Cowles Foundation

for Research in Economics at Yale University

Cowles Foundation Discussion Paper No. 1515

May 2005

GMM WITH MANY MOMENT CONDITIONS

Chirok Han and Peter C.B. Phillips

GMM with Many Moment Conditions*

Chirok Han
Victoria University
of Wellington

Peter C. B. Phillips
Cowles Foundation, Yale University
University of Auckland & University of York

September, 2004

Abstract

This paper provides a first order asymptotic theory for generalized method of moments (GMM) estimators when the number of moment conditions is allowed to increase with the sample size and the moment conditions may be weak. Examples in which these asymptotics are relevant include instrumental variable (IV) estimation with many (possibly weak or uninformed) instruments and some panel data models covering moderate time spans and with correspondingly large numbers of instruments. Under certain regularity conditions, the GMM estimators are shown to converge in probability but not necessarily to the true parameter, and conditions for consistent GMM estimation are given. A general framework for the GMM limit distribution theory is developed based on epiconvergence methods. Some illustrations are provided, including consistent GMM estimation of a panel model with time varying individual effects, consistent LIML estimation as a continuously updated GMM estimator, and consistent IV structural estimation using large numbers of weak or irrelevant instruments. Some simulations are reported.

Keywords: Epiconvergence, GMM, Irrelevant instruments, IV, Large numbers of instruments, LIML estimation, Panel models, Pseudo true value, Signal, Signal Variability, Weak instrumentation.

JEL Classification: C22 & C23.

*A preliminary version of this paper was presented at the New Zealand Econometric Study Group meeting at the University of Otago, August 2002. Phillips thanks the NSF for research support under grant nos. SES 00-92509 & SES 04-142254.

1 Introduction

Generalized method of moments (GMM) provides an attractive estimation methodology that has been widely used in empirical research and is well suited to situations where economic information is given in terms of moment conditions. The approach has several well known advantages including an easily implemented asymptotic theory. On the other hand, GMM asymptotics depend on regularity conditions that are not always satisfied and which can affect finite sample performance adversely. One example that is discussed in Stock and Wright (2000) occurs when the moment conditions are weak in the sense that their expected first derivatives vanish at the specific rate $O(n^{-1/2})$ where n is the sample size. Under such weak moment conditions, GMM estimates are not consistent but converge weakly to a nondegenerate limit distribution.

Lack of consistency can be understood in terms of the relative weakness of the 'signal' delivered by the moment conditions compared to the 'noise' component. To fix ideas, consider the moment restrictions

$$(1) \quad Eg(w_i, \theta_0) = 0, \ i = 1, ..., n$$

where the w_i are here assumed to be iid for expository purposes, θ_0 is the 'true' parameter, and g is a vector function of fixed dimension. The GMM estimator $\hat{\theta}$ minimizes $\bar{g}(\theta)'\bar{g}(\theta)$, where $\bar{g}(\theta)$ is the sample moment function defined as $\bar{g}(\theta) = n^{-1}\sum_{i=1}^{n} g(w_i, \theta)$, and which we decompose as $\bar{g}(\theta) = E\bar{g}(\theta) + [\bar{g}(\theta) - E\bar{g}(\theta)]$. The point θ_0 in the parameter space is identified as the parameter value satisfying (1). So, information on θ_0 is contained in the function $Eg(w_i, \theta)$. Correspondingly, information on θ_0 produced by the sample $w^n = (w_i)_1^n$ is contained in $E\bar{g}(\theta)$, whereas the component involving $\bar{g}(\theta) - E\bar{g}(\theta)$ is noise that disturbs this information. We may therefore regard $E\bar{g}(\theta)$ as the *signal* and $\bar{g}(\theta) - E\bar{g}(\theta)$ as the *noise* of the sample moment function. When $E\bar{g}(\theta)$ is flat and zero (respectively, close to zero) throughout a neighborhood of θ_0, the sample function will fail to identify (clearly identify) the unique parameter value θ_0, and we may say that the signal is uninformative (weak).

In conventional GMM asymptotics the signal from the sample moment function is taken to be strong in the sense that $E\bar{g}(\theta)$ is zero uniquely at θ_0 and this identifying information does not diminish as $n \to \infty$. Further, the noise is eliminated asymptotically by the action of a uniform law of large numbers. In consequence, the signal dominates the noise asymptotically and consistency is obtained. By contrast, in weak moment condition asymptotics like those in Stock and Wright (2000), the signal of the sample moment function is permitted to diminish to zero at a controlled $\sqrt{n}$ rate. Since the noise also vanishes asymptotically in this case again at a $\sqrt{n}$ rate by virtue of a functional limit law, the signal does not dominate the noise, GMM is inconsistent and its limit behavior is governed by weak convergence to a limiting functional of the sample components. In effect, at the $\sqrt{n}$ rate, the noise component is retained asymptotically and the GMM estimate has a non-degenerate limit distribution. The effect is entirely analogous to that shown originally in Phillips (1984, 1989) where, in the case of a totally unidentified structural system, the uncertainty that is inherent in lack of identification is retained in the limit by way of the estimator converging to a random variable.

The present paper reconsiders the GMM limit theory by allowing the number (q) of moment conditions to be large while at the same time permitting the moment conditions

to be weak. Here, the signal from a fixed number of moment functions again does not dominate the noise but, as q increases, variation over the moment conditions accumulates and provides an alternate route by which the totality of the signal and its cumulative variability can dominate random noise asymptotically. In such a situation, the GMM estimator has a nonrandom probability limit. However, the contribution from the variation over the sample moment conditions (or signal variability) as q grows does not necessarily reinforce the effect of the true signal coming from the mean of the sample moment conditions. Therefore, the probability limit of the GMM estimator may or may not equal the true parameter θ_0. When there is a failure in the consistency of GMM, the extent of the inconsistency depends on how weak the moment conditions are in relation to their degree of variation. This relationship is made explicit in the asymptotic theory developed here. In this regard the results of the present paper differ from those of Donald, Imbens and Newey (2003) and Koenker and Machado (1999), who consider GMM and empirical likelihood estimation under strong moment conditions and explore conditions on q that permit usual asymptotic theory and statistical testing.

A primary contribution of the paper therefore lies in the generality with which the GMM asymptotics are obtained. The treatment of nonlinearities in the GMM context introduces substantial complications in the theoretical development over linear IV estimation with many (possibly weak) instruments and leads to new results that extend existing findings in that literature. In making these extensions, the paper provides an analytical framework for confronting problems of the type where there may be a multiplicity of conditions or instruments of varying quality that relate to the underlying econometric model. As discussed in Koenker and Machado (1999) and elaborated below, there are many examples of such situations that arise in practical work.

A further contribution of the paper is to study some of these examples in detail. For instance, in considering linear structural equation estimation within our framework, we show how consistent estimation is possible in the presence of an increasing collection of weak and even apparently irrelevant instruments. This asymptotic theory highlights the role that is played in consistent estimation between the quality of the instruments in their totality on the one hand and the degree of endogeneity in the system on the other. In the finite instrument case, this is a feature of simultaneous equations estimation that is well-known to be of central importance from exact finite sample distribution theory (Phillips, 1980, 1983, 1989; Hillier, 2004). Its nature in the case of increasing numbers of instruments becomes manifest in the limit theory here, which shows that consistent structural estimation with many irrelevant instruments is possible when the degree of endogeneity is local to zero.

As another illustration, we explore some of the effects of proliferating moment conditions in panel data modeling. Here, the phenomenon of moment condition proliferation is far from being a theoretical construct and arises in a natural way in many empirical econometric settings. Some striking examples are: Angrist (1990), who generates 884 dummy instruments by natural experiments; Angrist and Krueger (1991), who use two stage least squares (2SLS) estimation with 180 instruments; Han, Orea and Schmidt (2003), who consider a panel data model with time varying individual effects in which $O(T^2)$ moment conditions are exploited, where T is the time span of the panel; and Ahn and Schmidt (1995), who estimate a dynamic panel data model with $O(T^2)$ moment conditions. In general, this type of moment condition

proliferation is endemic to dynamic panel data models in which instruments are generated from lagged dependent variables and the time span of the panel is of moderate size.

An example of this phenomenon that we investigate arises in the following simple panel data model with time varying individual effects

$$y_{it} = \beta_0 + \lambda_t(\theta_0)\alpha_i + n^{1/2}\varepsilon_{it}, \quad \lambda_t(\theta) = \exp\{\theta(t-1)/(T-1)\},$$

where β_0 and θ_0 are to be estimated from panel observations of y_{it} for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. In this model, the intercept and the (weak) time trend become negligible compared with the idiosyncratic error as $n \to \infty$. As a result, the information content in y_{it} that is useful for estimation of β_0 and θ_0 is dominated by the random disturbances, and so the derived moment conditions for these parameters are weak in the Stock and Wright (2000)'s sense. It follows that when T is finite, the GMM estimator is not consistent. But if $T \to \infty$, then a specifically weighted GMM estimator turns out to be consistent and is an application of the method that is developed later in this paper. This illustrative panel model is further pursued as Example 18 in Section 5.

Important precedents of the present paper are contained in Bekker (1994) and Chao and Swanson (2002), as well as the early research by Anderson (1977) and Morimune (1983) who considered linear instrumental variable (IV) structural equation estimation in which the number of instruments increases with the sample size. Just as Stock and Wright (2000) is a natural generalization to the GMM situation of the weak IV asymptotics of Staiger and Stock (1997) and Phillips (1989), so the present paper builds on Bekker (1994) and Chao and Swanson (2002) by providing a natural GMM extension of asymptotic theories with large numbers of instruments. Chao and Swanson made the important departure of combining the effect of large numbers of instruments with allowance for weak instrumentation, thereby extending the framework of both Staiger and Stock (1997) and Bekker (1994). That extension made it possible to study large sample effects in which the concentration parameter was allowed to grow at rates different from the sample size, so that different degrees of instrument weakness could be analyzed. While there is no immediate analogue of the concentration parameter in the GMM set up, our asymptotic theory similarly allows for varying degrees of instrument weakness (including the interesting special case of uninformative instruments) and in this respect our work is most closely related to Chao and Swanson (2002). Our limit distribution theory also complements ongoing independent work by Chao and Swanson (2003) for the linear structural equation case. Other recent work on IV limit theory with many instruments has been done by Stock and Yogo (2004) using sequential and joint limit arguments based on the methods of Phillips and Moon (1999). A review of the weak instrument literature is given in Stock, Wright and Yogo (2002).

The plan of the paper is as follows. The model, assumptions and GMM set up are given in Section 2. Section 3 provides results on the convergence property, giving the probability limits of GMM estimators in a general form that is applicable to cases where there are many weakly identifying moment conditions. Section 4 develops some general conditions for a limit distribution theory of GMM estimation where the number of moment conditions increases with the sample size and explores some specific examples, including some cases where the instruments are irrelevant. Section 5 provides some extensions to the weighted GMM case, consideration of continuous updating estimators (CUEs) and a limited information maximum

likelihood (LIML) and panel model application. Concluding remarks are made in Section 6. Section 7 provides a notational summary and proofs and technical derivations are given in the Appendices.

2 Moment Conditions and GMM Estimation

Suppose we have an array of observable random vectors w_{ni}, $i = 1, \ldots, n$, whose dimension may vary with the sample size n and whose elements include dependent variables, independent variables and instrumental variables. One reason the variable dimension of w_{ni} is useful in practice is that we may wish to allow the number of potential instrumental variables to grow with n. Accordingly, we presume the existence of an array of real (nonrandom) functions $\{g_{nk};\, k = 1, \ldots, q_n,\, n = 1, 2, \ldots\}$ of these variables and parameters whose mean values constitute moment conditions of the form

$$(2) \quad Eg_{nk}(w_{ni}, \theta_0) = 0,\ i = 1, \ldots, n,\ k = 1, \ldots, q_n,\ n = 1, 2, \ldots$$

where θ_0 is a fixed p-vector of parameters to be estimated. In (2), q_n prescribes the dimension of the moment condition vector, which depends on n, as it does for example when the moment conditions correspond to increasing numbers of instrumental variables.

In some cases it is convenient and appropriate to assume that the observables w_{ni} are independently distributed across i for all n, in which case the $g_{nk}(w_{ni}, \theta)$ are independently distributed across i for each n. But even when there is independence in the observations over i, the functions $g_{nk}(w_{ni}, \theta)$ are not necessarily independent across k.

This framework is intended to be fairly general so that it encompasses many existing GMM models. It includes, for example, linear structural equation systems with increasing numbers of instrument variables such as the model of Bekker (1994). In this case, we have a linear model $y_i = x_i'\theta + \varepsilon_i$ relating the jointly dependent variables (y_i, x_i) and a potential array of instrumental variables z_{ki}^n that are uncorrelated with ε_i. The moment functions are then the cross products $z_{ki}^n(y_i - x_i'\theta)$, $k = 1, \ldots, q_n$, where q_n is the number of instruments, the variables w_{ni} in (2) are (y_i, x_i, z_{ki}^n) and we have

$$g_{nk}(w_{ni}, \theta) = z_{ki}^n(y_i - x_i'\theta),\ k = 1, \ldots, q_n.$$

In conventional GMM asymptotics q_n is fixed and the functional form of $g_{nk}(\cdot, \theta)$ and its moments do not depend on n. In Staiger and Stock (1997) the covariance structure of w_{ni} allows for local to zero correlation at a specific $\sqrt{n}$ rate between the regressors and the instruments, thereby accommodating weak instrumentation. In an analogous way, the GMM asymptotics in Stock and Wright (2000) permit the functions $g_{nk}(\cdot, \theta)$ to flatten out at the $\sqrt{n}$ rate, so that the moment conditions are weakly identifying. In addition, the approach used in Chao and Swanson (2002), Bekker (1994), Morimune (1983) and Anderson (1977) all allow for q_n to increase with n, as explained above. This brief synopsis of earlier research in relation to the present work is summarized for convenience in Table 1, where some of the notation is defined later.

