Generalized Method of Moments and Macroeconomics

Bruce E. Hansen and Kenneth D. West
Department of Economics, University of Wisconsin, Madison, WI 53706 (*bhansen@ssc.wisc.edu*) (*kdwest@facstaff.wisc.edu*)

We consider the contribution to the analysis of economic time series of the generalized method-of-moments estimator introduced by Hansen. We outline the theoretical contribution, conduct a small-scale literature survey, and discuss some ongoing theoretical research.

KEY WORDS: Bootstrap; Efficiency bounds; HAC estimation; Heteroscedasticity; Serial correlation; Time-series.

1. INTRODUCTION

The three most important developments in time series econometrics in the last 25 years arguably are generalized method-of-moments (GMM) estimation, vector autoregressions (VARs), and the analysis of nonstationary time series (unit roots and cointegration). This article surveys the role of GMM in macroeconomic time series.

The seminal contribution to the literature on GMM was made by Lars Peter Hansen (1982); his work is the focal point of our survey. Hansen's article had important antecedents in the econometrics literature. Two-stage least squares was developed independently by Theil (1953) and Basmann (1957). Basmann (1960) provided an alternative derivation similar to minimum chi-squared estimation (see also Rothenberg 1973). The formulation of two-stage least squares as an optimal instrumental variable (IV) estimator under conditional homoscedasticity and a test for overidentifying restrictions was proposed by Sargan (1958, 1959). These methods were extended to nonlinear models by Amemiya (1974, 1977), Jorgenson and Laffont (1974), Gallant (1977), and Gallant and Jorgenson (1979), Gallant and Jorgenson also proposed a test statistic that ties naturally to Hansen's (1982) test of overidentifying restrictions. In addition, several articles by White (1980, 1982a, 1982b) can be viewed as GMM applied to cross-section linear regression.

The starting point for this article is not this earlier literature, however, but rather Hansen's (1982) contribution. Sections 2–4 exposit the rationale, structure, and impact on applied work of Hansen's article. Section 2 presents a simple rational forecasting example to illustrate why in many applications generalized least squares (GLS) is not an alternative to GMM. Section 3 defines notation and illustrates the use of GMM to estimate a nonlinear time series model (the consumption-based capital asset pricing model), and then outlines a linear dynamic panel model. Section 4 reports a small survey of economics journals, examining the prevalence of GMM and other estimation methods in empirical time series work.

Section 5 turns to subsequent literature that builds on the work of Hansen (1982). This section reviews some current issues of research interest for time series GMM: efficiency bounds, feasible attainment of efficiency bounds, weight matrix estimation, the time series bootstrap, and empirical likelihood methods. Section 6 concludes the article.

2. WHY GENERALIZED METHOD OF MOMENTS?

We use a simple example to motivate use of GMM in time series applications. Suppose that we wish to test the "rationality" of a scalar variable x_t as an n period ahead predictor of a variable q_{t+n}; the null is $E_t q_{t+n} = x_t$, where for the moment we leave unspecified the information set used in forming the expectation. The variable x_t might be the expectation of q_{t+n} reported by a survey. Alternatively, x_t might be a market-determined variable posited by economic theory to be the expectation of q_{t+n} (e.g., $q_{t+n}=$ spot rate, $x_t = n$ period-ahead-forward rate). Let u_t denote the expectational error, $u_t = q_{t+n} - E_t q_{t+n} = q_{t+n} - x_t$. (The expectational error u_t is dated t rather than $t+n$ for consistency with the dating of regression residuals in the main part of this article.) Under the null, $Ex_t u_t = 0$, and u_t follows a moving average process of order $n-1$.

The many implications of the hypothesis that x_t is the expectation of q_{t+n} can be tested in various ways. A particularly simple and natural approach is to regress q_{t+n} on x_t, to see whether the coefficient on x_t, call it β, is 1:

$$q_{t+n} = \beta x_t + u_t, \quad \mathrm{H}_0\colon \beta = 1. \tag{1}$$

(To keep the algebra relatively uncluttered, we omit the constant that in practice likely would be included in this regression.) Suppose that $n > 1$. A question is how to deal with the moving average disturbance u_t.

We note that a GLS transformation generally will *not* be appropriate. (In the presence of a serially correlated disturbance, GLS generally requires that x_t be strictly exogenous. This is unlikely to be true in dynamic models and cannot be true when x_t is a predictor for q_{t+n}.) This has been observed before (e.g., Hansen and Hodrick 1980), but may not be generally appreciated. We illustrate the problem with a particularly simple example (although even this "simple" example inevitably involves some tedious algebra). Suppose that $n = 2$. Let v_{1t} and v_{2t} be two iid random variables that are mutually

Journal of Business & Economic Statistics
October 2002, Vol. 20, No. 4
DOI 10.1198/073500102288618603

independent and have finite variance. Suppose that q_t is generated by

$$q_t = \phi q_{t-1} + v_{1t} + v_{2t-1}, \qquad |\phi| < 1, \qquad \phi \neq 0. \quad (2)$$

Let $E_t q_{t+2} \equiv E(q_{t+2} \mid v_{1t}, v_{2t}, v_{1t-1}, v_{2t-1}, \ldots)$; that is, the expectation is formed using observations on both v's and not just the sum $v_{1t} + v_{2t-1}$. Then $x_t \equiv E_t q_{t+2} = \phi^2 q_t + \phi v_{2t}$, $u_t \equiv q_{t+2} - E_t q_{t+2} = v_{1t+2} + \phi v_{1t+1} + v_{2t+1} \sim \text{MA}(1)$. Because $u_t \sim \text{MA}(1)$, it has a Wold representation of, say,

$$u_t = \epsilon_t - \theta\epsilon_{t-1}, \qquad |\theta| < 1, \quad \text{with } \theta \text{ the smaller root to the quadratic } (1+\theta^2)Eu_t u_{t-1} - \theta E u_t^2 = 0. \quad (3)$$

For $t \le 0$, let us define $q_t = 0, x_t = 0$. If we abstract from error in estimation of θ, then the GLS transformation involves applying the transformation $(1-\theta L)^{-1}$ to both sides of (1) and then estimating by least squares. The result of the transformation is

$$\sum_{j=0}^{\infty} \theta^j q_{t+2-j} = \beta\left[\sum_{j=0}^{\infty} \theta^j x_{t-j}\right] + \epsilon_t = \beta \overset{o}{x}_t + \epsilon_t (\text{say}). \quad (4)$$

A standard condition for consistency of the least squares estimator of β in (4) is that $E\overset{o}{x}_t \epsilon_t = 0$. But this condition typically will not hold. Observe that both ϵ_t and $\overset{o}{x}_t$ depend on lagged v_{1t}'s and v_{2t}'s,

$$\begin{aligned}
\epsilon_t = \sum_{j=0}^{\infty} \theta^j u_{t-j} = {} & \text{terms in } v_{1t+2}, v_{1t+1}, \text{and } v_{2t+1} \\
& + \theta[(\theta+\phi)v_{1t} + v_{2t}] \\
& + \theta^2[(\theta+\phi)v_{1t-1} + v_{2t-1}] + \cdots \\
\overset{o}{x}_t = \sum_{j=0}^{\infty} \theta^j x_{t-j} = {} & \sum_{j=0}^{\infty} \theta^j(\phi^2 q_{t-j} + \phi v_{2t-j}) \\
= {} & \phi^2 v_{1t} + \phi v_{2t} + (\phi^3 + \theta\phi^2)v_{1t-1} \\
& + (\phi^2 + \theta\phi)v_{2t-1} + \cdots \qquad (5)
\end{aligned}$$

It is clear from (5) that in general, both $\overset{o}{x}_t$ and ϵ_t depend on v_1's and v_2's dated period t and earlier. Thus $\overset{o}{x}_t$ and ϵ_t will be correlated with one another. An exception to this general result is if v_{2t} is shut down. If $q_t = \phi q_{t-1} + v_{1t}$, then $\theta = -\phi$, and $\epsilon_t = v_{1t+2}$: $\overset{o}{x}_t$ depends on past v's, but ϵ does not, and $E\overset{o}{x}_t \epsilon_t = 0$. In the present context, $v_{2t} \equiv 0$ means that the expectation of future q's is set using only data on past q's; more generally, in tests of rationality the condition for consistency of GLS is that the econometrician sees and uses all data used in setting the expectation. In many contexts, this condition seems implausibly strong. (Ka-fu Wong has suggested to us that the assumption be tested with a Hausman (1978) test comparing GLS and GMM estimates.)

