GMM INFERENCE WHEN THE NUMBER OF MOMENT CONDITIONS IS LARGE

ROGER KOENKER AND JOSÉ A. F. MACHADO

ABSTRACT. Asymptotic theory typically presumes that the dimensionality of econometric models is independent of the sample size even though this presumption is often quite unrealistic. In GMM estimation, whenever optimal instruments are not available, it can frequently be shown that adding over-identifying restrictions (moment conditions) will increase asymptotic precision. However, the conventional asymptotics which underlies this view insists that the number of moment conditions remain finite even though the number of available moment conditions may grow without bound. We consider the explicit dependence of the number of moment conditions (or instruments), q_n, on the sample size, n, and establish that, under conventional regularity conditions for the estimation of a linear model with general heteroscedasticity, $q_n^3/n \to 0$ is a sufficient condition for the validity of conventional asymptotic inference about the GMM estimator.

1. INTRODUCTION

Increasingly in econometrics we face problems of estimating models whose parametric dimension increases inexorably with the size of the sample. In some cases this occurs simply because we wish to refine the questions we would like to address as the available data becomes more abundant. In other circumstances, we begin with a more explicit sequence of models of increasing dimensionality, and estimate higher order models as n tends to infinity as a bias reduction device. This is quite explicit in recent work on series estimators, regression splines and a variety of other semiparametric models.

A crucial practical question in such circumstances is: how quickly can the parametric dimension of the model, p, increase with the sample size, n, and still retain the asymptotically Gaussian theory of classical inference. Huber (1973) showed that a *necessary* condition for consistency of the fitted values in least squares regression is that $p/n \to 0$, and Bickel(1976) reformulated this as: estimable functions of the least squares estimator, $\alpha'\hat{\beta}_n$ are asymptotically Gaussian with natural parameters if and only if the fitted values are consistent, again necessitating $p/n \to 0$. On the basis

Version: June 24, 1997. The authors would like to thank Joe Altonji, Manuel Arellano, Miguel Delgado, Essie Maasoumi, and Steve Portnoy for helpful comments, and two referees whose assiduous criticism of our regularity conditions led to a significant reformulation and extention of the results. They also gratefully acknowledge research support from NSF Grant SBR 93-20555 and FEDER (Praxis XXI program under grant 2/21/ECO/94) respectively.

of informal expansions and Monte Carlo experience, Huber(1973) conjectured that $p^2/n \to 0$, was needed to achieve a uniform normal approximation for smooth M-estimators in regression. Subsequently, this result has been improved, with e.g. Portnoy(1985) showing that if $(p \log p)^{3/2}/n \to 0$ then a uniform normal approximation is possible for M-estimators with sufficiently smooth influence functions. Further refinements may be found in Mammon(1989) and recent work by Shen and Wong(1994) extending the regression results to more general sieve estimators including penalized likelihood. In an empirical analysis of the wage-equation literature, Koenker(1988) shows that the number of estimated parameters, say p, grows roughly like $n^{1/4}$.

A related problem, but one which has attracted less attention, has arisen with the rapid diffusion of GMM estimation in econometrics. Even in problems where the parametric dimension of the model is fixed, the GMM proponent often faces the problem of choosing among a potentially very large class of admissible moment conditions and it would seem valuable to have some theoretical guidance as to how many moments are reasonable for GMM. In particular, how many moment conditions can be used and still expect to be able to obtain valid inference via GMM about the structural parameters of the model. As numerous authors have noted, traditional first-order asymptotics are not very helpful in answering these questions since they seem to imply the answer "the more the merrier" which even casual empiricism must recognize as foolish.

A vivid indication that these questions are of serious practical concern among proponents of GMM methods may be seen in the collection of papers assembled by Christiano (1996) on small sample properties of GMM. Several of these papers report disturbing Monte Carlo experience with GMM inference in a variety of contexts. The paper by Anderson and Sørenson (1996) explicitly examines, again via Monte Carlo, the effect of the dimensionality of the moment conditions on GMM performance. They conclude:

> As the sample expands, one should exploit additional moment restrictions. In small samples, however, the inclusion of an excessive number of moments results in more pronounced biases and larger RMSE. Thus, the use of additional information can be harmful. We conjecture that this occurs due to the need to obtain an estimate of the weighting matrix used in the GMM procedure.

The objective of this paper is to derive explicit rates on growth of the number of moment conditions, q_n, required to assure validity of the usual asymptotically Gaussian theory of GMM inference. These results complement the recent work of Newey (1990, 1993), de Jong and Bierens (1994) and Hong and White(1995) which emphasize specification testing based on the over-identifying restrictions. Also complementary is the work of Koenker, Machado, Skeels, and Welsh (1994), hereafter KMSW, which provides second order moment expansions for GMM estimators in a framework with q_n fixed. The $O(n^{-1})$ correction term derived there for the asymptotic variance of the normalized GMM estimator reflected the consequences of estimating the optimal GMM weighting matrix. Here too we shall see that with $q_n \to \infty$, estimation of this matrix can have serious implications, even jeopardizing the validity of conventional asymptotic inference based on the GMM estimator.

