G. S. Maddala, C. R. Rao and H. D. Vinod, eds., *Handbook of Statistics, Vol. 11*
© 1993 Elsevier Science Publishers B.V. All rights reserved.

Some Aspects of Generalized Method of Moments Estimation

Alastair Hall

1. Introduction

Since its introduction by Hansen (1982), generalized method of moments (GMM) estimation has had a considerable impact on econometrics. It provides a unifying framework for the analysis of many familiar estimators and includes least squares, instrumental variables and maximum likelihood as special cases. It also offers a convenient method of estimation in certain models which were computationally very burdensome to estimate by more traditional methods. GMM has been applied to a wide variety of models including the estimation of probit models with panel data (Avery, Hansen and Hotz, 1983) and nonlinear rational expectations models (e.g., Hansen and Singleton, 1982).

Due to the generality of the GMM estimation principle, many theoretical treatments of the estimator are at an advanced level. While this generality is very desirable for practical purposes, it can make it more difficult for the less technical reader to understand the intuition behind the estimation procedure. In this paper we provide an introduction to GMM which is designed to acquaint the applied researcher with the basic ideas behind GMM and its statistical properties. The emphasis is placed on intuition and not mathematical rigor. There are many other excellent sources which provide a rigorous treatment of GMM, e.g., see Hansen (1982); Gallant (1987); Gallant and White (1988).

Many researchers are aware that there is a close relationship between GMM and instrumental variables (IV) estimation, but are uncertain about the exact nature of the connection. This stems from the fact that GMM is most frequently applied to the estimation of nonlinear dynamic rational expectations models using a generalized IV procedure introduced by Hansen and Singleton (1982). In fact, as Hansen (1982) and Hansen and Singleton (1982) observe, IV is a special case of GMM. Although GMM can be applied to nonlinear dynamic models, many of the statistical issues which arise in this framework can be conveniently illustrated using the IV estimator of the familiar fixed regressor linear regression model. Sargan (1958, 1959) presents the first rigorous treatment of IV in linear systems, and although he concentrated on

the linear model, his work was the first to address and solve many of these issues. In fact parts of Hansen's analysis can be viewed as direct extensions of Sargan's work to a more general class of models (see Hansen, 1982, for further discussion). Although, it should be noted that Hansen (1982) was the first paper to develop the complete framework we describe and our discussion of IV relies heavily on his work.

In this paper we use the IV estimator in the fixed linear regression model under the classical assumptions to introduce the general principle behind GMM. In many textbooks the IV estimator is derived in an ad hoc manner (e.g., see Johnston 1984, pp. 363–366). However, we consider an alternative derivation in which the estimator is obtained by minimizing a quadratic form in a vector of sample moments which are functions of the parameters β and the data (e.g., see White, 1984a, pp. 7–9). It becomes immediately apparent from this framework that IV has desirable asymptotic statistical properties provided the analogous population moments equal zero when evaluated at the true value of β, β_0. In this particular application the population moment condition is that the instrument vector is orthogonal to the errors, and so our analysis delivers the familiar condition that IV is consistent provided the instruments are valid. This is the essential structure of GMM. To estimate β all one needs are a set of population moments[1] whose expectation is zero when evaluated at β_0. The GMM estimate is obtained by minimizing a quadratic form in the analogous sample moments.

The linear model also provides a convenient framework for introducing various statistical issues which arise in the GMM inference framework. We examine (i) the consistency of the estimator; (ii) asymptotic normality of the estimator; (iii) the optimal choice of weighting matrix in the quadratic form; (iv) the overidentifying restrictions test. It is hoped that the ideas behind GMM are more readily accessible when first encountered within the context of the linear model because its algebra is familiar to most researchers.

We then show how these ideas are extended to nonlinear dynamic models to produce a very general principle of estimation. Our discussion inevitably relies heavily on Hansen (1982) but we place particular emphasis on showing that the structure of the arguments is identical to those employed for the analysis of IV in the linear model.

To illustrate the power of the GMM estimation principle, we consider its application to the estimation of Euler equation models. Hansen and Singleton (1982) showed how one can use the GMM estimation principle to construct generalized IV estimators of the parameters of these models. Following Hansen and Singleton (1982) we discuss this technique in the context of a consumption based asset pricing model. In comparison to maximum likelihood estimation of these models, GMM is seen to have considerable computational and statistical advantages.

[1] There must be at least as many moment conditions as parameters. This is discussed in detail below.

The original papers of Hansen (1982) and Hansen and Singleton (1982) have had such an impact on econometrics that they have prompted a number of studies both analyzing the properties of GMM and also extending GMM based inference. In this paper we offer a selective review of this recent research, and this discussion helps to illuminate both the strengths and weaknesses of the GMM framework.

An outline of the paper is as follows. In Section 2 we consider IV estimation in the classical linear regression model. In Section 3 we introduce Hansen's (1982) GMM estimator for nonlinear dynamic models. In Section 4 we discuss the application of this technique by Hansen and Singleton (1982) to the estimation of Euler equation models. In Section 5 we summarize recent research on GMM based estimation and inference. Finally in Section 6 we offer some concluding remarks.

2. Instrumental variables estimation in the linear model

Consider the classical linear regression model

$$y = X\beta_0 + u\,, \tag{2.1}$$

where y is an $(n \times 1)$ vector of observations on the dependent variable, X is an $(n \times q)$ matrix of fixed in repeated sample regressors with $\operatorname{rank}(X) = q$; u is an $(n \times 1)$ vector of observations on the error process with $\mathrm{E}(u) = 0$, $\operatorname{var}(u) = \sigma^2 I_n$; and β_0 is the true value of β, an $(q \times 1)$ vector of unknown parameters. Let Z be an $(n \times r)$ matrix of fixed in repeated sample instruments which satisfy $\mathrm{E}(z_t u_t) = 0$ for any t, where z_t is the t-th row of Z. We define $Q_n(\beta)$ to be the following quadratic form:

$$Q_n(\beta) = [n^{-1}u(\beta)'Z]W_n[n^{-1}Z'u(\beta)]\,, \tag{2.2}$$

where W_n is a positive definite symmetric matrix which converges in probability to a positive definite symmetric matrix W and $u(\beta) = y - X\beta$. For the present we assume $r = q$. Let $\tilde{\beta}$ be the value of β which minimizes $Q_n(\beta)$. If we substitute $u(\beta) = y - X\beta$ into (2.2), multiply out the quadratic form and differentiate with respect to β then the first-order conditions are

$$-2X'ZW_nZ'y + 2X'ZW_nZ'X\tilde{\beta} = 0$$

which in turn implies

$$X'ZW_nZ'y = X'ZW_nZ'X\tilde{\beta}\,. \tag{2.3}$$

Therefore if $Z'X$ is nonsingular the solution for $\tilde{\beta}$ is

$$\tilde{\beta} = (Z'X)^{-1}Z'y\,.$$

Within this framework it becomes immediately obvious why IV estimators are consistent. A heuristic explanation is as follows. The population moment

condition $E[z_t u_t(\beta_0)] = 0$ (combined with the Weak Law of Large Numbers) implies $n^{-1}Z'u(\beta_0) \xrightarrow{p} 0$. Now W_n is positive definite for all n and so $Q_n(\beta) \geqslant 0$ with the equality only obtained for $n^{-1}Z'u(\beta) = 0$. Therefore if $\tilde{\beta}$ minimizes $Q_n(\beta)$ and $n^{-1}Z'u(\beta_0) \xrightarrow{p} 0$, it follows that $\tilde{\beta} \xrightarrow{p} \beta_0$.[2] Of course if $\mathrm{E}[z_t u_t(\beta_0)] \neq 0$ then this argument breaks down and in general $\tilde{\beta}$ is not consistent. This corresponds to the case where at least some elements of z_t are invalid instruments. It is interesting to compare this derivation with the more familiar textbook treatment of IV. Note that because $X'Z$ and W_n are nonsingular, (2.3) implies $Z'u(\tilde{\beta}) = 0$, and these equations are the starting point for many textbook derivations of $\tilde{\beta}$ (e.g., see Johnston, 1984, p. 364).

It is important to emphasize the structure of our derivation of $\tilde{\beta}$. The estimator can be obtained by minimizing a quadratic loss function in a sample moment (i.e., $n^{-1}Z'u(\beta)$). The resulting estimators are consistent because the population moment equals zero when evaluated at β_0 (i.e., $\mathrm{E}[z_t u_t(\beta_0)] = 0$). This is the basic principle that lies behind the generalized method of moments (GMM) estimator. However, before considering GMM in its full generality it is instructive to consider further the instrumental variables estimator in the basic linear model.

