e_t)$, and x_t is p-dimensional vector of random variables that are independently distributed of e_t. Since e_t and x_t are independent, the MLE of β, β_T, would maximize the log likelihood

(9.2) $$L = \sum \log \phi(y_t - x_t'\beta),$$

and would solve

(9.3) $$\sum d_t(\beta_T) = 0$$

if ϕ were known, where $d = \partial \log\phi(y_t - x_t\beta)/\partial\beta$ is the score of β. An efficient semi-parametric estimator is formed by estimating the score by a non-parametric method and emulating the MLE.

On the other hand, GMM estimators can be formed from moment restrictions that are implied by the assumption that e_t is distributed symmetrically distributed conditional upon x_t: $E(x_t e_t)=0$, $E(x_t e_t^3)=0$, etc. Noting that the score is of the form $x_t\xi(e_t)$ for a function $\xi(\cdot)$, the GMM estimator with these moment restrictions approximates $\xi(e_t)$ with a polynomial in e_t. Because the density of e_t is assumed to be symmetric, $\xi(e_t)$ is an odd function of e_t and thus odd functions are used to approximate $\xi(e_t)$. Intuitively, with a sufficiently high order polynomial, the unknown score is approximated well enough that the GMM estimator will be as efficient as the efficient semi-parametric estimator.

9.3. Small Sample Properties

Unfortunately, there have not been much work done on small sample properties of GMM estimators. Tauchen (1986) shows that GMM estimators and test statistics have reasonable small sample properties for data produced by simulations for a C-CAPM. Ferson and Foerster (1991) find similar results for a model of expected returns of assets as long as GMM is iterated for estimation of Ω. Kocherlakota (1990) uses preference parameter values of $\delta=1.139$ and $\alpha=13.7$ (in the notations used for (8.1) and (8.2) in the present paper) in his simulations for a C-CAPM that is similar to Tauchen's model. These parameter values are theoretically valid in a model sense that an equilibrium exists but are much larger than the estimates of these preference parameters by Hansen and Singleton (1982) and others. Kocherlakota shows that GMM estimators for these parameters are biased downward and the chi-square test for the overidentifying restrictions tend to reject the null too frequently compared with its asymptotic size. Mao (1990) reports that the chi-square test overrejects for more conventional values of these preference parameters in his Monte Carlo simulations.

Tauchen (1985) investigates small sample properties of Hansen's (1985) optimal instrumental variables GMM estimators. He finds optimal estimators do not perform well in small samples compared with GMM estimators with ad hoc instruments. Tauchen (1985) and Kocherlakota (1985) recommend small number of instruments rather when ad hoc instruments are used.

Nelson and Startz (1990) perform Monte Carlo simulations to investigate properties of t-ratios and the chi-square test for the overidentifying restrictions in the context of linear instrumental variables regressions. Their work is concerned with small sample properties of these statistics when the instruments are poor (in the sense that it is weakly correlated with explanatory variables). They find that the chi-square test tends to reject the null too frequently compared with its asymptotic distribution and that t-ratios tend to be too large when the instrument is poor. Their results for t-ratios may be counterintuitive because one may expect the consequence of having a poor instrument is a large standard error and a low t-ratio. Their results may be expected to carry over to NLIV estimation. Some of the findings by Kocherlakota (1990) and Mao (1990) that are apparently conflincting with those of Tauchen (1986) may be related to this problem of poor instruments (see Canova, Finn, and Pagan (1991) for a related discussion).

10. Concluding Remarks

In this paper, I have explained statistical properties of GMM estimators and test statistics and important regularity conditions underlying the GMM based inference that applied researchers should be aware of. I have discussed estimation methods for Ω have in detail, recently developed statistics for hypothesis testing and specification tests, and empirical applications. It seems that there is much more room for research on optimal choice of instrumental variables and small sample properties of GMM estimators and test statistics.

References

Altug, S. and R.A. Miller (1990): "Household Choices in Equilibrium," *Econometrica*, 58, 543-570.

Andrews, D.W.K. (1990): "Tests for Parameter Instability and Structural Change with Unknown Change Point, Cowles Foundation Discussion Paper No.943, New Haven, CT: Yale University.

Andrews, D.W.K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59, 817-858.

Andrews, D.W.K. and R. Fair (1988): "Inference in Econometric Models with Structural Change, *Review of Economic Studies*, 55, 615-640.

Andrews, D.W.K., and J.C. Monahan (1990): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator." Manuscript. New Haven, Connecticut: Yale University.