Although GLS is inconsistent, other estimators are not. One could specify a time series process for q_t and x_t, and apply maximum likelihood. Or one could estimate by least squares, because an expectational error (u_t) is uncorrelated with the expectation (x_t). This approach is problematic, because serial correlation in u_t invalidates conventional least squares standard errors—a fatal shortcoming in an application whose purpose is to test the hypothesis that $\beta = 1$. One possible solution is to eliminate the serial correlation by creating a subsample of every nth observation and then apply least squares.

This is patently an unattractive approach. A GMM solution is to estimate by least squares, using every observation and adjusting the covariance matrix of the estimator. The adjustment to the covariance matrix accounts not only for the moving average aspect of the disturbance, but also for heteroscedasticity of u_t conditional on x_t (if any). This was the estimation strategy of the Hansen and Hodrick (1980) study of efficiency of forward exchange rates. These authors used weekly observations on 13-week-ahead-forward rates; in the notation of this section, $n = 13$, q_{t+n} = difference between exchange rate in $t+13$ and exchange rate in t, x_t = difference between 13-week-ahead-forward rate in t and exchange rate in t.

This estimation technique allowed Hansen and Hodrick to estimate:

- efficiently
- under weak assumptions (6)
- with a technique that is computationally convenient.

(These points were made explicitly in Hansen and Hodrick's article.) Indeed, these benefits are what Hansen's subsequent (1982) article provided for a wide range of applications. These include models that, unlike the one presented here, are multiple-equation, nonlinear, and overidentified, with multiple endogenous variables in a single equation. To illustrate such models, we first review the setup of Hansen (1982).

3. GENERALIZED METHOD-OF-MOMENTS ESTIMATION

The first part of this section serves mainly to define notation. The second part illustrates two applications of GMM, one a nonlinear model and the second a panel data application.

Let β_0 denote the $k \times 1$ parameter vector of interest, and let $g_t(\beta)$ denote an $m \times 1$ vector of moments that depends on data through β, with $m \ge k$. The vector of moments is stationary and satisfies the orthogonality condition, $Eg_t(\beta_0) = 0$. We generally consider systems of a equations with additive regression errors, writing the orthogonality condition as

$$EW_t u_t = 0 \quad (7)$$

In (7), W_t is an $m \times a$ matrix of instruments and u_t is an $a \times 1$ vector of regression errors from the a equations in the system. We suppress dependence of u_t on the parameter vector in (7) and, subsequently, when we can do so without confusion.

Let $\widehat{D}$ be an $m \times m$ positive definite weighting matrix, with the "^" emphasizing that $\widehat{D}$ may be sample dependent. Let T be the sample size. Hansen's (1982) GMM estimator chooses $\hat{\beta}$ to minimize

$$\left[T^{-1}\sum_{t=1}^{T} g_t(\beta)\right]' \widehat{D} \left[T^{-1}\sum_{t=1}^{T} g_t(\beta)\right]. \quad (8)$$

Hansen (1982) showed that under general conditions, $\hat{\beta}$ is $\sqrt{T}$ consistent and asymptotically normal.

Let Ω be the long-run covariance of $g_t(\beta_0), \Omega = \sum_{j=-\infty}^{\infty} Eg_t(\beta_0)g_{t-j}(\beta_0)'$. The efficient GMM estimator chooses $\widehat{D}$ so that $\widehat{D}^{-1} \to_p \Omega$. Assume efficient estimation, and accordingly call the weighting matrix $\widehat{\Omega}^{-1}$. Let G_t denote the $m \times k$ matrix of derivatives of the orthogonality condition, evaluated at β_0: $G_t = \partial g_t(\beta_0)/\partial\beta$. Similarly, define $\widehat{G}$ as the sample counterpart evaluated at the sample estimate of β_0, $\widehat{G} = T^1 \sum_{t=1}^{T} \partial g_t(\hat{\beta})/\partial\beta$. The first-order condition satisfied by $\hat{\beta}$ is $\widehat{G}'\widehat{\Omega}^{-1}[T^{-1}\sum_{t=1}^{T} g_t(\hat{\beta})] = 0$. For future reference, it will be helpful to observe that because $g_t = W_t u_t$, this first-order condition may also be written as

$$\left[T^{-1}\sum_{t=1}^{T}\widehat{Z}_t\hat{u}_t\right] = 0, \qquad \widehat{Z}_t = \widehat{G}'\widehat{\Omega}^{-1}W_t, \quad \hat{u}_t = u_t(\hat{\beta}). \quad (9)$$

Thus, if there are more moment conditions than parameters ($m > k$), then GMM proceeds in a fashion familiar from IV estimation. It takes a linear combination of the instruments—a linear combination chosen to minimize the asymptotic variance of the estimator—and ensures zero sample correlation between this linear combination and the residual.

The asymptotic variance of the GMM estimator is $[(EG_t')\Omega^{-1}(EG_t)]^{-1}$. The criterion function (8), evaluated at the estimated parameter vector and suitably normalized by sample size, is asymptotically chi-squared,

$$J \equiv T\left[T^{-1}\sum_{t=1}^{T} g_t(\hat{\beta})\right]'\widehat{D}\left[T^{-1}\sum_{t=1}^{T} g_t(\hat{\beta})\right] \sim_A \chi^2(m-k). \quad (10)$$

We refer to the use of (9) as the "J test." It can also be called a test of overidentifying restrictions. Evidently, it requires that there be more moment conditions (m) than parameters (k)—that is, that the model be overidentified. This test ties naturally to criterion function–based tests of parametric hypotheses (Gallant 1987; Newey and West 1987b).

We now return to illustrating GMM in application. We do so with a widely used model, the consumption-based capital asset pricing model. Let C_t be consumption, $U(C_t)$ be per-period utility from consumption, and b be a subjective discount factor, $0 < b < 1$. A representative consumer maximizes $E_t\sum_{j=0}^{\infty} b^j U(C_{t+j})$ subject to a budget constraint that allows the consumer to invest in any of a securities. Period t investment in the ith security pays off in period $t+1$, with return R_{it+1}. A set of a first-order conditions for the maximization is $U'(C_t) = E_t[bR_{it+1}U'(C_{t+1})], i = 1, \ldots, a$. On rearrangement, these first-order conditions can be written as $1 = E_t[bR_{it+1}U'(C_{t+1})/U'(C_t)], i = 1, \ldots, a$.