Consider the problem of estimating a finite dimensional parameter $\theta \in \Theta \subseteq R^p$ defined by the conditional moment condition $E\{u_t(z_t, \theta)|\Omega_t\} = 0$ for observations at $t = 1, \dots, n$, where Ω_t is the information set corresponding to the observation t. Let w_t be a vector of q_n random variables in Ω_t and set $m_t(\theta) = w_t u_t(z_t, \theta)$, so the q_n unconditional moment conditions may be written as, $Em_t(\theta) = 0$. Our practical problem is the choice of variables to select for w_t. Since Ω_t typically contains the entire past history of the z's and possibly other random variables, and since any measurable function of a variable in Ω_t is also an admissible instrument, it is easy to imagine suffering an embarrassment of riches. Somewhat paradoxically, as we have already noted, first-order asymptotics *which treat q_n as fixed* imply that the asymptotic precision of GMM estimator cannot be worsened by including additional moment conditions representing further over-identifying restrictions. Efficiency gains of this sort have been explored by Chamberlain (1982) and Cragg (1983), among others, who showed how over-identifying moment conditions could be used to construct estimators for the linear model which were more efficient than OLS in the presence of heteroscedasticity or serial correlation of unknown form. Chamberlain (1987) established the existence of an asymptotic efficiency bound for GMM estimators in this context. In principle it is possible to choose the instrumental variables, or more generally the moment conditions optimally i.e. in such a way that the efficiency bound is achieved. Furthermore, the optimal instruments are shown to imply no over-identifying restrictions. Unfortunately, optimal instruments are usually difficult or even impossible to compute since they require estimation of the conditional expectation of the gradient of the conditional moment conditions ($E\{\partial u_t/\partial\theta|\Omega_t\}$) as well as estimation of the conditional covariance matrix of the u_t's. For example, the optimal estimator in the standard linear regression model with heteroscedasticity or serial correlation of unknown form is the GLS which, of course, is not feasible. Whenever optimal instruments are not available, over-identifying restrictions will increase asymptotic precision and yet, in practice, no guidance is offered by the conventional theory as to how many moment conditions are actually beneficial. As Hansen and Singleton(1982) observe: "the asymptotic covariance matrix will not increase as more orthogonality conditions are used", but "using more orthogonality conditions may, at some point, lead to estimators with less desirable small sample properties." Our task is to try to make this sentiment more precise.

The remainder of the paper is organized as follows: the next section presents a stylized version of the linear GMM problem, provides some examples of this structure, and introduces our formal regularity conditions for the asymptotics.

Section 3 presents our asymptotic results: beginning with the benchmark case in which the covariance, or weighting, matrix of the moment conditions is known. The fundamental role of the estimation of the GMM covariance matrix is then addressed. In contrast to the situation with q_n fixed, as analyzed in KMSW(1994), where the effect of estimating the GMM covariance matrix appears only in the second-order asymptotic terms, with $q_n \to \infty$ the stochastic nature of this matrix has a potential impact on the first order asymptotics as well. We analyse the impact of the estimation of this optimal weighting matrix and show that under conventional regularity conditions the condition $q_n^3/n \to 0$ on the number of moments suffices to insure the validity of conventional GMM asymptotic inference. Section 4 offers some further interpretation of the findings. Proofs of all results are collected in the appendix.

2. Framework

In this Section we introduce the linear GMM model, discuss some examples of the model, and set out the formal conditions we will employ for the asymptotics in Section 3.

2.1. **A Simple GMM Problem.**

We consider the extremely simple, stylized linear GMM problem of estimating a vector α of p parameters in the model

$$y_i = x_i'\beta + u_i \tag{1}$$

$$\beta = G\alpha \tag{2}$$

where the sequence $\{u_i\}$ is independent with mean zero and variances $\{\sigma_i^2\}$ and G is a known $q_n \times p$ matrix. We wish to investigate the consequences of allowing the dimension of β to grow with n, while holding the dimension of α fixed.

Let $\hat{\beta}$ denote the OLS estimator of β in (1), and write

$$\hat{\theta} = (X'X)^{1/2}(\hat{\beta} - \beta)$$

whose covariance matrix is $V_n = \sum_{i=1}^n z_i z_i' \sigma_i^2$ with $z_i = (X'X)^{-1/2} x_i$.

The optimal GMM estimator of α is based on the consistent estimator of V_n

$$\hat{V}_n = \sum_{i=1}^n z_i z_i' \hat{u_i}^2$$

where $\hat{u_i} = y_i - x_i'\hat{\beta}$, and can be written as

$$\hat{\hat{\alpha}} = \alpha + (F_n'\hat{V}_n^{-1}F_n)^{-1}F_n'\hat{V}_n^{-1}\hat{\theta} \tag{3}$$

where

$$F_n = (X'X)^{1/2}G.$$

When the dimension of β is assumed fixed, conventional asymptotic results establish that $\hat{\hat{\alpha}}$ is asymptotically equivalent to the GMM estimator based on the true V_n, that is to

$$\hat{\alpha} = \alpha + (F_n' V_n^{-1} F_n)^{-1} F_n' V_n^{-1} \hat{\theta}. \tag{4}$$

In this paper we analyse the rates at which q_n may grow so that this asymptotic equivalence remains valid.

2.2. **Two Examples.** The GMM problem presented above is extremely simple but not, we hope, overly simplistic. We will assume thoughout that p is fixed while $q_n \to \infty$ as $n \to \infty$. To illustrate this approach we will briefly consider two influential examples of the theory.