In our earlier discussion it was assumed that $r = q$ and so there were exactly the same number of moment conditions as parameters to be estimated. One can think of this as being the case where β is 'just identified' because we have just enough information (i.e., moment conditions) to estimate β. An immediate consequence of this was that $\tilde{\beta}$ does not depend on W_n and so the same estimator is obtained regardless of the choice of W_n. We now consider the statistical implications of $r \neq q$.

If $r < q$ then the number of parameters to be estimated exceeds the number of moment conditions. In this case β is 'underidentified' because there is insufficient information (i.e., moment conditions) from which to estimate β uniquely. The previous derivation of $\tilde{\beta}$ breaks down because if $r < q$ then the $(q \times q)$ matrix $X'ZW_nZ'X$ has rank less than or equal to r and so is singular. Consequently (2.3) cannot be uniquely solved for $\tilde{\beta}$.

If $r > q$ then the number of moment conditions exceeds the number of parameters to be estimated, and so β is 'overidentified' because there is more information (i.e., moment conditions) than is necessary to obtain an estimate of β. In this case the choice of W_n affects $\tilde{\beta}$. To see this we rewrite (2.3) as

$$X'ZW_nZ'u(\tilde{\beta}) = 0\,. \tag{2.4}$$

Now $X'ZW_n$ is $q \times r$ and so (2.4) does not imply $Z'u(\tilde{\beta}) = 0$. Rather $\tilde{\beta}$ is chosen to be the value which sets q linear combinations of the r moment conditions $Z'u(\beta)$ equal to zero, and the weights of these linear combinations

[2] We must also assume $\mathrm{E}[z_t u_t(\bar{\beta})] = \mu \neq 0$ for any $\bar{\beta} \neq \beta_0$, to ensure $\tilde{\beta}$ is uniquely defined asymptotically.

are given by $X'ZW_n$. In this case the solution for $\tilde{\beta}$ is

$$\tilde{\beta} = (X'ZW_nZ'X)^{-1}X'ZW_nZ'y \tag{2.5}$$

which clearly depends on W_n. This naturally raises the question of which W_n leads to the most efficient choice of $\tilde{\beta}$. Recall that we have made no assumptions about the distribution of u_t and so the finite sample distribution of $\tilde{\beta}$ is unknown. Therefore our efficiency comparison is based on asymptotic behavior and so we must first characterize the asymptotic distribution of $\tilde{\beta}$.

In our discussion of the asymptotic distribution of $\tilde{\beta}$ we set $r \geqslant q$ and so include the 'just identified' case as well. Note that if $r = q$ then (2.5) reduces to $\tilde{\beta} = (Z'X)^{-1}Z'y$, and so (2.5) trivially holds for $r \geqslant q$. If we substitute for y from (2.1) into (2.5) and rearrange the equation we can obtain

$$n^{1/2}(\tilde{\beta} - \beta_0) = (n^{-1}X'ZW_nn^{-1}Z'X)^{-1}n^{-1}X'ZW_nn^{-1/2}Z'u\,. \tag{2.6}$$

We deduce the asymptotic distribution of $\tilde{\beta}$ by considering the behavior of the cross product matrices on the right-hand side of (2.6). Under suitable regularity conditions we can assume

(C1): $\lim_{n\to\infty} n^{-1}Z'Z = M_{zz}$, a finite nonsingular matrix of constants;

(C2): $\lim_{n\to\infty} n^{-1}X'Z = M_{xz}$, a finite matrix of constants with rank q;

(C3): $\lim_{n\to\infty} n^{-1/2}Z'u \xrightarrow{d} N(0, \sigma^2 M_{zz})$.

Each of these conditions places different restrictions on the asymptotic behavior of the instruments. (C1) requires that each instrument provides some unique information in the sense that no one instrument z_{ti} can be written as a linear combination of the others infinitely often as $n \to \infty$. Condition (C2) requires that there must be at least q of the r instruments which are correlated with x_t asymptotically, and so β is identified asymptotically. Finally (C3) requires that the sample moment vector evaluated at β_0 converges to a normal distribution when scaled by $n^{1/2}$. Note that the mean of this distribution is zero and so this condition restates the requirement that z_t be a valid instrument. Using (C1)–(C3) we can deduce from (2.6) that

$$n^{1/2}(\tilde{\beta} - \beta_0) \xrightarrow{d} \mathrm{N}(0, \sigma^2 V)\,, \tag{2.7}$$

where $V = (M_{xz}WM_{zx})^{-1}M_{xz}WM_{zz}WM_{zx}(M_{xz}WM_{zx})^{-1'}$ and $M_{zx} = M'_{xz}$. Therefore we can perform asymptotic inference about β_0 based on (2.7) using either confidence intervals or using hypothesis tests (the latter will be discussed in more detail in Section 5). A consistent estimator of $\sigma^2 V$ is easily constructed from the data using W_n, the sample moment matrices such as $n^{-1}Z'X$ and $\tilde{\sigma}^2 = (y - X\tilde{\beta})'(y - X\tilde{\beta})/n$. In the just identified case, V reduces to $M_{zx}^{-1}M_{zz}M_{xz}^{-1}$.

We now return to the issue of characterizing the optimal choice of W_n. This characterization is based on finding the choice of W, the probability limit of W_n, which yields the most efficient IV estimator asymptotically. From Hansen (1982) Theorem 3.2 it follows that for this model the optimal choice of W is

M_{zz}^{-1}, in which case V reduces to

$$V^* = [M_{zx}M_{zz}^{-1}M_{zx}]^{-1} . \tag{2.8}$$

Note that the optimal choice of W is proportional to the inverse of asymptotic covariance matrix of $n^{-1/2}Z'u$. One can gain some intuition into this choice of W by considering the role of the weighting matrix in $Q_n(\beta)$. The elements of W_n control the relative importance of each term of the form $(n^{-1}\Sigma z_{ti}u_t)\times(n^{-1}\Sigma z_{tj}u_t)$ in $Q_n(\beta)$. Due to the inherent variability in z_t each element of the moment condition contains different information about β and these elements (or pieces of information) are related to each other. This is reflected in the asymptotic covariance matrix $n^{-1/2}Z'u$ which from (C3) is $\sigma^2 M_{zz}$. The optimal weighting matrix is the one which takes account of this covariance structure in the construction of $Q_n(\beta)$. An obvious choice of W_n which converges in probability to M_{zz}^{-1} is $(n^{-1}Z'Z)^{-1}$, and from (2.5) the associated sequence of IV estimators is characterized by

$$\begin{aligned}\tilde{\beta}^* &= [(X'Z(Z'Z)^{-1}Z'X)]^{-1}X'Z(Z'Z)^{-1}Z'y \\ &= (\hat{X}'\hat{X})^{-1}\hat{X}'y ,\end{aligned} \tag{2.9}$$

where $\hat{X} = Z(Z'Z)^{-1}Z'X$ namely, the predicted value of X from regression on Z.

It was noted that the asymptotic normality of $\tilde{\beta}$ facilitates inference about β. However, such inference is based on the assumption that the model is correctly specified. Typically it is desired to test whether the data is consistent with the model before focusing on inference within the context of that model. For the regression model we have discussed above, many tests of model adequacy have been proposed, e.g., tests for serial correlation or heteroscedasticity in u_t. However, in the more general models examined below, we typically do not assume $\mathrm{E}(uu') = \sigma^2 I_n$ and so these tests do not usefully extend. Instead we focus on a test principle which can be applied in nonlinear dynamic models. This test is often referred to as Hansen's (1982) overidentifying restrictions test. It is clear from our earlier discussion that the desirable statistical properties of $\tilde{\beta}$ rely crucially on the validity of the moment condition $\mathrm{E}[z_t u_t(\beta_0)] = 0$. It is therefore important to test whether the data are consistent with this moment condition. Now in the 'just identified' case $Z'u(\tilde{\beta}) = 0$ by construction regardless of whether or not the population moment condition is true and so one cannot derive a useful test from the sample moment. However, if β is overidentified, then from (2.4) it is clear that only q linear combinations of $Z'u(\tilde{\beta})$ are set equal to zero and so $Z'u(\tilde{\beta}) \neq 0$ by construction. However, if the population moment condition is true then one would expect $Z'u(\tilde{\beta}) \approx 0$. This provides a basis for a test of the model specification, and we base this inference on the asymptotically efficient estimator $\tilde{\beta}^*$. From Hansen (1982) Lemma 4.2 it follows that in this model if $\mathrm{E}[z_t u_t(\beta_0)] = 0$, then the statistic

$$\tau_n = n^{-1/2}u(\tilde{\beta}^*)'Z(n^{-1}Z'Z)^{-1}n^{-1/2}Z'u(\tilde{\beta}^*)/\tilde{\sigma}^2$$

converges in distribution to χ^2_{r-q} where $\tilde{\sigma}^2$ is a consistent estimator of σ^2. Note that the degrees of freedom are the number of moment conditions minus the number of parameters to be estimated. This is the case because only $r-q$ of the sample moment conditions $Z'u(\tilde{\beta}^*)$ are free on account of the q restrictions implied by the first-order conditions for $\tilde{\beta}^*$ in (2.4).[3] The statistic τ_n provides a simple test of the model and we will discuss its properties in greater detail in Section 5.

