Generalized Method of Moments Estimation

Lars Peter Hansen[1]
Department of Economics
University of Chicago
email: l-hansen@uchicago.edu

June 17, 2007

1 Introduction

Generalized Method of Moments (GMM) refers to a class of estimators which are constructed from exploiting the sample moment counterparts of population moment conditions (sometimes known as orthogonality conditions) of the data generating model. GMM estimators have become widely used, for the following reasons:

- GMM estimators have large sample properties that are easy to characterize in ways that facilitate comparison. A family of such estimators can be studied *a priori* in ways that make asymptotic efficiency comparisons easy. The method also provides a natural way to construct tests which take account of both sampling and estimation error.

- In practice, researchers find it useful that GMM estimators can be constructed without specifying the full data generating process (which would be required to write down the maximum likelihood estimator.) This characteristic has been exploited in analyzing partially specified economic models, in studying potentially misspecified dynamic models designed to match target moments, and in constructing stochastic discount factor models that link asset pricing to sources of macroeconomic risk.

Books with good discussions of GMM estimation with a wide array of applications include: Cochrane (2001), Arellano (2003), Hall (2005), and Singleton (2006). For a theoretical treatment of this method see Hansen (1982) along with the self contained discussions in the books. See also Ogaki (1993) for a general discussion of GMM estimation and applications, and see Hansen (2001) for a complementary entry that, among other things, links GMM estimation to related literatures in statistics. For a collection of recent methodological advances related to GMM estimation see Ghysels and Hall (2002). While some of these other references explore the range of substantive applications, in what follows we focus more on the methodology.

2 Setup

As we will see, formally there are two alternative ways to specify GMM estimators, but they have a common starting point. Data are a finite number of realizations of the process $\{x_t : t = 1, 2, ...\}$. The model is specified as a vector of moment conditions:

$$Ef(x_t, \beta_o) = 0$$

where f has r coordinates and β_o is an unknown vector in a parameter space $\mathbb{P} \subset \mathbb{R}^k$. To achieve identification we assume that on the parameter space $\mathbb{P}$

$$Ef(x_t, \beta) = 0 \quad \text{if, and only if} \quad \beta = \beta_o. \tag{1}$$

The parameter β_o is typically not sufficient to write down a likelihood function. Other parameters are needed to specify fully the probability model that underlies the data generation. In other words, the model is only partially specified.

Examples include:

i) linear and nonlinear versions of instrumental variables estimators as in Sargan (1958), Sargan (1959), Amemiya (1974);

ii) rational expectations models as in Hansen and Singleton (1982), Cumby et al. (1983), and Hayashi and Sims (1983)

iii) security market pricing of aggregate risks as described, for example, by Cochrane (2001), Singleton (2006) and Hansen et al. (2007);

iv) matching and testing target moments of possibly misspecified models as described by, for example, Christiano and Eichenbaum (1992) and Hansen and Heckman (1996).

Regarding example iv, many related methods have been developed for estimating correctly specified models, dating back to some of the original applications in statistics of method-of-moments type estimators. The motivation for such methods was computational. See Hansen (2001) for a discussion of this literature and how it relates to GMM estimation. With advances in numerical methods, the fully efficient maximum likelihood method and Bayesian counterparts have become much more tractable. On the other hand, there continues to be an interest in the study of dynamic stochastic economic models that are misspecified because of their purposeful simplicity. Thus moment matching remains an interesting application for the methods described here. Testing target moments remains valuable even when maximum likelihood estimation is possible (for example, see Bontemps and Meddahi (2005)).

2.1 Central Limit Theory and Martingale approximation

The parameter dependent average

$$g_N(\beta) = \frac{1}{N} \sum_{t=1}^{N} f(x_t, \beta)$$

is featured in the construction of estimators and tests. When the Law of Large Numbers is applicable, this average converges to the $Ef(x_t, \beta)$. As a refinement of the identification condition:

$$\sqrt{N} g_N(\beta_0) \Longrightarrow \text{Normal}(0, V) \tag{2}$$

where $\Longrightarrow$ denotes convergence in distribution and V is a covariance matrix assumed to be nonsingular. In an iid data setting, V is the covariance matrix of the random vector $f(x_t, \beta_o)$. In a time series setting:

$$V = \lim_{N \to \infty} NE\left[g_N(\beta_o) g_N(\beta_o)'\right], \tag{3}$$

which is the long run counterpart to a covariance matrix.