Further suppose that utility is isoelastic, $U(C_t) = C_t^{\gamma+1}$, $\gamma < 0$. With this utility function, $U'(C_{t+1})/U'(C_t) = (C_{t+1}/C_t)^\gamma$. Define consumption growth, $x_{t+1} = C_{t+1}/C_t$. Then the set of first-order conditions may be written as

$$E_t u_t = 0_{a\times 1}, \qquad u_t \equiv \begin{bmatrix} bR_{1t+1}x_{t+1}^{\gamma} - 1 \\ \cdots \\ bR_{at+1}x_{t+1}^{\gamma} - 1 \end{bmatrix} \quad (11)$$

The aim is to estimate the 2×1 parameter vector $\beta_0 \equiv (b, \gamma)'$.

A set of orthogonality conditions is obtained by noting that for any vector of variables in the consumer's period t information set, say $W_t, E_t W_t u_t = 0$, and thus by the law of iterated expectations, $EW_t u_t = 0$. Natural candidates for elements of W_t include lags of returns and consumption growth and the product of these two. For a single equation system ($a = 1$, $u_t = bR_{1t+1}x_{t+1}^{\gamma} - 1$), Hansen and Singleton (1982) set the orthogonality condition to $g_t = W_t u_t$, $W_t' = (1 R_{1t}x_t \ldots R_{1t-m+2}x_{t-m+2})$, for various choices of $m \geq 2$. Because the dimension of β_0 is $k = 2$, this system is overidentified if $m > 2$. Evidently this model is nonlinear, with multiple (specifically, two) endogenous variables in the equation. If $a > 1$, it is multiequation as well.

We briefly expand on the three points in (6) for the present example. Given a vector of moment conditions $W_t u_t$, GMM with weighting matrix $\widehat{\Omega}^{-1}$ is efficient in the class of estimators that exploit $EW_t u_t = 0$. Because much financial data, including the monthly stock return series used by Hansen and Singleton (1982), seem to display conditional heteroscedasticity, nonlinear two-stage least squares would be less efficient. The GMM estimator maintains weak assumptions, for example, not requiring a parametric model for conditional heteroscedasticity or any distributional assumption. Finally, the GMM estimator probably is no more computationally involved than maximum likelihood.

To illustrate panel data with dynamics, we modify this example as follows. First, for simplicity, we suppose that there is single security with a time-invariant return R. The Euler equation developed earlier can be written as $bR(C_{t+1}/C_t)^\gamma = 1 + \epsilon_{t+1}$, where $E_t\epsilon_{t+1} = 0$. Taking the natural logarithm, we find $\Delta\log C_{t+1} = \mu + e_{t+1}$, where μ is a constant and e_{t+1} is a mean-zero transformation of ϵ_{t+1} (and conditionally mean zero under normality). Next, we introduce habit formation by specifying the consumer's utility to be a function of $C_t - \alpha\, C_{t-1}$ rather than C_t for a parameter α. Under a series of approximations (see Dynan 2000), the first-order condition may be written as $\Delta\log C_t = \mu + \alpha\Delta\log C_{t-1} + e_t$. Finally, we suppose that we have data from a panel with an error-component model of the form $e_{it} = X_{it}'\gamma + \mu_i + \nu_{it}$. (Here X_{it} is a vector of individual specific variables and is not related to the variable x_t defined in the preceding example.) The regression equation is thus

$$y_{it} = \alpha y_{it-1} + X_{it}'\gamma + \mu_i + \nu_{it}, \quad (12)$$

where $y_{it} = \Delta\log C_{it}$.

Equation (12) is known as a dynamic panel. The complication is the joint presence of the lagged dependent variable y_{it-1} and the individual-specific effect μ_i. It is convenient to treat μ_i as a fixed effect, because y_{it-1} is predetermined but not strictly exogenous. However, the classic fixed-effects estimator [ordinary least squares (OLS) applied to (12) after removing individual-specific means] is inconsistent, because time averages of the lagged dependent variable are correlated with time averages of the error, as shown by Nickell (1981). A robust solution is to take first differences of (12),

$$\Delta y_{it} = \alpha\Delta y_{it-1} + \Delta X_{it}'\gamma + \Delta\nu_{it}, \quad (13)$$

which eliminates the fixed effect μ_i. Clearly, OLS on (13) is inappropriate, because $E(\Delta y_{it-1}\Delta\nu_{it}) \neq 0$. Anderson and Hsiao (1981, 1982) observed that lags of the data (such as Δy_{it-2} and y_{it-2}) are valid instruments, so IV estimation of (13) is appropriate. The number of valid instruments is proportional to the number of available lags, however, motivating Arellano and Bond (1991) and Arellano and Bover (1995) to formulate a GMM estimator that exploits all lags as instruments. Once again this estimator is efficient, maintains weak assumptions (i.e., does not require specification of the process followed by the X's), and is computationally convenient (the likelihood can be quite complicated when X is nontrivial). For these reasons, it is quite popular in the growing body of recent applied work that estimates dynamic panel models.

4. LITERATURE SURVEY

This section presents the results of a small-scale survey of some prominent journals that categorizes empirical time series articles in terms of the techniques used (e.g., GMM, maximum likelihood). Our aim is to get a sense (admittedly crude) of how prominent is the use of GMM in empirical work, and to get a feel for the role that GMM plays in articles that do use it. We recognize that our survey may have yielded unrepresentative set of articles, that others might disagree with our categorization of some articles, and that a simple tabulation does not reveal whether articles using one technique are, on average, more influential than those using another. We leave these as tasks for future research.

To conduct the survey, we looked for articles on empirical time series published in 1990 and 2000 in seven journals: *American Economic Review; Econometrica; Journal of Political Economy; Journal of Monetary Economics; Journal of Money, Credit and Banking; Quarterly Journal of Economics;* and *Review of Economics and Statistics*. We judged those journals to contain 84 such articles in 1990 and 103 in 2000.

In each article we attempted to identify the main technique or techniques used. Categories of techniques were "VAR," referring to estimation of vector autoregressions; "parametric," referring to non-VAR articles relying on parametric estimation by least squares or IV techniques that either assume iid errors or correct parametrically for serial correlation or conditional heteroscedasticity of disturbances; "unit root," for articles in which unit root or cointegration tests were focal; "maximum likelihood" (no explanation needed!); "nonparametric," referring to nonparametric estimation of a regression function, a category that does *not* include papers whose only nonparametric element was related to covariance matrix estimation; "calibration and other" (again, no explanation needed); and "GMM." The GMM category includes linear regressions that correct nonparametrically for serial correlation of disturbances, or that cite Hansen (1982) in presenting a test of overidentifying restrictions. But it typically excludes OLS and two- and three-stage least squares estimators of linear models whose disturbances are assumed to be iid, even though these are special cases of the work of Hansen (1982).

In the handful of articles that made heavy use of two techniques, we attributed a .5 share to each technique. An example is the article by Kocherlakota (1990), which received a weight of .5 in calibration and .5 in GMM. Such joint attribution was rare—we repeat that an article was counted in a given category only if it made central use of the relevant technique. For example, an article that performed unit root tests as a precursor to a detailed VAR analysis (e.g., Fackler 1990) would be counted as "VAR" but not as "unit root."

The results are given in Table 1. Column 1 indicates that these journals included 187 empirical time series articles in these 2 years. Column 9 indicates that 35 of these articles made heavy use of GMM. Of these 35 articles, several also made heavy use of a second technique, so the figures in column 8 are slightly smaller than those in column 9 (see notes 4 to the Table for further details). Columns 2–7 present totals for other techniques computed similarly to those for GMM in column 8.