3. Generalized method of moments

In the previous section it was shown that the familiar instrumental variables estimator can be derived by minimizing a quadratic form in a sample moment which was a function of β. Within this framework it became apparent that the desirable properties of the IV estimator depended on three properties of this condition, namely: (i) the validity of the population moment condition $\mathrm{E}[z_t u_t(\beta_0)] = 0$; (ii) there were at least as many moment conditions as parameters to be estimated; (iii) $n^{1/2}$ times the sample moment evaluated at β_0 converged to a normal distribution as $n \to \infty$, i.e., $n^{-1/2}Z'u(\beta_0) \xrightarrow{d} \mathrm{N}(0, \sigma^2 M_{zz})$. It is important to realize that we could have performed a similar analysis with any moment conditions which satisfied (i)–(iii). For example we could have used a different set of instruments, r_t say, and although this estimator would be numerically different from $\tilde{\beta}$ in (2.5), it would have similar statistical properties. It would be consistent and asymptotically normal, the optimal choice of W_n would be $n^{-1}R'R$ and one can perform inference about the model using an analogous statistic to τ_n. Furthermore, the intuition behind our discussion was not specific to a static linear regression model. Rather to estimate the parameters of any model, we can proceed by finding moment conditions involving the parameter vector which satisfy (i)–(iii) and minimize a quadratic form in the analogous sample moment. This is the essence of generalized method of moments estimation.

In this section we discuss the extension of our analysis for IV estimators in static linear models to nonlinear dynamic models. We adopt the statistical framework in Hansen (1982) and delay discussion of its possible extensions until Section 5.

Consider the case where we wish to estimate the $(q \times 1)$ parameter vector β which indexes a statistical model. Suppose we have the following $(r \times 1)$

[3] Note that τ_n is not asymptotically equivalent to the well-known Hausman (1978) test for the endogeneity of regressors. Hausman's test compares $\tilde{\beta}$ with $\hat{\beta}$, the ordinary least squares estimator. Although Hausman's test can be set up using moment conditions (see White, 1982, 1984b) neither of these versions corresponds to τ_n. In particular the asymptotic properties of this Hausman test depend on $n^{-1/2}X'u$ which is irrelevant to τ_n. Newey (1985) discusses the connections between Hausman type and conditional moment tests.

moment conditions[4]

$$\mathrm{E}[f(x_t, \beta_0)] = 0\,, \tag{3.1}$$

where x_t is a $(p \times 1)$ stationary ergodic random vector $f(x, \beta)$ is a $(r \times 1)$ vector of continuous functions of β for every x, and β_0 is the true value of β. Before proceeding with our discussion we wish to address an issue of terminology. Hansen (1982) refers to the moment conditions in (3.1) as 'orthogonality conditions'. This term arises naturally in the IV framework because as we have seen the moment conditions are based on the orthogonality of two sets of variables, i.e., u_t and z_t in Section 2. Although one can use the two terms interchangeably, in this paper we refer to (3.1) as moment conditions.

If $r \geqslant q$ then these moment conditions provide sufficient information to estimate β. If we let $g_n(\beta)$ be the sample moments

$$g_n(\beta) = n^{-1} \sum_{t=1}^{n} f(x_t, \beta)$$

then the intuition from Section 2 leads us to consider the quadratic form

$$Q_n(\beta) = g_n(\beta)' W_n g_n(\beta)\,,$$

where W_n is defined as before. The generalized method of moments estimator of β is $\tilde{\beta}$ the value which minimizes $Q_n(\beta)$ with respect to β. Given our earlier discussion one would anticipate that $\tilde{\beta}$ is consistent and asymptotically normally distributed just like the IV estimator in the linear model. In fact all our analysis of the IV estimator extends very naturally to the GMM estimator and in the remainder of this section we discuss the generalization of those arguments to this more general setting.

The first-order conditions for minimizing $Q_n(\beta)$ imply $\tilde{\beta}$ is the solution to

$$G_n(\tilde{\beta})' W_n g_n(\tilde{\beta}) = 0\,, \tag{3.2}$$

where $G_n(\beta)$ is the $(r \times q)$ matrix with (i, j)-th element $[G_n(\beta)]_{ij} = \partial g_{ni}(\beta)/\partial \beta_j$ and, $g_{ni}(\beta)$ is the i-th element of $g_n(\beta)$. We assume $G_n(\tilde{\beta})$ is of full rank. Just as in the linear case[5] the first-order conditions set q linear combinations of the r sample moment conditions $g_n(\tilde{\beta})$ equal to zero. If β is just identified then

[4] In this section we follow Hansen's (1982) notation to expedite the readers progression from our discussion to the original work. However, this does create a slight notational conflict with Section 2. In this section x_t represents the vector of all variables appearing in the moment condition, which for the linear model are the regressors, the instruments and the dependent variable. With an obvious abuse of notation (3.1) reduces to the moment condition in Section 2 if we set $f(x_t, \beta_0) = z_t(y_t - x_t'\beta_0)$. Note that our assumptions in Section 2 did not include stationarity and so our linear model is not encompassed by Hansen's original framework. However, one can apply GMM under much weaker assumptions about x_t which would include the model of the previous section. These are discussed in Section 5.

[5] Note that in the linear model $G_n(\tilde{\beta}) = n^{-1}X'Z$, and $g_n(\tilde{\beta}) = n^{-1}Z'u(\tilde{\beta})$ and so (3.2) is identical to (2.4).

$G_n(\tilde{\beta})$ and W_n are nonsingular, in which case $\tilde{\beta}$ is the solution to $g_n(\tilde{\beta}) = 0$ and so does not depend on W_n. However, if β is overidentified then $\tilde{\beta}$ does depend on W_n in a similar fashion to the linear model. Notice that unlike the linear case, (3.2) does not imply an explicit solution and so one must use a numerical optimization routine to solve for $\tilde{\beta}$ (e.g., see Gallant, 1987, pp. 26–46).

Most commonly used statistical packages do not include a convenient procedure for GMM estimation of nonlinear dynamic models. However, in the case where the moment condition arises from the orthogonality of some function $u_t(\beta)$ and an instrument vector, then one can use RATS (Doan, 1990, pp. 5-22, 5-23) and TSP (Hall and Cummins, 1991, pp. 67–68). One can also use matrix computer languages to perform GMM estimation. For example the Hansen/Heaton/Ogaki GMM GAUSS package (see Ogaki, 1992) contains suitable programs in GAUSS and Gallant (1987, pp. 448–449) presents a SAS PROC MATRIX (SAS, 1982) program which estimates the parameters of the Hansen and Singleton (1982) consumption-based asset pricing model and also calculates the other statistics described below. The latter provides a useful starting place for the structure of a numerical optimization procedure for more general GMM estimation or if it is desired to use an alternative matrix computer language.

Although we cannot obtain an explicit solution for $\tilde{\beta}$ one can show that it is a consistent estimator. The heuristic argument presented for the IV estimator in the linear model extends to this more general setting. The population moment condition $\mathrm{E}[f(x_t, \beta_0)] = 0$ combined with the Weak Law of Large Numbers implies $g_n(\beta_0) \xrightarrow{p} 0$. Now W_n is positive definite for all n and so $Q_n(\beta) \geqslant 0$ with the equality only obtained for $g_n(\beta) = 0$ for all n. Therefore if $\tilde{\beta}$ minimizes $Q_n(\beta)$ and $g_n(\beta_0) \xrightarrow{p} 0$, it follows that $\tilde{\beta} \xrightarrow{p} \beta_0$.[6]

We now derive the asymptotic distribution of $\tilde{\beta}$. Due to the absence of an explicit solution for $\tilde{\beta}$ from (3.2), it is necessary to develop an asymptotic approximation to $n^{1/2}(\tilde{\beta} - \beta_0)$ using a mean value theorem expansion of the sample moment conditions. By the mean value theorem we have

$$g_n(\tilde{\beta}) = g_n(\beta_0) + G_n(\bar{\beta})(\tilde{\beta} - \beta_0)\,, \tag{3.3}$$

where $\|\tilde{\beta} - \bar{\beta}\| \leqslant \|\tilde{\beta} - \beta_0\|$. If we premultiply both sides of (3.3) by $G_n(\tilde{\beta})'W_n$ then from (3.2) the left-hand side of (3.3) is zero and so we obtain

$$(\tilde{\beta} - \beta_0) = -[G_n(\tilde{\beta})'W_nG_n(\tilde{\beta})]^{-1}G_n(\tilde{\beta})'W_ng_n(\beta_0)\,. \tag{3.4}$$

Now $\tilde{\beta} \xrightarrow{p} \beta_0$ implies $\bar{\beta} \xrightarrow{p} \beta_0$ and so by Slutsky's theorem we obtain

$$n^{1/2}(\tilde{\beta} - \beta_0) = -[G_n(\beta_0)'W_nG_n(\beta_0)]^{-1}G_n(\beta_0)'W_nn^{1/2}g_n(\beta_0) + \mathrm{o_p}(1)\,, \tag{3.5}$$

[6] We must also assume that $\mathrm{E}[f(x_t, \bar{\beta})] = \mu \neq 0$ for $\bar{\beta} \neq \beta_0$ to ensure that $\tilde{\beta}$ is uniquely defined asymptotically.