Central limit theory for time series is typically built on martingale approximation. (See Gordin (1969) or Hall and Heyde (1980)). For many time series models, the martingale approximators can be constructed directly and there is specific structure to the V matrix.

A leading example is when $f(x_t, \beta_o)$ defines a conditional moment restriction. Suppose that $x_t, t = 0, 1, ...$ generates a sigma algebra $\mathcal{F}_t$, $E\left[|f(x_t, \beta_0)|^2\right] < \infty$ and

$$E\left[f(x_{t+\ell}, \beta_0)|\mathcal{F}_t\right] = 0$$

for some $\ell \geq 1$. This restriction is satisfied in models of multi-period security market pricing and in models that restrict multi-period forecasting. If $\ell = 1$, then g_N is itself a martingale; but when $\ell > 1$ it is straightforward to find a martingale m_N with stationary increments and finite second moments such that

$$\lim_{N \to \infty} E\left[|g_N(\beta_0) - m_N(\beta_0)|^2\right] = 0,$$

where $|\cdot|$ is the standard Euclidean norm. Moreover, the lag structure may be exploited to show that the limit in (3) is[1]

$$V = \sum_{j=-\ell+1}^{\ell-1} E\left[f(x_t, \beta_0)f(x_{t+j}, \beta_0)'\right]. \tag{4}$$

When there is no exploitable structure to the martingale approximator, the matrix V is the spectral density at frequency zero.

$$V = \sum_{j=-\infty}^{\infty} E\left[f(x_t, \beta_0)f(x_{t+j}, \beta_0)'\right]$$

2.2 Minimizing a Quadratic Form

One approach for constructing a GMM estimator is to minimize the quadratic form:

$$b_N = \arg\min_{\beta \in \mathbb{P}} g_N(\beta)'W g_N(\beta)$$

for some positive definite *weighting matrix* W. Alternative weighting matrices W are associated with alternative estimators. Part of the justification for this approach is that

$$\beta_0 = \arg\min_{\beta \in \mathbb{P}} Ef(x_t, \beta)'W Ef(x_t, \beta).$$

The GMM estimator mimics this identification scheme by using a sample counterpart.

There are a variety of ways to prove consistency of GMM estimators. Hansen (1982) established a uniform law of large numbers for random functions when the data generation is stationary and ergodic. This uniformity is applied to show that

$$\sup_{\beta \in \mathbb{P}} |g_N(\beta) - E\left[f(x_t, \beta)\right]| = 0$$

---

[1]The sample counterpart to this formula is not guaranteed to be positive semidefinite. There are a variety of ways to exploit this dependence structure in estimation in constructing a positive semidefinite estimate. See Eichenbaum et al. (1988) for an example.

and presumes a compact parameter space. The uniformity in the approximation carries over directly the GMM criterion function $g_N(\beta)'Wg_N(\beta)$. See Newey and McFadden (1994) for a more complete catalog of approaches of this type.

The compactness of the parameter space is often not ignored in applications, and this commonly invoked result is therefore less useful than it might seem. Instead the compactness restriction is a substitute for checking behavior of the approximating function far away from β_o to make sure that spurious optimizers are not induced by approximation error. This tail behavior can be important in practice, so a direct investigation of it can be fruitful. For models with parameter separation:

$$f(x, \beta) = Xh(\beta)$$

where X is a $r \times m$ matrix constructed from x and h is a one-to-one function mapping $\mathbb{P}$ into subset of $\mathbb{R}^m$, there is an alternative way to establish consistency. See Hansen (1982) for details. Models that are either linear in the variables or models based on matching moments that are nonlinear functions of the underlying parameters can be written in this separable form.