	Dimension of w_{ni}	q_n	**Moment properties of $g_{nk}(w_{ni},\theta)$**	**Limit criterion function**
Conventional	fixed	fixed	fixed	nonrandom
Phillips (1989)	fixed	fixed	$\text{cov}(z_{ki}^n, x_i) = 0$	random
Staiger & Stock (1997)	fixed	fixed	local to zero: $\text{cov}(z_{ki}^n, x_i) = O(n^{-1/2})$	random
Stock & Wright (2000)	fixed	fixed	local to zero: $Eg_{nk}(w_{ni},\theta) = O\left(n^{-1/2}\right)$	random
Bekker (1994)	$O(q_n)$	$\frac{q_n}{n} \to \alpha$	Concentration parameter $\pi' z' z \pi = O(n)$	nonrandom
Chao & Swanson (2002)	$O(q_n)$	$q_n \to \infty$	$\pi' z' z \pi = O(r_n)$ $\frac{\sqrt{q_n}}{r_n} \to 0$	nonrandom
Han & Phillips (2004)	$O(q_n)$	$q_n \to \infty$ $\frac{q_n}{nc_n} \to \alpha$	Strength measured by c_n $\pi' z' z \pi \to \infty$,constant or zero	nonrandom & random

Table 1: Comparison of Different GMM Asymptotics (notation defined in Sections 2 and 3)

Unweighted GMM is defined by minimizing the criterion function

$$(3) \quad G_n(\theta) = \sum_{k=1}^{q_n} \bar{g}_{nk}(\theta)^2 = \bar{g}_n(\theta)' \bar{g}_n(\theta),$$

where $\bar{g}_{nk}(\theta) = n^{-1} \sum_{i=1}^n g_{nk}(w_{ni},\theta)$ and $\bar{g}_n(\theta) = (\bar{g}_{n1}(\theta), \ldots, \bar{g}_{nq_n}(\theta))'$. In what follows, we use notation like $\bar{g}_n(\theta)$, where the subscript k is eliminated, to signify the q_n-vector formed by taking the column vector of the relevant elements, such as $\bar{g}_{nk}(\theta)$. Later in the paper it will be convenient to use the scaled GMM objective function

$$(4) \quad f_n(\theta) = c_n^{-1} G_n(\theta) = c_n^{-1} \bar{g}_n(\theta)' \bar{g}_n(\theta),$$

which is constructed with a normalizing sequence c_n that is introduced in Assumption 2.

3 Convergence in Probability

It is helpful to fix some aspect of the moment conditions as a standard for analytical purposes. Accordingly, we find it useful on occasions to normalize the moment conditions by dividing by the square-root of the quantity

$$(5) \quad q_n^{-1} \sum_{k=1}^{q_n} E\left[n^{-1/2} \sum_{i=1}^n g_{nk}(w_{ni},\theta_0) \right]^2,$$

which is the average (over k) long run variance of the moment conditions evaluated at the true parameter θ_0. We assume that (5) is nonzero, which will be so unless all the moment conditions have zero (long run) variance at θ_0, which seems of little practical or theoretical interest. Of greater interest is the case where (5) converges to zero or diverges

to infinity. The normalization then forces the moment conditions to have unit average (long run) variance evaluated at the true parameter. We emphasize that this scaling does not have to be performed in practical applications, nor does it alter any aspects of the extremum estimation problem. It is also unnecessary for the theoretical development when the quantity (5) converges to a nonzero value. From now on, therefore, we assume that the moment functions either need no normalization or have already been normalized by this quantity. The implications of this convention are discussed further below.

We also assume that the random variables w_{ni} are *independent* across i. This assumption allows us to put regularity conditions in a convenient and conventional form. If the assumption is relaxed, then the regularity conditions and proofs change correspondingly, but the essential ideas developed in what follows remain unaffected.

Let $\xi_{nk}(w_{ni},\theta) = g_{nk}(w_{ni},\theta) - Eg_{nk}(w_{ni},\theta)$ and $\xi_n(w_{ni},\theta)$ be the column q_n-vector of $\xi_{nk}(w_{ni},\theta)$, $k = 1,\ldots,q_n$. Define

$$\zeta_{nk}(\theta) := n^{-1/2}\sum_{i=1}^{n}\xi_{nk}(w_{ni},\theta),$$

and let $\zeta_n(\theta)$ be the corresponding column q_n-vector. Under usual circumstances, $\zeta_{nk}(\theta)$ would follow a (functional) central limit theorem for each k. Let Θ be a subset of $\mathbb{R}^p$.

Our focus of interest is mainly on the role of the number and the weakness of the moment conditions in the asymptotics and the following assumption about the stochastic behavior of the sample moment functions helps in the development of the limit theory. A particularly important role is played by the variability of the signal over k, as measured by $E\zeta_n(\theta)'\zeta_n(\theta)/q_n$.

Assumption 1 (Signal Variability)

(*i*) *The eigenvalues of* $E\xi_n(w_{ni},\theta)\xi_n(w_{ni},\theta)'$ *are bounded from above by a universal constant for all* $\theta\in\Theta$*, for all* i*, and for all* q_n *and* n*;*

(*ii*) $E\xi_{nk}(w_{ni},\theta)^4 \le B < \infty$*, for all* $k = 1,\ldots,q_n$*, for all* $\theta\in\Theta$*, for all* i*, and for all* n*;*

(*iii*) $\delta_n(\theta) := q_n^{-1}E\zeta_n(\theta)'\zeta_n(\theta)$ *converges to* $\delta(\theta)$ *uniformly in* $\theta\in\Theta$*;*

(*iv*) *The sequence of random processes* $\max_{1\le k\le q_n}|\zeta_{nk}(\theta)|$ *is tight.*

Conditions (*i*) and (*ii*) require that fourth moments of $\xi_n(w_{ni},\theta)$ exist and that the second moment matrix has bounded eigenvalues. Note that condition (*i*) is not necessarily implied by (*ii*) and requires that the moment conditions $\xi_{nk}(w_{ni},\theta)$ be not too closely correlated across k. A sufficient condition for (*i*) is that the covariance structure is dominated as in

$$\sup_{\theta\in\Theta}E\xi_n(w_{ni},\theta)\xi_n(w_{ni},\theta)' \le aI_{q_n} + b_{q_n}b_{q_n}',$$

where a is some (possibly large) constant and the elements of the vector b_{q_n} are square summable, viz., $b_{q_n}'b_{q_n} = \sum_{k=1}^{q_n}b_k^2 \le \sum_{k=1}^{\infty}b_k^2 < \infty$.

When the moment conditions are rescaled by (5), we have $\delta_n(\theta_0)\equiv 1$ in condition (*iii*). In case the rescaling is unnecessary, we again have $\delta_n(\theta_0)$ converging to a nonzero value. So

if the $E\xi_{nk}(w_{ni},\theta)$ are uniformly continuous, then $\delta_n(\theta)$ is nicely uniformly bounded, and (*iii*) adds that it also converges uniformly. When the random variables are independent as assumed here, $\delta_n(\theta)$ equals $q_n^{-1}\sum_{j=1}^{q_n} n^{-1}\sum_{i=1}^{n} E\xi_{nk}(w_{ni},\theta)^2$, so the condition loosely means that the average long-run variance function of the moment conditions converges uniformly. Condition (*iv*) means that the centered and rescaled moment functions are uniformly tight, and it is satisfied if $\zeta_{nk}(\theta)$, as a sequence (indexed by n) of processes indexed by k and θ, follows a functional cental limit theorem. An important implication of this tightness is that $\max_{1\leq k\leq q_n}\zeta_{nk}(\theta)^2$ is also tight, and as a result, so is $q_n^{-1}\zeta_n(\theta)'\zeta_n(\theta)$ (see Lemma 19 in the Appendix).

As discussed in the Introduction, two sources of signal emanate from the moment conditions. The first of these, which we call the *main signal,* arises from the squared sum of the expected sample moment functions. It is comparable to the usual sample moment matrix in a regression model, which leads to the conventional persistent excitation condition for consistency in regression (Lai and Wei, 1982, theorem 1). The second signal source, which we call the *signal variability***,** arises from variation over the sample moment functions and is measured by $\delta_n(\theta)$. The strength of the main signal depends on how well the expected sample moment functions $\bar{m}_{nk}(\theta) := E\bar{g}_{nk}(\theta)$ separate the true parameter θ_0 from other parameter values and its role is well known (e.g., Stock and Wright, 2000). Stock and Wright work with a fixed number of moment conditions and embody the weakness of the signal strength in terms of the *individual* moment functions.

Our treatment allows for an increasing number of moment conditions and we therefore need a tool to express the *totality* of the strength of the main signal. The quantity c_n that is introduced next fulfils this role. Let $\bar{m}_n(\theta)$ be the q_n-vector of the $\bar{m}_{nk}(\theta)$.

Assumption 2 (Main Signal) *There is a sequence of positive numbers c_n such that*

(*i*) $\gamma_n(\theta) := c_n^{-1}\bar{m}_n(\theta)'\bar{m}_n(\theta) \to \gamma(\theta)$ *uniformly in* $\theta\in\Theta$*;*

(*ii*) $\alpha_n := q_n/nc_n \to \alpha\in[0,\infty)$ *and* $nc_n\to\infty$ *as* $n\to\infty$*;*

In view of (2), the function $\gamma_n(\theta)$ is minimized at θ_0 and its minimal value is zero. If the moment conditions alone identify θ_0, then θ_0 uniquely minimizes $\gamma_n(\theta)$. Whether these properties are retained as $n\to\infty$, is instrumental in determining the usual asymptotic behavior of the GMM estimator. Loosely speaking, c_n measures the order of magnitude of the main signal. Of course, $\{c_n\}$ is unique only up to a sequence that converges to a positive constant, but rescaling c_n does not alter any aspect of the convergence property of the GMM estimator.

The functions $\gamma_n(\theta)$ and $\delta_n(\theta)$ in Assumptions 2 and 1, respectively, incorporate the main signal from the expected sample moment functions and the collective signal variability over individual moment conditions up to some scaling factor. The main signal in $\gamma_n(\theta)$ obviously conveys what sample information there is (if any) on the true parameter θ_0 arising directly in the moment conditions (2), while information in $\delta_n(\theta)$ may or may not correspond to θ_0, in the sense that $\gamma_n(\theta)$ is minimized at θ_0 (at least when the specification is correct), but $\delta_n(\theta)$ may not be.

The decomposition

$$\begin{aligned} E\left[n^{1/2}\bar{g}_n(\theta)\right]'\left[n^{1/2}\bar{g}_n(\theta)\right] &= \left[n^{1/2}\bar{m}_n(\theta)\right]'\left[n^{1/2}\bar{m}_n(\theta)\right] + E\zeta_n(\theta)'\zeta_n(\theta) \\ &= nc_n\gamma_n(\theta) + q_n\delta_n(\theta) \\ &= nc_n\left\{\gamma_n(\theta) + \alpha_n\delta_n(\theta)\right\}, \quad \alpha_n = \frac{q_n}{nc_n} \end{aligned} \tag{6}$$

makes it possible to compare the relative amount of information in the main signal and the signal variability through the quantities $nc_n\gamma_n(\theta)$ and $q_n\delta_n(\theta)$. So, roughly speaking, the quantity α_n measures the relative strength of the signal variability compared with the main signal. Of course, as c_n may be scaled up or down by any sequence converging to a strictly positive value, the ratio α_n would be rescaled correspondingly and so the ratio is not an absolute measure.

The ratio q_n/nc_n may diverge to infinity in some situations, as when nc_n increases more slowly than q_n and $q_n \to \infty$. When this happens, the main signal (whose order of magnitude is nc_n, if $\gamma(\theta) = 0$ at $\theta = \theta_0$ uniquely) is asymptotically dominated by the signal variability (of order q_n), and the result is the same as the case $\gamma_n(\theta) \to 0$ and $\alpha_n \to \alpha > 0$. So the condition that the sequence α_n converges is not actually binding, because c_n can always be chosen to be large enough for α_n to converge and for $c_n^{-1}\bar{m}_n(\theta)'\bar{m}_n(\theta) \to 0$.

Note that $\{c_n\}$ may be chosen to diverge so fast that the above conditions are trivially satisfied in such a way that $\gamma_n(\theta) \to 0$, $\alpha_n \to 0$ and $nc_n \to \infty$. The asymptotic identification condition given next excludes this trivial possibility.

Following the decomposition (6), we define $\bar{f}_n(\theta) := \gamma_n(\theta) + \alpha_n\delta_n(\theta)$ and let $f_\infty(\theta) := \lim_{n\to\infty}\bar{f}_n(\theta) = \gamma(\theta) + \alpha\delta(\theta)$.

