G. S. Maddala, C. R. Rao and H. D. Vinod, eds., *Handbook of Statistics, Vol. 11*

17

Generalized Method of Moments: Econometric Applications

Masao Ogaki

1. Introduction

The purpose of this 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 statistical properties of GMM estimators and test statistics are discussed. This paper also 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, and some statistics that are often used for hypothesis testing. In explaining empirical applications, the present paper emphasizes some of the pitfalls that researchers have encountered, and how they have dealt with them.[2]

The rest of this paper is organized as follows. Section 2 presents the basic GMM framework. Section 3 illustrates how ordinary least squares and linear and nonlinear instrumental variables estimation are embedded in the GMM framework as special cases. Section 4 presents some GMM related statistical procedures that extend the basic GMM framework. 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, methods for covariance matrix estimation are explained. These methods are necessary for calculating standard errors of GMM estimators and for using the optimal distance matrix for GMM estimation. Section 7 explains Wald, Lagrange multiplier and likelihood ratio type statistics for hypothesis testing and recently developed specification tests. In Section 8, empirical applications are discussed. Section 9 examines the optimal choice of instrumental variables, the relation between GMM and

[1] Hall (1993) provides a nontechnical introduction to GMM that offers the basic intuition behind GMM.

[2] In a companion paper, Ogaki (1993a), I describe the use of the Hansen/Heaton/Ogaki GAUSS GMM package for implementing GMM estimation and for forming test statistics.

semi-parametric estimation, and small sample properties of GMM estimators and test statistics. Section 10 concludes.

2. 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, 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 time series context is maintained throughout this paper, except for Section 4.3, where applications of GMM to cross sectional data and panel data are discussed. Assume that X_t is (strictly) stationary.[3] We refer to $u_t = f(X_t, \beta_0)$ as the disturbance of GMM. Consider the (unconditional) moment restrictions

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

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

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

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

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

with respect to β; where W_T is a positive semidefinite matrix, which satisfies

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

with probability one for a positive definite matrix W_0. The matrices W_T and W_0 are both referred to as the distance or weighting matrix. The GMM estimator β_T is the solution of the minimization problem (2.3). Under fairly general regularity conditions, the GMM estimator β_T is a consistent estimator for arbitrary distance matrices.[4] The selection of the distance matrix which yields an (asymptotically) efficient GMM estimator is discussed below in Section 2.3.

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

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

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 $(1/\sqrt{T})\,\Sigma_{t=1}^{T} u_t$ has an (asymptotic) normal distribution with mean zero and the covariance matrix Ω in large samples.[5] If u_t is serially uncorrelated, $\Omega = E(u_t u_t')$. If u_t is serially correlated,

$$\Omega = \lim_{j\to\infty} \sum_{-j}^{j} \mathrm{E}(u_t u_{t-j}') . \tag{2.5}$$

Some authors refer to Ω as the long run covariance matrix of u_t. Let $\Gamma = \mathrm{E}(\partial f(X_t, \beta)/\partial\beta')$ 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 suitable regularity conditions, $\sqrt{T}(\beta_T - \beta_0)$ approximately has a normal distribution with mean zero and the covariance matrix

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

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 a just identified system, the GMM estimator does not depend on the choice of distance matrix. When $q > p$, there exist overidentifying restrictions and different GMM estimators are obtained for different distance matrices. In this case, one may choose the distance matrix that results in (asymptotically) efficient GMM estimator. Hansen (1982) shows that the covariance matrix (6) is minimized when $W_0 = \Omega^{-1}$.[6] With this choice of the distance matrix, $\sqrt{T}(\beta_T - \beta_0)$ approximately has a normal distribution with mean zero and the covariance matrix

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

in large samples.

Let Ω_T be a 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, that it is optimal given $f(X_t, \beta)$. In the context of instrumental variable estimation, this means that instrumental variables are given. The optimal selection of instrumental variables is discussed in Section 9. Let Γ_T be a consistent estimator of Γ. Then the standard errors of the optimal GMM estimator β_T are calculated as square roots of the diagonal elements of

[5] An advantage of the GMM estimation is that a strong distributional assumption such that u_t is normally distributed is not necessary.

[6] 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.

$T^{-1}\{\Gamma_T'\Omega_T^{-1}\Gamma_T\}^{-1}$. The appropriate method for estimating Ω depends on the model. This problem is discussed in Section 6. It is usually easier to estimate Γ by $\Gamma_T = (1/T)\,\Sigma_{t=1}^{T}\,(\partial f(X_t, \beta_T)/\partial\beta')$ 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 are overidentifying restrictions ($q > p$), a chi-square statistic can be used to test the overidentifying restrictions. One application of this test is to test the validity of moment conditions implied by Euler equations for optimizing problems of economic agents. This application is discussed in Section 8. Hansen (1982) shows that T times the minimized value of the objective function $TJ_T(\beta_T)$, has an (asymptotic) chi-square distribution with $q - p$ degree of freedom if $W_0 = \Omega^{-1}$ in large samples. This test is sometimes called Hansen's J test.[7]

3. Special cases

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

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 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:

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

Since this is the case of a just identified system, the distance matrix W_0 does not matter. Note that the OLS estimator minimizes $\Sigma_{t=1}^{T}\,(y_t - x_t'\beta)^2$ while the GMM estimator minimizes $(\Sigma_{t=1}^{T}\,(x_t(y_t - x_t'\beta)))'(\Sigma_{t=1}^{T}\,(x_t(y_t - x_t'\beta)))$. It turns out that the GMM estimator coincides with the OLS estimator in this case. To see this, note that $(\Sigma_{t=1}^{T}\,(x_t(y_t - x_t'\beta)))'(\Sigma_{t=1}^{T}\,(x_t(y_t - x_t'\beta)))$ can be minimized by setting β_T so that $\Sigma_{t=1}^{T}\,f(X_t, \beta) = 0$ in the case of a just identified system. This implies that $\Sigma_{t=1}^{T}\,x_ty_t = \{\Sigma_{t=1}^{T}\,x_tx_t'\}\beta_T$. Thus as long as $\{\Sigma_{t=1}^{T}\,x_tx_t'\}$ is invertible, $\beta_T = \{\Sigma_{t=1}^{T}\,x_tx_t'\}^{-1}\{\Sigma_{t=1}^{T}\,x_ty_t\}$. Hence the GMM estimator β_T coincides with the OLS estimator.

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

3.2. *Linear instrumental variables regressions*

Consider the linear model (3.1) and 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 $u_t = z_t e_t$. Thus the moment conditions become the orthogonality conditions

$$\mathrm{E}(z_t e_t) = 0\,. \tag{3.3}$$

In the case of a just identified system ($q = p$), the instrumental variable regression estimator $\{\Sigma_{t=1}^{T} z_t x'_t\}^{-1}\{\Sigma_{t=1}^{T} z_t y_t\}$ coincides with the GMM estimator. For the case of an overidentifying system ($q > p$), that Sargan's (1958) generalized instrumental variables estimators, the two-stage least-squares estimators, and the three-stage least-squares estimators (for multiple regressions) can be interpreted as optimal GMM estimators when e_t is serially uncorrelated and conditionally homoskedastic.[8]

3.3. *Nonlinear instrumental variables estimation*

GMM is often used in the context of nonlinear instrumental variable estimation (NLIV). Section 8 presents some examples of applications based on the 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, $\mathrm{E}[e_t \mid I_t] = 0$, where $\mathrm{E}[\cdot \mid I_t]$ 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 $q \times k$ matrix of random variables that are in the information set I_t.[9] Then by the law of iterative expectations, we obtain unconditional moment restrictions:

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

Thus we let $X_t = (x'_t, z'_t)'$ and $f(X_t, \beta) = z_t g(x_t, \beta)$ in this case. Hansen (1982) points out that the NLIV estimators discussed by Amemiya (1974), Jorgenson and Laffont (1974), and Gallant (1977) can be interpreted as optimal GMM estimators when e_t is serially uncorrelated and conditionally homoskedastic.

4. Extensions

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

[8] This interpretation can be seen by examining the first order condition for the minimization problem (2.3).

[9] 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.