The choice of $W = V^{-1}$ receives special attention, in part because

$$Ng_N(\beta)'V^{-1}g_N(\beta) \implies \chi^2(r).$$

While the matrix V is typically not known, it can be replaced by a consistent estimator without altering the large sample properties of b_N. When using martingale approximation, the implied structure of V can often be exploited as in formula (4). When there is no such exploitable structure, the method of Newey and West (1987b) and others can be employed that are based on frequency-domain methods for time series data.

For asset pricing models there are other choices of a weighting matrix motivated by considerations of misspecification. In these models with parameterized stochastic discount factors, the sample moment conditions $g_N(\beta)$ can be interpreted as a vector of pricing errors associated with the parameter vector β. A feature of $W = V^{-1}$ is that if the sample moment conditions (the sample counterpart to a vector pricing errors) happened to be the same for two models (two choices of β), the one for which the implied asymptotic covariance matrix is larger will have a smaller objective. Thus there is a *reward* for parameter choices that imply variability in the underlying central limit approximation. To avoid such a reward, it is also useful to compare models or parameter values in other ways. An alternative weighting matrix is constructed by minimizing the least squares distance between the parameterized stochastic discount factor and one among the family of discount factors that correctly price the assets. Equivalently, parameters or models are selected on the basis of the maximum pricing error among constant weighted portfolios with payoffs that have common magnitude (a unit second moment). See Hansen and Jagannathan (1997) and Hansen et al. (1995) for this and related approaches.

2.3   Selection Matrices

An alternative depiction is to introduce a *selection matrix* A that has dimension $k \times r$ and to solve the equation system:

$$Ag_N(\beta) = 0$$

for some choice of β, which we denote b_N. The selection matrix A reduces the number of equations to be solved from r to k. Alternative selection matrices are associated with alternative GMM estimators. By relating estimators to their corresponding selection matrices, we have a convenient device for studying simultaneously an entire family of GMM estimators. Specifically, we explore the consequence of using alternative subsets of moment equations or more generally alternative linear combinations of the moment equation system. This approach builds on an approach of Sargan (1958, 1959) and is most useful for characterizing limiting distributions. The aim is to study simultaneously the behavior of a family of estimators. When the matrix A is replaced by a consistent estimator, the asymptotic properties of the estimator are preserved. This option expands considerably the range of applicability, and, as we will see, is important for implementation.

Since alternative choices of A may give rise to alternative GMM estimators, index alternative estimators by the choice of A. In what follows, replacing A by a consistent estimator does not alter the limiting distribution. For instance, the first-order conditions from minimizing a quadratic form can be representing using a selection matrix that converges to a limiting matrix A. Let

$$D = E\left[\frac{\partial f(x_t, \beta_o)}{\partial \beta}\right].$$

Two results are central to the study of GMM estimators:

$$\sqrt{N}\left(b_N - \beta_o\right) \approx -(AD)^{-1}A\sqrt{N}g_N(\beta_0) \tag{5}$$

and

$$\frac{1}{\sqrt{N}}g_N(b_N) \approx \left[I - D(AD)^{-1}D\right]\sqrt{N}g_N(\beta_0). \tag{6}$$

Both approximation results are expressed in terms of $\sqrt{N}g_N(\beta_0)$, which obeys a Central Limit Theorem, see (2). These approximation results are obtained by standard local methods. They require the square matrix AD to be nonsingular. Thus for there to exist a valid selection matrix, D must have full column rank k. Notice from (6) that the sample moment conditions evaluated at b_N have a degenerate distribution. Premultiplying by A makes the right-hand side zero. This is to be expected because linear combinations of the sample moment conditions are set to zero in estimation.

In addition to assess the accuracy of the estimator (approximation (5)) and to validate the moment conditions (approximation (6)), Newey and West (1987a) and Eichenbaum et al. (1988) show how to use these and related approximations to devise tests of parameter restrictions.[2]

---

[2]Their tests imitate the construction of the likelihood ratio, Lagrange multiplier and the Wald tests familiar from likelihood inference methods.