The share of GMM in techniques used is 14 of 84 in 1990 and 18 of 103 in 2000, or about one-sixth. Unsurprisingly, the modal technique is parametric regression, accounting for about 40% of all techniques in 1990 ($.40 \approx 36.5$ of 84) and about 30% in 2000 ($.30 \approx 31$ of 103). Had VARs not been broken out as a separate category, the parametric share would have continued to be about 40% in 1990. Equally unsurprisingly, calibration has risen to a (distant) second in 2000, although it and GMM are roughly equally common.

That our sample size is small, and our categorization rough, are suggested by the figures for VARs. These figures show a big jump between 1990 and 2000, whereas our sense is that VARs enjoyed more or less equal popularity in the 2 years.

Table 1. Empirical Techniques Used in Published Articles, 1990 and 2000

		Major techniques used							(9)
	(1) No. of articles	(2) VAR	(3) Parametric	(4) Unit root	(5) MLE	(6) Nonparametric	(7) Calibration + other	(8) GMM	Reference: no. of articles using GMM
1990	84	1.5	36.5	4	14.5	8	5.5	14	16
2000	103	9	31	4.5	10.5	7	23	18	19
Total	187	10.5	67.5	8.5	25	15	28.5	32	35

NOTE: We surveyed the following journals in 1990 and 2000: *American Economic Review; Econometrica; Journal of Political Economy; Journal of Monetary Economics; Journal of Money, Credit and Banking; Quarterly Journal of Economics; and Review of Economics and Statistics*. We judged these journals to contain 84 empirical time series articles in 1990 and 103 empirical time series articles in 2000. In each article we attempted to identify the main technique or techniques used. These techniques are listed in columns 2–8. VAR refers to estimation of vector autoregressions. "Parametric" refers to estimation by least squares or IV techniques that either assume iid errors or correct parametrically for serial correlation or conditional heteroscedasticity of disturbances; VAR articles are not counted in this category. An article falls in the "unit root" category only if unit root or cointegration tests seemed focal to the study. The nonparametric estimation category does not include articles whose only nonparametric element was related to covariance matrix estimation. GMM includes nonlinear overidentified models; linear models are included only if there is nonparametric correction for serial correlation of disturbances. Columns 2–8 sum to the total number of articles in column 1. If an article used two main techniques, then we gave each technique a count of .5. This explains the fractional figures in some columns. It also explains the discrepancy between the figures in columns 8 and 9. The totals in column 9 are larger than those in column 8 because 2 of the 16 GMM articles from 1990, and 4 of the 19 GMM articles from 2000, received a .5 weight in column 8.

Table 2. Aspects of Published Articles That Use GMM, 1990 and 2000

		Type of study				Aspects of study		
	(1) No. of articles	(2) Dynamic optimization	(3) Forecasting study	(4) Other		(5) Panel	(6) Nonlinear restrictions	(7) J test
1990	16	6	9	1		1	7	5
2000	19	10	1	8		7	7	11
Total	35	16	10	9		8	14	16

NOTE: The articles considered here are those defined as using GMM in column (9) of Table 1. Column (1) is the same as column (9) of Table 1. The categories in "Type of Study" are mutually exclusive and exhaustive; those in "Aspects of Study" are neither.

Similar caution applies to interpretation of the figures for the maximum likelihood estimator and nonparametric columns.

What sort of economic problem led to the use of GMM in these 35 articles? Most of the articles fall into one of three categories. The first category is those articles that estimated a first-order condition or decision rule from a dynamic optimization problem. The leading example is an Euler equation for consumption (e.g., Dynan 2000). Column 2 in Table 2 indicates that such articles account for about 40% of the 1990 GMM articles and 50% of the 2000 articles. The second category includes articles that examined forecasting ability over a multiperiod horizon, of either survey data or of a financial variable (e.g., Mishkin 1990). Least squares was the regression technique, with nonparametric computation of the variance-covariance matrix. Column 3 Table 2 indicates that such studies were very common in 1990 and less so in 2000. The third category comprises articles describing a setup in which there were efficiency gains came from the use of many moments (e.g., Attanasio, Picci, and Scorcu 2000); most of these involved dynamic panels [see (12)]. Column 4 in Table 2 includes these articles, as well as a couple of others that used GMM for computational convenience or as a check on maximum likelihood estimation (e.g., McConnell and Perez-Quiros 2000).

The remainder of Table 2 describes some technical characteristics of these articles. A given article can display more than one of these characteristics. Column 5 shows that panel studies using GMM have become much more common. A significant fraction of the articles estimate or test models with nonlinear restrictions (Column 6), and many use the test of overidentifying restrictions (Column 7).

We conclude, as expected, that parametric linear models dominate time series work. In terms of frequency of use, the GMM techniques introduced by Hansen (1982) are tied for second place, perhaps along with calibration and maximum likelihood. In recent work, GMM techniques are most commonly used in nonlinear and panel data studies; in 1990, forecasting applications were also common.

5. SOME TOPICS OF CURRENT RESEARCH

In this section we touch on some active areas of theoretical research. Much of this research has yet to have a noticeable impact on empirical practice. But this research may be motivated by three related practical problems. The first problem is that much simulation evidence indicates that the first-order asymptotic approximations for $\hat{\beta}$, and for t tests and J tests, work poorly in samples of typical size. The second problem is that in practice, the J test often rejects at traditional significance levels. The third problem is that minor changes in specification, weight matrix, or choice of instruments sometimes have major effects on estimates and p values. These problems have led to a desire to find improved GMM estimators, practical methods for selecting instruments and weight matrices, improved distributional approximations, and the desire to better understand the behavior of GMM estimation under misspecification. We discuss in turn theoretical results on the choice of instruments and efficiency bounds, feasible selection of optimal instruments, weight matrix estimation, the bootstrap, and estimation by empirical likelihood.

5.1 Efficiency Bounds

In our presentation of results of Hansen (1982) in Sections 2 and 3, we assumed that we were given an $m \times a$ matrix of instruments W_t that was uncorrelated with our $a \times 1$ vector of regression disturbances, $EW_t u_t = 0$. We illustrated where that matrix might come from. In practice, there is often a surplus of possible instruments. In the forecasting example of Section 2, lags of the expectation x_t satisfy $Ex_{t-j}u_t = 0$ for any $j \geq 0$, as does any other variable available for use when the expectation was formed. In the consumption-returns example of Section 3, lags of returns, lags of consumption growth, and nonlinear transformations of these are uncorrelated with u_t defined in (11), provided that consumers used these efficiently when making their consumption decisions. In the panel data example of that same section, lags of consumption growth beyond the first are uncorrelated with $\Delta\nu_{it}$ defined in (13).

This leads to both theoretical and practical questions about instrument choice. In practice, choice of instruments involves judgment and generally cannot be reduced to a mechanical rule. Good empirical practice generally calls for some experimentation with alternative instrument lists, to ensure that results are not sensitive to exact choice of moments. But there are also some useful formal results that can help guide practitioners. In this section we summarize theoretical results on efficiency bounds when there is an infinite set of potential instruments (see Hansen 1985; Hansen, Heaton, and Ogaki 1988; Bates and White 1993). Related results when there is a finite set of potential instruments have been given by Breusch, Qian, Schmidt, and Wyhowski (1999). Section 5.2 discusses feasible estimation.