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 $\beta = (\beta^1, \beta^{2\prime})'$, β^i is a p_i-dimensional vector of parameters, and f_i is a q_i-dimensional vector of functions. Although it is possible to estimate β^1 and β^2 simultaneously, it may be computationally convenient to estimate β^1 from $f_1(X_t, \beta^1)$ first, and then estimate β^2 from $f_2(X_t, \beta^1, \beta^2)$ in a second step (see, e.g., Barro, 1977, and Atkeson and Ogaki, 1991, for examples of empirical applications). In general, the asymptotic distribution of the estimator of β^2 is affected by estimation of β^1 (see, e.g., Newey, 1984, and Pagan, 1984, 1986). A GMM computer program for a sequential estimation can be used to calculate the correct standard errors that take into account of these effects from estimating β^1. If there are overidentifying restrictions in the system, an econometrician may wish to choose the second step distance matrix in an efficient way. The choice of the second step distance matrix is analyzed by Hansen, Heaton and Ogaki (1992).

Suppose that the first step estimator β^1_T minimizes

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

and that the second step estimator minimizes

$$J_{2T}(\beta^2) = \left\{\frac{1}{T}\sum_{t=1}^{T} f_2(X_t, \beta^1_T, \beta^2)\right\}' W_{2T}\left\{\frac{1}{T}\sum_{t=1}^{T} f_2(X_t, \beta^1_T, \beta^2)\right\}, \tag{4.3}$$

where W_{iT} is a positive semidefinite matrix that converges to W_{i0} with probability one. Let Γ_{ij} be the $q_i \times p_j$ matrix $\mathrm{E}(\partial f_i / \partial \beta_j')$ 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

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

With this choice of W_{20}, $(1/\sqrt{T})\,\Sigma_{t=1}^{T}(\beta^2_T - \beta^2_0)$ has an (asymptotic) normal distribution with mean zero and the covariance matrix

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

and $TJ_{2T}(\beta^2_T)$ has an (asymptotic) chi-square distribution with $q_2 - p_2$ degrees of freedom. 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 empirical examples). 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

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

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

The GMM procedure has been applied to cross-sectional data and panel data. Empirical examples include the work of Holtz-Eakin, Newey and Rosen (1988), Hotz, Kydland and Sedlacek (1988), Hotz and Miller (1988), Shaw (1989), Altug and Miller (1990, 1991), Engen (1991) and Runkle (1991). These authors also discuss many of the econometrics issues. Avery, Hansen and Hotz (1983) develop a method to apply GMM to probit models in panel data. Chamberlain's (1992) comment on Keane and Runkle (1992) discusses applications of GMM to obtain efficient estimators in panel data. Arellano and Bond (1991) discuss a GMM estimation method in panel data and propose a test of serial correlation. The reader who is interested in econometric issues that are not treated in this subsection is referred to these papers and references therein. This section explains a simple method which allows for both a general form of serial correlation and different serial correlation structures for different groups in panel data (see, e.g., Atkeson and Ogaki, 1991, and Ogaki and Atkeson, 1991, for empirical examples).

Consider a panel data set in which there exist H groups, indexed by $h = 1, \ldots, H$ (for example, H villages). Suppose that group h consists of N_h individuals and that the data set contain T periods of observations. Let $N = \Sigma_{h=1}^{H} N_h$ be the total number of individuals. It is assumed that N is large relative to 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 group 1 and $i = N_1 + 1, \ldots, N_1 + N_2$ are in group 2, etc. It is assumed that $\lim_{N\to\infty} N_h / N = \delta_h$ exists. 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. Let $q = \Sigma_{t=1}^{T} q_t$, $X_i = (x_{i1}', \ldots, x_{iT}')'$ and $f(X_i, \beta) = (f_1(x_{i1})', \ldots, f_T(x_{iT})')'$.

Thus a general form of serial correlation is allowed by stacking disturbance terms with different dates as different disturbance terms rather than treating them as different observations of one disturbance term. It is assumed that X_i is identically and independently distributed over the individuals. Assume that there exist q moment restrictions:

$$\mathrm{E}_N(f(X_i, \beta_0)) = 0\,, \tag{4.8}$$

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

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=M_{h-1}+1}^{M_h} f(X_i, \beta) = \delta_h \mathrm{E}_n(f(X_i, \beta)) \tag{4.9}$$

for each $h = 1, \ldots, H$, where $M_h = N_1 + \cdots + N_h$ and $M_0 = 0$. Let $g_N(X_i, \beta) = (\Sigma_{i=1}^{N_1} f(X_i, \beta)', \ldots, \Sigma_{i=M_{H-1}+1}^{M_H} f(X_i, \beta)')'$. Then the GMM estimator β_N minimizes a quadratic form

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

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 $(1/\sqrt{N})\,\Sigma_{i=M_{h-1}+1}^{M_h} u_i$ 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 group h. Let Ω be a matrix that has Ω_h in the h-th diagonal block for $h = 1, \ldots, H$ and the 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.

In estimating rational expectations models with panel data, it is important to recognize that Euler equation of the form $\mathrm{E}(u_{it} \mid I_t) = 0$ does not imply $(1/N)\,\Sigma_{i=1}^{N} u_{it}$ converges in probability to zero as N is increased. The sample counterpart of $\mathrm{E}(u_{it} \mid I_t)$ converges to zero as the number of time periods increases, but not as the number of households increases (see, e.g., Chamberlain, 1984; Keane and Runkle, 1990; Runkle, 1991, and Hayashi, 1992). This problem is especially severe when idiosyncratic shocks are insured away and only aggregate shocks are present as implied by models with complete markets (see, e.g., Altug and Miller, 1990). One way to deal with this problem is to use a panel data set with relatively large number of time periods and to use asymptotic theories as $T \to \infty$. In this case, issues of nonstationarity must be addressed as in applications to time series data (see Section 5 below).

4.4. *Minimum distance estimation*

Minimum distance estimation (MDE) provides a convenient way of obtaining an efficient estimator that imposes nonlinear restrictions (see, e.g., Chiang, 1956; Ferguson, 1958, and Chamberlain, 1982, 1984) and a test statistic for these restrictions. See Altug and Miller (1991) and Atkeson and Ogaki (1991) for examples of empirical applications. The MDE is closely related to GMM and a GMM program can be used to implement the MDE (see, e.g., Ogaki, 1993a). 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

$$\phi(\beta_0) = \theta_0 , \tag{4.12}$$

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

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

for a positive semidefinite matrix W_T that converges to a positive definite matrix W_0 with probability one. As with 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 (4.12) is rejected when this statistic is larger 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 care is taken.

5.1. *Stationarity*

In Hansen (1982), X_t is assumed to be (strictly) stationary.[10] A time series $\{X_t\colon -\infty < t < \infty\}$ 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 t, τ and s. Among other things,

[10] In the context of cross sectional data discussed in Section 4.3, this assumption corresponds with the assumption that X_i is identically distributed over individuals in cross sectional data.

this implies that when they exist, the unconditional moments $\mathrm{E}(X_t)$ and $\mathrm{E}(X_t X'_{t+\tau})$ cannot depend on t for any τ. Thus this assumption rules out deterministic trends, autoregressive unit roots, and unconditional heteroskedasticity.[11] On the other hand, conditional moments $\mathrm{E}(X_{t+\tau} \mid I_t)$ and $\mathrm{E}(X_{t+\tau} X'_{t+\tau+s} \mid I_t)$ 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 all admissible β, not just for $\beta = \beta_0$ (see Section 8.1.4 for an example in which $f(X_t, \beta)$ 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 Section 4.2, nonstationarity in the form of trend stationarity can be treated with ease. In order to treat 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 identification. Let

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

The identification assumption is that β_0 is the unique minimizer of $J_0(\beta)$.[12] Since $J_0(\beta) \geqslant 0$ and $J_0(\beta_0) = 0$, β_0 is a minimizer. Hence this assumption requires $J_0(\beta)$ to be strictly positive for any other β. This assumption is obviously violated if $f(X_t, \beta) \equiv 0$ for some β which did not have any economic meaning (see Section 8 for examples).

6. Covariance matrix estimation

An estimate of Ω is necessary to calculate asymptotic standard errors for the GMM estimator from (2.6) and to utilize the optimal distance matrix Ω^{-1}.