Assumption 3 (Asymptotic Identification) *The moment functions and the sequence (c_n) are such that the limit function $f_\infty(\theta) = \gamma(\theta) + \alpha\delta(\theta)$ is minimized at some unique parameter value $\theta^* \in \Theta$, i.e., for any $\varepsilon > 0$,*

$$\inf_{|\theta-\theta^*|>\epsilon} f_\infty(\theta) > f_\infty(\theta^*). \tag{7}$$

Condition (7) implies that θ^* uniquely minimizes $f_\infty(\theta)$ and is well separated from other parameter values, thereby implying a form of asymptotic identification. An important element in this condition is the potential contribution of the limiting signal variability, $\delta(\theta)$, which may be influential or even decisive in asymptotic identification. For example, in an extreme case where the expected sample moment functions have no information, i.e., when $\gamma(\theta) \equiv 0$, asymptotic identification (of θ^* not θ_0) may still be achieved through the presence of the $\delta(\theta)$ term.

It is worth noting that if there is no sequence $\{c_n\}$ that satisfies Assumption 3 and only a subset of parameters are asymptotically identified by $\gamma(\theta) + \alpha\delta(\theta)$ for some sequence $\{c_n\}$, then we would have an asymptotically partially identified model. In this case a nondegenerate limit for the unidentified parameters is expected, similar to the case Phillips (1989) analyzed for IV estimation of linear models.

A final set of conventional regularity assumptions are introduced to ensure the existence of the GMM estimator and to simplify derivations.

Assumption 4 (Standard Conditions)

(*i*) Θ *is a compact subset of* $\mathbb{R}^p$*;*

(*ii*) *For all* n *and for all* k, $E[g_{nk}(w_i,\theta)]$ *and* $\text{var}[g_{nk}(w_i,\theta)]$ *are continuous in* $\theta \in \Theta$*, and the criterion function (3) attains its minimum in* Θ*.*

When θ_0 is estimated under constraints, the Θ set may be regarded as the parameter set satisfying the constraints.

The first result states that, under these assumptions, the aggregate signal (i.e., the main signal plus the contribution from the signal variability) dominates the noise in extremum estimation and convergence in probability to a nonrandom limit occurs. Some modifications to the result (discussed in case II below) allow for weak identification under a finite number of weak moment conditions and corresponding weak convergence to a random limit. Proofs of this and later results are contained in the Appendix.

Theorem 5 (Convergence) *Under Assumptions 1, 2, 3 and 4,* $\hat{\theta} \to_p \theta^*$*, where* θ^* *is the unique minimizer of* $\gamma(\theta) + \alpha\delta(\theta)$ *on* Θ*.*

We consider the following three leading cases.

Case I: Conventional GMM Asymptotics. When q_n is fixed and the moment conditions are strong in the usual sense, we can choose $c_n \equiv 1$ (or any constant converging to a positive number). The strength of the signal carried by the expected sample moment function $\bar{m}_n(\theta) = E\bar{g}_n(\theta)$ is then, loosely speaking, fixed, while the contribution of the signal variability as well as the noise diminishes at a $\sqrt{n}$ rate. The assumptions for Theorem 5 are then satisfied with $\alpha = 0$, which implies that the GMM estimator converges in probability to the true parameter .

Case II: Weak Identification with a Fixed Number of Moment Conditions. In the usual weak moment condition case (c.f., Stock and Wright (2000)), it is assumed that $\bar{m}_{nk}(\theta)$ diminishes at a $\sqrt{n}$ rate and that q_n is fixed. In this case, a natural choice of c_n would seem to be n^{-1} (c.f. Assumption 6 below). But this sequence of c_n does not satisfy the second part of Assumption 2 (*ii*). When $c_n > O(n^{-1})$ instead, Assumptions 1 and 2 are all satisfied in such a way that $\gamma_n(\theta) \to 0$ and $\alpha_n \to 0$, but then the identification assumption 3 is violated. The aggregate signal does not dominate the noise and Theorem 5 fails. In this case, as is well known, the criterion function has a nondegenerate weak limit and the GMM estimator converges weakly to a nondegenerate distribution.

This case of weak moment identification with a fixed number of moment conditions, $q_n = q$, can be covered in the above theory by making some simple modifications to the assumptions and formulation. In particular, we can proceed as follows: (i) replace Assumption 1 (*iii*) by condition 6 (*i*) below; (ii) replace the second part of Assumption 2 (*ii*) by condition 6 (*ii*); and (iii) replace Assumption 2 (*i*) with condition 6 (*iii*) below. These changes allow for weak convergence of a standardized version of the GMM extremum condition.

Assumption 6 (Weak Identification)

(*i*) $\zeta_n(\theta) \to \zeta(\theta)$ *on* Θ*, where* $\zeta(\theta)$ *is a* q*-vector Gaussian stochastic process.*

(*ii*) $\alpha_n := q_n/nc_n \to \alpha \in (0,\infty)$ *and* $nc_n \to c \in (0,\infty)$ *as* $n \to \infty$*;*

(*iii*) $\frac{1}{\sqrt{c_n}}\bar{m}_n(\theta) \to m(\theta)$ *uniformly on* Θ*..*

Then

$$\begin{aligned} c_n^{-1}G_n(\theta) &= c_n^{-1}\bar{g}_n(\theta)'\bar{g}_n(\theta) \\ &= c_n^{-1}\left\{\bar{m}_n(\theta)+n^{-1/2}\zeta_n(\theta)\right\}'\left\{\bar{m}_n(\theta)+n^{-1/2}\zeta_n(\theta)\right\}) \\ &\to \left\{m(\theta)+c^{-1/2}\zeta(\theta)\right\}'\left\{m(\theta)+c^{-1/2}\zeta(\theta)\right\} := G(\theta), \end{aligned}$$

and by standard weak convergence arguments (e.g. van de Vaart and Wellner, 1996, theorem 3.2.2, p. 286) we obtain

$$\hat{\theta} \to \theta^* = \underset{\theta}{\operatorname{argmin}}\, G(\theta).$$

The unidentified case (c.f. Phillips, 1989) occurs as the special case where $m(\theta) = 0$. A further simple extension along the above lines covers the case where the parameter vector is partitioned as $\theta = (\theta_1', \theta_2')'$ and only the subvector θ_1 is weakly identified (c.f. Stock and Wright, 2000).

Case III: Many Weak Moment Conditions. When there are many (q_n) moment conditions each of which is weak (as in Stock and Wright, 2000), we can set $c_n = q_n/n$. A special case occurs when each moment condition delivers no signal, i.e., if $\bar{m}_{nk}(\theta) = 0$ for all k. In that case, the moment conditions may be said to be irrelevant, and any non-zero sequence c_n satisfies Assumption 2 (*i*) with limit $\gamma(\theta) = 0$. However, Assumption 2 (*ii*) requires that $c_n = O(q_n/n)$ and, in view of Assumption 3, c_n should be chosen such that α_n converges to a nonzero constant, so that $\delta(\theta)$ enters the limit function $f_\infty(\theta)$. The simplest such sequence would be $c_n = q_n/n$. In this event, we have $\alpha_n \equiv 1$, $\gamma_n(\theta) = 0$ and the limit function $\gamma(\theta) = 0$. In the case where the moment conditions are weak but not irrelevant, i.e., $\bar{m}_{nk}(\theta) \neq 0$, we have $\gamma_n(\theta) \neq 0$ and possibly $\gamma(\theta) \neq 0$. Then, the aggregate signal from the two alternative sources asymptotically dominates the noise and convergence in probability to a nonrandom limit is obtained. Of course, the limit may not equal the true parameter identified by the expected moment functions because the main signal is contaminated in the limit by the alternative signal which may not reinforce the main signal.

The following Corollary states sufficient conditions under which the GMM estimator is consistent and is a straightforward consequence of the above result.

Corollary 7 *Under the assumptions of Theorem 5,* $\hat{\theta} \to_p \theta_0$ *if* $\gamma(\theta)$ *is minimized uniquely at* θ_0 *and* $\alpha = 0$*, or if* $\delta(\theta)$ *is minimized at* θ_0*.*

The following examples help to illustrate the above results and some of the many possibilities that can arise in modeling with weak instrumentation.

Example 8 (Location Model and Weak Instrumentation) Let y_i be an *i.i.d.* sequence with mean θ_0 and variance σ^2. Suppose that θ_0 is estimated by IV estimation using instruments z_{ki}^n, which are *i.i.d.* for each n with $Ez_{ki}^n = n^{-1/2}r$ and $var(z_{ki}^n) = \sigma_z^2$ for all n. The mean

of z_{ki}^n is assumed to be local-to-zero, which suggests that the instruments are weak (for an intercept), and the z_{ki}^n are assumed to be independent of y_i. In this simple location model, of course, OLS estimation (i.e., the sample mean of y_i) is consistent and IV estimation is not needed. Nonetheless, this example serves to illustrate some interesting features of estimation with large numbers of instruments, including the possibility of consistent estimation with apparently irrelevant instruments (when $r = 0$). This example of weak instrumentation is intriguing because it shows that instruments can carry useful information in unexpected ways.

Consider the moment conditions

$$(8) \quad g_{nk}(w_{ni}, \theta) = z_{ki}^n(y_i - \theta),\ k = 1, \ldots, q_n,$$

where $w_{ni} = (y_i, z_{1i}^n, \ldots, z_{q_n i}^n)$. All the above assumptions are satisfied and we have

$$Eg_{nk}(w_{ni}, \theta) = -n^{-1/2} r(\theta - \theta_0),$$

and

$$(9) \quad \text{var}\left(g_{nk}(w_{ni}, \theta)\right) = \sigma^2\sigma_z^2 + \sigma_z^2(\theta - \theta_0)^2 + n^{-1}r^2\sigma^2.$$

Therefore, with the choice $c_n = q_n/n$, we have $\alpha_n \equiv 1 = \alpha$, and

$$\gamma_n(\theta) = r^2(\theta - \theta_0)^2, \quad \delta_n(\theta) = \sigma^2\sigma_z^2 + \sigma_z^2(\theta - \theta_0)^2 + n^{-1}r^2\sigma^2.$$

Hence,

$$(10) \quad \gamma(\theta) + \alpha\delta(\theta) = r^2(\theta - \theta_0)^2 + \sigma^2\sigma_z^2 + \sigma_z^2(\theta - \theta_0)^2,$$

which is minimized uniquely at $\theta = \theta_0$ regardless of the value of r (including $r = 0$) if $\sigma_z^2 > 0$. Therefore, the GMM estimator using (8) is consistent whenever $\sigma_z^2 > 0$. Interestingly, even if $\sigma_z^2 = 0$, we still have consistency if $r \neq 0$. With $\sigma_z^2 = 0$ and $r \neq 0$, we have $z_{ki}^n \equiv Ez_{ki}^n = n^{-1/2}r$ and the q_n moment functions reduce to the single moment function $n^{-1/2}r(y_i - \theta)$, corresponding to a constant instrument. GMM estimation using this single moment condition is equivalent to estimation using $r(y_i - \theta)$ as the moment condition, which is handled in our framework according to conventional GMM asymptotics. In effect, the GMM estimator is $\hat{\theta} = \bar{y}$, the sample mean of y_i, and is consistent for $Ey_i = \theta_0$. If both $\sigma_z^2 = 0$ and $r = 0$, the moment conditions are empty and the GMM estimator is not defined.

To understand how $\hat{\theta}$ is consistent for θ_0 when the instruments are independent random quantities, let $y_{nk}^* = \frac{1}{\sqrt{n}}\sum_{i=1}^n z_{ki}^n y_i$, $x_{nk}^* = \frac{1}{\sqrt{n}}\sum_{i=1}^n z_{ki}^n$ and $u_{nk}^* = \frac{1}{\sqrt{n}}\sum_{i=1}^n z_{ki}^n(y_i - \theta_0)$. Then clearly,

$$(11) \quad y_{nk}^* = \theta_0 x_{nk}^* + u_{nk}^*,\ k = 1, \ldots, q_n.$$

Note that $Ex_{nk}^* = r$, $Ex_{nk}^{*\,2} = r^2 + \sigma_z^2$, $Eu_{nk}^* = 0$, $Eu_{nk}^{*\,2} = \sigma^2\sigma_z^2$ and $Ex_{nk}^* u_{nk}^* = 0$, and that the x_{nk}^* are independent across k but the u_{nk}^* are not. We can easily verify that

$$\hat{\theta} = (x_n^{*\prime}x_n^*)^{-1}x_n^{*\prime}y_n^* = \theta_0 + (q_n^{-1}x_n^{*\prime}x_n^*)^{-1}q_n^{-1}x_n^{*\prime}u_n^*$$

where x_n^*, y_n^* and u_n^* are q_n-vectors of x_{nk}^*, y_{nk}^* and u_{nk}^* respectively, i.e., $\hat{\theta}$ is the OLS estimator of θ_0 in (11). It can be shown that $q_n^{-1}x_n^{*\prime}x_n^* \to_p r^2 + \sigma_z^2$ and $q_n^{-1}x_n^{*\prime}u_n^* \to_p 0$. Therefore, $\hat{\theta} \to_p \theta_0$ as long as $r^2 + \sigma_z^2 > 0$ as $q_n \to \infty$. Furthermore, this analysis appears to suggest that the rate of convergence is $\sqrt{q_n}$ because q_n serves as the effective "sample size". However, the $\sqrt{q_n}$ rate is not entirely obvious because the u_{nk}^* are not independent across k and the development of a limit distribution theory is more challenging. See the next section for more discussion on this point.