Throughout, we are informal in our statement of conditions. Moreover, for expositional ease, most of our discussion takes place in the model like that in Section 2: a linear regression with a single right-side variable that is uncorrelated with the MA(1) disturbance,

$$y_t = x_t\beta + u_t, \qquad Eu_t = E(u_t|x_t, x_{t-1}, \dots) = 0,$$
$$Ex_t u_t x_{t-j} u_{t-j} = 0 \qquad \text{for } |j| > 1. \tag{14}$$

Assume first that u_t is homoscedastic conditional on current and lagged x's: $E(u_t u_{t-j}|$ current and lagged x's$)= Eu_t u_{t-j}$ for all j. As in (3), let θ be the MA parameter for u_t. Also assume that the GMM estimators allowed are those that use current and lagged values of x_t as instruments. We know from Hansen (1982) how to choose the optimal linear combination if given a finite set of lagged x's. But what linear combination of current and lagged values of x_t results in the optimal instrument, say z_t^*, when an infinite set of lags of x may be used?

To derive z_t^*, follow Hansen and Sargent (1982) and Hayashi and Sims (1983) and *forward filter* (14) with the filter $1/(1-\theta L^{-1})$. Define $\tilde{y}_t \equiv y_t/(1-\theta L^{-1}) = \sum_{j=0}^{\infty}\theta^j y_{t+j}$, $\tilde{x}_t \equiv x_t/(1-\theta L^{-1}) = \sum_{j=0}^{\infty}\theta^j x_{t+j}$, and $\tilde{u}_t \equiv u_t/(1-\theta L^{-1}) = \sum_{j=0}^{\infty}\theta^j u_{t+j}$. Then (14) becomes $\tilde{y}_t = \beta\tilde{x}_t + \tilde{u}_t$. In contrast to the residual that results from the usual GLS transformation, defined in (4) and (5), $\tilde{u}_t$ is uncorrelated with all potential instruments because it is a linear combination of *future* u's, whereas the GLS residual is a linear combination of past u's. Moreover, $\tilde{u}_t$ is serially uncorrelated. Because it is also conditionally homoscedastic by assumption, an optimal estimator of β in the equation $\tilde{y}_t = \beta\tilde{x}_t + \tilde{u}_t$ is one that uses an instrument $\tilde{z}_t^* = E(\tilde{x}_t|$ instruments$)$. In terms of the initial equation (14), an asymptotically equivalent estimator relies on an instrument and the estimator

$$z_t^* = \theta z_{t-1}^* + c\tilde{z}_t^*, \hat{\beta}^* = \left(\sum_{t=1}^{T} z_t^* x_t\right)^{-1}\left(\sum_{t=1}^{T} z_t^* y_t\right);$$
$$\tilde{z}_t^* = E(\text{forward filtered } x_t|\text{ instruments})$$
$$= E\left(\sum_{j=0}^{\infty}\theta^j x_{t+j}|x_t, x_{t-1}, \dots\right). \tag{15}$$

Here c is an arbitrary nonzero scalar. Thus $z_t^* = c\sum_{j=0}^{\infty}\theta^j\tilde{z}_{t-j}^*$; the optimal instrument puts nonzero weight on *all* lags of x_t. For example, if $x_t \sim$ AR(1) with parameter ϕ, $\tilde{z}_t^* = (1-\theta\phi)^{-1}x_t$ and $z_t^* = c\sum_{j=0}^{\infty}\theta^j\tilde{z}_{t-j}^*$. Hansen (1985) provided the generalization of this result to multiple-equation, possibly nonlinear systems with possibly high-order serial correlation (see the final paragraph of this section).

Now suppose that the regression disturbance is heteroscedastic conditional on the instruments. The literature on efficient GMM estimation under such conditional heteroscedasticity is less well developed. Suppose first that the GMM estimators allowed are those that exploit the infinite set of moment conditions $Ex_{t-j}u_t = 0$ for $j \geq 0$. Then in some special cases, most notably that of a serially uncorrelated disturbance, closed forms for the optimal instruments have been derived (Kuersteiner 2000; West 2001).

But with conditional heteroscedasticity, it is natural to aim to exploit moments associated with the conditional variance of the regression disturbance, in example (14) allowing for instruments that are measurable with respect to the sigma algebra generated by current and lagged x's. In the special case when the disturbance is serially uncorrelated ($Ex_t u_t x_{t-j} u_{t-j} = 0$ for all $j \neq 0$), an analog to weighted least squares is optimal, $z_t^* = x_t/\sigma_t^2$, where $\sigma_t^2 = E(u_t^2|x_t, x_{t-1}, \dots)$ (Hansen et al. 1988).

Hansen et al. (1988) described the efficiency bound for when the regression disturbance follows an MA process and is possibly conditionally heteroscedastic. This bound may be described in a series of steps that parallel the argument for the conditionally homoscedastic MA(1) model given earlier. First, forward filter the equation with a filter that, when applied to u_t, yields a serially uncorrelated and conditionally homoscedastic random variable. Let $\tilde{x}_t$ denote the random variable that results when the filter is applied to x_t. The minimum asymptotic variance is obtained by instrumenting the resulting equation with the instrument $E(\tilde{x}_t|x_t, x_{t-1}, \dots)$. The efficiency bound may be obtained with estimation of (14) by an instrument that results from filtering $E(\tilde{x}_t|x_t, x_{t-1}, \dots)$ in a certain way (see Hansen et al. 1988; Anatolyev 2002).

5.2 Feasible Estimation

How does one construct a feasible estimator that asymptotically attains the efficiency bound described in the previous section? Hayashi and Sims (1983) and Hansen and Singleton (1991) noted that one possibility is to estimate using the techniques of Hansen (1982) with an instrument list that grows with the sample size. In our example (14), this might mean estimating with an instrument vector $W_t = (x_t, x_{t-1}, \dots, x_{t-n})'$, where n grows with T at a suitable rate. Koenker and Machado (1999) established the rate for a cross-sectional model with errors that are independent across observations. To our knowledge, the rate has not been established for time series models.

In addition, there is some simulation evidence that in samples of typical size, a GMM estimator that uses sufficient number of instruments to come close to the efficiency bound will tend to have poor properties in practice (West and Wilcox 1996; West, Wong, and Anatolyev 2001). In conditionally homoscedastic models, an alternative route to attaining the efficiency bound is to obtain the necessary quantities with parametric models, letting the dimension of these models increase with sample size. To our knowledge, the relevant rate results have not been established. But simulation evidence of West and Wilcox (1996) and West et al. (2001) suggests that such estimators will work well in practice (see also Keane and Runkle 1992; Schmidt, Ahn, and Wyhowski 1992).

Fewer results are available in conditionally heteroscedastic models. In an autoregression with a serially uncorrelated disturbance, Kuersteiner (2000) showed how to obtain the efficiency bound in the space of estimators that used lagged values as instruments. A major roadblock is that we typically do not know how to construct the forward filter referenced in the last paragraph of Section 5.1. Tauchen (1986) and Anatolyev (2000) used approximations and nonfeasible estimators to evaluate GMM estimators that efficiently use information on conditional heteroscedasticity in the regression disturbance. They often find that the finite-sample distribution is poorly approximated by the asymptotic distribution.