Atkeson, A., and M. Ogaki (1991): "Wealth-Varying Intertemporal Elasticity of Substitution: Evidence from Panel and Aggregate Data." RCER Working Paper No.303. Rochester, New York: University of Rochester.

Barro, R.J. (1977): "Unanticipated Money Growth and Unemployment in the United States," *American Economic Review* 67, 101-115.

Bohn, H. (1991): "On Cash-in Advance Models of Money Demand and Asset Pricing," *Journal of Money, Credit, and Banking* 23, 224-242.

Bossaerts, P. (1989): "The Asymptotic Normality of Method of Simulated Moments Estimators of Option Pricing Models," manuscript.

Braun, R. A. (1990): "The Dynamic Interaction of Distortionary Taxes and Aggregate Variables in Postwar U.S. Data," manuscript, University of Virginia.

Braun, R.A. and C.L. Evans (1991): "Seasonal Solow Residuals and Christmas: A Case for Labor Hoarding and Increasing Returns," manuscript.

Burnside, C., M. Eichenbaum, and S. Rebelo (1990): "Labor Hoarding and the Business Cycle," manuscript.

Canova, F., M. Finn, and A. Pagan (1991): "Econometric Issues in the Analysis of Equilibrium Models," manuscript.

Chamberlain, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5-46.

——— (1984): "Panel Data" in *Handbook of Econometrics*, vol.II edited by Z. Griliches and M. D. Intriligator, Amsterdam: North Holland.

——— (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305-334.

Chiang, C.L. (1956): "On Regular Best Asymptotically Normal Estimates," *Annals of Mathematical Statistics*, 27, 336-351.

Christiano, L.J. and M. Eichenbaum (1990): "Current Real Business Cycle Theories and Aggregate Labor Market Fluctuations," forthcoming in *American Economic Review*.

Christiano, L.J., M. Eichenbaum, and D. Marshall (1991): "The Permanent Income Hypothesis Revisited," *Econometrica* 59, 397-423.

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

Cooley, T.F., and M. Ogaki (1991): "A Time Series Analysis of Real Wages, Consumption, and Asset Returns under Optimal Labor Contracting: A Cointegration-Euler Equation Approach." RCER Working Paper No.285, Rochester, New York: University of Rochester.

Duffie, D. and K.J. Singleton (1989): "Simulated Moments Estimation of Markov Models of Asset Prices," manuscript, Graduate School of Business, Stanford University.

Dunn, K.B. and K.J. Singleton (1986): "Modeling the Term Structure of Interest Rates Under Non-Separable Utility and Durability of Goods," *Journal of Financial Economics* 17, 27-55.

Eckstein, Z. and L. Leiderman (1989): "Estimating an Intertemporal Model of Consumption, Money Demand, and Seignorage, manuscript, Tel-Aviv: Tel-Aviv University.

Eichenbaum, M. (1984): "Rational Expectations and the Smoothing Properties of Inventories of Finished Goods," *Journal of Monetary Economics*, 14, 71-96.

Eichenbaum, M. (1990): "Some Empirical Evidence on the Production Level and Production Cost Smoothing Models of Inventory Investment," *American Economic Review*, 79, 853-864.

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

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

Engle, R.F., and C.W.J. Granger (1987): "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251-276.

Epstein, L. and S. Zin (1991): "Substitution, Risk Aversion and the Temporal Behavior of Asset Returns," *Journal of Political Economy* 99, 263-286.

Ferguson, T.S. (1958): "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities," *Annals of Mathematical Statistics*, 29, 336-351.

Ferson, W.E., and G.M. Constantinides (1991): "Habit Persistence and Durability in Aggregate Consumption," *Journal of Financial Economics* 29, 199-240.

Ferson, W.E., and S.R. Foerster (1991): "Finite Sample Properties of the Generalized Method of Moments in Tests of Conditional Asset Pricing Models," manuscript, University of Chicago and University of Western Ontario.

Ferson, W.E., and C.R. Harvey (1991): "Seasonality and Consumption-Based Asset Pricing," forthcoming in *Journal of Finance*.

Finn, M.G. (1989): "An Econometric Analysis of the Intertemporal General Equilibrium Approach to Exchange Rate and Current Account Determination," *Journal of International Money and Finance* 8, 467-486.

Finn, M.G., D.L. Hoffman, and D.E. Schlagenhauf (1990): "Intertemporal Asset-Pricing Relationships in Barter and Monetary Economies: An Example Analysis," *Journal of Monetary Economics* 25, 431-451.