Next we derive a sharp lower bound on the asymptotic distribution of a family of GMM estimators indexed by the selection matrix A. For a given A, the asymptotic covariance matrix for a GMM estimator constructed using this selection is:

$$cov(A) = (AD)^{-1}AVA'(D'A')^{-1},$$

A selection matrix in effect over-parameterizes a GMM estimator, as can be seen from this formula. Two such estimators with selection matrices of the form A and BA for a nonsingular matrix B imply

$$cov(BA) = cov(A)$$

because the same linear combinations of moment conditions are being used in estimation. Thus without loss of generality we may assume that $AD = I$. With this restriction we may imitate the proof of the famed Gauss-Markov Theorem to show that

$$D'V^{-1}D \leq cov(A) \tag{7}$$

and that the lower bound on left is attained by any $\tilde{A}$ such that $\tilde{A} = BD'V^{-1}$ for some nonsingular B. The quadratic form version of a GMM estimator typically satisfies this restriction when W_N is a consistent estimator of V^{-1}. This follows from the first-order conditions of the minimization problem.

To explore further the implications of this choice, factor the inverse covariance matrix V^{-1} as $V^{-1} = \Lambda'\Lambda$ and form $\Delta = \Lambda D$. Then

$$V^{-1}D(D'V^{-1}D)^{-1}D'V^{-1} = \Lambda'[\Delta(\Delta'\Delta)^{-1}\Delta']\Lambda$$

The matrices $\Delta(\Delta'\Delta)^{-1}\Delta'$ and $I - \Delta(\Delta'\Delta)^{-1}\Delta'$ are each idempotent and

$$\begin{bmatrix} [I - \Delta(\Delta'\Delta)^{-1}\Delta'] \\ \Delta(\Delta'\Delta)^{-1}\Delta' \end{bmatrix} \sqrt{N}\Lambda g_N(\beta_0) \longrightarrow \text{Normal}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I - \Delta(\Delta'\Delta)^{-1}\Delta' & 0 \\ 0 & \Delta(\Delta'\Delta)^{-1}\Delta' \end{bmatrix} \right).$$

The first coordinate block is an approximation for $\sqrt{N}\Lambda g_N(b_N)$ and the sum of the two coordinate blocks is $\sqrt{N}\Lambda g_N(\beta_o)$. Thus we may decompose the quadratic form

$$N[g_N(\beta_o)]'V^{-1}g_N(\beta_o) \approx N[g_N(b_N)]'V^{-1}g_N(b_N) + N[g_N(\beta_o)]'V^{-1}D(D'V^{-1}D)^{-1}D'V^{-1}g_N(\beta_o). \tag{8}$$

where the two terms on the right-hand side are distributed as independent chi-square. The first has r degrees of freedom and the second one has $r - k$ degrees of freedom.

3 Implementation using the Objective Function Curvature

While the formulas just produced can be used directly using consistent estimators of V and D in conjunction with the relevant normal distributions, looking directly at the curvature of

the GMM objective function based on a quadratic form is also revealing. Approximations (5) and (6) give guidance on how to do this.

For a parameter vector β let $V_N(\beta)$ denote an estimator of the long run covariance matrix. Given an initial consistent estimator b_N, suppose that $V_N(b_N)$ is a consistent estimator of V and

$$D_N = \frac{1}{N} \sum_{t=1}^{N} \frac{\partial f(x_t, b_N)}{\partial \beta}.$$

Then use of the selection $A_N = D_N'[V_N(b_N)]^{-1}$ attains the efficiency bound for GMM estimators. This is the so-called two step approach to GMM estimation. Repeating this procedure, we obtain the so-called iterative estimator.[3] In the remainder of this section we focus on a third approach resulting in what we call the continuous-updating estimator. This is obtained by solving:

$$\min_{\beta \in \mathbb{P}} L_N(\beta)$$

where

$$L_N(\beta) = N[g_N(\beta)]'[V_N(\beta)]^{-1}g_N(\beta).$$

Let b_N denote the minimized value. Here the weighting matrix varies with β.