Under certain regularity conditions (see Hansen, 1982),

$$n^{1/2} g_n(\beta_0) \xrightarrow{d} \mathrm{N}(0, S_w) ,$$

where

$$S_w = \lim_{n\to\infty} E\left[n^{-1} \left(\sum_{t=1}^{n} f_t(\beta_0) \right) \left(\sum_{t=1}^{n} f_t(\beta_0) \right)' \right] ,$$

is a finite positive definite matrix. Therefore from (3.5)

$$n^{1/2}(\tilde{\beta} - \beta_0) \xrightarrow{d} \mathrm{N}(0, V_G) , \tag{3.6}$$

where $V_G = [G_0' W G_0]^{-1} G_0' W S_w W G_0 [G_0' W G_0]^{-1}$, and $G_0 = \mathrm{E}[\partial f(x_t, \beta_0)/\partial\beta]$. Notice that the structure of V_G is identical to V in (2.7).[7]

To perform inference about β_0 based on (3.6), it is necessary to construct a consistent estimator of V_G. This can be obtained by replacing the component matrices of V_G by their consistent estimators. It is easily verified that $G_n(\tilde{\beta}) \xrightarrow{p} G_0$ and by definition $W_n \xrightarrow{p} W$, but some care is needed in constructing a consistent estimator of S_w. There have recently been a number of studies proposing positive definite consistent estimators of covariance matrices like S_w (see inter alia Gallant, 1987, pp. 445 and 532–539; Newey and West, 1987a; Andrews, 1991). One such covariance matrix estimator originally proposed by Gallant (1987) is

$$\hat{S}_w = \sum_{m=-l(n)}^{l(n)} \omega_m \hat{S}_{nm} , \tag{3.7}$$

where

$$\omega_m = \begin{cases} 1 - 6m^2 + 6|m|^3 , & 0 \leq m \leq 0.5l(n) , \\ 2(1 - |m|)^3 , & 0.5l(n) \leq m \leq 1 , \end{cases}$$

$$\hat{S}_{nm} = n^{-1} \sum_{t=m+1}^{n} f(x_t, \tilde{\beta}) f(x_{t-m}, \tilde{\beta})'$$

and $l(n) = \mathrm{O}(n^{1/5})$. Therefore using $G_n(\tilde{\beta})$, W_n and $\hat{S}_w$ one can obtain a consistent estimator of V_G and so perform inference about β_0 using the asymptotic distribution of $\tilde{\beta}$.

In our discussion of IV in the linear model it was argued that the optimal choice of W_n is the sequence of matrices which yields the asymptotically most efficient estimator as $n \to \infty$. This criterion is still appropriate for GMM estimators in nonlinear dynamic models. It was noted above that the structure of V_G is identical to V in (2.7), and so we can use analogous arguments to characterize the optimal W. This time the analysis shows that the optimal W, is

[7] To see this put $G_0 = M_{zx}$, $S_w = \sigma^2 M_{zz}$.

S_w^{-1}, which implies the asymptotically efficient GMM estimator $\tilde{\beta}^*$ has covariance matrix $V_G^* = [G_0' S_w^{-1} G_0]^{-1}$. To construct $\tilde{\beta}^*$ we require a weighting matrix W which is both consistent for S_w^{-1} and is also positive definite. This presents a problem which we did not encounter in the linear model because in this more general setting such an estimator will inevitably depend on β because of the definition of S_w. The solution is to adopt a multi-step estimation procedure. For example one could estimate $\tilde{\beta}^*$ in two steps. In the first step we use a suboptimal choice of W_n (which does not depend on β) to obtain $\tilde{\beta}$. Although $\tilde{\beta}$ is inefficient it is still consistent and so we can use $\tilde{\beta}$ to construct $\hat{S}_w$ using (3.7). In the second step $\tilde{\beta}^*$ is obtained from (3.2) using $W_n = \hat{S}_w^{-1}$. While this two-step estimator is asymptotically efficient, one need not stop here. The second step estimator of β can be used to form a new estimator of S_w and the model reestimated to produce a new estimate of β which in turn yields a new estimator of S_w etc. This iterative procedure could be continued until convergence and the resulting estimator has the same asymptotic properties as the two-step estimator. However the finite sample properties may be different and we discuss this further in Section 5.

Finally we consider the extension of the overidentifying restrictions test to nonlinear dynamic models. It was observed above that if β is overidentified then $g_n(\tilde{\beta}) \neq 0$ in general. Following the intuition from the linear model, it would be anticipated that if the population moment condition $\mathrm{E}[f(x_t, \beta_0)] = 0$ is correct then $g_n(\tilde{\beta}) \approx 0$. Therefore the sample moment provides a convenient test of the model specification. Consider the test statistic

$$\tau_n = n g_n(\tilde{\beta}^*)' \hat{S}_w^{-1} g_n(\tilde{\beta}^*) .$$

Under the null hypothesis that $\mathrm{E}[f(x_t, \beta_0)] = 0$, τ_n converges in distribution to χ^2_{r-q}.

In this section it has been demonstrated that the GMM estimator can be used to perform inference about the parameters of a nonlinear dynamic econometric model. At this stage we wish to emphasize that we have actually assumed very little about the model. Apart from certain regularity conditions, we have only made two key assumptions. Firstly we assumed x_t is stationary and ergodic to facilitate the application of the Weak Law of Large Numbers and central limit theorem to various functions of x_t. This can be relaxed as we discuss in Section 5. Secondly and more importantly we assumed that there are sufficient moment conditions to identity β. However, it is important to realize that we did not make an explicit assumption about the parametric form of the data generation process for x_t. In some cases a model for x_t may be required to obtain the moment conditions. Our analysis of the IV estimator in the linear model was an example of this case. However, in many situations, one can deduce moment conditions without completely specifying a model for x_t. This makes GMM a very attractive estimation principle for two reasons. Firstly the data generation process for x_t may be a large (possibly nonlinear) system of equations and so GMM offers the potential to estimate β without the

computational burden of estimating the complete system.[8] Secondly economic theory may imply a set of moment conditions but not a suitable model for x_t. This creates a problem for the implementation of a number of traditional estimation methods (e.g., maximum likelihood) which require an explicit model for x_t. In this case we could assume specification for x_t and estimate the complete system. However, if the model for x_t is misspecified, then this may bias our inference about β.[9] Both of these reasons are particularly relevant in the estimation of the parameters of discrete dynamic programming models and in the next section we describe how GMM can be applied to greatly simplify the estimation of these models.

4. GMM and Euler equation models

In his famous policy critique Lucas (1976) argued that econometric policy evaluations based on traditional dynamic simultaneous equation models are seriously flawed. Most of these analyses assume that the model's parameters are invariant across policy regimes. Lucas argued that rational agents take account of policy changes in their decision making and so one cannot expect the same marginal response to a change in a policy instrument in different policy regimes. Consequently it is more appropriate to view these parameters as functions of the economic environment and a set of underlying fundamental parameters which govern peoples preferences and the technology in the economy: 'taste and technology' parameters. This criticism has sparked a considerable research program formulating rational expectational models which are explicitly parameterized in terms of the 'taste and technology' parameters. These models are typically nonlinear and are computationally burdensome to estimate using standard methods such as maximum likelihood. However, Hansen and Singleton (1982) demonstrated that GMM can be applied in these models relatively easily and so facilitates inference about these parameters. In this section we summarize Hansen and Singleton's analysis and follow their approach of using a consumption based asset pricing model to illustrate the arguments. However, we consider a simplified version of their model to expedite the exposition.

Consider the case where a representative economic agent chooses consumption and investment to maximize his/her expected discounted utility

$$\mathrm{E}_0\left[\sum_{t=0}^{\infty} \delta^t U(C_t)\right], \tag{4.1}$$

[8] Note that we have made no assumption about the relationship between p, the dimension of x_t, and q the dimension of β.