As indicated, $\hat{\theta}$ is consistent even when $r = 0$, provided $\sigma_z^2 > 0$. This case is especially interesting because the instruments z_{ki}^n have zero mean and therefore appear to be totally irrelevant for estimating an intercept parameter like θ. In fact, the GMM estimator is inconsistent if $q_n = q$ is fixed (see below). However, when $\sigma_z^2 > 0$, the instruments z_{ki}^n take non zero values with positive probability and, since there are an infinite number of them as $q_n \to \infty$, the instruments end up providing enough leverage for consistent estimation (although the resulting estimator is infinitely deficient in comparison to the sample mean). The leverage is revealed by the fact that the variance of the moment conditions (9) has value $\sigma^2\sigma_z^2 + \sigma_z^2(\theta - \theta_0)^2$ when $r = 0$, which is informative about θ. As we have seen earlier, this variance figures in the limiting value of the objective function (10) because of role played by the variability of the signal over k, as measured by $E\zeta_n(\theta)'\zeta_n(\theta)/q_n$.

On the other hand, suppose that $q_n = q$ is fixed. Then, by standard central limit theory, the q collection (x_{nk}^*, u_{nk}^*), $k = 1, \ldots, q$, converges in distribution to joint normal random variables $(\tilde{x}_k, \tilde{u}_k)$, $k = 1, \ldots, q$, as $n \to \infty$, where $E\tilde{x}_k = r$, $\text{var}(\tilde{x}_k) = \sigma_z^2$, $E\tilde{x}_k\tilde{x}_j = 0$ for $k \neq j$, $E\tilde{u}_k = 0$, $E\tilde{u}_k^2 = \sigma^2\sigma_z^2$, $E\tilde{u}_k\tilde{u}_j = 0$ for $k \neq j$, and $E\tilde{x}_k\tilde{u}_k = 0$. Therefore, by continuous mapping we have

$$\hat{\theta} \to \theta_0 + (\tilde{x}'\tilde{x})^{-1}\tilde{x}'\tilde{u},$$

where $\tilde{x}$ and $\tilde{u}$ are q-vectors of the $\tilde{x}_k$ and $\tilde{u}_k$, respectively. When $r = 0$, this is a scaled and translated t distribution, as shown in Phillips (1989). ■

In the next example, the GMM estimator is generally (but not always) inconsistent for θ_0 because the moment conditions are too weak in relation to the endogeneity of the structural equation. The example highlights the trade-off between the degree of endogeneity and the quality of the instrumentation that is needed for the successful estimation of a structural equation. This trade-off was mentioned in the Introduction and plays a role in the exact finite sample distribution theory of the simultaneous equations model. The present example shows that the trade-off is also manifest in the limit theory and, moreover, that consistent estimation of structural systems even with irrelevant instruments (if there are increasing numbers of them) is possible when the degree of endogeneity is local to zero.

Example 9 (Linear Structural Equation Estimation) Consider the linear model

$$(12) \quad y_i = \theta x_i + \varepsilon_i, \; i = 1, \ldots, n,$$

where the regressors x_i may be correlated with ε_i and the (x_i, ε_i) are iid across i. For simplicity, set $Ex_i = 0$ and $E\varepsilon_i = 0$. Denote $\sigma_x^2 = Ex_i^2$, and $\sigma_\varepsilon^2 = E\varepsilon_i^2$. Suppose that there are available q_n instrumental variables z_{ki}^n which are valid in the sense that they satisfy the

orthogonality condition $E(\varepsilon_i|z_{ki}^n) = 0$, and which are *iid* over both i and k with $Ez_{ki}^n = 0$, $Ez_{ki}^{n\,4} < \infty$ and $Ez_{ki}^{n\,2} = \sigma_z^2$ for all n. Let $E(x_i|z_{ki}^n) = \pi_k^n z_{ki}^n$ for all n and j, and then for all n, $E(x_i|z_{ki}^n, k = 1, \ldots, q_n) = \sum_{k=1}^{q_n} \pi_k^n z_{ki}^n$. Let

$$v_i^n = x_i - E(x_i|z_{ki}^n, j = 1, \ldots, q_n) = x_i - \sum_{k=1}^{q_n} \pi_k^n z_{ki}^n.$$

Suppose that for all n, $E(v_i^{n2}|z_{ki}^n, j = 1, \ldots, q_n) = Ev_i^{n2} = \sigma_{v,n}^2$, which means that the expected variation of x_i around its conditional mean does not depend on the conditioning variables though it may depend on the totality of them (for instance, by way of their moments). Assume also that

$$(13)\quad E(x_i\varepsilon_i|z_{ki}^n, k = 1, \ldots, q_n) = E(v_i^n\varepsilon_i|z_{ki}^n, k = 1, \ldots, q_n) = \rho_n,$$

and $E(\varepsilon_i^2|z_{ki}^n, k = 1, \ldots, q_n) = \sigma_\varepsilon^2$ for all n. With no surprise, we then have $\sigma_{v,n}^2 = \sigma_x^2 - \sum_{k=1}^{q_n} \pi_k^{n2}\sigma_z^2$.

In the weak instrument case, we assume that the coefficients π_k^n in the reduced form behave in such a way that

$$(14)\quad A_n^2 := \frac{n}{q_n}\sum_{k=1}^{q_n} \pi_k^{n2} \to A^2 > 0$$

(so that the instruments z_{ki}^n are "mildly weak" in the sense of Chao and Swanson, 2002). Then, $\sigma_{v,n}^2 = \sigma_x^2 - c_n A_n^2 \sigma_z^2$. We also assume that the endogeneity measure ρ_n in (13) satisfies $\rho_n \to_{a.s.} \rho$ as $n \to \infty$. In doing so, we allow for the (conventional) case where $\rho > 0$ as well as the case where $\rho = 0$, in which the endogeneity of the system may be described as being local to zero.

Now consider the q_n moment conditions

$$g_{nk}(w_{ni}, \theta) = z_{ki}^n(y_i - \theta x_i) = z_{ki}^n\varepsilon_i - (\theta - \theta_0)z_{ki}^n x_i,$$

where w_{ni} contains all the observable variables y_i, x_i, z_{ki}^n, $k = 1, \ldots, q_n$. Clearly,

$$Eg_{nk}(w_{ni}, \theta) = -(\theta - \theta_0)\pi_k^n\sigma_z^2,$$

so when we choose $c_n = q_n/n$, we have $\alpha_n \equiv 1$ and $\gamma_n(\theta) = (\theta - \theta_0)^2\sigma_z^4 A_n^2 \to \gamma(\theta) = (\theta - \theta_0)^2\sigma_z^4 A^2$. Further,

$$\begin{aligned}\operatorname{var} g_{nk}(w_{ni}, \theta) &= \operatorname{var}(z_{ki}^n\varepsilon_i) - 2(\theta - \theta_0)\operatorname{cov}(z_{ki}^n\varepsilon_i, z_{ki}^n x_i) \\ &\quad + (\theta - \theta_0)^2\operatorname{var}(z_{ki}^n x_i) \\ &= \sigma_z^2\{\sigma_\varepsilon^2 - 2(\theta - \theta_0)\rho_n + (\theta - \theta_0)^2[\sigma_x^2 + \pi_k^{n2}(\kappa_z^4 - 1)\sigma_z^2]\},\end{aligned}$$

where $\kappa_z^4 = E(z_{ki}^n/\sigma_z)^4 < \infty$ is assumed, and we have $\delta(\theta) = \sigma_z^2[\sigma_\varepsilon^2 - 2(\theta - \theta_0)\rho + (\theta - \theta_0)^2\sigma_x^2]$ because $q_n^{-1}\sum_{k=1}^{q_n} \pi_k^{n2} \to 0$ as $q_n \to \infty$ by virtue of the weak instrument condition (14). It follows that $\gamma(\theta) + \delta(\theta) = \sigma_z^4 A^2(\theta - \theta_0)^2 + \sigma_z^2[\sigma_\varepsilon^2 - 2(\theta - \theta_0)\rho + (\theta - \theta_0)^2\sigma_x^2]$, and therefore

$$(15)\quad \hat{\theta} \to_p \theta_0 + \frac{\rho}{\sigma_z^2 A^2 + \sigma_x^2},$$

which is the minimizer of $\gamma^*(\theta) + \delta^*(\theta)$, if the limit is in Θ.

To reiterate the argument in another form, we may use the variables y^*_{nk}, x^*_{nk}, and u^*_{nk} that were introduced at the end of Example 8, and correspondingly change the previous definition of x^*_{nk} to $n^{-1/2}\sum_{i=1}^n z^n_{ki}x_i$. Due to the correlation between x_i and ε_i, we then have $Ex^*_{nk}u^*_{nk} = E({z^n_{ki}}^2 x_i\varepsilon_i) = \sigma_z^2\rho_n$, which leads naturally to the inconsistency of the GMM estimator.

The inconsistency $\rho/\left(\sigma_z^2A^2 + \sigma_x^2\right)$ of $\hat{\theta}$ in (15) is a ratio that depends on ρ, A^2 and the variance parameters. Roughly speaking, this ratio measures the limiting endogeneity in the system (ρ) relative to the strength of the instruments (A^2). In doing so, the ratio quantifies the trade-off between the extent of the endogeneity in a simultaneous system and the quality of the instruments needed for a 'good' estimate asymptotically. Observe that if $\rho = 0$, then $\hat{\theta}$ is consistent and will be so even if $A^2 = 0$. Thus, even when the instruments are totally irrelevant for the regressors x_i (that is, when $\pi^n_k = 0$ and z^n_{ki} is independent of x_i), these instruments still produce a consistent estimate of θ_0 provided there are increasing numbers of them and provided the endogeneity in the system is local to zero in the sense that $\rho_n \to 0$ as $n \to \infty$. This result is provides an interesting structural equation extension of the consistency result for many irrelevant instruments given in Example 8. Of course, consistency also holds when $A^2 = \infty$ and the quality of the instruments is high relative to the limiting endogeneity in the system. ■

4 Asymptotic Distribution Theory

Define the normalized objective function $f_n(\theta) = c_n^{-1}\bar{g}_n(\theta)'\bar{g}_n(\theta)$ as in (4) and let $\bar{f}_n(\theta) = \gamma_n(\theta) + \alpha_n\delta_n(\theta)$, as before. The previous section established that under suitable conditions

$$\begin{aligned} &f_n(\theta) - \bar{f}_n(\theta) \to_p 0, \text{ uniformly in } \theta, \\ &\bar{f}_n(\theta) \to f_\infty(\theta) = \gamma(\theta) + \alpha\delta(\theta), \text{ uniformly in } \theta, \end{aligned}$$

and the minimizing value $\hat{\theta} = \arg\min_\theta f_n(\theta)$ converges in probability to the minimizing value $\theta^* = \arg\min_\theta f_\infty(\theta)$. This section examines the asymptotic distribution of $\hat{\theta}$. Since the limit θ^* is not necessarily equal to the true parameter value $\theta_0 = \arg\min_\theta \gamma(\theta)$, appropriate centering as well as rescaling is required for a complete development. The situation is analogous in this respect to the development of limit theory under conditions that allow for possible misspecification. However, in the present context, the recentering arises from the fact that the moment conditions may not be sufficiently informative about the parameters of interest to secure consistency even though there may be an increasing number of such conditions. As the examples just discussed and those considered below in Section 5 illustrate, there are many situations of this type in structural equation and panel data modeling where the moment conditions are plentiful but may or may not be strong enough to secure consistent estimation to the true parameter. To handle such cases at a reasonable level of generality, we need a framework that will allow for increasing numbers of moment conditions and potential inconsistencies arising from the weakness of these conditions. The development below uses epiconvergence techniques to help achieve this level of generality in a reasonably straightforward manner.

A common starting point in developing an asymptotic distribution theory for an extremum estimator is to define a centred stochastic process based on the objective function and study its limit behavior. In the present case, we define

$$(16)\quad h_n(t) := r_n^2[f_n(\bar{\theta} + t/r_n) - f_n(\bar{\theta})]$$

constructed from the objective function f_n with an appropriate centering parameter $\bar{\theta}$ and a suitable rate of convergence sequence r_n. In his development of constrained M-estimators, Geyer (1994) uses this approach with $\sqrt{n}$ for r_n and θ_0 for $\bar{\theta}$. (In general, the leading factor r_n^2 in (16) does not have to be the square of the rate of convergence r_n, but it is so in our limit theory.) In conventional GMM, $\bar{\theta}$ is θ_0 and r_n is $\sqrt{n}$, just as in Geyer (1994). However, to achieve the required degree of generality in our framework, we have to allow for both the centering and the convergence rate to be nonstandard. Once (16) is formulated, the usual approach is to invoke a general result on the weak convergence of argmax or argmin functionals for which epiconvergence is helpful. The following theorem of Knight (2003, theorem 1) is particularly useful in this regard.

Proposition 10 (Knight, 2003) *Suppose that $h_n(t)$ epiconverges in distribution to $h(t)$ and*

(i) $\tilde{t}_n$ is such that $h_n(\tilde{t}_n) \leq \inf_t h_n(t) + o_p(1)$;

(ii) $\tilde{t}_n = O_p(1)$;

(iii) $h(t)$ has an almost sure unique minimizer $\tilde{t}$.

Then $\tilde{t}_n \to \tilde{t}$.

The theorem is proved and the meaning of epiconvergence in distribution is fully discussed in Knight (2003) (see also Geyer, 1994). Epiconvergence and similar concepts in a nonstochastic environment are considered in Rockafellar and Wets (1998). Epiconvergence is a form of convergence that is particularly useful in the context of function optimization, making it well suited to extremum estimation problems. Thus (see Geyer, 1994, Proposition 3.1), if f_n epiconverges to f, $x_n \to x$, and $f_n(x_n) = \inf_y f_n(y) + o(1)$, then $f(x) = \inf_y f(y) = \lim_{n\to\infty} f_n(x_n)$, thereby preserving optimization in the limit.