[11] Gallant (1987) and Gallant and White (1988) show that the GMM strict stationarity assumption can be relaxed to allow for unconditional heteroskedasticity. This does *not* mean that X_t can exhibit nonstationarity by having deterministic trends or autoregressive unit roots. Some of their regularity conditions are violated by these popular forms of nonstationarity. It is still necessary to detrend X_t if it is trend stationary. For this reason, the strict stationarity assumption is emphasized in the context of time series applications rather than the fact that this assumption can be relaxed.

[12] Hansen, Heaton and Ogaki (1992) show that without the identification assumption, a sequence of sets of minimizers for (2.3) converges to the set of minimizers with probability one when all other regularity conditions hold.

This section discusses estimation methods for Ω. In the following, it is assumed 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, it is recommended that the third stage GMM be used because the gains from further iterations may be small.

6.1. *Serially uncorrelated disturbance*

This subsection treats the case where $\mathrm{E}(u_t u_{t+\tau}) = 0$ for $\tau \neq 0$.[13] In this case, Ω can be estimated by $(1/T)\,\Sigma_{t=1}^{T} f(X_t, \beta_T)f(X_t, \beta_T)'$. In the models considered in Section 3, this is White's (1980) heteroskedasticity consistent estimator. For example, consider the NLIV model. In this model, $u_t = z_t g(X_t, \beta_0)$ and

$$\frac{1}{T}\sum_{t=1}^{T} f(X_t, \beta_T)f(X_t, \beta_T)' = \frac{1}{T}\sum_{t=1}^{T} z_t g(X_t, \beta_T)g(X_t, \beta_T)'z_t' .$$

Note that u_t is serially uncorrelated if $e_t = g(X_t, \beta_0)$ is in the information set I_{t+1} because

$$E(u_t u_{t+j}') = E(E(u_t u_{t+j}' \mid I_{t+1})) = E(u_t E(u_{t+j}') \mid I_{t+1})) = 0 \quad \text{for } j \geq 1 .$$

In some cases, conditional homoskedasticity is assumed and an econometrician may wish to impose this on his estimate for Ω. Then

$$\frac{1}{T}\sum_{t=1}^{T} z_t \left\{ \frac{1}{T}\sum_{t=1}^{T} g(X_t, \beta_T)g(X_t, \beta_T)' \right\} z_t'$$

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 $\Phi(\tau) =$

[13] In the context of the cross sectional model of Section 4.3, this means that the disturbance is uncorrelated across households, even though it can be serially correlated.

$E(u_t u'_{t-\tau})$,

$$\Phi_T(\tau) = \frac{1}{T} \sum_{t=\tau+1}^{T} f(X_t, \beta_T) f(X_{t-\tau}, \beta_T)' \quad \text{for } \tau \geq 0 , \tag{6.1}$$

and $\Phi_T(\tau) = \Phi_T(-\tau)'$ for $\tau < 0$. Many estimators for Ω in the literature have the form

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

where $k(\cdot)$ is a real-valued kernel, and S_T is a bandwidth parameter. The factor $T/(T-p)$ is a small sample degrees of freedom adjustment.[14] See Andrews (1991) for example of kernels. The estimators of Hansen (1982) and White (1984, p. 152) use the truncated kernel; the Newey and West (1987a) estimator uses the Bartlett kernel; and the estimator of Gallant (1987, p. 533) uses Parzen kernel. The estimators corresponding to these kernels place zero weights on $\Phi(\tau)$ for $\tau \geq S_T$, 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 on any $\Phi(\tau)$ for $|\tau| \leq T-1$.[15]

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. These formulas include unknown parameters and 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 to estimate the law of motion of the disturbance u_t. The second step is to calculate the parameters for the optimal bandwidth parameter from the estimated law of motion. In his Monte Carlo simulations, Andrew uses an AR(1) parameterization for each term of the disturbance. This seems to work well in the models he considers.

Another issue is the choice of the kernel. One serious problem with the truncated kernel is that the corresponding estimator is not guaranteed to be positive semidefinite. Andrews (1991) shows 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 positive semidefinite. His Monte Carlo simulations show that the QS kernel and the Parzen kernel work better than the Bartlett kernel in most of the models he considers. 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

[14] Some other forms of small sample adjustments have been used (see, e.g., Ferson and Foerster, 1991).

[15] Hansen (1992) relaxes an assumption made by these authors to show the consistency of the kernel estimators.

and Monahan (1992) propose an estimator based on a VAR prewhitening. The intuition behind this is that the estimators of the form (6.2) only take 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

$$u_t = A_1 u_{t-1} + A_2 u_{t-2} + \cdots + A_n u_{t-n} + v_t \,. \tag{6.3}$$

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 of 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 applied to v_t for example. Then the sample counterpart of the formula

$$\Omega = \left[I - \sum_{\tau=1}^{n} A_\tau\right]^{-1} \Omega^* \left[I - \sum_{\tau=1}^{n} A'_\tau\right]^{-1} \tag{6.4}$$

is used to form an estimate of Ω. Andrews and Monahan use 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.[16]

In sum, existing Monte Carlo evidence for estimation of Ω recommends VAR prewhitening and either the QS or Parzen kernel estimator together with Andrew's (1991) automatic bandwidth parameter. Though the QS kernel estimator may be preferred to the Parzen kernel estimator 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 of 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 $\mathrm{E}(u_t u'_{t+\tau}) = \mathrm{E}(\mathrm{E}(u_t u'_{t+\tau} \mid I_{t+s+1})) = \mathrm{E}(u_t \mathrm{E}(u'_{t+\tau}) \mid I_{t+s+1})) = 0$ for $\tau \geqslant s+1$. Thus the order of serial correlation of u_t is s and u_t has an MA(s) structure in this case.

In this case, there exist the zero restrictions on the autocovariances that $\Phi(\tau) = 0$ for $|\tau| > s$. Imposing these zero restrictions on the estimator of Ω leads to a more efficient estimator.[17] Since $\Omega = \Sigma_{\tau=-s}^{s} \Phi(\tau)$ in this case, a

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

[17] 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.

natural estimator is

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

which is the truncated kernel estimator. Hansen and Hodrick (1980) study a multi-period forecasting model that leads to $s \geq 1$. They use (6.5) with conditional homoskedasticity imposed (as discussed at the end of Section 6.1). Their method of calculating the standard errors for linear regressions is known as Hansen–Hodrick correction.

A possible problem with the estimator (6.5) is that Ω_T is not guaranteed to be positive semidefinite if $s \geq 1$. 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 (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

$$u_t = B_1 v_{t-1} + \cdots + B_s v_{t-s} + e_t , \tag{6.6}$$

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

$$\Omega = (I + B_1 + \cdots + B_s)\mathrm{E}(e_t e_t')(I + B_1 + \cdots + B_s)' \tag{6.7}$$

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 too many parameters in (6.3) need to be estimated. The number of elements in $u(t)$ needs 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 a simple example, consider the case where $s = 1$ and Newey–West's (1987a) Bartlett kernel estimator is used. Then

$$\Omega_T = \frac{T}{T-p} \sum_{\tau=-l}^{l} \frac{T-|\tau|}{T} \Phi_T(\tau) , \tag{6.8}$$

where $l = S_T - 1$ is the lag truncation number. If $l = 1$ is used to impose the zero restrictions, then Ω_T converges to $\Phi(0) + \frac{1}{2}\Phi(1) + \frac{1}{2}\Phi(-1)$, 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 $l > 1$ is used and the zero restrictions are imposed by setting $\Phi_T(\tau)$ in (6.8) to zero for $|\tau| > 1$, then the resulting estimator is no longer guaranteed to be positive semidefinite.

7. Hypothesis testing and specification tests

This section discusses 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 Eichenbaum, Hansen and Singleton (1988) considered the likelihood ratio type test for GMM (or a 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 an $s \times 1$ vector of functions. The null hypothesis H_0 is tested against the alternative that $R(\beta_0) \neq r$. Let $\Lambda = \partial R / \partial \beta'$ and Λ_T be a 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

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

where Ω_T, Γ_T, and Λ_T are estimated from β_T^r. Note that in linear models LM_T is equal to (7.2), where Ω_T, Γ_T, and Λ_T are estimated from β_T^r rather than β_T^u. The likelihood ratio type test statistic is

$$T(J_T(\beta_T^r) - J_T(\beta_T^u)) , \tag{7.4}$$

which is T times the difference between the minimized value of 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 likelihood ratio type 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 in general, the test result for nonlinear restrictions depends on the parameterization (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 (1989), Ghysels and Hall (1990a,b,c), Hansen (1990), Dufour, Ghysels and Hall (1991), and Andrews (1993). Some of these tests are discussed by Hall (1993).