[9] Gallant and Tauchen (1989) propose using maximum likelihood estimation using a semi-nonparametric approximation to the probability density function of x_t which is estimated from the data. This approach may avoid this bias and is discussed in the next section.

where C_t is consumption in period t, $U(\cdot)$ is a strictly concave utility function, δ is a constant discount factor and $\mathrm{E}_t[\cdot]$ denotes conditional expectation given Ψ_t, the information set available to the agent at time t, for $t = 0, 1, \ldots$. In any period t the agent can choose to buy either consumption goods or purchase an asset which yields payoff R_{t+1} in the next period. Let P_t and Q_t be the price and quantity purchased of the asset in period t respectively. The agent is assumed to receive income from assets purchased in the previous period and labor income W_t. All prices are denominated in terms of the consumption good.[10] Therefore the agent's budget constraint is

$$C_t + P_t Q_t \leq R_t Q_{t-1} + W_t \,. \tag{4.2}$$

The maximization of (4.1) subject to (4.2) yields the first-order condition

$$P_t U'(C_t) = \beta \mathrm{E}_t[R_{t+1} U'(C_{t+1})] \,, \tag{4.3}$$

where $U'(\cdot) = \partial U / \partial C$. It is convenient to rewrite (4.3) as

$$\mathrm{E}_t\left[\delta \frac{R_{t+1}}{P_t} \frac{U'(C_{t+1})}{U'(C_t)} - 1\right] = 0 \tag{4.4}$$

and we refer to (4.4) (or equivalently (4.3)) as the Euler equation of the system. To estimate this model it is necessary to specify the utility function. Following Hansen and Singleton (1982) we define

$$U(C_t) = C_t^{\gamma}/\gamma \,, \quad \gamma < 1 \,. \tag{4.5}$$

Therefore the Euler condition becomes

$$\mathrm{E}_t\left[\delta \frac{R_{t+1}}{P_t}\left[\frac{C_{t+1}}{C_t}\right]^{\alpha} - 1\right] = 0 \,, \tag{4.6}$$

where $\alpha = \gamma - 1$.

Now consider the problem of estimating (α, δ). Given the consumption and asset data customarily used in these models, it is reasonable to model R_{t+1}/P_t and C_{t+1}/C_t as stationary series. Therefore, to estimate this model by maximum likelihood it is necessary to specify the conditional distribution of $lw_{1t+1} = \log[R_{t+1}/P_t]$ and $lw_{2t+1} = \log[C_{t+1}/C_t]$ given Ψ_t.[11] The maximum likelihood estimates of the parameters of the model are obtained by maximizing the likelihood subject to the Euler equation constraint given in (4.6) for each t. In some cases the imposition of this constraint may reduce to restrictions on the parameters (e.g., see Hansen and Singleton, 1982, pp. 1279–1280). However, most often it is necessary to use a constrained optimization technique in which case the imposition of (4.6) requires numerical integration (see Gallant and Tauchen, 1989). Therefore the maximum likelihood method has both the

[10] In other words P_t is the price of the asset in dollars ÷ price of the consumption good in dollars.

[11] The log transformation is taken due to the nonnegativity of R_{t+1}/P_t and C_{t+1}/C_t.

problems discussed at the end of Section 3 namely: computational burden and potentially biased inference if the conditional distribution is misspecified. One can protect oneself from the second problem by adopting a flexible functional form for the joint conditional probability density function of (lw_{1t+1}, lw_{2t+1}). This is the strategy advocated by Gallant and Tauchen (1989) who estimate a semi-nonparametric approximation to the true density function. However, such methods are inevitably computationally more burdensome.

In contrast, now consider GMM estimation of (α, δ). To use GMM it is necessary to find at least two population moment conditions. The Euler condition yields one directly because by the law of iterated expectations

$$\mathrm{E}\left\{\delta \frac{R_{t+1}}{P_t}\left[\frac{C_{t+1}}{C_t}\right]^{\alpha} - 1\right\} = \mathrm{E}\left\{\mathrm{E}_t\left\{\delta \frac{R_{t+1}}{P_t}\left[\frac{C_{t+1}}{C_t}\right]^{\alpha} - 1\right\}\right\} = 0$$

from (4.6). However, we can combine the Euler equation with the rational expectations hypothesis to obtain additional moment conditions. If agents are rational, then they use all available information at time t (i.e., Ψ_t) to form their expectations. It follows that if $y_{t+1} \notin \Psi_t$ but $z_t \in \Psi_t$ then

$$\mathrm{E}_t[y_{t+1} z_t] = \{\mathrm{E}_t[y_{t+1}]\} z_t \,.$$

Now if $\mathrm{E}_t[y_{t+1}] = 0$ then by the law of iterated expectations

$$\mathrm{E}[y_{t+1} z_t] = 0 \,.$$

Using the same argument in the consumption based asset pricing model, the Euler equation implies

$$\mathrm{E}[u_{t+1}(\alpha, \delta) z_t] = 0 \,, \tag{4.7}$$

where $u_{t+1}(\alpha, \delta) = \delta(R_{t+1}/P_t)(C_{t+1}/C_t)^{\alpha} - 1$ and z_t is an $(r \times 1)$ vector of variables contained in Ψ_t. In practice there is no shortage of candidates for z_t. One can use C_{t-i}, R_{t-i}, P_{t-i}, $i \geqslant 0$, any function of these variables or any other macroeconomic variables, such as lagged money supply, which are known to the agent at time t. This abundance can create a problem because the parameter estimates will vary with the choice of z_t. This issue is discussed further in Section 5.

Notice that the structure of (4.7) is identical to that of the population moment condition for IV in the linear model. For this reason z_t is referred to as a vector of instruments. The key difference between the moment conditions in the two models is that in the linear model we had to explicitly specify the data generation process for y_t to construct $u_t(\beta)$, whereas in the Euler equation model $u_t(\alpha, \delta)$ is just a function of the data with conditional expectation zero. Due to this close relationship between the estimators,

Hansen and Singleton (1982) refer to the GMM estimator based on (4.7) as a generalized instrumental variables estimator.[12]

From this discussion it is clear that GMM has two advantages over ML estimation in these models.[13] Firstly it is computationally more convenient and secondly it avoids potential bias due to misspecification of the distribution of (lw_{1t+1}, lw_{2t+1}). However, GMM is not unambiguously better, because ML based on a correctly specified model yields the asymptotically most efficient estimators. The size of these efficiency gains depends on the model in question. This potential provides the incentive for the approach taken by Gallant and Tauchen (1989). However, to our knowledge no formal efficiency comparison has been made between GMM and MLE based on a semi-nonparametric likelihood function in the context of time series models.[14]

Due to its computational convenience Hansen and Singleton's approach has been used by many other researchers to estimate rational expectations models. In each case the Euler equation provides the basis for the moment conditions in the same way as we described above. Examples of these studies are (i) asset pricing: Mark (1985), Dunn and Singleton (1986), Epstein and Zin (1991); (ii) factor demand: Pindyck and Rotemberg (1983); (iii) labor supply: Mankiw, Rotemberg and Summers (1985), Eichenbaum, Hansen and Singleton (1988); (iv) inventory holdings: West (1986), Miron and Zeldes (1988), Fair (1989).

5. Further issues concerning GMM based inference

The discussion in Sections 3 and 4 was confined to the original articles about GMM by Hansen (1982) and Hansen and Singleton (1982). Subsequently there have been a number of studies both evaluating the performance of GMM and also further developing the GMM inference framework. In this section we provide a selective review of this literature.

So far the discussion has focussed on the asymptotic properties of GMM. For practical purposes, it is important to have evidence on its finite sample properties. Tauchen (1986) simulated data from a small artificial economy and examined the finite sample behavior of the GMM estimators of (α, δ) in the consumption based asset pricing model described in Section 4. He used the instrument vector $z_t = (1, w_{1t}, \ldots, w_{1t-L+1}, w_{2t}, \ldots, w_{2t-L+1})$ where $w_{1t} = C_t/C_{t-1}$, $w_{2t} = R_t/P_{t-1}$ for $L = 1, 2, 3, 4$. The sample sizes were set at 50 and 75.

[12] In this example we had only one asset and so there is only one Euler condition. If there are m assets then there will be m Euler equations each of which can be used to construct moment conditions in the way described above (see Hansen and Singleton, 1982).

[13] Hansen and Sargent (1982) provide a more detailed comparison of GMM and MLE for linear rational expectations models.

[14] However, in the case where the variables are iid, Chamberlain (1987) shows that GMM can be equally efficient asymptotically as maximum likelihood estimation in which the probability density function of x_t is estimated nonparametrically along with the parameters of interest (see Section 5 for further discussion).