The above proposition is more flexible in application than well-known results such as theorem 3.2.2 of van der Vaart and Wellner (1996) or Theorem 2.7 of Kim and Pollard (1990) on the weak convergence of argmax/argmin functionals which have been used in the econometric literature in the past. In particular, Knight's result readily accommodates restrictions on the parameter space Θ. It is therefore particularly appropriate for developing a limit theory in a general moment condition context.

To fix ideas, first assume that a second order Taylor series expansion of $f_n(\theta)$ is permissable and takes the form

$$(17)\quad f_n(\theta) = f_n(\bar{\theta}) + \nabla f_n(\bar{\theta})'(\theta - \bar{\theta}) + \tfrac{1}{2}(\theta - \bar{\theta})'\nabla^2 f_n(\bar{\theta})(\theta - \bar{\theta}) + R_{2n}(\theta, \bar{\theta}),$$

where $\nabla f_n(\cdot)$ and $\nabla^2 f_n(\cdot)$ are the first and the second derivative arrays of $f_n(\cdot)$ respectively, and $R_{2n}(\theta, \bar{\theta})$ is a remainder term satisfying

$$\lim_{\theta \to \bar{\theta}} \limsup_{n \to \infty} P\left(\frac{R_{2n}(\theta, \bar{\theta})}{|\theta - \bar{\theta}|^2} > \epsilon \right) = 0,$$

for any $\epsilon > 0$. This expansion leads to the expression

$$h_n(t) = r_n \nabla f_n(\bar{\theta})' t + \tfrac{1}{2} t' \nabla^2 f_n(\bar{\theta}) t + o_p(1).$$

Looking at the hessian matrix first, we observe that convergence of $\nabla^2 f_n(\bar{\theta})$ in an appropriate mode is generally a consequence of the smoothness of $f_n(\theta)$, since $f_n(\cdot) \to_p f_\infty(\cdot)$ uniformly. However, the term involving $r_n \nabla f_n(\bar{\theta})$ usually converges in distribution as a result of some central limit theory or weak convergence argument and presents much more difficulty in the present context because of the nonstandard convergence rate, which relies on the number of moment conditions, and the dependence properties among the moment conditions.

In conventional GMM, we have $E\nabla f_n(\theta_0) \equiv 0$, where standard regularity conditions allow the interchange of integration and differentiation. Consequently, $r_n \nabla f_n(\bar{\theta})$, which becomes $\sqrt{n} \nabla f_n(\theta_0)$ upon setting $r_n = \sqrt{n}$ and $\bar{\theta} = \theta_0$, constitutes a sequence of *zero mean* random variables that are stochastically bounded, and the term converges in distribution under regularity conditions that permit the use of standard central limit theory. In the present case, however, $E\nabla f_n(\theta_n^*)$ rather than $E\nabla f_n(\theta^*)$ is zero, and, for any fixed $\bar{\theta}$, $E\nabla f_n(\bar{\theta})$ may be different from 0. Though the discrepancy diminishes to 0 as n increases with the choice of $\bar{\theta} = \theta^*$, the amplified mean $Er_n \nabla f_n(\bar{\theta})$ may be significantly different from 0 or even divergent so that the sequence $r_n \nabla f_n(\bar{\theta})$ may not be tight.

To address this difficulty, we choose the sequence $\theta_n^* = \arg\min_\theta \bar{f}_n(\theta)$ as the centering sequence for $\hat{\theta}$ and correspondingly define $h_n(t)$ as

$$h_n(t) := r_n^2 [f_n(\theta_n^* + t/r_n) - f_n(\theta_n^*)], \tag{18}$$

for some r_n, in place of (16). Naturally $h_n(t)$ is minimized at $\tilde{t}_n := r_n(\hat{\theta} - \theta_n^*)$ and $\theta_n^* \to \theta^*$. We establish the limit distribution of $r_n(\hat{\theta} - \theta_n^*)$, whose center drifts with n in a form of *moving center asymptotics* like that of local power asymptotics. Unlike $r_n \nabla f_n(\theta^*)$, the term $r_n \nabla f_n(\theta_n^*)$ has zero mean and under suitable regularity conditions may converge in distribution.

The rate of convergence r_n is closely related to the convergence rate of $f_n(\theta)$ and the modulus of continuity of the properly rescaled function. We simplify and modify Theorem 5.52 of van der Vaart (1998) to accommodate the moving center asymptotics in this application. Let $W_n(\cdot) = f_n(\cdot) - \bar{f}_n(\cdot)$. The notation '$A(\theta) \lesssim B(\theta)$' means that $A(\theta)$ is bounded from above by $B(\theta)$ up to a finite universal constant (and $B(\theta) \gtrsim A(\theta)$ has the same meaning).

Proposition 11 (Rate of Convergence) *Suppose that there exist an $n_0 < \infty$ and a neighborhood of θ^* such that for every θ in the neighborhood,*

$$\bar{f}_n(\theta) - \bar{f}_n(\theta_n^*) \gtrsim |\theta - \theta_n^*|^2, \quad n > n_0. \tag{19}$$

Suppose also that there exists some $\delta_0 > 0$ such that for all n and for all $0 < \delta \leq \delta_0$, $r_n W_n(\cdot)$ has the following modulus of continuity:

$$(20) \quad E \sup_{|\theta - \theta_n^*| < \delta} |r_n W_n(\theta) - r_n W_n(\theta_n^*)| \lesssim \delta.$$

Then

$$r_n(\hat{\theta} - \theta_n^*) = O_p(1).$$

Condition (19) is satisfied if $\bar{f}_n(\cdot)$ has a nonsingular second derivative matrix and its curvature is uniformly bounded away from singularity for $n > n_0$. It is not necessarily implied by the corresponding condition for $f_\infty(\cdot)$, i.e., the condition that $f_\infty(\theta) - f_\infty(\theta^*) \gtrsim |\theta - \theta^*|^2$, even though $\bar{f}_n(\cdot) \to f_\infty(\cdot)$ uniformly and $\theta_n^* \to \theta^*$. To show this, a simple counter example defined on $\Theta = [-1, 1]$ is given by the following sequence $\bar{f}_n(\theta)$ and limit function $f_\infty(\theta)$:

$$\bar{f}_n(\theta) = \theta^2\{|\theta| > n^{-1}\} + (\theta^4 + n^{-2} - n^{-4})\{|\theta| \leq n^{-1}\}, \quad f_\infty(\theta) = \theta^2.$$

In this case, both $\bar{f}_n(\cdot)$ and $f_\infty(\cdot)$ are continuous, $\theta^* = 0$, $\theta_n^* \equiv 0$ and $\bar{f}_n(\cdot) \to f_\infty(\cdot)$ uniformly (the uniform distance is $n^{-2} - n^{-4}$). Furthermore, the limit function $f_\infty(\cdot)$ satisfies that $f_\infty(\theta) - f_\infty(0) \geq |\theta - 0|^2$ in any open neighborhood of 0. Nevertheless, $\bar{f}_n(\cdot)$ does not satisfy (19) in any neighborhood of 0.

If the quadratic order on the right of (19) changes to another order, then the rate of convergence will change accordingly. That is, a lower order gives faster convergence.

Regarding condition (20), the scale factor r_n can be chosen such that $r_n W_n(\cdot)$ remains tight and its distribution is non-degenerate. Condition (20) means that the modulus of continuity of the noise function rescaled this way is 'Lipschitz in parameters' with exponent unity in the stochastic sense. Of course, this result is not as general as Theorem 5.52 of van der Vaart (1998), but it is useful for our development here and is implied by further low-level smoothness assumptions, especially, the first order Taylor series expandability combined with other regularities. Again, if the order of the right hand side of (20) (the modulus of continuity) changes, then the rate of convergence correspondingly changes. As the order increases, the rate of convergence becomes faster.

Note that centering at θ_n^* and defining $W_n(\cdot)$ as the residual around the mean of $f_n(\cdot)$ in the above proposition are essential. For, if we re-define $W_n(\cdot) = f_n(\cdot) - f_\infty(\cdot)$, then condition (20) is not generally satisfied (unless $|\bar{f}_n(\theta) - f_\infty(\theta)|$ shrinks to zero faster than r_n); or, if we kept the current definition of $W_n(\cdot)$ and centered at θ^* instead, then condition (19) is unlikely to hold; or if we assumed (19) and (20) and centered at θ^* then the proof breaks down (in particular, equation (50) of the Appendix A fails).

Naturally, similar results could be established by more traditional methods such as use of the Taylor expansion (17), scaling by nc_n and with θ_n^* replacing $\bar{\theta}$. But we choose epi-convergence arguments to secure the most generality and convenience, so that the present method is applicable with little modification to specific problems possibly involving non-smooth functions or constraints.

With these tools in hand, we make the following 'high level' assumptions on $f_n(\cdot)$. Of course, it is also of interest to find 'low level' regularity conditions which are sufficient for

this asymptotic development in particular cases. But the issues involved are sufficiently complicated to make a general treatment difficult, and it seems more appropriate to leave such developments to later work where further specific applications of the present method are considered.

Assumption 12 *The $f_n(\cdot)$ function satisfies the following conditions.*

(*i*) *$\bar{f}_n(\cdot)$ permits a local quadratic approximation of the form*

$$\bar{f}_n(\theta) = \bar{f}_n(\theta_n^*) + \tfrac{1}{2}(\theta - \theta_n^*)' V_n(\theta_n^*)(\theta - \theta_n^*) + \bar{R}_{2n}(\theta), \tag{21}$$

where $V_n(\theta_n^*)$ converges to a positive definite matrix V and the residuals $\bar{R}_{2n}(\cdot)$ are such that for any $\epsilon > 0$ there exists a neighborhood B of θ^* such that

$$\limsup_{n\to\infty} \sup_{\theta\in B} \frac{|\bar{R}_{2n}(\theta)|}{|\theta - \theta_n^*|^2} < \epsilon; \tag{22}$$

(*ii*) *$W_n(\cdot)$ permits a local linear approximation of the form*

$$W_n(\theta) = W_n(\theta_n^*) + \Delta_n(\theta_n^*)'(\theta - \theta_n^*) + R_{1n}(\theta), \tag{23}$$

where

$$(nc_n)^{1/2}\Delta_n(\theta_n^*) \to Z, \tag{24}$$

for some random vector Z; and, for any $\epsilon > 0$ and $\eta > 0$, there exists a neighborhood B of θ^* such that

$$\limsup_{n\to\infty} P\left(\sup_{\theta\in B} \frac{|(nc_n)^{1/2} R_{1n}(\theta)|}{|\theta - \theta_n^*|} \geq \eta\right) \leq \epsilon. \tag{25}$$

Also, there is a neighborhood B_1 and a finite number $\tilde{M}_1$ such that

$$E \sup_{\theta\in B_1} (nc_n)^{1/2} |W_n(\theta) - W_n(\theta_n^*)| \leq \tilde{M}_1 \text{ for all } n; \tag{26}$$

(*iii*) *The sequence of sets $(nc_n)^{1/2}(\Theta - \theta_n^*)$ converges to a closed set $T(\theta^*)$ in the sense of Painlevé-Kuratowski set convergence.*

Some comments on these conditions are in order. First, we remark that condition (*i*) is analogous to Assumption A of Geyer (1994). The only difference is that in our case the regularity is imposed on the sequence of the mean processes, which is necessary because the number of moment conditions is possibly increasing and their distributions are unlikely to be identical. The restriction (22) on the remainder term is a natural generalization because of the increasing number of moment conditions and its further complication relates to the presence of moving center asymptotics around θ_n^*.

Condition (i) requires twice continuous differentiability of $\bar{f}_n(\theta)$, which seems mild especially since $\bar{f}_n(\theta) = \gamma_n(\theta) + \alpha_n\delta_n(\theta)$ and the components $\gamma_n(\theta)$ and $\delta_n(\theta)$ involve functions of expectations. Also, Taylor's theorem implies that for any $\epsilon > 0$, there is a neighborhood B_n such that

$$\sup_{\theta\in B_n} |\bar{R}_{2n}(\theta)|/|\theta - \theta_n^*|^2 \leq \epsilon,$$

for each n. Equation (22) goes further and requires that the continuity be uniform, which again seems mild because $\bar{f}_n(\cdot)$ converges uniformly. It should be noted, as in Geyer (1994), that the first derivative term of this expansion is set to zero as θ_n^* (in Θ) satisfies the first order conditions for minimizing $\bar{f}_n(\theta)$.

The nonsingularity requirement on V in condition (i) is conventional. If V were singular, then higher order expansions would be needed and the rate of convergence and other related asymptotic properties would be affected correspondingly.

Condition 12 (ii) is more easily understood by considering the form of $(nc_n)^{1/2}W_n(\cdot)$, viz.,

$$(27)\quad (nc_n)^{1/2}W_n(\theta) = -2c_n^{-1/2}\bar{m}_n(\theta)'\zeta_n(\theta) + \alpha_n^{1/2}q_n^{-1/2}\sum_{k=1}^{q_n}[\zeta_{nk}(\theta)^2 - E\zeta_{nk}(\theta)^2].$$

For each θ, if the moment conditions are not too closely correlated in the sense that

$$(28)\quad \left|\frac{1}{q_n}\sum_{k\neq l}\operatorname{cov}\left[\zeta_{nk}(\theta)^2, \zeta_{nl}(\theta)^2\right]\right| < \infty,$$

then (27) has a bounded second moment under Assumption 1. We can go a step further and assume that (27) is (uniformly) tight. Then it seems natural for $W_n(\cdot)$ to have an $(nc_n)^{1/2}$ rate of convergence. If the left hand side of (28) is not bounded and/or Assumption 1 is violated, then $W_n(\cdot)$ will have a different rate of convergence (e.g., if there are $I(1)$ variables in the instrument set).