8. Empirical applications

The GMM estimation has been frequently applied to rational expectations models. This section discusses examples of these applications.[18] The main purpose is not to provide a survey of the literature but to illustrate applications.[19] Problems that researchers have encountered in applying GMM and procedures they have used to address these problems are discussed. In this section, the notations for the NLIV model of Section 3 will be used.

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 to a consumption-based capital asset pricing model (C-CAPM). Consider an economy in which a representative agent maximizes

$$\sum_{t=1}^{\infty} \delta^t \mathrm{E}(U(t) \mid I_0) \tag{8.1}$$

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

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

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 optimi-

[18] Some other empirical examples are mentioned in Section 4.

[19] See Cochrane and Hansen (1992) for a survey on asset pricing puzzles.

zation problem is

$$\frac{\mathrm{E}[\delta C_{t+1}^{-\alpha} R_{t+1} \mid I_t]}{C_t^{-\alpha}} = 1 , \tag{8.3}$$

where R_{t+1} is the (gross) real return of any asset.[20] The observed C_t they use is obviously nonstationary, although the specific form of nonstationarity is not clear (difference stationary or trend stationary, for example). Hansen and Singleton use C_{t+1}/C_t in their econometric formulation, which is assumed to be stationary.[21] 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.[22] 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 rejects their model especially when nominal risk free bond returns and stock returns are used simultaneously.[23] 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 diagnostic that could provide such guidance.[24]

8.1.2. *Time aggregation*

The use of consumption data for the C-CAPM is subject to a time aggregation problem (see, e.g., Hansen and Sargent, 1983a,b) 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.

In linear models, the time aggregation means that the disturbance has an MA(1) structure and the instrumental variables need to be lagged an additional period. See, e.g., Grossman, Melino and Shiller (1987), Hall (1988), and Hansen and Singleton (1988) for applications to C-CAPM and Heaton (1990)

[20] This asset pricing equation can be applied to any asset returns. For example, Mark (1985) applies the Hansen–Singleton model in asset returns in foreign exchange markets.

[21] In the following, assumptions about trend properties of equilibrium consumption are made. The simplest model in which these assumptions are satisfied is a pure exchange economy, with the trend assumptions imposed on endowments.

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

[23] Cochrane (1989) points out that the utility that the representative consumer loses by deviating from the optimal consumption path is very small in the Hansen–Eingleton model and in Hall's (1978) model. In this sense, the Hansen–Singleton test and Hall's test may be too sensitive to economically small deviations caused by small costs of information and transactions.

[24] Garber and King (1983) criticize Hansen and Singleton's methodology by pointing out that their estimators for nonlinear models are not consistent when unknown preference shocks are present. Nason (1991) applies GMM to his linear permanent income model with stochastic preference shocks.

and Christiano, Eichenbaum and Marshall (1991) for applications to Hall (1978) type permanent income models.

It is sometimes not possible for GMM to take into account the effect of time aggregation. For example, Heaton (1991) uses the method of simulated moments (MSM) for his nonlinear asset pricing model with time-nonseparable preferences in taking time aggregation into account. Bossaerts (1989), Duffie and Singleton (1989), MacFadden (1989), Pakes and Pollard (1989), Lee and Ingram (1991), and Pearson (1991), among others, have studied asymptotic properties of MSM.

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

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

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

$$S_t = a_0 C_t + a_1 C_{t-1} + a_2 C_{t-2} + \cdots . \tag{8.5}$$

This type of specification for the 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 (1991, 1993), Cooley and Ogaki (1991), Ferson and Constantinides (1991), Ferson and Harvey (1991) and Ogaki and Park (1993) among others.[25] Depending on the values of the a_τ, the model (8.4) leads to a model with habit formation and/or durability. Constantinides (1990) argues that habit formation could help solve the equity premium puzzle. He shows how the intertemporal elasticity of substitution and the relative risk aversion coefficient depend on the a_τ and α parameters 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 (1993) to illustrate econometric formulations.[26] In their models, it is assumed that $a_\tau = 0$ for $\tau \geq 2$. Let us normalize a_0 to be one, so that $\beta = (\delta, \alpha, a_1)$. The asset pricing equation takes the form

$$\frac{\mathrm{E}[\delta\{S_{t+1}^{-\alpha} + \delta a_1 S_{t+2}^{-\alpha}\}R_{t+1} \mid I_t]}{\mathrm{E}[S_t^{-\alpha} + \delta a_1 S_{t+1}^{-\alpha} \mid I_t]} = 1 . \tag{8.6}$$

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

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

[26] Eichenbaum, Hansen and Singleton (1988) and Eichenbaum and Hansen (1990) consider similar models with nonseparable preferences across goods.

(8.6) implies that $\mathrm{E}(e_t^0 \mid I_t) = 0$, 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 the nonstationarity of C_t and by the 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 α positive. Ferson and Constantinides (1991) solve both of these problems by defining $e_t = e_t^0 / \{S_t^{-\alpha}(1 + \delta a_1)\}$. Since $S_t^{-\alpha}$ is in I_t, $\mathrm{E}(e_t \mid I_t) = 0$. 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 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 $C_{t+1} + a_1 C_t$ may be negative when a_1 is close to minus one. In a nonlinear search for β_T or in calculating numerical derivatives, a GMM computer program will stall if it tries a value of a_1 that makes $C_{t+1} + a_1 C_t$ negative for any t. 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 nonadmissible parameter values are used. However, it is necessary to ignore these large values of $f(X_t, \beta)$ when calculating numerical derivatives. This can be done by suitably modifying programs that calculate numerical derivatives.[27]

8.1.4. *Multiple-good models*

Mankiw, Rotemberg and Summers (1985), Dunn and Singleton (1986), Eichenbaum, Hansen and Singleton (1988), Eichenbaum and Hansen (1990) and Osano and Inoue (1991), among others, have estimated versions of multiple-good C-CAPM. Basic economic formulations in these multiple-good models will be illustrated in the context of a simple model with one durable good and one nondurable good.

Let us replace (8.2) by Houthakker's (1960) addilog utility function that Miron (1986), Ogaki (1988, 1989), and Osano and Inoue (1991) among others have estimated:

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

where C_t is nondurable consumption and K_t is household capital stock from purchases of durable consumption good D_t.[28] The stock of durables is assumed to depreciate at a constant rate $1 - a$, where $0 \leq a < 1$:

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

[27] Ogaki (1993a) explains these modifications for Hansen/Heaton/Ogaki GMM package.

[28] 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.

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 practice, the data for K_t is constructed from data for an initial stock K_0, and for D_t for $t = 1, \ldots, T$. Let P_t be the intratemporal relative price of durable and nondurable consumption. Then the intraperiod first order condition that equates the relative price with the marginal rate of substitution is

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

Assume that D_{t+1}/D_t is stationary. Then $K_{t+\tau}/D_t$ is stationary for any τ because $K_{t+\tau}/D_t = \Sigma_{\tau=0}^{\infty} a^\tau D_{t+\tau}/D_t$. From (8.9),

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

Assume that the variables in I_t are stationary.[29] Then (8.10) implies that the $P_t C_t^{-\alpha}/D_t^{-\eta}$ 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

$$e_t^0 = P_t C_t^{-\alpha} - (1 - \delta a F)^{-1} \theta K_t^{-\eta}, \tag{8.11}$$

where F is the forward operator. The first order condition (8.9) implies that $\mathrm{E}(e_t^0 \mid I_t) = 0$. 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 $e_t = (1 - \delta a F) e_t^0$. Note that e_t involves only C_t, C_{t+1}, and K_t and that $\mathrm{E}[e_t \mid I_t] = 0$. 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, rather it is also 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).[30] Ogaki (1988, 1989) assumes that C_t and D_t are trend stationary and applies the method of Section 4 above 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 imposed on the GMM estimators. Imposing the stationarity restrictions also leads to more reasonable point estimates for α and η.