Cragg (1983) showed that one can improve upon the OLS estimator of α in the model $y_i = w_i'\alpha + u_i$ where u_i is independent with mean zero and variance σ_i^2, by using the estimator,

$$\tilde{\alpha} = [W'Q(Q'\Sigma Q)^{-1}Q'W]^{-1}W'Q(Q'\Sigma Q)^{-1}Q'Y,$$

where $\Sigma = Euu'$ and Q is a matrix obtained by augmenting W with observations on p auxiliary variables (typically powers and cross-products of powers of the variables in W). It is easy to cast Cragg's estimator in our GMM framework: estimate α in

$$\begin{aligned} Y &= Q\delta + u \\ \delta &= \begin{bmatrix} I_p \\ O \end{bmatrix} \alpha \end{aligned}$$

where O is a $q_n \times p$ matrix of zeros and p is the dimension of α. Then $\tilde{\alpha}$ is exactly equal to the GMM estimator given by (4). Consequently our results qualify Cragg's, in the sense of providing sufficient conditions on the rate of growth of the number of auxiliary variables for the usual first-order asymptotics to provide a relevant benchmark to gauge the efficiency gains of his estimator. This example is closely connected to the discussion of optimal moments versus sequences of feasible moments in Chamberlain (1987).

Our second example is the panel data model of Arellano and Bond (1991) who consider the dynamic specification,

$$y_{it} = \alpha y_{i,t-1} + \eta_i + u_{it} \quad t = 1, \ldots, T; i = 1, \ldots, n.$$

Differencing to remove the fixed effect yields,

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta u_{it}.$$

If the $\{u_{it}\}$ are assumed iid, then we have the moment conditions,

$$E[(\Delta y_{it} - \alpha \Delta y_{i,t-1}) y_{i,t-s}] = 0$$

for $t = 3, \ldots, T$ and $s = 2, \ldots, t-1$. Thus, y_{i1} serves as a valid instrument for Δy_{i3}, (y_{i1}, y_{i2}) for Δy_{i4}, and so forth, implying that for a panel of length T we have $q_T = 1 + 2 + \ldots + (T-2) = (T-2)(T-1)/2$ valid instruments, or moment conditions. With $T = 6$ we already have $q_T = 10$, while for $T = 20$ we have $q_T = 171$.

The linear instrumental variable model exemplified by the Arellano-Bond specification can be represented in the form (1-2) for appropriate choice of β and G. If $\{w_i\}$ denotes a vector of q_T available instruments, we may set $\hat{\beta} = (W'W)^{-1}W'y$ and $G = (W'W)^{-1}W'X$ and consider moment conditions $\hat{\beta} = G\alpha$. The optimal instrumental variable estimator is thus a variant of (4) in which both $\hat{\beta}$ and G depend upon the instruments $\{w_i\}$, and $\{x_i\}$ is allowed to be correlated with $\{u_i\}$ in (1). Related examples of the application of mimimum distance estimation to panel data may be found in Chamberlain (1984).

2.3. **Assumptions.** Unless stated otherwise all the results in the next section assume that the following regularity conditions (U), (X) and (G) hold. The error term $\{u_i\}$ is a independent sequence satisfying

$$Eu_i = 0,\ Eu_i^2 = \sigma_i^2 \text{ and } Eu_i^3 = 0; \tag{U1}$$

$$Eu_i^4 \leq \Delta < \infty\ \forall i; \tag{U2}$$

$$\sigma_i^2 \geq \delta > 0\ \forall i. \tag{U3}$$

The regressor sequence $\{x_i\}$ belongs to $\Re^{q_n}$ and there exist sequences of real numbers $\{\nu_{n1}\}$, $\{\nu_{n2}\}$ tending to infinity with n, and constants $0 < C_1 < C_2 < \infty$, satisfying (*a.s.* in the case of random regressors)

$$\lambda_{max}(X'X) \leq \nu_{n2}C_2 \text{ and } \lambda_{min}(X'X) \geq \nu_{n1}C_1 \tag{X1}$$

where λ_{max}, and λ_{min} denote the maximum and minimum eigenvalues; furthermore,

$$\max_{1\leq i\leq n} |x_i'(X'X)^{-1}x_i| \equiv \max_{1\leq i\leq n} ||z_i||^2 \leq (q_n/\nu_{n1})C_3 \text{ for some } C_3 < \infty. \tag{X2}$$

The matrix G of constraints defining the structural parameters satisfies

$$\lambda_{max}(G'G) \leq \kappa_n C_5 \text{ and } \lambda_{min}(G'G) \geq \kappa_n C_4 \tag{G}$$

for a nondecreasing sequence of real numbers κ_n and constants $0 < C_4 < C_5 < \infty$.

Remarks. Several remarks on the regularity conditions may be useful at this point. To focus the interpretation of our results we will stress the special status of conditions (X1-2) with $\nu_{n1} = \nu_{n2} = n$. By stressing this case, we are adopting a relatively optimistic view about the informational content of the moment sequence. Portnoy (1984,1985,1987) uses closely related conditions and has shown that they hold *in probability* for a wide class of random designs. While this special case is certainly conventional within the broader literature on the large sample theory of the linear model it should not be considered totally innocuous. Some applications may necessitate even more stringent control over the growth of q_n due to the slower growth of the eigenvalues in (X1). But the general form of the results can obviously acommodate a wide variety of growth rates different than the conventional rate n. In Theorem 3 we will impose an additional requirement that the growth in the eigenvalues of the design sequence generating the moment conditions is "balanced", more specifically, that $\nu_{n2}^2/\nu_{n1}^3 \to 0$.

One should, perhaps, be less sanguine about applications in which moment conditions are generated by series expansions of a fixed set of initial covariates. However, in certain cases moment sequences generated from series expansions may also be shown to satisfy the conventional rate n form of (X1-2). Andrews (1991), for example, considers trignometric series defined for independent random regressors $\{X_i \in [0, 2\pi]^d, \quad i = 1, 2, ..., n\}$ with (Lebesgue) densities $f_i(x)$ whose averages are bounded above and bounded away from zero from below for n sufficiently large. He establishes that a rate n form of the eigenvalue condition (X1) holds, and given (X1), (X2) follows immediately from the boundedness of the trignometric series. Other series examples are considered by Andrews(1991) as well as Cox(1988) and Newey (1990,1993), which would require alternative rates for the ν's.