5.3 Weight Matrix

In Section 3 we stated that the efficient GMM estimator chooses the weight matrix $\widehat{D} = \widehat{\Omega}^{-1}$, where $\widehat{\Omega}$ is a consistent estimator of Ω, the long-run covariance of $g_t(\beta_0)$. Because this matrix is defined as the sum of autocovariances of all lag orders, $\widehat{\Omega}$ is typically a kernel-weighted average of sample autocovariances of the form

$$\widehat{\Omega} = \widehat{\Gamma}(0) + \sum_{j=1}^{h} w_{jh}[\widehat{\Gamma}(j) + \widehat{\Gamma}(j)']$$

$$\widehat{\Gamma}(j) = T^{-1} \sum_{t=1}^{T-j} g_{t+j}(\beta^*) g_t(\beta^*)'. \tag{16}$$

where β^* is a preliminary estimate of β. This heteroscedasticity and autocorrelation consistent (HAC) estimator depends critically on the kernel weights w_{jh} and the bandwidth h (sometimes called the lag truncation).

Which kernel weights should be used? The early literature (e.g., Hansen 1982; White and Domowitz 1984) presented an estimator that is equivalent to setting $w_{jh} = 1$, the so-called truncated or uniform kernel, which weights all estimated autocovariances equally. One problem with this choice is that the estimate $\widehat{\Omega}$ is not necessarily positive definite (which it should be to be a proper weight matrix). This motivated Newey and West (1987a) to suggest using kernel weights that guarantee a positive definite $\widehat{\Omega}$. Popular choices in this class of kernels includes the Bartlett $w_{jh} = 1 - j/(h+1)$, Parzen, and quadratic spectral.

How should the bandwidth h be selected in practice? One guideline is to select h to minimize the mean squared error of the estimator $\widehat{\Omega}$. Optimal rates and plug-in rules for h were derived by Andrews (1991) and Newey and West (1994).

An alternative to kernel HAC estimation is parametric estimation based on a fitted VAR. Some properties of such a covariance matrix estimator have been analyzed by den Haan and Levin (2000). An estimator that fits an MA model, under the assumption that the autocovariances are known a priori to be zero after a lag, was analyzed by West (1997).

As noted earlier, the estimator $\widehat{\Omega}$ is a function of a preliminary estimate β^*. We can make this dependence explicit by writing $\widehat{\Omega} = \widehat{\Omega}(\beta^*)$. Thus the conventional GMM estimator minimizes

$$\left[T^{-1}\sum_{t=1}^{T} g_t(\beta)\right]' \widehat{\Omega}(\beta^*)^{-1} \left[T^{-1}\sum_{t=1}^{T} g_t(\beta)\right]. \tag{17}$$

Alternatively, we can allow $\widehat{\Omega}(\beta)$ to be freely varying in β in the construction of the GMM criterion

$$\left[T^{-1}\sum_{t=1}^{T} g_t(\beta)\right]' \widehat{\Omega}(\beta)^{-1} \left[T^{-1}\sum_{t=1}^{T} g_t(\beta)\right]. \tag{18}$$

The estimator that minimizes (18), known as the continuously updated GMM (CU-GMM) estimator, was introduced by Hansen, Heaton, and Yaron (1996).

Focusing on the test for overidentifying restrictions, Hall (2000) observed that whereas $g_t(\beta_0)$ is mean zero under the null hypothesis of correct specification, it has a non-zero mean under the alternative hypothesis of misspecification. Thus conventional HAC estimators $\widehat{\Omega}^{-1}$, which are based on autocovariances of $g_t(\beta^*)$ without recentering, will be asymptotically singular under the alternative hypothesis reducing the power of the overidentification test. A simple solution is to use a HAC estimator $\widehat{\Omega}$ based on the autocovariances of demeaned $g_t(\beta^*)$.

Because efficient GMM requires that the inverse weight matrix be consistent for Ω, the conventional asymptotic approximation specifies that $h/T \to 0$. An alternative asymptotic framework has been proposed by Kiefer and Vogelsang (2002). Working with the "large bandwidth" approximation $h/T \to c > 0$, they derive asymptotic distributions that are nonstandard and reflect the randomness in the estimation of the weight matrix. They find that larger values of c (e.g., larger bandwidths h) result in tests with better finite-sample size performance (when the nonstandard critical values are used) but with reduced power. In some sense, this result is not surprising. The large bandwidth decreases the bias of the estimated weight matrix but increases its variance. The expansions of Inoue and Shintani (2001) discussed in the next section suggest that bias is important for distributional approximations, which is consistent with the improved size performance. And an inconsistently estimated optimal weight matrix necessarily results in inefficient parameter estimates and reduced power. However, the new asymptotic distribution of Kiefer and Vogelsang (2002) may prove useful as an alternative approximation device, especially for analysis of the bootstrap.

5.4 Bootstrap

Monte Carlo simulations have shown that asymptotic first-order approximations for GMM estimators and tests have low accuracy in moderate sample sizes. An important alternative to conventional asymptotic approximations are bootstrap distributions, which calculate the finite-sample distribution of GMM test statistics based on an approximate data generating process. Because GMM is inherently semiparametric (i.e., the full distribution of the data is unspecified), it is necessary to use a nonparametric bootstrap (rather than a parametric bootstrap).

The primary bootstrap method appropriate for nonparametric time series data is the block bootstrap, which works as follows. Let y_t denote the vector of observations (including lags) used to construct the moment $g_t(\beta)$ of Section 3. (Note that this is a change in notation from earlier sections.) For some block length ℓ (which grows at a slower rate than T) construct blocks of data of the form $(y_t, y_{t-1}, \ldots, y_{t-\ell})$. Then construct a random time series y_t^* by drawing random data blocks and pasting them together, until a time series of length T is created. GMM estimators and tests can be constructed on this bootstrap data, and their distribution calculated by simulation to construct confidence intervals and significance levels.

One important issue is block construction and length. Hall and Horowitz (1996) considered nonoverlapping blocks, whereas Andrews (2002) allowed for overlapping blocks. Hall, Horowitz, and Jing (1995) discussed the optimal choice of block length ℓ. They found that $\ell \propto T^{1/4}$ for a one-sided significance test (or equal-tailed confidence interval) and $\ell \propto T^{1/5}$ for a two-sided significance test (or symmetric confidence interval). A disappointing message delivered by Zvingelis (2001) is that the best possible error rates for confidence intervals constructed using the block bootstrap are $O(T^{-3/4})$ and $O(T^{-4/3})$ for equal-tailed and symmetric confidence intervals, which are noticeably larger than the $O(T^{-1})$ and $O(T^{-2})$ rates obtained for independent sampling. Thus the block bootstrap achieves a low asymptotic refinement—the improvement in the error rate relative to a conventional test.

It is probably well known that bootstrap test statistics (e.g., a t ratio) should be centered at the sample estimate of the parameter, rather than at the value specified under the null hypothesis. (From the perspective of the bootstrap, the sample estimate is the true value.) What is probably less well understood is that overidentified GMM estimation requires additional recentering. Because the number of moments m exceeds the number of parameters k, the sample moments cannot be set to 0. The problem is that the "true" moments of the bootstrap sample will equal the non-zero sample values, which clearly violates the orthogonality condition (7). As a result, naive bootstrap estimates and tests will be biased. A solution proposed by Hall and Horowitz (1996) is to alter the bootstrap criterion function so that the moments are centered at the sample moments. An alternative solution based on empirical likelihood weights has been suggested by Brown and Newey (2002) in the context of independent data, but their method does not immediately apply to the block bootstrap.