Eichenbaum, Hansen and Singleton (1988) and Eichenbaum and Hansen

[29] If I_t includes 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.

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

(1990) use the Cobb–Douglas utility function, so that α and η are known to be one.[31] 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 transformation does not involve any unknown parameters. 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 as defined by Engle and Granger (1987). Ogaki (1988) and Ogaki and Park (1993) propose to estimate the curvature parameters α and η of the addilog utility function, using a cointegrating regression.[32] Cooley and Ogaki (1991) combine this cointegration approach with the 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.[33] This two-step procedure does not alter the asymptotic distributions of GMM 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) propose a specification test like Hausman's (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 power against the factors that make the two estimates different, such as nonseparability in preferences across goods, measurement errors, and liquidity constraints.

8.1.6. Seasonality

Miron (1986) augments Hansen and Singleton's (1982) model by including deterministic 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.[34] Although this is theoretically possible, English, Miron and Wilcox (1989) find that seasonally unadjusted quarterly

[31] Also see Ogaki (1992) for a discussion of the stationarity restriction implied by the Cobb–Douglas utility function.

[32] Ogaki and Park (1993) use Park's (1992) canonical cointegrating regressions and Park and Ogaki's (1991a) seemingly unrelated canonical regressions (also see Ogaki, 1993b,c).

[33] In the applications of Cooley and Ogaki (1991) and Ogaki and Park (1993), time-non-separability in preferences is allowed for nondurable consumption and the asset pricing equation (8.6) is used to estimate δ and a_1. In their applications, the first and second goods are assumed to be separable in preferences over time. See Ogaki (1992) for an application of the cointegration approach without this separability assumption.

[34] It should be noted that a deterministic seasonal dummy can be viewed as an artificial stationary and ergodic stochastic process (see, e.g., Ogaki, 1988, pp. 26–27). Hence GMM can be applied to models with deterministic seasonal taste shifts.

data reject asset pricing equations at least as strongly as seasonally adjusted data.[35] 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.

Singleton (1988) argues that the inclusion of taste shifters in C-CAPM is essentially equivalent to studying directly consumption data with deterministic seasonality removed. This is because we do not obtain much identifying information from seasonal fluctuations about preferences if most of seasonal fluctuations come from seasonal taste shifts.[36] On the other hand, seasonal fluctuations may contain useful identifying information about production functions if production functions are relatively stable over the seasonal cycle. Braun and Evans (1991b) utilize such identifying information.

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 significant. Heaton (1993) also finds evidence for seasonal habit formation in Hall (1978) type permanent income models.[37]

8.1.7. State-nonseparable preferences

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

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

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

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

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_{t+1}^e is the (gross) return of the optimal portfolio (R_{t+1}^e 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_{t+1}^e. Thus Roll's (1977) critique of CAPM is relevant here as Epstein and Zin discuss.

Even though (8.13) holds for $R_{t+1} = R_{t+1}^e$, 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_{t+1}^e are used simultaneously, then the whole system can satisfy the identification assumption but the GMM estimators for this partially

[35] Hoffman and Pagan (1989) obtain similar results.

[36] Beaulieu and Miron (1991) cast doubt on the view that negative output growth in the first quarter (see, e.g., Barskey and Miron, 1989) is caused by negative technology seasonal by observing negative output growth in the southern hemisphere.

[37] See Ghysels (1990, especially Section I.3) for a survey of the economic and econometric issues of seasonality.

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 includes 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. Monetary models

In some applications, monetary models are estimated by applying GMM to Euler equations and/or intratemporal first order conditions. Singleton (1985), Ogaki (1988), Finn, Hoffman and Schlagenhauf (1990), Bohn (1991) and Sill (1992) estimate cash-in-advance models, Poterba and Rotemberg (1987), Finn, Hoffman and Schlagenhauf (1990), Imrohoroglu (1991), and Eckstein and Leiderman (1992) estimate money-in-the-utility-function (MIUF) models, and Marshall (1992) estimates a transactions–cost monetary model.

It turns out that cash-in-advance models involve only minor variations on the asset pricing equation (8.3) 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 than in the Hansen–Singleton model, and u_t has an MA(1) structure (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 does not find monetary distortions in the U.S. data he examines.

8.3. Linear rational expectations models

There are two alternative methods to apply GMM to linear rational expectations models. The first method applies GMM directly to linear Euler equations implied by the optimization problems of economic agents. There are many empirical applications of this method, including those of Pindyck and Rotemberg (1983), Fair (1989), and Eichenbaum (1990). In linear models, two-stage least squares estimators and three-stage least squares estimators can be considered as special cases of GMM estimators when the disturbances of the regressions are serially uncorrelated and conditionally homoskedastic.

The second method is developed by Hansen and Sargent (1982), which applies GMM to Hansen and Sargent's (1980, 1981a) linear rational expectations models. This method imposes nonlinear restrictions implied by Wiener–Kolmogorov prediction formulas (see, e.g., Hansen and Sargent, 1981b) on a

VAR representation. Compared with the first method, this second method requires more assumptions about the stochastic law of motion of economic variables but utilizes more restrictions and is (asymptotically) more efficient when these assumptions are valid. West (1989) extends Hansen and Sargent's (1980) formulas to include deterministic terms, and Hansen and Sargent (1991, Chapters 7, 8, 9) provide the continuous time counterparts of these formulas. Maximum likelihood estimation has been used more frequently for the Hansen–Sargent type linear rational expectations model than GMM (see, e.g., Sargent, 1978, 1981a,b; Eichenbaum, 1984; Finn, 1989; and Giovannini and Rotemberg, 1989) even though Hansen and Sargent's (1982) method can be applied to these models. West (1987, 1988a) does not impose the nonlinear restrictions on his estimates but tests these restrictions. West (1987, 1988a) uses West's (1988b) results when difference stationary variables are involved.[38]

8.4. *Calculating standard errors for estimates of standard deviation, correlation and autocorrelation*

In many macroeconomic applications, researchers report estimates of standard deviations, correlations, and autocorrelations of economic variables. It is possible to use a GMM program to calculate standard errors for these estimates, in which the serial correlation of the economic variables is taken into account (see, e.g., Backus, Gregory and Zin, 1989, and Backus and Kehoe, 1992).

For example, let x_t and y_t be economic variables of interest which are assumed to be stationary and let $X_t = (x_t, y_t)$ and $f(X_t, \beta) = (x_t, x_t^2, y_t, y_t^2, x_t y_t, x_t x_{t-1})' - \beta$. Then the parameters to be estimated are the population moments: $\beta_0 = (E(x_t),\ E(x_t^2),\ E(y_t),\ E(y_t^2),\ E(x_t y_t),\ E(x_t x_{t-1}))$. Applying GMM to $f(X_t, \beta)$, one can obtain an estimate of β_0, β_T, and an estimate of the covariance matrix of $T^{1/2}(\beta_T - \beta_0)$, Σ_T. In most applications, the order of serial correlation of $(x_t, x_t^2, y_t, y_t^2, x_t y_t)'$ is unknown, and its long-run covariance matrix, Ω, can be estimated by any of the methods of Section 6 (such as Andrews and Monahan's prewhitened QS kernel estimation method).

Standard deviations, correlations, and autocorrelations are nonlinear functions of β_0. Hence one can use the delta method to calculate the standard errors of estimates of these statistics. Let $g(\beta_0)$ be the statistic of interest. For example, $g(\beta_0) = (\beta_{02} - \beta_{01}^2)^{1/2}$ for the standard deviation of x_t. Then $g(\beta_T)$ is a consistent estimator of $g(\beta_0)$. By the mean value theorem, $g(\beta_T) = g(\beta_0) + \Lambda_T(\beta_T - \beta_0)$, where Λ_T is the derivative of $g(\cdot)$ evaluated at an intermediate point between β_T and β_0. Since Λ_T converges in probability to $\Lambda_0 = \partial g(\beta_0)/\partial\beta'$, $(g(\beta_T) - g(\beta_0))$ has an approximate normal distribution with the variance $(1/T)\Lambda_T \Sigma_T \Lambda_T'$ in large samples.

[38] It should be noted that West (1986b) 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).

8.5. *Other empirical applications*

Though I have focused on consumption-based pricing models that relate asset returns to the intertemporal decisions of consumers, GMM can be applied to production-based asset pricing models that relate asset returns to the intertemporal investment decisions (see, e.g., Braun, 1991, and Cochrane, 1991, 1992).