Condition (X2) is, in effect, the standard Lindeberg condition for regression. Without it we could certainly not expect valid asymptotic inference based on normal approximations. Clearly it holds in the case of bounded regressors but boundedness is hardly necessary as the following example illustrates. Suppose the rows $\{x_i\}$ of X are generated randomly from $\mathcal{N}(0, I_{q_n})$, then as $n \to \infty$, $v_i = nx_i'(X'X)^{-1}x_i$, behaves like a random sample from $\chi^2_{q_n}$. Consequently, the standardized random variables $w_i = (v_i - q_n)/\sqrt{2q_n}$ behave, for n and q_n sufficiently large, like a random sample from the standard normal distribution. For such a sample,

$$\max_i\{w_i\} = (2 \log n)^{1/2} + o_p(1),$$

so (X2) holds provided $\log n/q_n \to 0$. Portnoy(1985,1987) extends the heuristics of this simple example to consider a general class of random designs generated by scale mixtures of normals, rigorously establishing that (X1) and (X2) hold in probability with $\nu_{n1} = \nu_{n2} = n$.

Condition (G) ensures that the sequence of moment conditions serves to identify the structural parameter, α. If $\kappa_n \to \infty$ we obtain an improvement in the *rate* of convergence from allowing the number of moment conditions to tend to infinity; more typically, κ_n would be bounded and we would hope to obtain an improvment in the limiting covariance matrix of $\sqrt{n}(\hat{\alpha} - \alpha)$ by letting $q_n \to \infty$ but no alteration in the rate of convergence. Cragg's example clearly satisfies (G) with $\kappa_n \equiv 1$.

3. Large-q_n Asymptotics

This section contains the main results of the paper. We begin by stating a result for a pseudo-estimator which uses the unknown, optimal GMM weighting matrix, and then turn our attention to the more realistic case of the consequences of estimating this matrix.

3.1. **Optimal Weights.**

Our first result is of little practical interest since the individual error variances are assumed to be known, but it provides a useful benchmark for subsequent results. It establishes the asymptotic distribution of $\hat{\alpha}$ when q_n grows with n.

Theorem 1. *If $q_n/\nu_{n1} \to 0$ as $n \to \infty$, then $(F_n' V_n^{-1} F_n)^{1/2}(\hat{\alpha} - \alpha) \xrightarrow{D} N(0, I_p)$.*

In what we regard as the conventional case that $\nu_{n1} = \nu_{n2} = n$ and $\kappa_n = 1$, Theorem 1 establishes that for $q_n = o(n)$, $\sqrt{n}(\hat{\alpha} - \alpha) = O_p(1)$ and, therefore, suggests that $(\hat{\hat{\alpha}} - \alpha)$ should also be $O_p(1/\sqrt{n})$. The condition $q_n = o(n)$ is arguably the weakest conceivable, i.e. the fastest permissible rate of growth of q_n. It coincides with the rate found for the OLS estimator in linear regression with i.i.d. errors, a case in which it can be shown to be also necessary for asymptotic normality. See, e.g., Huber (1981), Chapter 7.

3.2. **Feasible Estimation.** When the individual error variances are unknown the estimator $\hat{\alpha}$ is obviously not feasible. We now turn to the more realistic case in which the covariance matrix of the empirical moment conditions must be estimated. Using the identity

$$\hat{A}^{-1} = A^{-1} - \hat{A}^{-1}(\hat{A} - A)A^{-1} \tag{5}$$

one can write

$$(F_n' \hat{V}_n^{-1} F_n)^{1/2}(\hat{\hat{\alpha}} - \alpha) = (C_n + A_n)[(F_n' V_n^{-1} F_n)^{1/2}(\hat{\alpha} - \alpha) + B_n] \tag{6}$$

where

$$C_n = (F_n' \hat{V}_n^{-1} F_n)^{1/2}(F_n' V_n^{-1} F_n)^{-1/2} \tag{7}$$

$$A_n = (F_n' \hat{V}_n^{-1} F_n)^{-1/2} F_n' \hat{V}_n^{-1}(\hat{V}_n - V_n)V_n^{-1} F_n (F_n' V_n^{-1} F_n)^{-1/2} \tag{8}$$

$$B_n = (F_n' V_n^{-1} F_n)^{-1/2} F_n' \hat{V}_n^{-1}(\hat{V}_n - V_n)V_n^{-1}\hat{\theta} \tag{9}$$

Given the results in Theorem 1 our goal is to establish the conditions under which $C_n - I_p$, A_n, and B_n are all $o_p(1)$. The crux of the argument is the consistent estimation of the optimal weighting matrix. Theorem 2, establishes conditions for the weak (norm) consistency of $\hat{V}_n$. We will use $\| A \|$ to denote the Frobenious matrix norm ($\| A \| = \sqrt{Tr(AA')}$). In a sense, this is the crucial result of the paper for it fundamentally determines the rate at q_n may grow for the asymptotic equivalence to hold.

Theorem 2. *If $q_n^2/\nu_{n1} \to 0$ as $n \to \infty$, $\| \hat{V}_n - V_n \| = O_p(q_n/\sqrt{\nu_{n1}})$.*

It can be shown that for consistent estimation of individual elements of V_n a faster rate of increase in q_n would be permitted. However, unlike the fixed dimension case, with $q_n \to \infty$ consistent estimation of each individual element of a matrix is not equivalent to consistent estimation of its norm. We can now state the main result.