Consider three alternative methods of inference that look at the global properties of the GMM objective $L_N(\beta)$:

a) $\{\beta \in \mathbb{P} : L_N(\beta) \leq C\}$ where C is a critical value from a $\chi^2(r)$ distribution.

b) $\{\beta \in \mathbb{P} : L_N(\beta) - L_N(b_N) \leq C\}$ where C is critical value from a $\chi^2(k)$ distribution.

c) Choose a *prior* π. Mechanically, treat $-\frac{1}{2}L_N(\beta)$ as a log-likelihood and compute

$$\frac{\exp\left[-\frac{1}{2}L_N(\beta)\right]\pi(\beta)}{\int \exp\left[-\frac{1}{2}L_N(\beta)\right]\pi(\tilde{\beta})d\tilde{\beta}}.$$

Method a) is based on the left-hand side of (8). It was suggested and studied in Hansen et al. (1995) and Stock and Wright (2000). As emphasized by Stock and Wright, it avoids using a local identification condition (a condition that the matrix D have full column rank). On the other hand, it combines evidence about the parameter as reflected by the curvature of the objective with overall evidence about the model. A misspecified model will be reflected as an empty confidence interval.

Method b) is based on the second term on right-hand side of (8). By translating the objective function, evidence against the model is netted out. Of course it remains important to consider such evidence because parameter inference may be hard to interpret for a misspecified model. The advantage of b) is that the degrees of freedom of the chi-square distribution are reduced from r to k. Extensions of this approach to accommodate nuisance parameters

---

[3]There is no general argument that repeated iteration will converge.

were used by Hansen and Singleton (1996) and Hansen et al. (1995). The decomposition on the right-hand side of (8) presumes that the parameter is identified locally in the sense that D has full column rank, guaranteeing that the $D'V^{-1}D$ is nonsingular. Kleibergen (2005) constructs an alternative decomposition based on a weaker notion of identification that can be used in making statistical inferences.

Method c) was suggested by Chernozhukov and Hong (2003). It requires an integrability condition which will be satisfied by specifying a uniform distribution π over a compact parameter space. The resulting histograms can be sensitive to this choice of this set or more generally to the choice of π. All three methods explore the global shape of the objective function when making inferences.[4]

4 Backing off from Efficiency

In what follows we give two types of applications that are not based on efficient GMM estimation.

4.1 Calibration-Verification

An efficient GMM estimator selects the best linear combination among a set of moment restrictions. Implicitly a test of the over-identifying momment conditions examines whatever moment conditions are not used in estimation. This complicates the interpretation of the resulting outcome. Suppose instead there is one set of moment conditions for which we have more confidence and are willing to impose for the purposes and calibration or estimation. The remaining set of moment conditions are used for the purposes of verification or testing. The decision to use only a subset of the available moment conditions for purposes of estimation implies a corresponding loss in efficiency. See Christiano and Eichenbaum (1992) and Hansen and Heckman (1996) for a discussion of such methods for testing macroeconomic models.

To consider this estimation problem formally, partition the function f as:

$$f(x, \beta) = \begin{bmatrix} f^{[1]}(x, \beta) \\ f^{[2]}(x, \beta) \end{bmatrix}$$

where $f^{[1]}$ has r_1 coordinates and $f^{[2]}$ has $r - r_1$ coordinates. Suppose that $r_1 \geq k$ and that β is estimated using an A matrix of the form:

$$A = \begin{bmatrix} A_1 & 0 \end{bmatrix}.$$

and hence identification is based only on

$$A_1 E f^{[1]}(x_t, \beta) = 0.$$

This is the so-called calibration step. Let b_N be the resulting estimator.

---

[4]The large sample justification remains local, however.

To verify or test the model we check whether $g_N^{[2]}(b_N)$ is close to zero as predicted by the moment implication:

$$Ef^{[2]}(x_t, \beta_0) = 0.$$

Partition the matrix D of expected partial derivatives as:

$$D = \begin{bmatrix} D_1 \\ D_2 \end{bmatrix}$$

where D_1 is r_1 by k and D_2 is $r - r_1$ by k. Here we use limit approximation (6) to conclude that

$$\sqrt{N} g_N^{[2]}(b_N) \approx \begin{bmatrix} -D_2(A_1 D_1)^{-1} A_1 & I \end{bmatrix} \sqrt{N} g_N(\beta_0),$$

which has a limiting normal distribution. A chi-square test can be constructed by building a corresponding quadratic form of $r - r_1$ asymptotically independent standard normally distributed random variables.[5]

4.2 Sequential Estimation

Sequential estimation methods have a variety of econometric applications. For models of sample selection see Heckman (1976) and related methods with generated regressors see Pagan (1984). For testing asset pricing models see Cochrane (2001) (chapters 12 and 13).