Singleton (1988) discusses the use of GMM in estimating real business cycle models. Christiano and Eichenbaum (1992) develop a method to estimate real business cycle models, using GMM. They apply their method to U.S. data. Braun (1990), Burnside, Eichenbaum and Rebelo (1993), Braun and Evans (1991a,b) have estimated real business cycle models, among others.

There has not been much work to apply GMM to models of asymmetric information. An exception is an application of GMM to a model of moral hazard by Margiotta and Miller (1991).

9. Further issues

9.1. *Optimal choice of instrumental variables*

In the NLIV model discussed in Section 3, there are infinitely many possible instrumental variables because any variable in I_t can be used as an instrument. Hansen (1985) characterizes an efficiency bound (that is, a greatest lower bound) for the asymptotic covariance matrices of the alternative GMM estimators and optimal instruments that attain the bound. Since it can be time consuming to obtain optimal instruments, an econometrician may wish to compute an estimate of the efficiency bound to assess efficiency losses from using ad hoc instruments. Hansen (1985) also provides a method for calculating this bound for models with conditionally homoskedastic disturbance terms with an invertible MA representation.[39] Hansen, Heaton and Ogaki (1988) extend this method to models with conditionally heteroskedastic disturbances and models with an MA representation that is not invertible.[40] 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, 1992), Newey (1988), and Hansen (1988) among others have studied the relationship between GMM estimators and

[39] Hayashi and Sims' (1983) estimator is applicable to this example.

[40] 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.

efficient semi-parametric estimators in this environment. Chamberlain (1992) links optimal GMM estimators in panel data to recent semi-parametric work, such as that of Robinson (1987, 1991) and Newey (1990). Technically, Hansen (1988) 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 a model that Newey (1988) studies:[41]

$$y_t = x'_t\beta_0 + e_t\,, \tag{9.1}$$

where the disturbance e_t is a scalar iid random variable with unknown symmetric density $\phi(e_t)$, and x_t is p-dimensional vector of nonstochastic exogenous variables. The MLE of β, β_T, would maximize the log likelihood

$$L = \sum \log \phi(y_t - x'_t\beta)\,, \tag{9.2}$$

and would solve

$$\sum d_t(\beta_T) = 0 \tag{9.3}$$

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 nonparametric 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: $\mathrm{E}(x_te_t) = 0$, $\mathrm{E}(x_te_t^3) = 0$, 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 $\xi(e_t)$. With a sufficiently high order polynomial, the unknown score is well approximated and the GMM estimator is also efficient.

9.3. Small sample properties

Unfortunately, there has 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 (8.1) and (8.2) above) in his simulations for a C-CAPM that is similar to Tauchen's model. While these parameter values do not violate any

[41] The material that follows in this subsection was suggested by Adrian Pagan.

theoretical restrictions for existence of an equilibrium, they 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 (1986) investigates small sample properties of Hansen's (1985) optimal instrumental variables GMM estimators. He finds that the optimal estimators do not perform well in small samples as compared to GMM estimators with ad hoc instruments. Tauchen (1986) and Kocherlakota (1990) recommend small number of instruments rather than large number of instruments when ad hoc instruments are used.

Nelson and Startz (1990) perform Monte Carlo simulations to investigate the 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 seem counterintuitive because one might expect that the consequence of having a poor instrument would be 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 conflicting with those of Tauchen (1986) may be related to this problem of poor instruments (see Canova, Finn and Pagan, 1991, for a related discussion).

Arellano and Bond (1991) report Monte Carlo results on GMM estimators for dynamic panel data models. They report that the GMM estimators have substantially smaller variances than commonly used Anderson and Hsiao's (1981) estimators in their Monte Carlo experiments. They also report that the small sample distributions of the serial-correlation tests they study are well approximated by their asymptotic distributions.

10. Concluding remarks

Many researchers have used GMM to estimate nonlinear rational expectations models with aggregate time series data. There are many other possible applications of GMM and GMM programs. We will probably see more applications of GMM to panel data and to models with asymmetric information in the near future. Other uses of a GMM program discussed here include the implementation of minimum distance estimation, the calculation of standard errors that take into account serially correlated disturbances, and generating inferences that take into account the effects of the first estimation step in sequential (or two-step) estimation. Pagan and Pak (1993) present a method for using a GMM program to calculate tests for heteroskedasticity.

Acknowledgment

This is a revised version of 'An Introduction to the Generalized Method of Moments'. I thank R. Anton Braun, Mario Crucini, Mark Dwyer, Eric Ghysels, Lars Peter Hansen, Esfandiar Maasoumi, and Adrian Pagan for their suggestions and/or clarifications, and Changyong Rhee for a conversation that motivated this work. All remaining shortcomings are mine.

References

Amemiya, T. (1974). The nonlinear two-stage least-squares estimator. *J. Econometrics* **2**, 105–110.

Altug, S. and R. A. Miller (1990). Household choices in equilibrium. *Econometrica* **58**, 543–570.

Altug, S. and R. A. Miller (1991). Human capital, aggregate shocks and panel data estimation. Economic Research Center NORC Working Paper No. 91-1.

Anderson, T. W. and C. Hsiao (1981). Estimation of dynamic models with error components. *J. Amer. Statist. Assoc.* **76**, 598–606.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–858.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, to appear.

Andrews, D. W. K. and R. Fair (1988). Inference in econometric models with structural change. *Rev. Econom. Stud.* **55**, 615–640.

Andrews, D. W. K. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* **60**, 953–966.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econom. Stud.* **58**, 277–297.

Atkeson, A. and M. Ogaki (1991). Wealth-varying intertemporal elasticity of substitution: Evidence from panel and aggregate data. RCER Working Paper No. 303, University of Rochester.

Avery, R., L. Hansen and V. Hotz (1983). Multiperiod probit models and orthogonality condition estimation. *Internat. Econom. Rev.* **24**, 21–36.

Backus, D. K., A. W. Gregory and S. E. Zin (1989). Risk premiums in the term structure: Evidence from artificial economies. *J. Monetary Econom.* **24**, 371–399.

Backus, D. K. and P. J. Kehoe (1992). International evidence on the historical properties of business cycles. *Amer. Econom. Rev.* **82**, 864–888.

Barro, R. J. (1977). Unanticipated money growth and unemployment in the United States. *Amer. Econom. Rev.* **67**, 101–115.

Barksey, R. B. and J. A. Miron (1989). The seasonal cycle and the business cycle. *J. Politic. Econom.* **97**, 503–534.

Beaulieu, J. and J. A. Miron (1991). A cross country comparison of seasonal cycles and business cycles. *Econom. J.* **102**, 772–788.

Bohn, H. (1991). On cash-in advance models of money demand and asset pricing. *J. Money Credit Banking* **23**, 224–242.

Bossaerts, P. (1989). The asymptotic normality of method of simulated moments estimators of option pricing models. Manuscript.