Theorem 3. *If $\max\{q_n^3 \nu_{n2}/\nu_{n1}^2, q_n^2 \nu_{n2}^2/\nu_{n1}^3\} \to 0$, $(F_n' \hat{V}_n^{-1} F_n)^{1/2}(\hat{\hat{\alpha}} - \alpha) \xrightarrow{D} N(0, I_p)$.*

Remark. Again, we would like to stress the conventional case $\nu_{n1} = \nu_{n2} = n$, where the hypothesis of Theorem 3 may be reduced to $q_n = o(n^{1/3})$. This rate is essentially determined by two factors: the rate of consistency of the normalized OLS estimator of β and the rate at which its asymptotic covariance matrix can be consistently estimated in the sense of Theorem 2. The fact that $\| \hat{\theta} \|^2 = O_p(q_n)$ is quite natural, see Portnoy (1985), Theorem 3.3, and it does not seem it can be improved upon. To see this it suffices to consider the case of homoscedastic, Gaussian errors where $\| \hat{\theta} \|^2$ has an exact $\chi^2(q_n)$ distribution. Therefore the asymptotic equivalence is critically determined by the rate established in Theorem 2, that is, the rate at which the optimal weights can be consistently estimated.

When $\nu_{n1} = \nu_{n2} = \nu_n$, so the eigenvalues are fully balanced, we see that $q_n = o(n^{\nu_n/3})$ is required. When the ν's are different the situation is somewhat more delicate. To illustrate, consider the case that $\nu_{n1} = n^a$ and $\nu_{n2} = n^b$ with $a \leq b$. Note that the hypothesis in Theorem 3 requires in addition that $b < \frac{3}{2}a$. This is the "balance" condition on the eigenvalues alluded to earlier. In this case if $1 \leq b/a < 5/4$ we have $q_n = o(n^{(2a-b)/3})$, while for $5/4 \leq b/a < 3/2$ we need $q_n = o(n^{(3a-2b)/2})$.

4. Conclusions

We have argued that the classical paradigm of asymptotic theory for GMM estimation which presumes a fixed number of moment conditions independent of the sample size is often quite unrealistic. Instead, we consider the explicit dependence of the number of unconditional moment conditions, or instruments, q_n, on the sample size n. We establish that, for a class of linear models with general heteroscedasticity, $q_n = o(n^{1/3})$ is a sufficient condition for the usual limiting distribution of the GMM estimator of a fixed number of parameters to remain valid. It is difficult to imagine that more complicated, nonlinear GMM models could sustain a faster rate of growth than the one derived here for the linear case. We should also again stress that our regularity conditions (X1-2) on the moment sequence may in some applications impose stringent further control on the permissible rate of growth of the number of moment conditions.

If, as is the case in some series expansion GMM settings for example, the rates proportional to n in Conditions (X1) and (X2) are inappropriate, Theorem 3 provides explicitly allowable rates of growth for q_n in terms of the eigenvalue rates ν_{n1} and ν_{n2} appearing in conditions (X1-2).

This investigation helps to qualify existing results on the efficiency gains associated with over-identifying moment conditions by providing sufficient conditions for the conventional fixed-q_n GMM asymptotics to provide a relevant benchmark to gauge those efficiency gains. Since the difficulties we have identified in assuring the validity of GMM inference stem from the potential failure of consistency in the estimation of V when $q_n \to \infty$ too quickly, it may appear attractive to circumvent the problem of estimating V entirely, replacing $\hat{V}_n$ by a fixed weighting matrix like I_{q_n}. This strategic retreat from "optimal GMM" to fixed "estimators" of V has the obvious advantage of simplifying the convergence of the corresponding estimator of α, and thereby justifying a faster rate of increase for the number of moment conditions. But it forges a delicate balance between the efficiency gains available by introducing new moment conditions and the efficiency gains from better reweighting of fewer of these conditions. This balance demands serious consideration in each particular application.

Our large-q_n asymptotics shed light on several aspects of GMM behavior, which may be of practical significance, and which are simply invisible with a fixed q_n approach. Paraphrasing Huber's ((1981), p.164) comment about M-estimators for the linear model: for GMM the point is that if 20 observations are needed to justify asymptotic inference for a single moment condition, then the use of 10 moment conditions would require 20,000 observations.

Department of Economics, University of Illinois, Champaign, IL 61820
Faculdade de Economia, Universidade Nova de Lisboa, P1070 Lisbon

Appendix

Proof: (**Theorem 1**) For an arbitrary c in $\Re^p$ with $\| c \| = 1$, we will write $a_n' \equiv c'(F_n'V_n^{-1}F_n)^{-1}F_n'V_n^{-1}$. Then,

$$c'(\hat{\alpha} - \alpha) = \sum_i a_n' z_i u_i.$$

Given conditions (U2) and (U3), to verify the Liapunov condition it suffices that

$$D_n = \frac{\sum_i (a_n' z_i)^4}{(\sum_i (a_n' z_i)^2)^2} \to 0.$$

Since $\sum z_i z_i' = I$, the denominator equals $(a_n' a_n)^2$. As, $(\gamma' A \gamma)^2 \leq (\gamma' A^2 \gamma)(\gamma'\gamma)$, for any symmetric A,

$$\sum_i (a_n' z_i)^4 \leq a_n' \sum_i z_i z_i' z_i z_i' a_n \leq (a_n' a_n)^2 \max_{1 \leq i \leq n} \| z_i \|^2 \leq (a_n' a_n)^2 (q/\nu_{n1}) C_3,$$

by X2. ∎

Proof: (**Theorem 2**) One may write,

$$\begin{aligned}\hat{V}_n - V_n =& \sum_i z_i z_i'(u_i^2 - \sigma_i^2) + \sum_i z_i z_i'[x_i'(\hat{\beta} - \beta)]^2 - 2\sum_i z_i z_i' u_i x_i'(\hat{\beta} - \beta) \\ \equiv& A_{1n} + A_{2n} + A_{3n}.\end{aligned}$$