Gallant, A.R. (1987): *Nonlinear Statistical Models*. New York: John Wiley and Sons.

Gallant, A.R. and H. White (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, New York: Basil Blackwell.

Giovannini, A. and J.J. Rotemberg (1989): "Exchange-Rate Dynamics With Sticky Prices: The Deutsche Mark, 1974-1982," *Journal of Business & Economic Statistics* 7, 169-178.

Gregory, A. and M. Veall (1985): "On Formulating Wald Tests of Nonlinear Restrictions," *Econometrica* 53, 1465-1468.

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

Hall, A. (1991): "An Introduction to Generalized Method of Moments Estimation," manuscript, North Carolina State University.

Hall, R. (1978): "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis," *Journal of Political Economy* 86, 971-987.

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

Hayashi, F. (1985): "The Permanent Income Hypothesis and Consumption Durability: Analysis Based on Japanese Panel Data," *Quarterly Journal of Economics*, 100, 1083-1113.

Hansen, B. (1990a): "Lagrange Multiplier Tests for Parameter Instability in Non-Linear Models," manuscript, University of Rochester.

——— (1990b): "Consistent Covariance Matrix Estimation for Dependent Heterogeneous Processes," manuscript, University of Rochester.

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

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

——— (1988): "Semi-Parametric Efficiency Bounds for Linear Time Series Models," manuscript.

Hansen, L.P., J. Heaton, and M. Ogaki (1988): "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions," *Journal of the American Statistical Association* 83, 863-871.

——— (1992): "Lecture Notes on Generalized Method of Moments Estimators," manuscript in progress.

Hansen, L.P. and R.J. Hodrick (1980): "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy* 88, 829-853.

——— (1983): "Risk Averse Speculation in the Forward Exchange Market: An Econometric Analysis of Linear Models," in J.A. Frenkel, ed., *Exchange Rates and International Macroeconomics*, Chicago: University of Chicago Press.

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

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

——— (1981a): "Linear Rational Expectations Models for Dynamically Interrelated Variables," in R.E. Lucas, Jr, and T.J. Sargent, eds, *Rational Expectations and Econometric Practice*, Minneapolis, MN: University of Minnesota Press.

——— (1981b): "A Note on Wiener-Kolmogorov Prediction Formulas for Rational Expectations Models," *Economics Letters*, 255-260.

——— (1982): "Instrumental Variables Procedures for Estimating Linear Rational Expectations Models, *Journal of Monetary Economics*, 9, 263-296.

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

——— (1984): "Errata," *Econometrica*, 52, 267-268.

——— (1988): "Efficient Estimation of Linear Asset Pricing Models with Moving Average Errors," manuscript.

Heaton, J.H. (1990): "The Interaction between Time-Nonseparable Preferences and Time Aggregation," Sloan School of Management Working Paper No.3181-90-EFA, Massachusetts Institute of Technology.

——— (1991): "An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications," Sloan School of Management Working Paper No.3245-91-EFA, Massachusetts Institute of Technology.

Heaton, J.H. and M. Ogaki (1991): "Efficiency Bound Calculations for a Time Series Model with Conditional Heteroskedasticity," *Economics Letters*, 35, 167-171.

Hoffman, D. and A. Pagan (1988): "Post-Sample Prediction Tests for Generalized Method of Moments," manuscript.

Hotz, V.J., F.E. Kydland, and G.L. Sedlacek (1988), "Intertemporal Preferences and Labor Supply," *Econometrica* 56, 335-360.

Imrohoroglu, S. (1991): "An Empirical Investigation of Currency Substitution," manuscript, University of Sourthern California.

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

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

Lucas, R.E., Jr. and N.L. Stokey (1987): "Money and Interest in a Cash-in-Advance Economy," *Econometrica*, 55, 5491-5513.

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

Mankiw, N.G. (1982): "Hall's Consumption Hypothesis and Durable Goods," *Journal of Monetary Economy*, 10, 1982, 417-425.

Mankiw, N.G., J. Rotemberg, and L. Summers (1985): "Intertemporal Substitution in Macroeconomics," *Quarterly Journal of Economics*, 100, 225-252.

Mao, C.S. (1990): Hypothesis Testing and Finite Sample Properties of Generalized Method of Moments Estimators: A Monte Carlo Study," manuscript, Federal Reserve Bank of Richmond.

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

Nelson, C. and R. Startz (1990): "The Distribution of the Instrumental Variables Estimator and Its t-Ratio When the Instrument is a Poor One," *Journal of Business*, 63, S125-S140.