Braun, P. A. (1991). Asset pricing and capital investment: Theory and evidence. Manuscript, Northwestern University.
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 (1991a). Seasonality in equilibrium business cycle theories. Institute for Empirical Macroeconomics Discussion Paper No. 45.
Braun, R. A. and C. L. Evans (1991b). Seasonal Solow residuals and Christmas: A case for labor hoarding and increasing returns. Manuscript.
Burnside, C., M. Eichenbaum and S. Rebelo (1993). Labor hoarding and the business cycle. *J. Politic. Econom.*, to appear.
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. *J. Econometrics* **18**, 5–46.
Chamberlain, G. (1984). Panel data. In: Z. Griliches and M. D. Intriligator, eds., *Handbook of Econometrics*, Vol. 2. North-Holland, Amsterdam,
Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *J. Econometrics* **34**, 305–334.
Chamberlain, G. (1992). Comment: Sequential moment restrictions in panel data. *J. Business Econom. Statist.* **10**, 20–26.
Chiang, C. L. (1956). On regular best asymptotically normal estimates. *Ann. Math. Statist.* **27**, 336–351.
Christiano, L. J. and M. Eichenbaum (1992). Current real business cycle theories and aggregate labor market fluctuations. *Amer. Econom. Rev.* **82**, 430–450.
Christiano, L. J., M. Eichenbaum and D. Marshall (1991). The permanent income hypothesis revisited. *Econometrica* **59**, 397–423.
Cochrane, J. H. (1989). The sensitivity of tests of the intertemporal allocation of consumption to near-rational alternatives. *Amer. Econom. Rev.* **79**, 319–337.
Cochrane, J. H. (1991). Production-based asset pricing and the ling between stock returns and economic fluctuations. *J. Finance* **46**, 207–234.
Cochrane, J. H. (1992). A cross-sectional test of a production-based asset pricing model. Manuscript, University of Chicago.
Cochrane, J. H. and L. P. Hansen (1992). Asset pricing explorations for macroeconomics. In: V. J. Blanchard and S. Fischer, eds., *NBER Macroeconomics Ann. 1992.* MIT Press, Cambridge, MA, 115–165.
Constantinides, G. (1990). Habit formation: A resolution of the equity premium puzzle. *J. Politic. Econom.* **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 approach. RCER Working Paper No. 285, 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.
Dufour, J.-M., E. Ghysels and A. Hall (1991). Generalized predictive tests and structural change analysis in econometrics. Manuscript, Université de Montréal.
Dunn, K. B. and K. J. Singleton (1986). Modeling the term structure of interest rates under non-separable utility and durability of goods. *J. Financ. Econom.* **17**, 27–55.
Eckstein, Z. and L. Leiderman (1992). Seigniorage and the welfare cost of inflation: Evidence from an intertemporal model of money and consumption. *J. Monetary Econom.* **29**, 389–410.
Eichenbaum, M. (1984). Rational expectations and the smoothing properties of inventories of finished goods. *J. Monetary Econom.* **14**, 71–96.
Eichenbaum, M. (1990). Some empirical evidence on the production level and production cost smoothing models of inventory investment. *Amer. Econom. Rev.* **79**, 853–864.
Eichenbaum, M. and L. P. Hansen (1990). Estimating models with intertemporal substitution using aggregate time series data. *J. Business Econom. Statist.* **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. *Quart. J. Econom.* **103**, 51–78.
Engen, E. (1991). Stochastic life cycle model with mortality risk: Estimation with panel data. Manuscript, University of California, Los Angeles, CA.
Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction, representation, estimation, and testing. *Econometrica* **55**, 251–276.
English, W. B., J. A. Miron and D. W. Wilcox (1989). Seasonal fluctuations and the life cycle-permanent income model of consumption: A correction. *J. Politic. Econom.* **97**, 988–991.
Epstein, L. and S. Zin (1991). Substitution, risk aversion and the temporal behavior of asset returns. *J. Politic. Econom.* **99**, 263–286.
Fair, R. C. (1989). The production-smoothing model is alive and well. *J. Monetary Econom.* **24**, 353–370.
Ferguson, T. S. (1958). A method of generating best asymptotically normal estimates with application to the estimation of bacterial densities. *Ann. Math. Statist.* **29**, 336–351.
Ferson, W. E. and G. M. Constantinides (1991). Habit persistence and durability in aggregate consumption. *J. Financ. Econom.* **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. *J. Finance* **417**, 511–552.
Finn, M. G. (1989). An econometric analysis of the intertemporal general equilibrium approach to exchange rate and current account determination. *J. Internat. Money 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. *J. Monetary Econom.* **25**, 431–451.
Gallant, A. R. (1977). Three-stage least-squares estimation for a system of simultaneous, nonlinear implicit equations. *J. Econometrics* **5**, 71–88.
Gallant, A. R. (1987). *Nonlinear Statistical Models*. Wiley, New York.
Gallant, A. R. and H. White (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell, New York.
Garber, P. M. and R. G. King (1983). Deep structural excavation? A critique of Euler equation methods. National Bureau of Economic Research Technical Paper No. 31, Cambridge, MA.
Ghysels, E. (1990). On the economics and econometrics of seasonality. Manuscript, University of Montreal.
Ghysels, E. and A. Hall (1990a). Are consumption-based intertemporal capital asset pricing models structural? *J. Econometrics* **45**, 121–139.
Ghysels, E. and A. Hall (1990b). A test for structural stability of Euler conditions parameters estimated via the generalized method of moments estimator. *Internat. Econom. Rev.* **31**, 355–364.
Ghysels, E. and A. Hall (1990c). Testing non-nested Euler conditions with quadrature-based methods of approximation. *J. Econometrics* **46**, 273–308.
Giovannini, A. and J. J. Rotemberg (1989). Exchange-rate dynamics with sticky prices: The Deutsche Mark 1974–1982, *J. Business Econom. Statist.* **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. *J. Business Econom. Statist.* **5**, 315–327.
Hall, A. (1993). Some aspects of generalized method of moments estimation. In: G. S. Maddala, C. R. Rao and H. D. Vinod, eds., *Handbook of Statistics*, Vol. 11. North-Holland, Amsterdam, Chapter 15.
Hall, R. (1978). Stochastic implications of the life cycle-permanent income hypothesis. *J. Politic. Econom.* **86**, 971–987.
Hall, R. (1988). Intertemporal substitution in consumption. *J. Politic. Econom.* **96**, 339–357.

Hansen, B. (1990). Lagrange multiplier tests for parameter instability in non-linear models. Manuscript, University of Rochester.

Hansen, B. (1992). Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica* **60**, 967–972.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029–1054.

Hansen, L. P. (1985). A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *J. Econometrics* **30**, 203–238.

Hansen, L. P. (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. *J. Amer. Statist. Assoc.* **83**, 863–871.

Hansen, L. P., J. Heaton and M. Ogaki (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. *J. Politic. Economy*. **88**, 829–853.

Hansen, L. P. and R. J. Hodrick (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*. Univ. of Chicago Press, Chicago, IL.

Hansen, L. P. and R. Jagannathan (1991). Implications of security market data for models of dynamic economies. *J. Politic. Economy* **99**, 225–262.

Hansen, L. P. and T. Sargent (1980). Formulating and estimating dynamic linear rational expectations models. *J. Econom. Dynamics Control* **2**, 7–46.

Hansen, L. P. and T. Sargent (1981a). Linear rational expectations models for dynamically interrelated variables. In: R. E. Lucas Jr and T. J. Sargent, eds., *Rational Expectations and Econometric Practice*. Univ. Minnesota Press, Minneapolis, MN, 127–156.

Hansen, L. P. and T. Sargent (1981b). A note on Wiener–Kolmogorov prediction formulas for rational expectations models. *Econom. Lett.* **8**, 255–260.

Hansen, L. P. and T. Sargent (1982). Instrumental variables procedures for estimating linear rational expectations models. *J. Monetary Econom.* **9**, 263–296.

Hansen, L. P. and T. Sargent (1983a). Aggregation over time and the inverse optimal predictor problem for adaptive expectations in continuous time. *Internat. Econom. Rev.* **24**, 1–20.

Hansen, L. P. and T. Sargent (1983b). The dimensionality of the aliasing problem in models with rational spectral densities. *Econometrica* **51**, 377–387.

Hansen, L. P. and T. Sargent (1991). *Rational Expectations Econometrics*. Westview Press, San Francisco, CA.

Hansen, L. P. and K. J. Singleton (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* **50**, 1269–1286.

Hansen, L. P. and K. J. Singleton (1984). Errata. *Econometrica* **52**, 267–268.

Hansen, L. P. and K. J. Singleton (1988). Efficient estimation of linear asset pricing models with moving average errors. Manuscript.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica* **46**, 1251–1271.

Hayashi, F. (1985). The permanent income hypothesis and consumption durability: Analysis based on Japanese panel data. *Quart. J. Econom.* **100**, 1083–1113.

Hayashi, F. (1992). Comment. *J. Business Econom. Statist.* **10**, 15–19.

Hayashi, F. and C. A. Sims (1983). Nearly efficient estimation of time series models with predetermined, but not exogenous instruments. *Econometrica* **51**, 783–798.

Heaton, J. H. (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, Cambridge, MA.

Heaton, J. H. (1993). The interaction between time-nonseparable preferences and time aggregation. *Econometrica* **61**, 353–385.

Heaton, J. H. and M. Ogaki (1991). Efficiency bound calculations for a time series model with conditional heteroskedasticity. *Econom. Lett.* **35**, 167–171.