By the triangle inequality it suffices to establish the rate of convergence of $\| A_{in} \|$for $i = 1, 2, 3$. ForA_{1n} one has,

$$A_{1n}^2 = \sum_i \sum_j z_i z_i' z_j z_j' (u_i^2 - \sigma_i^2)(u_j^2 - \sigma_j^2)$$

and, thus,

$$E(Tr(A_{1n}^2)) \leq 4\Delta \sum_i (z_i' z_i)^2 \leq 4\Delta \max_{1\leq i\leq n} ||z_i||^2 \sum_{i=1}^{n} (z_i' z_i) = O(\frac{q_n^2}{\nu_{n1}}).$$

Consequently,

$$\| A_{1n} \| = O_p(q_n/\sqrt{n}).$$

For the second term one has

$$Tr(A_{2n}^2) = \sum_i \sum_j (z_i' z_j)^2 [x_i'(\hat{\beta}-\beta)]^2 [x_j'(\hat{\beta}-\beta)]^2 \leq \max_i ||z_i||^2 \max_j ||z_j||^2 ||\hat{\theta}||^4$$

where, $||\hat{\theta}||^2 = \sum_i [x_i'(\hat{\beta}-\beta)]^2$. It is easy to see that $||\hat{\theta}||^2 = O_p(q_n)$, so, by (X2),

$$\| A_{2n} \| = O_p(q_n^2/\nu_{n1}).$$

It remains to analyse $\| A_{3n} \|$.

$$Tr(A_{3n}^2) = \sum_i \sum_j (z_i' z_j)^2 u_i u_j x_i'(\hat{\beta}-\beta) x_j'(\hat{\beta}-\beta) = \sum_{ijkl} (z_i' z_j)^2 z_i' z_k z_j' z_l u_i u_j u_k u_l$$

and thus,

$$\begin{aligned} E(Tr(A_{3n}^2)) &= \sum_i (z_i' z_i)^4 E(u_i^4) + \sum_{ij} (z_i' z_i)^2 (z_i' z_j)^2 \sigma_i^2 \sigma_j^2 \\ &\quad + \sum_{ij} (z_i' z_j)^2 (z_i' z_i)(z_j' z_j) \sigma_i^2 \sigma_j^2 + \sum_{ij} (z_i' z_j)^4 \sigma_i^2 \sigma_j^2. \end{aligned}$$

Using (X2), (U2) and (U3), repeated applications of the Cauchy-Schwartz inequality yield the upper bound

$$\begin{aligned} E(Tr(A_{3n}^2)) &\leq \Delta\{\sum_i (z_i' z_i)^4 + \sum_i (z_i' z_i)^3 \sum_j (z_j' z_j) + 2(\sum_i (z_i' z_i)^2)^2\} \\ &\leq \Delta\{q_n (C q_n/\nu_{n1})^3 + q_n^2 (C q_n/\nu_{n1})^2 + 2(C q_n^2/\nu_{n1})^2\} \\ &= O(\frac{q_n^4}{\nu_{n1}^2}). \end{aligned}$$

Therefore, $\| A_{3n} \| = O_p(q_n^2/\nu_{n1})$ and the result follows. ■

Corollary 1. *If* $q_n^2/\nu_{n1} \to 0$, $\max\{|\lambda_{max}(\hat{V}_n - V_n)|, |\lambda_{min}(\hat{V}_n - V_n)|\} = O_p(q_n/\sqrt{\nu_{n1}})$.

Proof: Both $|\lambda_{max}(\hat{V}_n - V_n)|$ and $|\lambda_{min}(\hat{V}_n - V_n)|$ are continuous functions in the topology induced by the Frobenious norm of $(\hat{V}_n - V_n)$. ■

The proof of Theorem 3 requires three preliminary results.

Lemma 1. *If* $q_n^2/\nu_{n1} \to 0$ $\lambda_{min}(\hat{V}_n)/\lambda_{min}(V_n) \xrightarrow{P} 1$ *and* $\lambda_{max}(\hat{V}_n)/\lambda_{max}(V_n) \xrightarrow{P} 1$.

Proof: The proof is analogous to that of Lemma 4d of De Jong and Bierens(1994). Let c be a unit vector defined by $c'V_nc = \lambda_{min}(V_n)$, if $\lambda_{min}(\hat{V}_n) > \lambda_{min}(V_n)$ and $c'\hat{V}_nc = \lambda_{min}(\hat{V}_n)$, otherwise. If $\lambda_{min}(\hat{V}_n) > \lambda_{min}(V_n)$

$$c'\hat{V}_nc - c'V_nc = c'\hat{V}_nc - \lambda_{min}(V_n) > \lambda_{min}(\hat{V}_n) - \lambda_{min}(V_n).$$

If $\lambda_{min}(\hat{V}_n) \leq \lambda_{min}(V_n)$

$$c'\hat{V}_nc - c'V_nc = \lambda_{min}(\hat{V}_n) - c'V_nc \leq \lambda_{min}(\hat{V}_n) - \lambda_{min}(V_n).$$

Therefore,

$$|\lambda_{min}(V_n) - \lambda_{min}(\hat{V}_n)| \leq |c'(\hat{V}_n - V_n)c| = O_p(q_n/\sqrt{\nu_{n1}}),$$

by Corollary 1, which is $o_p(1)$ provided $q_n^2/\nu_{n1} \to 0$.