To formulate this problem in a GMM setting, partition the parameter vector as $\beta = \begin{bmatrix} \beta^{[1]} \\ \beta^{[2]} \end{bmatrix}$ where $\beta^{[1]}$ has k_1 coordinates. Partition the function f as:

$$f(x, \beta) = \begin{bmatrix} f^{[1]}\left(x, \beta^{[1]}\right) \\ f^{[2]}\left(x, \beta\right) \end{bmatrix}$$

where $f^{[1]}$ has r_1 coordinates and $f^{[2]}$ has $r - r_1$ coordinates. Notice that the first coordinate block only depends on the first component of the parameter vector. Thus the matrix d is block lower triangular:

$$D = \begin{bmatrix} D_{11} & 0 \\ D_{21} & D_{22} \end{bmatrix}$$

where

$$D_{ij} = E\left[\frac{\partial f^{[i]}\left(x_t, \beta_o\right)}{\partial \beta^{[j]}} \right].$$

A sequential estimation approach exploits the triangular structure of the moment conditions as we now describe. The parameter $\beta_o^{[1]}$ is estimable from the first partition of moment conditions. Given such an estimator, $b_N^{[1]}$, $\beta_o^{[2]}$ is estimable from the second partition of

---

[5]When r_1 exceeds k it is possible to improve the asymptotic power by exploiting the long-run covariation between $f^{[2]}(x_t, \beta_o)$ and linear combination of $f^{[1]}(x_t, \beta_o)$ not used in estimation. This can be seen formally by introducing a new parameter $\gamma_o = E\left[f^{[2]}(x_t, \beta)\right]$ and using the GMM formulas for efficient estimation of β_o and γ_o.

moment conditions. Estimation error in the first stage alters the accuracy of the second stage estimation as I now illustrate.

Assume now that $r_1 \geq k_1$. Consider a selection matrix that is block diagonal:

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}$$

where A_{11} has dimension k_1 by r_1 and A_{22} has dimension $k - k_1$ by $r - r_1$. It is now possible to estimate $\beta_o^{[1]}$ using the equation system:

$$A_{11} g_N^{[1]}(\beta^{[1]}) = 0$$

or a method that is asymptotically equivalent to this. Let $b_N^{[1]}$ be the solution. This initial estimation may be done for simplicity or because these moment conditions are embraced with more confidence. Given this estimation of $\beta_o^{[1]}$, we seek an estimator $b_N^{[2]}$ of $\beta_0^{[2]}$ by solving:

$$A_{22} g_N^{[2]} \left(b_N^{[1]}, \beta^{[2]} \right) = 0.$$

To proceed, we use this partitioning and apply (5) to obtain the limiting distribution for the estimator $b_N^{[2]}$. Straightforward matrix calculations yield,

$$\sqrt{N} \left(b_N^{[2]} - \beta_o^{[2]} \right) \approx - \left(A_{22} D_{22} \right)^{-1} A_{22} \begin{bmatrix} -D_{21}(A_{11} D_{11})^{-1} A_{11} & I \end{bmatrix} \sqrt{N} g_N(\beta_0). \qquad (9)$$

This formula captures explicitly the impact of the initial estimation of $\beta_o^{[1]}$ on the subsequent estimation of $\beta_0^{[2]}$. When D_{21} is zero an adjustment is unnecessary.