He found that the estimator performed reasonably well but appeared sensitive to the choice of L. As L increased the estimators' variance decreased but there was an increase in bias. Tauchen concluded that it is advisable to choose L to be small. Finally he found that the finite sample distribution of τ_n was well approximated by its asymptotic distribution. Kocherlakota (1990) conducts a similar study to Tauchen (1986) but uses the iterated GMM estimator in which one reestimates with updated versions of $\hat{S}_w$ as the weighting matrix until the estimate converge (see Section 3). He considers exactly the same structure of artificial economy as Tauchen and also investigates the Hansen and Singleton (1982) consumption based asset pricing model. However, unlike Tauchen's study he also examines the model with multiple assets and different instrument sets all of which are indexed by t (i.e., $L = 0$ in Tauchen's notation). His simulations are for example size 90. In general he confirms many of Tauchen's findings although he reports evidence that τ_n tends to reject too frequently. Furthermore, he finds GMM performs worse when larger instrument sets are used. In particular some coefficients are biased and confidence intervals based on the asymptotic distribution are too narrow. His results suggest that Tauchen's finding about the sensitivity of GMM to L may in part occur because in Tauchen's framework an increase in L also increases the size of the instrument vector.

Both these studies concentrate on highly nonlinear but small systems of equations. Ferson and Foerster (1991) report evidence on the finite sample behavior of GMM in larger systems. Their simulation design is different from that of Tauchen (1986) and Kocherlakota (1990). Ferson and Foerster (1991) estimate a seemingly unrelated regression model with cross equation restrictions for asset returns using real financial data. The residuals from these equations are used to bootstrap artificial data on asset returns for the observed sequence of explanatory variables. They control the size of the system, the number of instruments and the sample size.[15] Unlike the other studies cited above they consider the behavior of both the two-step and the iterated GMM estimators. In their study GMM exhibits relatively small biases even in samples of size 60. However, the standard errors based on asymptotic theory understate the variability of the estimators and this bias increases as either the dimension of the system (e.g., number of assets) or the sample size decreases. They find that degrees of freedom adjustments to the asymptotic standard errors reduce but do not eliminate this bias. The behavior of the overidentifying restrictions test appears sensitive to the choice of estimator. When applied using the two-step estimator, the test rejects too often whereas with the iterated estimator the test is undersized but much closer to its nominal size.

All of these studies considered the generalized IV estimator proposed by Hansen and Singleton (1982) and described in the previous section. One

[15] The number of assets varies from 3 to 14, the number of instruments is 3 or 8 and the sample size varies from 60 to 720.

possible explanation for the small sample performance within these sampling designs may be the properties of the chosen instrument vector. In the context of a simple linear regression model, Nelson and Startz (1990) provide simulation evidence that the finite sample behavior of τ_n can be very sensitive to the quality of the instrument (i.e., the correlation between regressor and instrument). If the instrument is of very poor quality then τ_n tends to reject too frequently. They conjecture that a similar problem may be present in consumption based asset pricing models of the type described in Section 4 and may account for the frequent rejections of these models in practice. None of the other studies cited above explicitly control the correlation between $u_t(\beta_0)$ and z_t,[16] and so it is not clear to what extent instrument choice provides an explanation for the over rejections using τ_n reported in the simulation studies of Kocherlakota (1990) and Ferson and Foerster (1991). Further research is needed to explore the impact of instrument choice in nonlinear models. However this work is an important reminder that just as in linear models, the quality of instrument may be very important determinant of the finite sample properties of the 'generalized IV' estimator in nonlinear models.

Taken together these studies suggest that if the estimated model is correctly specified then the finite sample performance of GMM is sensitive to both the number of moment conditions and the sample size. In particular the asymptotic standard errors can understate the finite sample variances and the distribution of the overidentifying restriction test may not be well approximated by asymptotic theory in moderate size samples. Ferson and Foerster (1991) argue that the iterated GMM estimator exhibits better finite sample behavior than the two-step estimator and so should be used in practice.

Newey (1985) derived the asymptotic distribution of τ_n under local alternatives to the data generation process for x_t. Although in many cases this distribution is noncentral χ^2, he found that there was a class of local alternatives for which τ_n has power equal to size. This is an important reminder that τ_n should not be considered an 'omnibus' specification test. Clearly certain types of misspecification may not cause the moment condition to be violated and so cannot be detected using τ_n. More generally, Newey shows that there exist local alternatives for which both the assumed population moment condition is invalid and also $n^{1/2}(\tilde{\beta}^* - \beta_0)$ is asymptotically biased but τ_n still converges to χ^2_{r-q}. This can happen because τ_n is a function of $n^{1/2}g_n(\tilde{\beta}^*)$ and so depends on the behavior of both $n^{1/2}g_n(\beta_0)$ and $n^{1/2}(\tilde{\beta}^* - \beta_0)$. Although both these vectors may not converge to mean zero distributions it is possible for their asymptotic biases to offset in such a way that $n^{1/2}g_n(\beta^*)$ converges to a mean zero normal distribution.[17] Ferson and Foerster (1991) also report evidence on the finite sample behavior of GMM when the model is mis-

[16] Tauchen (1986) found no evidence that τ_n was sensitive to the autocorrelation in lw_{1t} and lw_{2t} for his simulation design.

[17] An example is where the local alternative is characterized by structural instability of the moment condition (see Ghysels and Hall, 1990b) and is discussed further below.

specified. They consider the case where the data are generated from two different models for asset returns but the linear model described above is estimated. Their results suggests that the power of the τ_n to detect model misspecification depends crucially on the relationship between the true and estimated models. The test exhibits reasonable power against one of their alternatives but low power against the other.

In all our discussion of GMM we treated the moment condition as given. In many cases there are a large number of potential moment conditions which can be chosen, and so it is important to consider which set yield the most asymptotically efficient estimators. This question has been analyzed for the situation were $f(x_t, \beta) = u_t(\beta) \otimes z_t$.[18] In this case the problem is to find the choice of instrument, z_t, which yields the asymptotically most efficient estimator *given* $u_t(\beta)$. Hansen (1985) provides a very general characterization of the asymptotic efficiency bound of GMM estimators. Hansen, Heaton and Ogaki (1988) use Hansen's (1985) method to characterize the bound when the estimation is based on multiperiod conditional moment conditions.[19] While it is possible to characterize this bound, it has proved difficult in nonlinear dynamic models to calculate the instrument vector which attains this bound. However, if we place more structure on the model, then the situation is more favorable. In the case where $u_t(\beta)$ is serially uncorrelated, then Tauchen (1986) shows that Hansen's (1985) bound implies that GMM based on $\mathrm{E}[u_t(\beta) \otimes z_t]$ with the optimal choice of instrument is asymptotically equivalent to GMM based on the population moment conditions

$$\mathrm{E}[Z_t^* u_t(\beta)] = 0\,, \tag{5.1}$$

where

$$Z_t^* = \mathrm{E}_t[\partial u_t(\beta_0)/\partial\beta]'\{\mathrm{E}_t[u_t(\beta_0)u_t(\beta_0)']\}^{-1}\,.$$

Typically these conditional expectations are not known and so Z_t^* is not a feasible instrument choice. Consequently interest has focused on the construction of feasible optimal GMM estimators which use an estimate of Z_t^* as instrument but are still asymptotically efficient. Most often these estimated instruments depend on β and so cannot be used in the first step estimation but can be calculated in subsequent estimations using $\tilde{\beta}$ from the previous step. Newey (1990) and Robinson (1991) demonstrate that under certain conditions on the dependence structure of x_t and $u_t(\beta)$, a feasible optimal GMM estimator can be constructed using certain nonparametric estimates of Z_t^*.[20]

[18] In this case $u_t(\beta)$ is $(m \times 1)$ and the Kronecker product notation denotes that the moment conditions are derived by interacting each element of $u_t(\beta)$ with each element of z_t.

[19] Also, see Heaton and Ogaki (1991) for an example of this bound in the context of a continuous time financial economics model.

[20] Newey (1990) restricts attention to the case where x_t, $u_t(\beta)$ are iid and Z_t^* is estimated by nonparametric regression using nearest neighbor or series approximation techniques. Robinson (1991) considers a bootstrap estimator of Z_t^* when (i) $u_t(\beta)$ is a function of contemporaneous and lagged values of the endogenous variables but $u_t(\beta)$ is itself uncorrelated; (ii) $u_t(\beta)$ is a function of only contemporaneous endogenous variables but may itself be serially correlated.

However, in more general models, the estimation of Z_t^* will probably require numerical integration and also the imposition of distributional assumptions about x_t. Furthermore, Tauchen (1986) found that a feasible optimal instrument GMM estimator often performed very poorly in finite samples for his simulation design. One case in which the optimal instrument is easily calculated is the linear regression model discussed in Section 2. For this model if we replace the fixed in repeated samples by the assumption by

$$x_{ti} = z_t' \gamma_i + v_{ti} ,$$

where x_{ti} is the i-th element of x_t, z_t is fixed in repeated samples, v_{ti} is an iid mean zero error and γ_i are unknown parameters, then the optimal instrument is $\hat{x}_t$. Therefore $\tilde{\beta}^*$ in (2.9) is the minimum asymptotic variance GMM estimator.