Let $\Sigma = Diag\{\sigma_1^2, \cdots, \sigma_n^2\}$. Since $X(X'X)^{-1/2}$ is semi-orthogonal, it follows from Poincaré's separation theorem (Magnus and Neudecker (1988), p.209) and from (U2) and (U3) that $\delta \leq \lambda_{min}(\Sigma) \leq \lambda_{min}(V_n)$ and $\lambda_{max}(V_n) \leq \lambda_{max}(\Sigma) \leq \Delta^{1/2}$. Thus,

$$|[\lambda_{min}(V_n) - \lambda_{min}(\hat{V}_n)]\lambda_{min}^{-1}(V_n)| = O_p(q_n/\sqrt{\nu_{n1}}) = o_p(1).$$

The proof for $\lambda_{max}(\hat{V}_n)/\lambda_{max}(V_n)$ is identical defining c now as $c'V_nc = \lambda_{max}(V_n)$, if $\lambda_{max}(\hat{V}_n) < \lambda_{max}(V_n)$ and $c'\hat{V}_nc = \lambda_{max}(\hat{V}_n)$, otherwise. ∎

Lemma 2. *If $q_n^2/\nu_{n1} \to 0$, there exist constants $0 < D_1 < D_2 < \infty$ such that, with probability approaching 1 as $n \to \infty$,*

$$\lambda_{max}(F_n'\hat{V}_n^{-1}F_n) \leq \nu_{n2}\kappa_n D_2 \text{ and } \lambda_{min}(F_n'\hat{V}_n^{-1}F_n) \geq \nu_{n1}\kappa_n D_1.$$

Proof: Using the eigenvalue bounds for V_n from the proof of Lemma 1, we have for any vector $c \in \Re^p$ with $\| c \|= 1$,

$$c'F_n'\hat{V}_n^{-1}F_nc \leq \frac{c'F_n'F_nc}{\lambda_{min}(\hat{V}_n)} \leq \frac{\lambda_{max}(F_n'F_n)}{\delta}\frac{\lambda_{min}(V_n)}{\lambda_{min}(\hat{V}_n)}$$

and

$$c'F_n'\hat{V}_n^{-1}F_nc \geq \frac{c'F_n'F_nc}{\lambda_{max}(\hat{V}_n)} \geq \frac{\lambda_{min}(F_n'F_n)}{\Delta^{1/2}}\frac{\lambda_{max}(V_n)}{\lambda_{max}(\hat{V}_n)}.$$

Using Lemma 1, the result then follows from conditions (X1) and (G). ∎

Lemma 3. *If $q_n^2\nu_{n2}^2/\nu_{n1}^3 \to 0$ as $n \to \infty$*

$$(F_n'\hat{V}_n^{-1}F_n)(F_n'V_n^{-1}F_n)^{-1} - I = o_p(1).$$

Proof: Using the matrix identity (5),

$$I - (F_n'\hat{V}_n^{-1}F_n)(F_n'V_n^{-1}F_n)^{-1} = F_n'\hat{V}_n^{-1}(\hat{V}_n - V_n)V_n^{-1}F_n(F_n'V_n^{-1}F_n)^{-1}.$$

Let e_i denote the i-th unit vector in $\Re^p$. Then

$$\begin{aligned}
|e_i' F_n' \hat{V}_n^{-1}(\hat{V}_n - V_n)V_n^{-1}F_n e_j|^2 &\leq (e_i' F_n' \hat{V}_n^{-2} F_n e_i)(e_j' F_n' V_n^{-1}(\hat{V}_n - V_n)^2 V_n^{-1} F_n e_j) \\
&\leq \frac{\lambda_{max}(F_n' F_n)}{\lambda_{min}(\hat{V}_n^2)}(e_j' F_n' V_n^{-2} F_n e_j)\lambda_{max}((\hat{V}_n - V_n)^2) \\
&\leq \frac{\lambda_{max}^2(F_n' F_n)}{(\lambda_{min}(\hat{V}_n)\lambda_{min}(V_n))^2}\lambda_{max}((\hat{V}_n - V_n)^2).
\end{aligned}$$

Since $(\hat{V}_n - V_n)$ is symmetric,

$$\lambda_{max}((\hat{V}_n - V_n)^2) = \max\{[\lambda_{min}(\hat{V}_n - V_n)]^2, [\lambda_{max}(\hat{V}_n - V_n)]^2\}.$$

From Theorem 2, both $|\lambda_{min}(\hat{V}_n - V_n)|$ and $|\lambda_{max}(\hat{V}_n - V_n)|$ are $O_p(q_n/\sqrt{\nu_{n1}})$. And, from Lemma 1, we have, $\lambda_{max}(F_n' F_n) = O(\nu_{n2}\kappa_n)$, $1/\lambda_{min}(\hat{V}_n) = O_p(1)$, and $1/\lambda_{min}(V_n) = O_p(1)$, so the result follows from

$$\lambda_{max}((F_n' V_n^{-1} F_n)^{-1}) = O_p(1/\nu_{n1}\kappa_n).$$

■

Proof: (**Theorem 3**) The same argument used to establish Lemma 3 can be used to show that $A_n = o_p(1)$ provided $q_n^2 \nu_{n2}^2/\nu_{n1}^3 = o(1)$. Thus, it remains to establish the conditions under which B_n is $o_p(1)$.