Consider next a (second best) efficient choice of selection matrix A_{22}. Formula (9) looks just like formula (5) with A_{22} replacing A, D_{22} replacing D and a particular linear combination of $g_N(\beta_0)$. The matrix used in this linear combination "corrects" for the estimation error associated with the use of an estimator $b_N^{[1]}$ instead of the unknown true value $\beta_o^{[1]}$. By imitating our previous construction of an asymptotically efficient estimator, we construct the (constrained) efficient choice of A_{22} given A_{11}:

$$A_{22} = B_{22}(D_{22})' \left(\begin{bmatrix} -D_{21} \left(A_{11} D_{11} \right)^{-1} A_{11} & I \end{bmatrix} V \begin{bmatrix} - \left[D_{21} \left(A_{11} D_{11} \right)^{-1} A_{11} \right]' \\ I \end{bmatrix} \right)^{-1}$$

for some nonsingular matrix B_{22}. An efficient estimator can be implemented in the second stage by solving:

$$\min_{\beta^{[2]}} g_N^{[2]} \left(b_N^{[1]}, \beta^{[2]} \right)' W_N g_N^{[2]} \left(b_N^{[1]}, \beta^{[2]} \right)$$

for $V_N^{[2]}$ given by a consistent estimator of

$$V^{[2]} = \left(\begin{bmatrix} -D_{21} \left(A_{11} D_{11} \right)^{-1} A_{11} & I \end{bmatrix} V \begin{bmatrix} - \left[D_{21} \left(A_{11} D_{11} \right)^{-1} A_{11} \right]' \\ I \end{bmatrix} \right)^{-1}$$

10

or by some other method that selects (at least asymptotically) the same set of moment conditions to use in estimation. Thus we have a method that adjusts for the initial estimation of $\beta^{[1]}$ while making efficient use of the moment conditions $E f^{[2]}(x_t, \beta) = 0$.

As an aside, notice the following. Given an estimate $b_N^{[1]}$, the criterion based methods of statistical inference described in section 3 can be adapted to making inferences in this second stage in a straightforward manner.

5 Conditional Moment Conditions

The bound (7) presumes a finite number of moment conditions and characterizes how to use these conditions efficiently. If we start from the conditional moment restriction:

$$E\left[f(x_{t+\ell}, \beta_0)|\mathcal{F}_t\right] = 0$$

then in fact there are many moment conditions at our disposal. Functions of variables in the conditioning information set can be used to extend the number of moment conditions. By allowing for these conditions, we can improve upon the asymptotic efficiency bound for GMM estimation. Analogous conditional moment restrictions arise in cross-sectional settings.

For a characterizations and implementations appropriate for cross sectional data see Chamberlain (1986) and Newey (1993), and for characterizations and implementations in a time series settings see Hansen (1985), Hansen (1993), and West (2001). The characterizations are conceptually interesting but reliable implementation is more challenging. A related GMM estimation problem is posed and studied by Carrasco and Florens (2000) in which there is a pre-specified continuum of moment conditions that are available for estimation.

6 Conclusion

GMM methods of estimation and inference are adaptable to a wide array of problems in economics. They are complementary to maximum likelihood methods and their Bayesian counterparts. Their large sample properties are easy to characterize. While their computational simplicity is sometimes a virtue, perhaps their most compelling use is in the estimation of partially specified models or of misspecified dynamic models designed to match a limited array of empirical targets.

References

Amemiya, T. 1974. The Nonlinear Two-stage Least-squares Estimator. *Journal of Econometrics* 2:105–110.

Arellano, M. 2003. *Panel Data Econometrics.* New York: Oxford University Press.

Bontemps, C. and N. Meddahi. 2005. Testing Normality: A GMM Approach. *Journal of Econometrics* 124:149–186.

Carrasco, M. and J. P. Florens. 2000. Generalization of GMM to a Continuum of Moment Conditions. *Econometric Theory* 20:797–834.

Chamberlain, G. 1986. Asymptotic Efficiency in Estimation with Conditional Moment Restrictions. *Journal of Econometrics* 34:305–334.

Chernozhukov, V. and H. Hong. 2003. An MCMC Approach to Classical Estimation. *Journal of Econometrics* 115:293–346.

Christiano, L. J. and M. Eichenbaum. 1992. Current Real Business Cycle Theories and Aggregate Labor Market Fluctuations. *American Economic Review* 82:430–450.