Given that it is possible to characterize an optimal instrument GMM estimator, it is natural to consider its asymptotic efficiency relative to other estimators. In the case where x_t is iid Chamberlain (1987) shows that GMM based on (5.1) is asymptotically as efficient as maximum likelihood in which one estimates nonparametrically the probability distribution of the data along with the parameters of interest. Furthermore, in linear time series models with moving average errors, Hansen and Singleton (1988) present a method of calculating the optimal instrument and show the resulting GMM estimator is asymptotically as efficient as maximum likelihood under normality.[21]

There has been considerable interest in extending the inference framework for GMM. Gallant and White (1988, Chapter 7) and Newey and West (1987b), and Andrews and Fair (1988) propose Wald, likelihood ratio type and Lagrange multiplier tests for the null hypothesis that β satisfies a nonlinear set of restrictions. In general the likelihood ratio type statistic is only appropriate under more restrictive conditions which are not satisfied in many GMM applications. These authors show that all three tests have the same asymptotic power against local alternatives.[22] However, Monte Carlo evidence reported in Gallant (1987) suggests the likelihood ratio and Lagrange multiplier tests are more reliable in finite samples. In Section 3 we observed that τ_n can be used to test all of the overidentifying restrictions of the model. If this test is statistically significant one may wish to test whether only a subset of these restrictions are satisfied with the aim of identifying which moment conditions are violated. Eichenbaum, Hansen and Singleton (1988) propose a likelihood ratio type test which can be used to test this hypothesis.

It was observed in Section 4 that Euler equation models arose as part of the response to the Lucas (1976) policy critique. These models are formulated in terms of 'taste and technology' parameters which are believed to be constant over time. Therefore a natural test for these models is to examine whether the

[21] Also see Hansen and Singleton (1991).

[22] This only applies to the likelihood ratio test in the cases where it is appropriate.

parameter estimates exhibit the constancy predicted by the theory. Furthermore, Ghysels and Hall (1990b) use Newey's (1985) analysis to show that τ_n has power equal to size against local alternatives characterized by structural instability. This suggests it may be wise to augment the test of overidentifying restrictions with structural stability tests. Andrews and Fair (1988) propose Wald, likelihood ratio type and Lagrange multiplier type tests for parameter constancy. The essential idea behind these tests is as follows. Suppose it is suspected that the parameters change their values after a particular point in the sample termed a breakpoint. Accordingly we split the sample into two subsamples which contain the observations from before and after the breakpoint. If the parameters remain constant over the sample, then estimates from either of the subsamples or the whole sample should be approximately the same. Ghysels and Hall (1990a) propose a predictive test for structural stability. In this approach parameter estimates from the first subsample are used to evaluate the moment conditions in the second subsample. If the model is structurally constant over the sample, then these predicted moment conditions should be statistically insignificantly different from zero. Ghysels and Hall (1990b) propose a test for parameter stability based on the likelihood ratio type procedure developed by Eichenbaum, Hansen and Singleton (1988). All of these tests have the twin drawbacks of requiring the breakpoint to be known[23] and also the subsamples must be asymptotically large. Andrews (1990) proposes a procedure for testing parameter stability when the breakpoint is not known. His strategy is to calculate the Wald statistic, say, for a set of possible breakpoints and to use the maximum as the test statistic. In other words we choose the test which maximizes the evidence against structural stability. Andrews derives the tabulates the distribution of this test. However, the subset of breakpoints must be restricted to ensure the subsamples are asymptotically large for each breakpoint chosen. Therefore like the other tests described above Andrews' test cannot be used to draw inference about parameter constancy at either end of the sample. Dufour, Ghysels and Hall (1991) propose a class of generalized predictive tests which permit inference at the ends of the sample. Within their framework, one must have an asymptotically large estimation sample from which to consistently estimate β. This estimate is used to evaluate $f(x_t, \beta)$ for every observation in the prediction sample, which can be of any size. Unlike the Ghysels and Hall (1990a) test which focussed on the average of the predicted moment conditions, the generalized predictive tests are based on the individual predictions $f(x_t, \tilde{\beta})$. Dufour, Ghysels and Hall discuss a number of methods for assessing whether these predictions residuals are statistically significantly different from zero.

All of these techniques permit inference within a given model. However, very often it is desired to discriminate between two economic models which imply nonnested Euler conditions for the same data. For example we may wish

[23] If the breakpoint is unknown a common strategy is to apply the above tests assuming the break occurs halfway through the sample. However, intuition suggests that in this case the tests will lose power as the distance between the true breakpoint and the middle of the sample increases.

to choose which of two different nonnested utility functions is more appropriate in a consumption based asset pricing model. Singleton (1985) proposes a simple test based on nesting the two Euler equations within a more general model whose Euler equation is a convex combination of the two. While computationally convenient, this approach may not always be appropriate because the convex combination of Euler equations may not be the Euler equation of an economically meaningful model (see Ghysels and Hall, 1990c). Ghysels and Hall (1990c) propose a test based on the encompassing principle (Mizon and Richard, 1986), which can be used to discriminate between two rational expectations models. However, to implement this test it is necessary to specify a model for x_t. This is undesirable because it requires much stronger assumptions that are necessary to estimate the individual Euler equation models by GMM.

Finally it should be noted that the stationarity and ergodicity assumption is not necessary for our analysis. Its role was to facilitate the application of laws of large numbers and central limit theorems to the data. As part of their development of a unified theory of dynamic models, Gallant and White (1988) generalize Hansen's (1982) results to the case where x_t is nonstationary but near epoch dependent on an α-mixing process (see Gallant and White, 1988, Chapter 2). This condition places sufficient restrictions on the memory of the process to allow the application of Laws of Large Numbers and Central Limit Theorems to functions of x_t. Many of the tests described above can be applied in this more general setting.

6. Concluding remarks

Generalized method of moments estimation has provided important contributions to both the theoretical and applied econometrics literature. In the theoretical literature it provides a unified framework for estimation theory. In the applied literature it provides a computationally convenient method of estimation in some models which were burdensome to estimate by other methods. We conclude our discussion of GMM by offering some final comments on both these contributions.

The first issue has not been explicitly addressed in this paper. In Section 2 we showed that IV is a special case of GMM and obviously the same result holds for OLS in this model because it is an IV estimator with $Z = X$. However, it may not be obvious how other familiar estimators can be nested within the GMM framework in more general models. To understand this connection we consider two estimators (i) $\hat{\beta}$ the value of β which optimizes $N_n(\beta)$; (ii) $\tilde{\beta}$ the GMM estimator obtained from minimizing $Q_n(\beta) = [\partial N(\beta)/\partial\beta]' W_n [\partial N(\beta)/\partial\beta]$. Now under certain regularity conditions $\hat{\beta}$ satisfies the first-order conditions $\partial N(\hat{\beta})/\partial\beta = 0$.[24] Similarly $\tilde{\beta}$ satisfies the first order conditions in (3.2)

[24] It is possible for $N(\beta)$ to have a unique optimum without the first-order conditions having a solution. Our discussion does not apply to these cases.

appropriate for this choice of $Q_n(\beta)$. Note that $\partial N(\beta)/\partial\beta$ is $(q \times 1)$ by construction and so β is just identified. In this case it follows from our earlier discussion that $\partial N(\tilde{\beta})/\partial\beta = 0$. Therefore, provided there is a unique solution to these equations $\hat{\beta} = \tilde{\beta}$. The advantage to viewing any estimator from a GMM perspective is that it focuses attention explicitly on the moment condition being exploited in estimation. For example if $N_n(\beta) = \Sigma_{t=1}^{n} \mathrm{LLF}_t(\beta)$ is the quasi log likelihood function (see White, 1982), then the condition for the consistency is that $\mathrm{E}[\partial \mathrm{LLF}_t(\beta_0)/\partial\beta)] = 0$, where $\mathrm{LLF}_t(\beta_0)$ is the conditional quasi log likelihood of the t-th observation. Therefore the GMM framework ties in very naturally with previous discussions of consistency in which this condition has received considerable attention, e.g., in the quasi maximum likelihood estimator literature (see White, 1982) and also in the context of nonlinear FIML estimation of simultaneous equation models (see Amemiya, 1977; Phillips 1982).