Since $W_n(\theta)$ has zero mean for all θ, it is natural to assume that $\Delta_n(\theta)$ also has zero mean. From the zero mean property and the $(nc_n)^{1/2}$ rate of convergence of $W_n(\cdot)$, the weak convergence (24) to some limit random variable Z seems a reasonable high-level condition. Equation (25) is a uniform stochastic version of the corresponding first order Taylor series expansion for each n. Uniform integrability in (26) is explicitly assumed to facilitate the derivation of the $(nc_n)^{1/2}$ convergence rate as an application of Proposition 11 above, and is a reasonable extension of the $(nc_n)^{1/2}$ convergence rate of $W_n(\cdot)$ in (23) and (24).

Condition 12 (ii) modifies Assumptions B and C of Geyer (1994) in a suitable way to allow for moving center asymptotics and nonstandard convergence rates. The linear approximation (23) and the stochastic equicontinuity condition (25) correspond to Geyer's Assumption B, and the convergence in (24) to his Assumption C. The probability in (25) may be replaced by outer probability if measurability of the supremum were in doubt.

Condition 12 (iii) puts a requirement on the sequence of sets $(nc_n)^{1/2}(\Theta - \theta_n^*)$. Here, a set $a + bC$ (or $bC + a$) for some real numbers a, and b and a given set $C \subset \mathbb{R}^p$ is defined in the usual way as the set $\{t \in \mathbb{R}^p : t = a + bc \text{ for some } c \in C\}$. Condition 12 (iii) is stronger than

Θ's being Chernoff regular at θ^*, as described in Geyer (1994). While Chernoff regularity is relevant when the estimates are centered at a fixed value, our case needs a requirement on the parameter set centered at a convergent sequence of points, rescaled appropriately. Clarke regularity is more appropriate here and condition (*iii*) is implied by it. This condition relaxes the conventional assumption that θ^* is an interior point of Θ. Discussion of this requirement is given in Knight (2003), Geyer (1994) and Rockafellar and Wets (1998) to which the reader is referred.

Under these conditions, we have the following result.

Theorem 13 (Asymptotic Distribution) *Define*

$$h(t) = Z't + \tfrac{1}{2}t'Vt, \quad t \in \mathbb{R}^p$$

where Z is defined in (24). Under Assumptions 1, 2, 3, 4 and 12,

$$(nc_n)^{1/2}(\hat{\theta} - \theta_n^*) \to \operatorname*{argmin}_{t \in T(\theta^*)} h(t)$$

where the minimizer is unique almost surely.

Theorem 13 implies that if θ^* is an interior point of Θ, so that $T(\theta^*)$ spans $\mathbb{R}^p$, then the minimizer of $h(t)$ is $-V^{-1}Z$, so that

$$(nc_n)^{1/2}(\hat{\theta} - \theta_n^*) \to -V^{-1}Z. \tag{29}$$

The limit distribution (29) may not be normal. Taking as an example the case where the series expansion (23) is satisfied with $\Delta_n(\cdot) = \nabla W_n(\cdot)$, we notice that

$$\begin{aligned}(nc_n)^{1/2}\nabla W_n(\theta_n^*) = 2n^{-1/2}\sum_{i=1}^{n} c_n^{-1/2}\nabla\{\bar{m}_n(\theta_n^*)'\xi_n(w_{ni},\theta_n^*)\} \\ + \alpha_n^{1/2}q_n^{-1/2}\sum_{k=1}^{q_n}\nabla\left\{\zeta_{nk}^2(\theta_n^*) - E\zeta_{nk}^2(\theta_n^*)\right\}.\end{aligned}$$

The first term of the right above is likely to converge to a normal distribution if the Lindeberg condition is satisfied. But the summands of the second term may not be independent of one another except for some special cases (e.g., Bekker, 1994), and consequently Z may be non-normal. If the second term diminishes to 0 as $n \to \infty$ (e.g., when $\alpha_n \to 0$), then this potential source of nonnormality vanishes and asymptotic normality will follow from the first term provided central limit theory regularity conditions hold for this term.

Corollary 14 *Suppose that the conditions for Theorem 13 hold with $\Delta_n(\cdot) = \nabla W_n(\cdot)$. Suppose further that $\alpha_n \to 0$, θ_0 is an interior point of Θ, and*

$$2n^{-1/2}\sum_{i=1}^{n}\nabla\{c_n^{-1/2}\bar{m}_n(\theta_n^*)'\xi_n(w_{ni},\theta_n^*)\} \to_d N(0,A).$$

Then

$$(nc_n)^{1/2}(\hat{\theta} - \theta_n^*) \to_d N(0, V^{-1}AV^{-1}).$$

If in addition to the requirement that $\alpha_n \to 0$, the moment conditions are sufficiently strong and the number of moment conditions is not too large, then we end up with a limit theory that is analogous to that of the 'usual' asymptotics centered at θ_0, albeit with a different rate of convergence.

Corollary 15 *Suppose that the assumptions for Corollary 14 hold. Suppose further that (i)* $\nabla \bar{f}_n(\cdot)$ *permits a linear approximation*

$$(30) \quad \nabla \bar{f}_n(\theta) = \nabla \bar{f}_n(\theta_0) + V_n(\tilde{\theta})(\theta - \theta_0),$$

with $V_n(\theta_n) \to V(\theta_0)$ *if* $\theta_n \to \theta_0$ *where* $V(\theta_0)$ *is nonsingular, and (ii)* $q_n/(nc_n)^{1/2} \to 0$. *Then*

$$(nc_n)^{1/2}(\hat{\theta} - \theta_0) \to_d N(0, V^{-1}AV^{-1}).$$

It is worth summarizing the two special cases of conventional GMM and GMM with weak moment conditions.

Case I: Conventional GMM asymptotics. When q_n is fixed and the moment conditions are strong, we choose $c_n \equiv 1$ as described in Section 3. By the second result of Corollary 14, we have asymptotic normality for $\sqrt{n}(\hat{\theta} - \theta_0)$ under regularity.

Case II: Weak moment conditions. If q_n is fixed, consistency does not apply and the case reduces to that considered in Stock and Wright (2000), as discussed earlier. If $q_n \to \infty$ instead, then the limit theory may be non-normal unless $\alpha_n \to 0$. If $\alpha_n \to 0$, asymptotic normality may hold, but the centering will still be θ_n^* rather than θ_0. In order for $(nc_n)^{1/2}(\hat{\theta} - \theta_0)$ to be asymptotically normal, the moment conditions need to be quite strong and the growth in the number of moment conditions should be quite slow, so that $q_n \to \infty$ but $q_n/(nc_n)^{1/2} \to 0$. Interestingly, for the last claim (i.e., asymptotic normality) to hold, the moment conditions should be stronger than in Stock and Wright (2000)'s setting ($c_n = q_n/n$) since if $c_n = q_n/n$ and $q_n \to \infty$ then $q_n/(nc_n)^{1/2} = q_n^{1/2} \to \infty$. More specifically, if the strength of the moment conditions is such that $c_n = q_n/n^b$ for some b, then in order for $(q_n n^{1-b})^{1/2}(\hat{\theta} - \theta_0)$ to have a proper asymptotic distribution it is necessary that $q_n \to \infty$ and $q_n = o(n^{1-b})$. If instead $c_n = 1$, then $q_n \to \infty$ and $q_n = o(n^{1/2})$ will guarantee an asymptotic distribution for $\sqrt{n}(\hat{\theta} - \theta_0)$; if $O(n^{1/2}) \leq q_n < O(n)$, then $\hat{\theta} \to_p \theta_0$ but the asymptotic distribution theory is established for $\sqrt{n}(\hat{\theta} - \theta_n^*)$, so that there is some bias in the limit theory.

Table 2 summarizes some of these results, detailing the convergence properties, the rate of convergence and the appropriate centering for the asymptotic distributions under various scenarios.

Returning to the general case, although asymptotic normality does not usually apply, the asymptotic variance can be calculated in the case where $\Delta_n(\cdot) = \nabla W_n(\cdot)$ and $V = \lim_{n\to\infty} \nabla^2 \bar{f}_n(\theta_n^*)$. The results for this case are provided in Appendix B. Moreover, if $\alpha = 0$, then the asymptotic variance of $(nc_n)^{1/2}(\hat{\theta} - \theta_n^*)$ has the 'usual' sandwich form

$$(31) \quad (c_n^{-1}D_n'D_n)^{-1}c_n^{-1}D_n'\Omega_n D_n(c_n^{-1}D_n'D_n)^{-1},$$

c_n	$q_n \to \infty$?	Order of q_n	$\hat{\theta} \to_p \theta_0$?	Rate of conv.	Centering
q_n/n	yes	any	no	$q_n^{1/2}$	θ_n^*
q_n/n^b	yes	$o(n^{1-b})$	yes	$(n^{1-b}q_n)^{1/2}$	θ_0
$0 < b < 1$		$\geq O(n^{1-b})$			θ_n^*
1	no	fixed†	yes	$n^{1/2}$	θ_0
	yes	$o(n^{1/2})$	yes		θ_0
		$\geq O(n^{1/2})$, $o(n)$	yes		θ_n^*
		$\sim n^\ddagger$	no		θ_n^*

† Conventional GMM. ‡ Bekker's (1994) result belongs here.

Table 2: Comparison of Limit Results by the Strength and the Number of the Moment Conditions.

where $D_n = \nabla E\bar{g}_n(\theta_0)$ and $\Omega_n = E\zeta_n(\theta_0)\zeta_n(\theta_0)'$. Accordingly, we may say that the usual large sample asymptotic variance is calculated under the implicit assumption that $\alpha = 0$. Note here that (31) makes sense only when $c_n^{-1}D_n'D_n$ converges to a nonsingular matrix. Note also that $E\bar{g}_n(\theta) \equiv 0$, which implies $\nabla E\bar{g}_n(\theta) \equiv 0$, necessarily implies $\alpha > 0$ and accordingly (31) can not be the asymptotic covariance matrix.

Example 16 (Continuation of Example 8) The minimizer of $\gamma_n(\theta) + \alpha_n\delta_n(\theta)$ in Example 8 is θ_0. The asymptotic variance of $q_n^{1/2}(\hat{\theta} - \theta_0)$ is then found to be

$$(32) \quad \frac{[cr^4 + (1+2c)r^2\sigma_z^2 + (1+c)\sigma_z^4]\sigma^2}{(r^2+\sigma_z^2)^2},$$

where $c = \lim q_n/n$, from the calculation given in Appendix B. The rate of convergence $q_n^{1/2}$ follows from the fact that $c_n = q_n/n$ and so $(nc_n)^{1/2} = (nq_n/n)^{1/2} = q_n^{1/2}$.

In the extreme case where the moment conditions are uninformative (in the sense that the mean functions are all 0) which corresponds to $r = 0$, the asymptotic variance of $q_n^{1/2}(\hat{\theta} - \theta_0)$ is simply $(1+c)\sigma^2$. In this specific case, we can also show that the asymptotic distribution is normal under certain regularity conditions (see Appendix B for details). Simulation results are provided in Figure 1 for this case, based on 2000 replications with $n = 100$, $q_n = 10$, $c = 0.1$ and Gaussian random variables. The dotted lines correspond to the normal distribution with mean 0 and variance $(1+c)\sigma^2$, which is $N(0, 1.1)$ in this case. The limit distribution looks close to normal.

Simulations not reported here show that the $N(0, 1+c)$ distribution is close to the empirical distribution of the estimates under other settings, even including those where q_n is as large as $q_n = 10n$. ■

5 The Matter of Weighting

In classical large sample GMM asymptotics, a weight matrix W is usually embodied in the criterion function by way of the quadratic form as in $\bar{g}(\theta)'W\bar{g}(\theta)$. The optimal weight

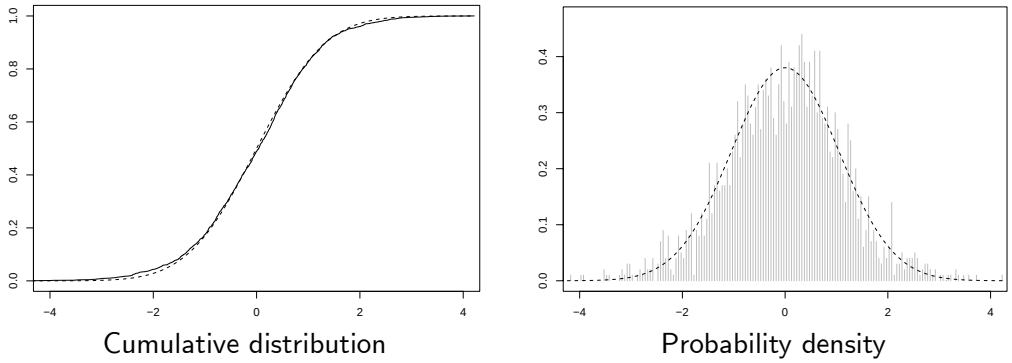

Figure 1: Example 8 with $\sigma^2 = 1$, $\sigma_z^2 = 1$, $r = 0$, $n = 100$ and $q_n/n = 0.1$

is the inverse of $Eg(w_i, \theta_0)g(w_i, \theta_0)'$. This optimal weighting is equivalent to transforming the moment conditions linearly so that each of the transformed moment conditions, when evaluated at the true parameter, has unit variance and is uncorrelated with the others.