The choice of kernel for HAC estimation when using the block bootstrap is not straightforward. Hall and Horowitz (1996) and Andrews (2002) assumed use of the truncated kernel. This may appear to be only a technicality, but it actually is an important simplification. Gotze and Kunsch (1996) showed that for one-sided confidence intervals, the block bootstrap does not provide refinement for the Bartlett kernel. Inoue and Shintani (2001) further showed that for symmetric confidence intervals and the J test, asymptotic refinements essentially require use of the truncated kernel. [The technical condition relates to the characteristic exponent of the kernel (Anderson 1971, chap. 9). Gotze and Kunsch require an exponent greater than 1; Inoue and Shintani, an exponent greater than 2.] Their insight is that bootstrap refinements focus on the bias of the variance estimator, which is minimized by the truncated kernel. The dilemma is that the truncated kernel does not guarantee a positive definite weight matrix. We are not aware of a constructive solution to this problem.

Another major difficulty imposed by the block bootstrap is that the serial correlation properties of the bootstrap data are altered by the blocking. This affects the variance of sample averages of bootstrap data. Regardless of the actual serial correlation properties of the data series y_t, the bootstrap series y_t^* is ℓ dependent. Thus the sample mean of the bootstrap series y_t^* has an exact variance that takes a complicated form involving the sample autocovariances of y_t. For example, if overlapping blocks are used,

$$\operatorname{var}\left(T^{-1/2}\sum_{t=1}^{T} y_t^*\right) = \ell^{-1}(T-\ell+1)^{-1}\sum_{i=0}^{T-\ell}\sum_{j=1}^{\ell}\sum_{k=1}^{\ell}(y_{i+j}-\bar{y})(y_{i+k}-\bar{y}). \quad (19)$$

Because the formula for the variance of the bootstrap sample is not the usual one, a naive application of the block bootstrap using a conventional variance formula to construct the bootstrap test statistic will be incorrectly normalized. Davison and Hall (1993) showed that the resulting error is of order $O(\ell/T)+O(\ell^{-1})$, which can be greater than or equal to the error of the first-order asymptotic approximation. Hall and Horowitz (1996) showed that this problem can be solved by rescaling the bootstrap t statistic by a correction factor that involves calculating (19). Andrews (2002) adopted the same solution. Inoue and Shintani (2001) went a step further, recommending using the corrected estimate (19) for the weight matrix used for the bootstrap GMM criterion function, parameter estimate, and test statistics, eliminating the need for the correction factor.

To summarize, regardless of the choice of kernel and bandwidth used in the actual estimation, when using the block bootstrap it is necessary to construct the bootstrap GMM criterion function, parameter estimates, and test statistics using the corrected variance (19). It is probably surprising to realize that this must be done even if the weight matrix used in estimation does not take the HAC form (e.g., in an autoregression with martingale difference errors.) Again, this is because the block bootstrap produces data with an altered serial correlation pattern, and this must be handled in the computation of the bootstrap variance estimators. Without these corrections, the block bootstrap may perform worse than conventional asymptotic approximations. Notice that there is a disconnect between estimation and the bootstrap, because different formulas are used for the HAC estimator in actual estimation and in bootstrap estimation. This dilemma might appear artificial, but it is the best recommendation currently offered by theory.

A recent development is the Markov conditional bootstrap (MCB) of Horowitz (2002), who suggested estimating the one-step-ahead conditional density of the time series using a multivariate nonparametric kernel density estimator, and using the estimated conditional density to construct the bootstrap time series. This method avoids many of the problems of the block bootstrap, including the need to modify the HAC matrix. Horowitz (2002) showed that in some cases the MCB achieves a better asymptotic refinement than the block bootstrap.

5.5 Empirical Likelihood Estimation

GMM is used in econometrics when estimation and testing is based solely on a set of unconditional moment equations. In this sense the method is semiparametric, because the other dimensions of the joint distribution are left unspecified. A closely related alternative estimation and testing framework, also well suited for semiparametric models defined by a set of unconditional moment conditions, was developed by Owen (1988, 2001) (see also Imbens 2002). Owen called this method "empirical likelihood" (EL), because it is based on the construction of a nonparametric likelihood using no information other than the sample and the moment conditions. An interesting feature is that unlike GMM estimation, EL does not require estimation of an optimal weight matrix. The criterion function naturally adapts to the data.

Qin and Lawless (1994) and Imbens, Spady, and Johnson (1998) developed EL estimation for independent observations and general overidentified nonlinear moment equations. For just-identified models, the parameter estimates are identical to GMM. For overidentified models, the estimator is a relative of the CU-GMM estimator (18). Testing and confidence intervals in the EL framework are naturally based on the EL criterion function. Kitamura (2001) suggested that such tests have a strong optimality property.

Kitamura (1997) and Kitamura and Stutzer (1997) extended EL estimation and testing to time series. When the moment conditions are martingale differences, estimation and inference are identical to the case of independent observations. Under serial correlation, however, the criterion must be modified. Kitamura (1997) suggested a blocking of the EL that stylistically resembles the block bootstrap, but plays a statistical role closer to that of HAC estimation with the Bartlett kernel.

6. CONCLUSION

GMM is essential to macroeconometrics. It has been central in a wide variety of applications. First, many dynamic optimizing models imply conditional moment restrictions (Euler equations) that can be used to construct unconditional moment equations. Second, many forecasting equations imply orthogonality relations that can be used for moment construction. Third, complicated dynamics, in dynamic panel models and elsewhere, can give rise to intractable likelihoods but feasible moment relationships. Finally, models that are linear in the variables but subject to nonlinear restrictions on the parameters are naturally estimated by GMM.

The generalized method of moments estimator has had a profound impact on our own research and the entire field of macroeconometrics. This article has partially surveyed this literature. We have no doubt that GMM will continue to be vital to macroeconometric research for the foreseeable future.

ACKNOWLEDGMENTS

We thank the National Science Foundation for financial support, and Art Goldberger, Taisuke Otsu, and the editors for attentive comments on the manuscript.

[Received May 2002. Revised June 2002.]

REFERENCES

Amemiya, T. (1974), "The Nonlinear Two-Stage Least-Squares Estimator," *Journal of Econometrics*, 2, 105–110.

——— (1977), "The Maximum Likelihood and Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model," *Econometrica*, 45, 955–968.

Anatolyev, S. (2000), "Approximately Optimal Nonlinear Instrumental Variables Estimation in Time Series Models," manuscript, New Economic School, Moscow.

——— (2002), "The Form of the Optimal Nonlinear Instrument for Multiperiod Conditional Moment Restrictions," manuscript, New Economic School, Moscow.

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, New York: Wiley.

Anderson, T. W., and Hsiao, C. (1981), "Estimation of Dynamic Models With Error Components," *Journal of the American Statistical Association*, 76, 598–606.

——— (1982), "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18, 47–82.

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

——— (2002), "Higher-Order Improvements of a Computationally Attractive *k*-Step Bootstrap for Extremum Estimators," *Econometrica*, 70, 119–162.

Arellano, M., and Bond, S. (1991), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and An Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

Arellano, M., and Bover, O. (1995), "Another Look at the Instrumental Variables Estimation of Error-Components Models," *Journal of Econometrics*, 68, 29–51.

Attanasio, O. P., Picci, L., and Scorcu, A. (2000), "Saving, Growth, and Investment: A Macroeconomic Analysis Using a Panel of Countries," *Review of Economics and Statistics*, 82, 182–211.