In this discussion we have implicitly concentrated on the situation in which all the elements of β are estimated simultaneously. Newey (1984) demonstrates that sequential estimators[25] also fit within the GMM framework. For example consider the following two-step estimation procedure. In the first step one estimates the just identified parameter vector β_1 from the sample moment conditions $n^{-1} \Sigma f_1(x_t, \beta_1)$. In the second step one estimates β_2 conditional on $\hat{\beta}_1$ from the sample moment conditions $n^{-1} \Sigma f_2(x_t, \hat{\beta}_1, \beta_2)$, which provides sufficient information so that β_2 is just identified given $\hat{\beta}_1$. Newey (1984) demonstrates that $\hat{\beta} = (\hat{\beta}_1', \hat{\beta}_2')'$ is identical to the GMM estimator obtained from the recursive sample moment $n^{-1} \Sigma f(x_t, \beta) = [n^{-1} \Sigma f_1(x_t, \beta_1)'$ $n^{-1} \sum f_2(x_t, \beta_1, \beta_2)']'$. An advantage of viewing sequential estimators from a GMM perspective is that one can use the results in equation (3.6) to facilitate calculation of the correct asymptotic covariance matrix of the estimators. One can also utilize the GMM framework to analyze the asymptotic behavior of certain simulation estimators. Lee and Ingram (1991) propose an estimation method for the time series models based on gathering artificial data. The idea is to estimate the parameters of the model by the values which yield sample moments for the artificial data that mimic the sample moments of the real data.[26] Lee and Ingram (1991) demonstrate that under certain regularity conditions their estimator is a special case of GMM and is therefore consistent and asymptotically normal.

The computational convenience of GMM has been illustrated using nonlinear rational expectations models. This convenience is achieved because for GMM it is only necessary to specify the Euler equation in order to obtain sufficient moment conditions to identify the parameters. In other words we gain computationally by only specifying part of the statistical model for x_t.

[25] For example stage least squares (e.g., see Johnston, 1984, p. 442–443) or the residual adjusted Aitken estimator for partial adjustment models with serially correlated errors (see Hatanaka, 1974).

[26] Also see Duffie and Singleton (1990) and Gourieroux and Monfort (1993).

However, as we have seen in Section 5, there are limits to the inference one can perform using a partially specified model. For example τ_n is not an 'omnibus' specification test and one needs to specify a model for x_t to calculate either the optimal instrument in general nonlinear dynamic models or non-nested hypothesis tests based on the encompassing principle. In spite of these limitations GMM continues to be an important method of estimation in econometrics. Many interesting questions remain and future work will no doubt provide us with a clearer understanding of the properties of GMM based inference.

Acknowledgment

This paper was originally prepared for presentation in an invited lecture entitled 'An Introduction to Generalized Method of Moments Estimation' at the Division of Research and Statistics, the Board of Governors of the Federal Reserve System, Washington, DC 20551, November 12, 1991. I am very grateful to those who attended this lecture for valuable comments which have helped improve the exposition. I am particularly indebted to Glenn Rudebusch for suggesting this topic and providing detailed comments on the paper. This paper has also benefitted from the comments of Eric Ghysels, G. S. Maddala, Whitney Newey, Masao Ogaki, and John Seater.

References

Amemiya, T. (1977). The maximum likelihood and the nonsingular three-stage least squares estimator in the general nonlinear simultaneous equation model. *Econometrica* **45**, 955–968.

Andrews, D. (1990). Tests for parameter instability and structural change with unknown breakpoint. Cowles Foundation Discussion Paper #943, Cowles Foundation for Research in Economics, Yale University, New Haven, CT.

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

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

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

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *J. Econometrics* **34**, 305–334.

Doan, T. (1990). *RATS, Version 3.10, Users Manual.* VAR Econometrics, Minneapolis, MN.

Duffie, D. and K. Singleton (1990). Simulated method of moments estimation of Markov models of asset prices. Discussion paper, Graduate School of Business, Stanford University, Stanford, CA.

Dufour, J.-M., E. Ghysels and A. Hall (1991). Generalized predictive tests and structural change analysis in econometrics. Unpublished mimeo, C.R.D.E., Université de Montréal, Quebec, Canada.

Dunn, K. and K. Singleton (1986). Modelling the term structure of interest rates under nonseparable utility and durable goods. *J. Financ. Econom.* **17**, 27–55.

Eichenbaum, M., L. Hansen and K. Singleton (1988). A time series analysis of representative agent models of consumption and leisure choice under certainty. *Quart. J. Econom.* **103**, 51–78.

Epstein, L. and S. Zin (1991). Substitution, risk aversion and the temporal behavior of asset returns. *J. Politic. Econom.* **99**, 263–286.

Fair, R. (1989). The production smoothing model is alive and well. *J. Monetary Econom.* **24**, 353–370.

Ferson, W. and E. Foerster (1991). Finite sample properties of the generalized method of moment in tests of conditional asset pricing models. Unpublished mimeo, Graduate School of Business, University of Chicago, Chicago, IL.

Gallant, A. R. (1987). *Nonlinear Statistical Models*. Wiley, New York.

Gallant, A. R. and G. Tauchen (1989). Seminonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica* **57**, 1091–1120.

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

Ghysels, E. and A. Hall (1990a). A test for structural stability of Euler parameters estimated via the generalized method of moments estimator. *Internat. Econom. Rev.* **31**, 355–364.

Ghysels, E. and A. Hall (1990b). Are consumption-based intertemporal capital asset pricing models structural? *J. Econometrics* **45**, 121–139.

Ghysels, E. and A. Hall (1990c). Testing nonnested Euler conditions with quadrature-based methods of approximation. *J. Econometrics* **46**, 273–308.

Gourieroux, C. and A. Monfort (1993). Simulation based inference: A survey with special reference to panel data models. *J. Econometrics*, to appear.

Hall, B. and C. Cummins (1991). *TSP Users Guide Version 4.2*. TSP International, Palo Alto, CA.

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

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

Hansen, L., J. Heaton and M. Ogaki (1988). Efficiency bounds implied by multi-period conditional moment restrictions, *J. Amer. Stat. Assoc.* **83**, 863–871.

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

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

Hansen, L. and K. Singleton (1988). Efficient estimation of linear asset pricing models with moving average errors. Unpublished mimeo, Department of Economics, University of Chicago, Chicago, IL.

Hansen, L. and K. Singleton (1991). Computing semi-parametric efficiency bounds for linear time series models. In: W. A. Barnett, J. L. Powell and G. Tauchen, eds., *Proc. 5th Internat. Sympos. in Economic Theory and Econometrics*.

Hatanaka, M. (1974). An efficient two-step estimator for the dynamic adjustment model with autoregressive errors. *J. Econometrics* **2**, 199–220.

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

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

Johnston, J. (1984). *Econometric Methods*. McGraw-Hill, New York.

Kocherlakota, N. (1990). On tests of representative consumer asset pricing models. *J. Monetary Econom.* **26**, 285–304.

Lee, B.-S. and B. Ingram (1991). Simulation estimation of time series models. *J. Econometrics* **47**, 197–205.

Lucas, R. (1976). Econometric policy evaluation: A critique. In: K. Brunner, ed., *The Phillips Curve and Labor Markets*, Carnegie–Rochester Conference Series on Public Economics, Vol. 1. North-Holland, Amsterdam.

Mankiw, N. G., J. Rotemberg and L. Summers (1985). Intertemporal substitution in macroeconomics. *Quart. J. Econom.* **100**, 225–252.

Mark, N. (1985). On time varying risk premia in foreign exchange market. *J. Monetary Econom.* **16**, 3–18.

Miron, J. and S. Zeldes (1988). Seasonality, cost shocks and the production smoothing model of inventories. *Econometrica* **56**, 877–908.

Mizon, G. and J.-F. Richard (1986). The encompassing principle and its application to testing non-nested hypotheses. *Econometrica* **54**, 657–678.

Nelson, C. and R. Startz (1990). The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *J. Business* **63**, 125–140.

Newey, W. (1984). A method of moments interpretation of sequential estimators. *Econom. Lett.* **14**, 201–206.

Newey, W. (1985). Generalized method of moments specification testing. *J. Econometrics* **29**, 229–256.

Newey, W. (1990). Efficient instrumental variables estimation of nonlinear models. *Econometrica* **58**, 809–837.

Newey, W. and K. West (1987a). A simple, positive definite, heteroscedasticity and autocorrelation covariance matrix. *Econometrica* **55**, 703–708.

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

Ogaki, M. (1992). GMM: A user's guide. Unpublished mimeo, Department of Economics, University of Rochester, Rochester, NY.

Phillips, P. (1982). On the consistency of nonlinear FIML. *Econometrica* **50**, 1307–1324.

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

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

SAS (1982). *SAS User Guide: Statistics*. SAS Institute Inc., Cary, NC.

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

Sargan, J. (1959). The estimation of relationships with autocorrelated residuals by the use of instrumental variables. *J. Roy. Statist. Soc. Ser. B* **21**, 91–105.

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

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

West, K. (1986). A variance bounds test of the linear quadratic inventory model. *J. Politic. Econom.* **94**, 374–401.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1–26.

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

White, H. (1984b). Comment on 'Tests of specification in econometrics' by P. Ruud. *Econometric Rev.* **3**, 261–267.