In GMM estimation with a large number of moment conditions, a weight matrix may be used in the same way. Thus, for some given sequence of symmetric and positive definite nonrandom matrices W_n, the criterion function becomes $\bar{g}_n(\theta)'W_n\bar{g}_n(\theta)$, which is equivalent to transforming the moment conditions as $g_n(w_{ni}, \theta) \longmapsto A_n g_n(w_{ni}, \theta)$, where A_n satisfies $A_n'A_n = W_n$. All the arguments in the proceeding sections hold if the transformed moment conditions $A_n g_n(w_{ni}, \theta)$ satisfy the stated regularity conditions. In the present case, it is interesting to observe that the probability limit of the weighted GMM estimator may differ from that of the unweighted GMM estimator.

In general, it is difficult to use the optimal weighting matrix because the asymptotic variance matrix is so complicated. An alternative is to use the weight matrix

$$[n^{-1}\sum_{i=1}^{n} Eg_n(w_{ni}, \theta^*)g_n(w_{ni}, \theta^*)']^{-1},$$

where θ^* is the probability limit of the unweighted GMM estimator, and minimize the corresponding criterion function. The simplicity of the optimal weighting scheme found in traditional (i.e., fixed "q") large sample asymptotics does not apply in general in the present case. But if $\alpha = 0$ then $\hat{\theta}$ is consistent**,** the asymptotic variance has the conventional form, and the weight matrix Ω_n^{-1} is optimal. In this case the asymptotic variance becomes

$$\left(\lim_{n\to\infty} c_n^{-1}D_n'\Omega_n^{-1}D_n\right)^{-1},$$

with D_n and Ω_n denoting $E\nabla\bar{g}_n(\theta_0)$ and $n^{-1}\sum_{i=1}^{n} Eg_n(w_{ni}, \theta_0)g_n(w_{ni}, \theta_0)'$.

One can also apply "continuous updating" procedures in estimation with the (nonrandom) weight function $W_n(\theta)$, which is equivalent to minimizing

$$c_n^{-1}\bar{g}_n(\theta)'W_n(\theta)\bar{g}_n(\theta).$$

When the weight function $W_n(\theta)$ is chosen such that the regularity assumptions given in the previous sections are satisfied, $\hat{\theta}$ converges in probability to the minimizer of the limit of

$$c_n^{-1}\bar{m}_n(\theta)'W_n(\theta)\bar{m}_n(\theta) + \alpha_n q_n^{-1}E\zeta_n(\theta)'W_n(\theta)\zeta_n(\theta). \quad (33)$$

The most interesting case is $W_n(\theta) = \Omega_n(\theta)^{-1}$ where $\Omega_n(\theta) = E\zeta_n(\theta)\zeta_n(\theta)'$. In that case, the center of the weighted criterion function (33) is

$$c_n^{-1}\bar{m}_n(\theta)'\Omega_n(\theta)^{-1}\bar{m}_n(\theta) + \alpha_n,$$

which is minimized at θ_0 provided $\bar{m}_n(\theta) = 0$ only at θ_0.[1] So the $\Omega_n(\theta)^{-1}$ weighted continuous updating estimator (CUE) is consistent as long as $\lim_{n\to\infty} c_n^{-1}\bar{m}_n(\theta)'\Omega_n(\theta)^{-1}\bar{m}_n(\theta)$ uniquely identifies θ_0 even when $\alpha_n \to \alpha > 0$.

It is remarkable here that Assumption 1 (i) is automatically satisfied with this use of weighting but in order for the CUE to be consistent, the strength of the moment conditions represented by c_n should be reasonably high so that the main signal identifies θ_0.

It seems very difficult (if it is even possible) to obtain a feasible consistent CUE in the completely general case, mainly because the remainder

$$(34)\quad q_n^{-1}\zeta_n(\theta)'\left[\hat{\Omega}_n(\theta)^{-1} - \Omega_n(\theta)^{-1}\right]\zeta_n(\theta)$$

does not seem to converge to zero in general even if $\hat{\Omega}_n(\theta)$ is uniformly consistent (uniformly in θ and for all elements). However, when the $q_n \times q_n$ matrix $\Omega_n(\theta)$ is a product of a consistently estimable scalar function of θ and an increasing number of known elements, a consistent feasible CUE is available. More specifically, let

$$(35)\quad \Omega_n(\theta) = \psi_n(\theta)Q_n(\theta),$$

where $\psi_n(\theta)$ is a scalar function and $Q_n(\theta)$ is a known matrix of increasing dimension. Suppose that $\psi_n(\theta)$ is uniformly consistently estimated by $\hat{\psi}_n(\theta)$, viz.,

$$\sup_{\theta\in\Theta} |\hat{\psi}_n(\theta) - \psi_n(\theta)| \to_p 0,$$

with $\liminf_{n\to\infty} \inf_{\theta\in\Theta} \psi_n(\theta) > 0$. Let $\hat{\Omega}_n(\theta) = \hat{\psi}_n(\theta)Q_n(\theta)$. Then (34) equals

$$(36)\quad \left[\frac{\psi_n(\theta)}{\hat{\psi}_n(\theta)} - 1\right] q_n^{-1}\zeta_n(\theta)'\Omega_n^{-1}(\theta)^{-1}\zeta_n(\theta),$$

which converges in probability uniformly to zero under regularity, and therefore the feasible CUE using $\hat{\Omega}_n(\theta)^{-1}$ as weight function shares the same limit as the infeasible CUE using $\Omega_n(\theta)^{-1}$ as weight and is therefore consistent whether or not $\alpha_n \to 0$.

Let $J_n(\theta) = \psi_n(\theta)/\hat{\psi}_n(\theta)$, and $f_n(\theta) = c_n^{-1}\bar{g}_n(\theta)'\Omega_n(\theta)^{-1}\bar{g}_n(\theta)$. Then, the criterion function for feasible CUE is $J_n(\theta)f_n(\theta)$. When all the regularity conditions hold for $f_n(\theta)$, the asymptotic distribution of the feasible CUE is determined by the limit of

$$r_n^2[J_n(\theta_0 + t/r_n)f_n(\theta_0 + t/r_n) - J_n(\theta_0)f_n(\theta_0)],$$

if its behavior is regular. Rewrite this expression as

$$(37)\quad J_n(\theta_0 + t/r_n)r_n^2[f_n(\theta_0 + t/r_n) - f_n(\theta_0)] + f_n(\theta_0)r_n^2[J_n(\theta_0 + t/r_n) - J_n(\theta_0)].$$

[1] We thank an anonymous referee who pointed this out.

The first term demonstrates conventional behavior since $J_n(\theta_0 + t/r_n)$ would normally converge to 1 and the factor $r_n^2[f_n(\theta_0 + t/r_n) - f_n(\theta_0)]$ is usual. So any unusual behavior may be ascribed to the second term of (37). Under regularity, $f_n(\theta_0) = \alpha_n + o_p(1)$, and when $J_n(\theta)$ allows a second order Taylor series expansion at $t = 0$, we have

$$(38)\quad f_n(\theta_0) r_n^2 \left[J_n(\theta_0 + t/r_n) - J_n(\theta_0)\right] = \alpha_n \left[r_n \nabla J_n(\theta_0)' t + \tfrac{1}{2} t' \nabla^2 J_n(\theta_0 + \tilde{t}/r_n) t\right] + o_p(1),$$

where $\tilde{t}$ lies in between 0 and t, if $r_n \nabla J_n(\theta_0) = O_p(1)$. In the special case where $\alpha_n \to 0$, the asymptotic distribution of the CUE should be equal to that of the feasible CUE (which substitutes $\hat{\psi}(\cdot)$ for $\psi(\cdot)$), and asymptotic normality for $\sqrt{nc_n}(\hat{\theta} - \theta_0)$ is expected. (Compare this result to Table 2, where the estimator needed centering at θ_n^* when $q_n \to \infty$, $\alpha_n \to 0$ and $q_n/\sqrt{nc_n} \to \infty$.) Limited information maximum likelihood (LIML) estimation for linear simultaneous equation models is an example, which we consider next.

Example 17 (LIML) Reconsider Example 9 but this time from the perspective of a feasible CUE. For simplicity, let $\theta_0 = 0$ and assume that the z_{ki}^n are independent of ε_i and v_i^n. Let x and y be the n-vectors of x_i and y_i respectively. Let z be the $n \times q_n$ observation matrix of the z_{ki}^n. Then, conditional on the instruments, we have $\Omega_n(\theta) = \psi(\theta)(\frac{1}{n} z'z)$, where $\psi(\theta) = \sigma_\varepsilon^2 - 2\sigma_{\varepsilon v}\theta + \sigma_v^2 \theta^2$, and by the law of large numbers,

$$\hat{\psi}(\theta) = n^{-1}(y - x\theta)'(y - x\theta) \to_p \psi(\theta).$$

Now denote $\hat{\Omega}_n(\theta) = \hat{\psi}(\theta)(\frac{1}{n} z'z)$. Then, the CUE using $\hat{\Omega}_n(\theta)^{-1}$ as the weight function solves

$$\min_{\theta \in \Theta} \frac{(y - x\theta)' P_z (y - x\theta)}{(y - x\theta)'(y - x\theta)},$$

which is the LIML objective function, where $P_z = z(z'z)^{-1}z'$ and $M_z = I - P_z$, as usual. LIML as a feasible CUE shares the same probability limit as the CUE using $\Omega_n(\theta)^{-1}$ as the weight matrix. It follows that LIML is consistent under this asymptotic setting because $\delta(\theta) = 1$, due to the standardization of the criterion function by $\hat{\psi}(\theta)$, and so $\delta(\theta)$ does not depend on θ. Note again that consistency requires that the parameter "A" in (14) be nonzero. Finally, the conditional consistency of LIML implies unconditional consistency by dominated convergence.

If we assume that $A_n^2 = \sum_{k=1}^{q_n} \pi_k^{n2} \to A^2 > 0$ (as in Donald and Newey, 2001) instead of the local-to-zero setting (14) for the strength of the instruments, we have $c_n = 1$, $r_n = n^{1/2}$, and if furthermore $\alpha_n = q_n/n \to 0$, then (38) is negligible because

$$r_n \nabla J_n(\theta_0) = 2 J_n(\theta_0) n^{1/2} \left(\frac{n^{-1} x'\varepsilon}{n^{-1}\varepsilon'\varepsilon} - \frac{\sigma_{\varepsilon v}}{\sigma_\varepsilon^2} \right) = O_p(1),$$

usually. So in this case $n^{1/2}(\hat{\theta} - \theta_0)$ would converge to a normal distribution. See Proposition 2 of Donald and Newey (2001). ■

Consistent CUE can be useful in practical applications as the following panel data example shows.

Example 18 (Time varying individual effects) We consider a panel data model with weak temporal variation in the individual effects and dominating disturbances,

$$y_{it} = \beta_0 + \lambda_t(\theta_0)\alpha_i + n^{1/2}\varepsilon_{it}, \quad \lambda_t(\theta) = \exp\{\theta(t-1)/(T-1)\}.$$

Let $T \to \infty$ as $n \to \infty$. Consider GMM estimation making use of the zero mean condition $E(\varepsilon_{it}) = 0$ for all t. Assume at the same time that the ε_{it} are white noise over time and *iid* across i, and that the α_i are *iid* across i. For identification, let $\mu_\alpha = E(\alpha_i) \neq 0$. Denote $\sigma^2 = E(\varepsilon_{it}^2)$ and $\sigma_\alpha^2 = E[(\alpha_i - \mu_\alpha)^2]$. Since $y_{i1} = \beta_0 + \alpha_i + \varepsilon_{it}$, the $(T-1)$ moment functions are

$$g_{it}^\circ(\beta, \theta) = [y_{it} - \lambda_t(\theta)y_{i1}] - [1 - \lambda_t(\theta)]\beta, \quad t = 2, \ldots, T.$$

(See Han, Orea and Schmidt, 2003, for the derivation of the moment conditions.) We have $g_{it}^\circ(\beta_0, \theta_0) - Eg_{it}^\circ(\beta_0, \theta_0) = n^{1/2}\varepsilon_{it} - \lambda_t(\theta_0)\varepsilon_{it}$, and divide the moment functions by $\sqrt{n}$ in view of (5). Then the moment functions are

$$g_{it}(\beta, \theta) = n^{-1/2} g_{it}^\circ(\beta, \theta), \quad t = 2, \ldots, T. \tag{39}$$

The mean functions are

$$Eg_{it}(\beta, \theta) = -n^{-1/2}\{[1 - \lambda_t(\theta)](\beta - \beta_0) + [\lambda_t(\theta) - \lambda_t(\theta_0)]\mu_\alpha\}, \quad t = 2, \ldots, T,$$

where $\mu_\alpha = E(\alpha_i)$, and the deviations $\xi_{it}(\beta, \theta) = g_{it}(\beta, \theta) - Eg_{it}(\beta, \theta)$ are

$$\xi_{it}(\beta, \theta) = [\varepsilon_{it} - \lambda_t(\theta)\varepsilon_{i1}] - n^{-1/2}[\lambda_t(\theta) - \lambda_t(\theta_0)](\alpha_i - \mu_\alpha),$$

implying that

$$\begin{aligned}\delta_n(\beta, \theta) &= q_n^{-1} \sum_{t=2}^{T} [1 + \lambda_t(\theta)^2]\sigma^2 + (nq_n)^{-1} \sum_{t=2}^{T} [\lambda_t(\theta) - \lambda_t(\theta_0)]^2 \sigma_\alpha^2 \\ &\to \sigma^2 \int_0^1 \left(1 + e^{2r\theta}\right) dr = \delta(\beta, \theta),\end{aligned}$$

where $q_n = T - 1$ is the number of moment conditions. Because $c_n = q_n/n$, we have $\alpha_n \equiv 1$, and the unweighted GMM estimator using (39) is likely to be inconsistent if convergent because $\delta(\beta, \theta)$ is minimized at $\theta = -\infty$. (In fact, the moment conditions violate Assumption 1 (i), and we expect the GMM to have a nondegenerate limit distribution. See the simulation below.)