Cochrane, John. 2001. *Asset Pricing.* Princeton University Press.

Cumby, R. E., J. Huizinga, and M. Obstfeld. 1983. Two-step Two-stage Least Squares Estimation in Models with Rational Expectations. *Journal of Econometrics* 21:333–335.

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

Ghysels, E. and A. Hall, eds. 2002. *Journal of Business and Economic Statistics*, vol. 20.

Gordin, M. I. 1969. The Central Limit Theorem for Stationary Processes. *Soviet Mathematics Doklady* 10:1174 – 1176.

Hall, A. R. 2005. *Generalized Method of Moments.* New York: Oxford University Press.

Hall, P. and C. C. Heyde. 1980. *Martingale Limit Theory and Its Application.* Boston: Academic Press.

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

———. 1985. A Method for Calculating Bound on Asymptotic Covariance Matrices of Generalized Method of Moments Estimators. *Journal of Econometrics* 30:203–238.

———. 1993. *Models, Methods and Applications of Econometrics: Essays in Honor of A. R. Bergstrom*, chap. Semiparametric Efficiency Bounds for Linear Time-Series Models, 253–271. Cambridge, MA: Blackwell.

———. 2001. *International Encyclopedia of the Social and Behavior Sciences*, chap. Method of Moments, 9743–9751. New York: Elsevier.

Hansen, L. P. and J. J. Heckman. 1996. The Empirical Foundations of Calibration. *Journal of Economic Perspectives* 10:87–104.

Hansen, L. P. and R. Jagannathan. 1997. Assessing Specification Errors in Stochastic Discount Factor Models. *Journal of Finance* 52:557–590.

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

———. 1996. Efficient Estimation of Linear Asset Pricing Models with Moving Average Errors. *Journal of Business and Economic Statistics* 14:53–68.

Hansen, L. P., J. Heaton, and E. Luttmer. 1995. Econometric Evaluation of Asset Pricing Models. *Review of Financial Studies* 8:237–274.

Hansen, L. P., J. C. Heaton, J. Lee, and N. Roussanov. 2007. Intertemporal Substitution and Risk Aversion. In *Handbook of Econonometrics*, vol. 6A, edited by J. Heckman and E. Leamer. New York: Elsevier.

Hayashi, F. and C. Sims. 1983. Nearly Efficient Estimation of Time-Series Models with Predetermined, but Not Exogenous, Instruments. *Econometrica* 51:783 – 798.

Heckman, J. J. 1976. The Common Structure of Statistical Methods of Trucation, Sample Selection, and Limited Dependent Variables and a Simple Estimator of Such Models. *Annals of Economic and Social Measurement* 5:475–492.

Kleibergen, F. 2005. Testing Parameters in GMM without Assuming that they are Identified. *Econometrica* 73:1103–1123.

Newey, W. 1993. Efficient Estimation of Models with Conditional Moment Restrictions. In *Handbook of Statistics*, vol. 11, edited by G. S. Maddala, C. R. Rao, and H. D. Vinod. Amsterdam: North Holland.

Newey, W. and D. McFadden. 1994. *Handbook of Econometrics*, vol. 4, chap. Large Sample Estimation and Hypothesis Testing, 2113–2148. Amsterdam: Elsevier.

Newey, W. K. and K. D. West. 1987a. Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review* 28:777–787.

———. 1987b. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55:703–708.

Ogaki, M. 1993. *Handbook of Statistics*, vol. 11, chap. Generalized Method of Moments: Econometric Applications, 455–486. Elsevier Science Publishers.

Pagan, A. R. 1984. Econometric Issues in the Analysis of Models with Generated Regressors. *International Economic Review* 25:221–247.

Sargan, J. 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: Series B* 21:91–105.

Singleton, K. J. 2006. *Empirical Dynamic Asset Pricing: Model Specification and Econometric Assessment.* Princeton University Press.

Stock, J. H. and J. H. Wright. 2000. GMM with Weak Identification. *Econometrica* 68:1055–1096.

West, K. D. 2001. On Optimal Instrumental Variables Estimation of Stationary Time Series Models. *International Economic Review* 42:1043–1050.