We have already seen in the proof of Lemma 3 that, $\lambda_{max}((F_n' V_n^{-1} F_n)^{-1}) = O_p(1/\nu_{n1}\kappa_n)$. Successive applications of the Cauchy-Schwartz inequality imply that if $q_n^3 \nu_{n2}\kappa_n/\nu_{n1} = o(1)$,

$$\begin{aligned}
|e_i' F_n' \hat{V}_n^{-1}(\hat{V}_n - V_n)V_n^{-1}\hat{\theta}|^2 &\leq (e_i' F_n' \hat{V}_n^{-2} F_n e_i)(\hat{\theta}' V_n^{-1}(\hat{V}_n - V_n)^2 V_n^{-1}\hat{\theta}) \\
&\leq \lambda_{max}(F_n' F_n)\lambda_{max}(\hat{V}_n^{-2})(\hat{\theta}' V_n^{-2}\hat{\theta})\lambda_{max}((\hat{V}_n - V_n)^2) \\
&\leq \lambda_{max}(F_n' F_n)(\hat{\theta}'\hat{\theta})\lambda_{max}(\hat{V}_n^{-2})\lambda_{max}(V_n^{-2})\lambda_{max}((\hat{V}_n - V_n)^2) \\
&= o_p(1)
\end{aligned}$$

The last step follows from the previously established facts that: $\lambda_{max}(F_n' F_n) = O_p(\nu_{n2}\kappa_n)$, $\lambda_{max}(V_n^{-2}) = (\lambda_{max}(V_n^{-1}))^2 = O_p(1)$, $\lambda_{max}(\hat{V}_n^{-2}) = (\lambda_{max}(\hat{V}_n^{-1}))^2 = O_p(1)$, $\lambda_{max}((\hat{V}_n - V_n)^2) = O_p(q_n^2/\nu_{n1})$ and $||\hat{\theta}|| = O_p(\sqrt{q_n})$. ■

REFERENCES

Anderson, T.G. and B.E. Sørenson (1996) GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics*, **14**, 328-352.

Andrews, D.W.K. (1991) Asymptotic normality of series estimators for nonparametric and semiparametric regression models, *Econometrica*, **59**, 307-346.

Arellano, M. and Bond, S. (1991) Some tests of specification for panel data: Monte Carlo evidence and an applications to employment equations, *Review of Economic Studies*, **58**, 277-297.

Bickel, P.J. (1976) Another look at robustness: A review of reviews and some new developments, *Scandinavian Journal of Statistics*, **3**, 145-168.

CHAMBERLAIN, G. (1982) Multivariate regression models for panel data, *Journal of Econometrics*, **18**, 5-46.

CHAMBERLAIN, G. (1984) Panel Data, *Handbook of Econometrics*, **2**, in Z. Griliches, and M. Intriligator, (eds.), North-Holland: New York.

CHAMBERLAIN, G. (1987) Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics*, **34**, 305-334.

CHRISTIANO, L.J. (1996) Associate Editor's Introduction to a Special Section on small-sample properties of GMM, *Journal of Business and Economic Statistics*, **14**, 261.

COX, D.D. (1988) Approximation of least squares regression on nested subspaces, *Annals of Statistics*, **16**, 713-732.

CRAGG, J. (1983) More efficient estimation in the presence of heteroscedasticity, *Econometrica*, **51**, 751-764.

DE JONG, R.M. AND BIERENS, H.J. (1994) On the limit behavior of a chi-square type test if the number of conditional moments tested approaches infinity, *Econometric Theory*, **9**, 70-90.

HANSEN, L. P. (1982) Large sample properties of generalized method of moment estimators, *Econometrica*, **50**, 1029-1054.

HANSEN, L. P.AND SINGLETON, K.J. (1982) Generalized instrumental variables estimation of nonlinear rational expectations models *Econometrica*, **50**, 1269-1286.

HONG Y. AND WHITE, H. (1995) Consistent specification testing via nonparametric series regression *Econometrica*, **63**, 1133-1159.

HUBER, P.J. (1973) Robust regression, asymptotics, conjectures and Monte-Carlo, *Annals of Statistics*, **1**, 799-821.

HUBER, P.J. (1981) *Robust Statistics*, Wiley: New York.

KOENKER, R. (1988) Asymptotic theory and econometric practice, *Journal of Applied Econometrics*, **3**, 139-147.

KOENKER, R., MACHADO J.A.F., SKEELS, C. , AND WELSH, A.H. (1994) Momentary Lapses: Moment expansions and the robustness of minimum distance estimators, *Econometric Theory*, **10**, 172-197.

MAGNUS, J.R AND NEUDECKER, H (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics,* Wiley: New York.

MAMMEN, E. (1989) Asymptotics with increasing dimension for robust regression with applications to the bootstrap, *The Annals of Statistics*, **17**, 382-400.

NEWEY, W.K. (1990) Efficient instrumental variables estimation of nonlinear models, *Econometrica*, **58**, 809-837.

NEWEY, W.K. (1993) Efficient estimation of models with conditional moment restrictions *Handbook of Statistics*, G.S. Maddala, et al. (eds.), **11**, 419-454.

PORTNOY, S. (1984) Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large II: Consistency, *Annals of Statistics*, **12**, 1298-1209.

PORTNOY, S. (1985) Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large II: Normal Approximation, *Annals of Statistics*, **13**, 1403-1417.

PORTNOY, S. (1987) A central limit theorem applicable to robust regression estimators, *Journal of Multivariate Analysis*, **22**, 24-50.

SHEN, X. AND WONG, W.H. (1994) Convergence rates of sieve estimators, *Annals of Statistics*, **22**, 580-615.