Newey, W.K. (1984): "A Method of Moments Interpretation of Sequential Estimators," *Economics Letters* 14, 201-206.

——— (1985): "Generalized Method of Moments Specification Testing," *Journal of Econometrics* 29, 229-256.

——— (1988): "Adaptive Estimation of Regression Models Via Moment Restrictions," *Journal of Econometrics* 38, 301-339.

Newey, W.K. and K.J. West (1987a): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

——— (1987b): "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review* 28, 777-787.

Ogaki, M. (1988): *Learning about Preferences from Time Trends*, unpublished Ph.D. dissertation. Chicago: University of Chicago.

——— (1989): "Information in Deterministic Trends about Preferences." Manuscript. Rochester, New York: University of Rochester.

——— (1990): "Aggregation of Intratemporal Preferences under Complete Markets," RCER Working Paper No. 249, Rochester, N.Y.: University of Rochester.

——— (1992a): "GMM: A User's Guide," manuscript. Rochester, New York: University of Rochester.

——— (1992b): "Applications of Cointegration to Models with Optimizing Agents," manuscript. Rochester, New York: University of Rochester.

——— (1992c): "CCR: A User's Guide," manuscript. Rochester, New York: University of Rochester.

Ogaki, M. and A. Atkeson (1991): "Estimating Subsistence Levels with Euler Equations from Panel Data," manuscript.

Ogaki, M. and J.Y. Park (1992): "A Cointegration Approach to Estimating Preference Parameters," manuscript.

Pakes, A. and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica* 57, 1027-1058.

Pagan, A. (1984): "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review* 25, 221-247.

——— (1986): "Two Stage and Related Estimators and Their Applications," *Review of Economic Studies* LIII, 517-538.

Park, J.Y. (1990): "Canonical Cointegrating Regressions," CAE Working Paper No.88-29R, Cornell University, forthcoming in *Econometrica*.

Park, J.Y., and Ogaki, M. (1991a): "Seemingly Unrelated Canonical Cointegrating Regressions," RCER Working Paper No. 280, Rochester, New York: University of Rochester.

——— (1991b): "Inference in Cointegrated Models Using VAR Prewhitening to Estimate Shortrun Dynamics," RCER Working Paper No. 281, Rochester, New York: University of Rochester.

Park, J.Y. and P.C.B. Phillips (1988): "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory* 4, 468-497.

Pearson, N. (1991): "A Simulated Moments Estimator of Discrete Time Asset Pricing Models, manuscript, University of Rochester.

Phillips, P.C.B. and J.Y. Park (1988): "On the Formulation of Wald Tests of Nonlinear Restrictions," *Econometrica* 56, 1065-1083.

Poterba, J. and J. Rotemberg (1987): "Money in the Utility Function: An Empirical Implementation," in W. Barnett and K. Singleton, eds.,*New Approaches to Monetary Economics, Proceedings of the Second International Symposium in Economic Theory and Econometrics*, Cambridge: Cambridge University Press, 219-240.

Roll, R. (1977): "A Critique of the Asset Pricing Theory's Tests," *Journal of Financial Economics*, 4, 129-176.

Sargent, T.J. (1978): "Rational Expectations, Econometric Exogeneity, and Consumption," *Journal of Political Economy* 86, 673-700.

——— (1981a): "The Demand for Money during Hyperinflations under Rational Expectations," in R.E. Lucas, Jr, and T.J. Sargent, eds, *Rational Expectations and Econometric Practice*, Minneapolis, MN: University of Minnesota Press.

——— (1981b): "Estimation of Dynamic Labor Demand Schedules under Rational Expectations," in R.E. Lucas, Jr, and T.J. Sargent, eds, *Rational Expectations and Econometric Practice*, Minneapolis, MN: University of Minnesota Press.

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

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

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

West, K.D. (1987): "A Specification Test for Speculative Bubbles," *Quarterly Journal of Economics*, 102, 553-580.

——— (1988a): "Dividend Innovations and Stock Price Volatility," *Econometrica*, 56, 37-61.

——— (1988b): "Asymptotic Normality, When Regressors Have a Unit Root," *Econometrica*, 56, 1397-1417.

——— (1989): "Estimation of Linear Rational Expectations Models, in the Presence of Deterministic Terms," *Journal of Monetary Economics*, 24, 437-442.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48, 817-838.

White, H. (1984): *Asymptotic Theory for Econometricians*, New York: Academic Press.