Hoffman, D. and A. Pagan (1989). Post-sample prediction tests for generalized method of moments. *Oxford Bull. Econom. Statist.* **51**, 333–344.
Holtz-Eakin, D., W. Newey and H. Rosen (1988). Estimating vector autoregressions with panel data. *Econometrica* **56**, 1371–1395.
Hotz, V. J., F. E. Kydland and G. L. Sedlacek (1988). Intertemporal preferences and labor supply. *Econometrica* **56**, 335–360.
Hotz, V. J. and R. A. Miller (1988). An emporical analysis of life cycle fertility and female labor supply. *Econometrica* **56**, 91–118.
Houthakker, H. S. (1960). Additive preferences. *Econometrica* **28**, 244–257.
Imrohoroglu, S. (1991). An empricial investigation of currency substitution. Manuscript, University of Southern California.
Jorgenson, D. W. and J. Laffont (1974). Efficient estimation of nonlinear simultaneous equations with additive disturbances. *Ann. Econom. Social Measurement* **3**, 615–640.
Keane, M. P. and D. E. Runkle (1990). Testing the rationality of price forecasts: New evidence from panel data. *Amer. Econom. Rev.* **80**, 714–735.
Keane, M. P. and D. E. Runkle (1992). On the estimation of panel-data models with serial correlation when instruments are not strictly exogenous. *J. Business Econom. Statist.* **10**, 1–9.
Kocherlakota, N. (1990). On tests of representative consumer asset pricing models. *J. Monetary Econom.* **26**, 285–304.
Lee, B. S. and B. F. Ingram (1991). Simulation estimation of time-series models. *J. 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. *J. Monetary Econom.* **10**, 417–425.
Mankiw, N. G., J. Rotemberg and L. Summers (1985). Intertemporal substitution in macroeconomics. *Quart. J. Econ.* **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.
Margiotta, M. M. and R. A. Miller (1991). Managerial compensation and the cost of moral hazard. Manuscript.
Mark, N. C. (1985). On time varying risk premia in the foreign exchange market: An econometric analysis. *J. Monetary Econom.* **16**, 3–18.
Marshall, D. (1992). Inflation and asset returns in a monetary economy. *J. Finance* **47**, 1315–1342.
Mehra, R. and E. C. Prescott (1985). The equity premium: A puzzle. *J. Monetary Econom.* **15**, 145–161.
Miron, J. A. (1986). Seasonal fluctuations and the life cycle-permanent income model of consumption. *J. Politic. Econom.* **94**, 1258–1279.
Nason, J. M. (1991). The permanent income hypothesis when the bliss point is stochastic. Manuscript, University of British Columbia.
Nelson, C. and R. Startz (1990). The distribution of the instrumental variables estimator and its *t*-ratio when the instrument is a poor one. *J. Business* **63**, S125–S140.
Newey, W. K. (1984). A method of moments interpretation of sequential estimators. *Econom. Lett.* **14**, 201–206.
Newey, W. K. (1985). Generalized method of moments specification testing. *J. Econometrics* **29**, 229–256.
Newey, W. K. (1988). Adaptive estimation of regression models via moment restrictions. *J. Econometrics* **38**, 301–339.
Newey, W. K. (1990). Efficient instrumental variables estimation of nonlinear models. *Econometrica* **58**, 809–837.
Newey, W. K. and K. J. West (1987a). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703–708.

Newey, W. K. and K. J. West (1987b). Hypothesis testing with efficient method of moments estimation. *Internat. Econom. Rev.* **28**, 777–787.

Ogaki, M. (1988). Learning about preferences from time trends. Ph.D. dissertation, University of Chicago.

Ogaki, M. (1989). Information in deterministic trends about preferences. Manuscript, University of Rochester.

Ogaki, M. (1990). Aggregation of intratemporal preferences under complete markets. RCER Working Paper No. 249, University of Rochester.

Ogaki, M. (1992). Engel's law and cointegration. *J. Politic. Econom.* **100**, 1027–1046.

Ogaki, M. (1993a). GMM: A user's guide. RCER Working Paper No. 348, University of Rochester.

Ogaki, M. (1993b). Unit roots in macroeconometrics: A survey. *Monetary Econom. Studies*, to appear.

Ogaki, M. (1993c). CCR: A user's guide. RCER Working Paper No. 349, University of Rochester.

Ogaki, M. and A. Atkeson (1991). Estimating subsistence levels with Euler equations from panel data. Manuscript, University of Rochester and University of Chicago.

Ogaki, M. and J. Y. Park (1993). A cointegration approach to estimating preference parameters. RCER Working Paper No. 209R, University of Rochester.

Osano, H. and T. Inoue (1991). Testing between competitive models of real business cycles. *Internat. Econom. Rev.* **32**, 669–688.

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. *Internat. Econom. Rev.* **25**, 221–247.

Pagan, A. (1986). Two stage and related estimators and their applications. *Rev. Econom. Stud.* **53**, 517–538.

Pagan, A. R. and Y. Pak (1993). Testing for heteroskedasticity. In: G. S. Maddala, C. R. Rao and H. D. Vinod, eds., *Handbook of Statistics*, Vol. 11. North-Holland, Amsterdam, Chapter 18.

Park, J. Y. (1990). Canonical cointegrating regressions. CAE Working Paper No. 88-29R, Cornell University. *Econometrica* **60**, 119–143.

Park, J. Y. and M. Ogaki (1991a). Seemingly unrelated canonical cointegrating regressions. RCER Working Paper No. 280, University of Rochester.

Park, J. Y. and M. Ogaki (1991b). Inference in cointegrated models using VAR prewhitening to estimate shortrun dynamics. RCER Working Paper No. 281, 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, Proc. 2nd Internat. Sympos. in Economic Theory and Econometrics*. Cambridge Univ. Press, Cambridge, 219–240.

Pindick, R. S. and J. J. Rotemberg (1983). Dynamic factor demands and the effects of energy price shocks. *Amer. Econom. Rev.* **73**, 1066–1079.

Robinson, P. (1987). Asymptotic efficient estimation of nonlinear models. *Econometrica* **55**, 875–891.

Robinson, P. (1991). Best nonlinear three-stage least squares estimation of certain econometric models. *Econometrica* **59**, 755–786.

Roll, R. (1977). A critique of the asset pricing theory's tests. *J. Financ. Econom.* **4**, 129–176.

Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica* **26**, 393–415.

Sargent, T. J. (1978). Rational expectations, econometric exogeneity, and consumption. *J. Politic. Economy* **86**, 673–700.

Sargent, T. J. (1981a). The demand for money during hyperinflations under rational expectations. In: R. E. Lucas Jr and T. J. Sargent, eds., *Rotational Expectations and Economic Practice*. Univ. of Minnesota Press, Minneapolis, MN, 429–452.

Sargent, T. J. (1981b). Estimation of dynamic labor demand schedules under rational expectations. In: R. E. Lucas Jr and T. J. Sargent eds., *Rotational Expectations and Econometric Practice*. Univ. of Minnesota Press, Minneapolis, MN, 463–499.

Shaw, K. L. (1989). Life-cycle labor supply with human capital accumulation. *Internat. Econom. Rev.* **30**, 431–456.

Sill, K. (1992). Money and cash-in-advance models: An empirical implementation. Ph.D. dissertation, University of Virginia.

Singleton, K. J. (1985). Testing specifications of economic agents' intertemporal optimum problems in the presence of alternative models. *J. Econometrics* **30**, 391–413.

Singleton, K. J. (1988). Econometric issues in the analysis of equilibrium business cycle models. *J. Monetary Econom.* **21**, 361–386.

Tauchen, G. (1986). Statistical properties of generalized method of moments estimators of structural parameters obtained from financial market data. *J. Business Econom. Statist.* **4**, 397–425.

West, K. D. (1987). A specification test for speculative bubbles. *Quart. J. Econom.* **102**, 553–580.

West, K. D. (1988a). Dividend innovations and stock price volatility. *Econometrica* **56**, 37–61.

West, K. D. (1988b). Asymptotic normality, when regressors have a unit root. *Econometrica* **56**, 1397–1417.

West, K. D. (1989). Estimation of linear rational expectations models, in the presence of deterministic terms. *J. Monetary Econom.* **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*. Academic Press, New York.