Basmann, R. L. (1957), "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," *Econometrica*, 25, 77–83.

——— (1960), "On the Asymptotic Distribution of Generalized Linear Estimators," *Econometrica*, 28, 97–107.

Bates, C. E., and White, H. (1993), "Determination of Estimators With Minimum Asymptotic Covariance Matrices," *Econometric Theory*, 9, 633–648.

Breusch, T., Qian, H., Schmidt, P., and Wyhowski, D. (1999), "Redundancy of Moment Conditions," *Journal of Econometrics*, 91, 89–111.

Brown, B. W., and Newey, W. K. (2002), "GMM, Efficient Bootstrapping, and Improved Inference," *Journal of Business & Economic Statistics*, 20, 507–517.

Davison, A. C., and Hall, P. (1993), "On Studentizing and Blocking Methods for Implementing the Bootstrap With Dependent Data," *Australian Journal of Statistics*, 35, 215–284.

den Haan, W. J., and Levin, A. T. (2000), "Inferences From Parametric and Non-Parametric Covariance Matrix Estimation Procedures," Technical Working Paper 195, National Bureau of Economic Research.

Dynan, K. E. (2000), "Habit Formation in Consumer Preferences: Evidence From Panel Data," *American Economic Review*, 90, 391–406.

Fackler, J. S. (1990), "Federal Credit, Private Credit, and Economic Activity," *Journal of Money, Credit, and Banking*, 22, 444–464.

Gallant, A. R. (1977), "Three-Stage Least-Squares Estimation for a System of Simultaneous, Nonlinear, Implicit Equations," *Journal of Econometrics*, 5, 71–88.

——— (1987), *Nonlinear Statistical Models*, New York: Wiley.

Gallant, A. R., and Jorgenson, D. W. (1979), "Statistical Inference for a System of Simultaneous, Non-Linear Implicit Equations in the Context of Instrumental Variable Estimation," *Journal of Econometrics*, 11, 275–302.

Gotze, F., and Kunsch, H. R. (1996), "Second Order Correctness of the Blockwise Bootstrap for Stationary Observations," *The Annals of Statistics*, 24, 1914–1933.

Hall, A. R. (2000), "Covariance Matrix Estimation and the Power of the Overidentifying Restrictions Test," *Econometrica*, 68, 1517–1527.

Hall, P., and Horowitz, J. L. (1996), "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators With Dependent Data," *Econometrica*, 64, 891–916.

Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995), "On Blocking Rules for the Bootstrap," *Biometrika*, 82, 561–574.

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

——— (1985), "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators," *Journal of Econometrics*, 30, 203–228.

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

Hansen, L. P., Heaton, J., and Yaron, A. (1996), "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 14, 262–280.

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

Hansen, L. P., and Sargent, T. J. (1982), "Instrumental Variables Procedures for Estimating Linear Rational Expectations Models," *Journal of Monetary Economics*, 9, 263–296.

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

——— (1991), "Computing Semi-Parametric Efficiency Bounds for Linear Time Series Models," *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, eds. W. Barnett, J. Powell, and G. Tauchen, Cambridge, UK: Cambridge University Press, pp. 387–412.

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

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

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

Horowitz, J. L. (2002), "Bootstrap Methods for Markov Processes," unpublished manuscript, Northwestern University.

Imbens, G. W. (2002), "Generalized Method of Moments and Empirical Likelihood," *Journal of Business & Economic Statistics*, 20, 493–506.

Imbens, G. W., Spady, R. H., and Johnson, P. (1998), "Information-Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333–357.

Inoue, A., and Shintani, M. (2001), "Bootstrapping GMM Estimators for Time Series," unpublished manuscript, North Carolina State University.

Jorgenson, D. W., and Laffont, J. (1974), "Efficient Estimation of Nonlinear Simultaneous Equations With Additive Disturbances," *Annals of Economic and Social Measurement*, 3, 615–640.

Keane, M. P., and Runkle, D. E. (1992), "On the Estimation of Panel Data Models With Serial Correlation when Instruments Are Not Strictly Exogenous," *Journal of Business & Economic Statistics*, 10, 1–9.

Kiefer, N. M., and Vogelsang, T. J. (2002), "A New Asymptotic Theory for Heteroskedasticity—Autocorrelation Robust Tests," unpublished manuscript, Cornell University.

Kitamura, Y. (1997), "Empirical Likelihood Methods With Weakly Dependent Processes," *The Annals of Statistics*, 25, 2084–2102.

——— (2001), "Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions," *Econometrica*, 69, 1661–1672.

Kitamura, Y., and Stutzer, M. (1997), "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861–874.

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

Koenker, R., and Machado, J. A. F. (1999), "GMM Inference When the Number of Moment Conditions Is Large," *Journal of Econometrics*, 93, 327–344.

Kuersteiner, G. M. (2000), "Efficient IV Estimation for Autoregressive Models With Conditional Heteroskedasticity," unpublished manuscript, Massachusetts Institute of Technology.

McConnell, M. M., and Perez-Quiros, G. (2000), "Output Fluctuations in the United States: What Has Changed Since the Early 1980s?" *American Economic Review*, 90, 1464–1476.

Mishkin, Fr. S. (1990), "What Does the Term Structure Tell Us About Future Inflation?" *Journal of Monetary Economics*, 25, 77–95.

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

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

——— (1994), "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.

Nickell, S. (1981), "Biases in Dynamic Models With Fixed Effects," *Econometrica*, 49, 1417–1426.

Owen, A. B. (1988), "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, 75, 237–249.

——— (2001), *Empirical Likelihood*, New York: Chapman and Hall.

Qin, J., and Lawless, J. (1994), "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22, 300–325.

Rothenberg, T. J. (1973), *Efficient Estimation With A Priori Information*, New Haven, CT: Yale University Press.

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

——— (1959), "The Estimation of Relationships With Autocorrelated Residuals by the Use of Instrumental Variables," *Journal of the Royal Statistical Society, Ser. B*, 21, 91–105.

Schmidt, P., Ahn, S. C., and Wyhowski, D. (1992), "Comment," *Journal of Business & Economic Statistics*, 10, 10–14.

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

Theil, H. (1953), "Repeated Least Squares Applied to Complete Equation Systems," unpublished manuscript, Central Planning Bureau, The Hague.

West, K. D. (1997), "Another Heteroskedasticity- and Autocorrelation-Consistent Covariance Matrix Estimator," *Journal of Econometrics*, 76, 171–191.

——— (2001), "On Optimal Instrumental Variables Estimation of Stationary Time Series Models," *International Economic Review*, 42, 1043–1050.

West, K. D., and Wilcox, D. W. (1996), "A Comparison of Alternative Instrumental Variables Estimators of a Dynamic Linear Model," *Journal of Business & Economic Statistics*, 14, 281–293.

West, K. D., Wong, K.-F., and Anatolyev, S. (2001), "Instrumental Variables Estimation of Heteroskedastic Linear Models Using All Lags of Instruments," unpublished manuscript, University of Wisconsin.

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

——— (1982a), "Instrumental Variables Regression With Independent Observations," *Econometrica*, 50, 483–500.

——— (1982b), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

White, H., and Domowitz, I. (1984), "Nonlinear Regression With Dependent Observations," *Econometrica*, 52, 143–162.

Zvingelis, J. J. (2001), "On Bootstrap Coverage Probability With Dependent Data," in *Computer-Aided Econometrics*, ed. D. Giles, New York: Marcel Dekker.