Let $\lambda(\theta)$ be the vector of $\lambda_t(\theta)$ for $t = 2, \ldots, T$ and $\lambda_*(\theta) = \lambda(\theta) - \lambda(\theta_0)$. Then from the calculation of

$$E\xi_{it}\xi_{is} = [1\{t = s\} + \lambda_t(\theta)\lambda_s(\theta)]\,\sigma^2 + n^{-1}\,[\lambda_t(\theta) - \lambda_t(\theta_0)]\,[\lambda_s(\theta) - \lambda_s(\theta_0)]\,\sigma_\alpha^2,$$

we get (suppressing the arguments)

$$\Omega_n = \sigma^2\,[I_{T-1} + \lambda\lambda'] + n^{-1}\sigma_\alpha^2\lambda_*\lambda_*' = \Omega_{0,n} + n^{-1}\sigma_\alpha^2\lambda_*\lambda_*'.$$

Next, we show that the CUE using Ω_n^{-1} as the weight function has the same limit as the CUE using $\Omega_{0,n}^{-1}$. Note that

$$(40)\quad q_n^{-1}\zeta_n'\left(\Omega_{0,n}^{-1}-\Omega_n^{-1}\right)\zeta_n = q_n^{-1}\tilde{\zeta}_n'\left(\Omega_n^{1/2}\Omega_{0,n}^{-1}\Omega_n^{1/2}-I_q\right)\tilde{\zeta}_n,$$

where $\tilde{\zeta}_n = \Omega_n^{-1/2}\zeta_n$ has zero mean and identity covariance matrix. Writing

$$\Omega_{0,n} = \sigma^2(I+\lambda\lambda') = \sigma^2[M_\lambda + (1+\lambda'\lambda)P_\lambda],$$

where $P_\lambda = \lambda(\lambda'\lambda)^{-1}\lambda'$ and $M_\lambda = I_{T-1} - P_\lambda$, we get

$$\Omega_{0,n}^{-1} = \sigma^{-2}[M_\lambda + (1+\lambda'\lambda)^{-1}P_\lambda],\ \text{and}\ \Omega_{0,n}^{-1/2} = \sigma^{-1}[M_\lambda + (1+\lambda'\lambda)^{-1/2}P_\lambda].$$

The eigenvalues of $\Omega_n^{1/2}\Omega_{0,n}^{-1}\Omega_n^{1/2} - I_{T-1}$ are equal to those of

$$(41)\quad \Omega_{0,n}^{-1}\Omega_n - I_{T-1} = \frac{1}{n}\left(\frac{\sigma_\alpha^2}{\sigma^2}\right)\left(\lambda_* - \frac{\lambda'\lambda_*}{1+\lambda'\lambda}\lambda\right)\lambda_*',$$

which comprise $(T-2)$ zeros and one non zero value, being the trace. Now, by Schur's decomposition, (40) equals the trace (i.e., the non zero eigenvalue) of (41) times $q_n^{-1}(b_n'\tilde{\zeta}_n)^2$ where b_n is the orthogonal eigenvector corresponding to the nonzero eigenvalue. Noting that $b_n'\tilde{\zeta}_n$ is a centered random function (of θ) with unit variance (for all θ), it is most likely that (40) is $O_p(q_n^{-1})$, which converges in probability to zero as $T\to\infty$. As a result, the CUE using $\Omega_{0,n}(\beta,\theta)^{-1}$ as weight function has the same probability limit as the CUE using $\Omega_n(\beta,\theta)^{-1}$, and therefore the former is also consistent. Finally, because σ^2 does not depend on the parameter values, the CUE using $\sigma^2\Omega_{0,n}(\beta,\theta)^{-1}$ is also consistent. This CUE equals the minimizer of

$$\bar{g}^\circ(\beta,\theta)'\Big(M_{\lambda(\theta)} + [1+\lambda(\theta)'\lambda(\theta)]^{-1}P_{\lambda(\theta)}\Big)\bar{g}^\circ(\beta,\theta)$$

where $\bar{g}_i^\circ$ is the vector of $n^{-1}\sum_{i=1}^n g_{it}^\circ$ for $t=2,\ldots,T$.

It is interesting that the unweighted GMM may have a nondegenerate limit distribution (see the simulation below) while the CUE is consistent. It is because the weighting forces the variance matrix of the transformed moment conditions to abide by the regularity conditions needed for convergence. In this case, the behavior of the two step GMM estimator is intriguing because the first step GMM estimator and the weighting matrix at the second step seem to be random while the second step objective function effectively eliminates asymptotic randomness. According to the simulation below, the two step GMM estimator seems to converge. A theoretical analysis of these properties is certainly interesting but is a topic to be pursued in separate work.

We report simulation results from 3000 iterations with $n=50$ and $n=100$ for $T/n=0.8$. The true parameters are $\beta_0=-2$ and $\theta_0=2$. Numerical minimizations are done by the R function `optim`. The starting parameter value for β in the CUE is set to the global average of y_{it}, i.e., the OLS estimator, and zero is used for θ. The resulting CUE for β and θ is used as the initial value in unweighted GMM estimation, which in turn is fed into the two

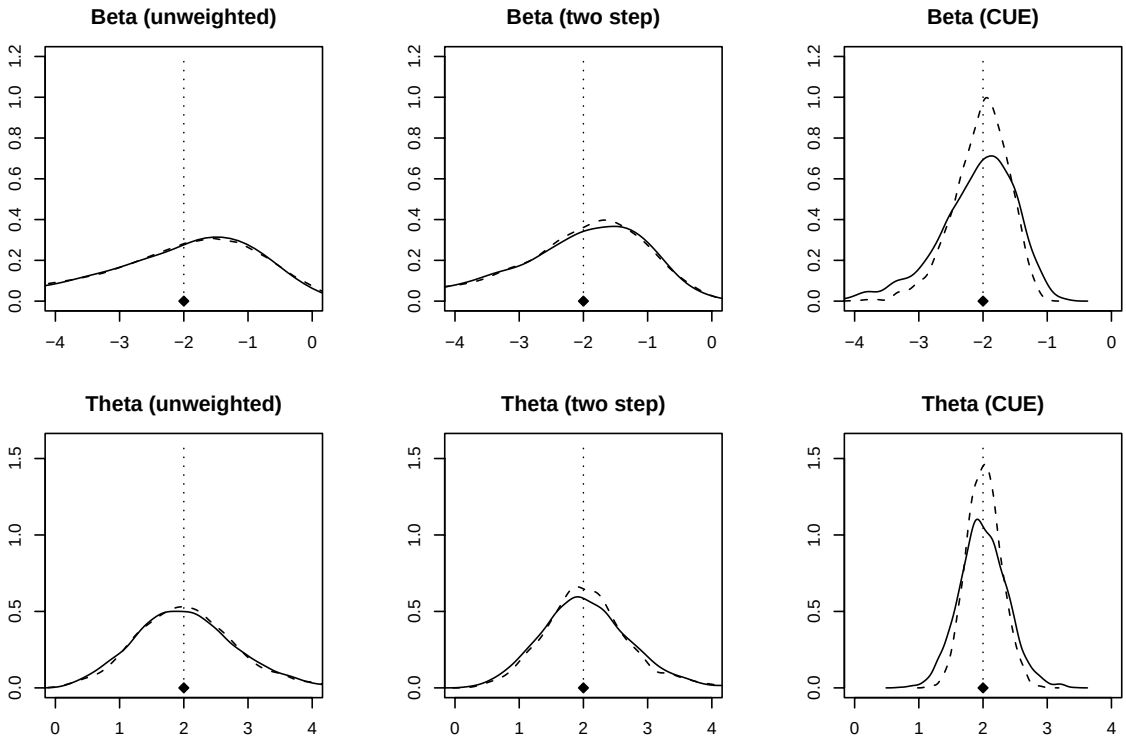

Figure 2: Kernel densities. The upper panel plots are for β estimates and the lower panel for θ estimates. From left to right are shown unweighted GMM, usual two step GMM, and CUE estimation. The solid lines are for $n = 50$ and $T = 40$, and dashed lines for $n = 100$ and $T = 80$. The solid diamonds signify the true parameters.

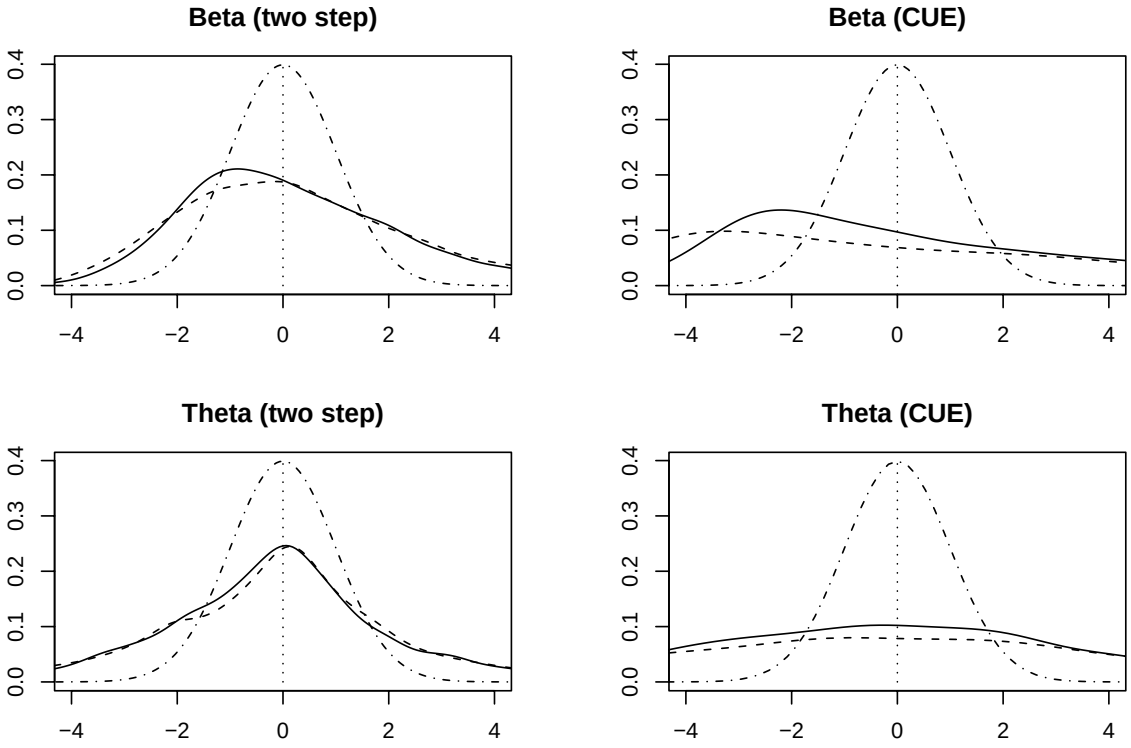

Figure 3: Simulation densities of conventional t-statistics. Solid lines show $n = 50$ and $T = 40$; dashed lines show $n = 100$ and $T = 80$; the $N(0, 1)$ distribution is shown in longer dashes.

n	T	Method	β estimate	θ estimate
50	40	Unweighted	-2.7474 (3.0804)	2.0364 (0.7970)
100	80	Unweighted	-2.8019 (5.1869)	2.0443 (0.7728)
50	40	Two step	-2.4800 (1.9943)	2.0526 (0.7169)
100	80	Two step	-2.3783 (1.7219)	2.0564 (0.6660)
50	40	CUE	-2.1124 (0.6455)	2.0046 (0.3728)
100	80	CUE	-2.0381 (0.4318)	2.0127 (0.2671)

Table 3: Simulated mean and variance. Simulated averages are provided together with the simulated standard deviations inside parentheses.

step GMM procedure. The individual effects α_i are generated from $N(1,1)$, and the errors ε_{it} are from $N(0, 0.25)$. So the variance of the idiosyncratic error is $n \times 0.25$. Out of 3000 iterations, 12 failed to attain numerical minima with $n = 50$ and $T = 40$, and 8 failed with $n = 100$ and $T = 80$.

Figure 2 shows kernel density estimates of the distribution of the various estimates from the simulation. The upper panel plots results for the β parameter and the lower panel for θ. The figures from left to right record unweighted GMM, usual two step GMM, and the CUE, respectively. The solid lines are for $n = 50$ and $T = 40$ and the dashed lines are for $n = 100$ and $T = 80$. The solid diamonds signify the true parameters. Observe that as n doubles, the variability of the unweighted GMM estimator (on the left) remains about the same, while the variability of the CUE (on the right) reduces by a similar factor. The outcome is not clear in the case of the two step efficient GMM estimator (in the middle). The simulated distributions of the estimates for β look skewed in smaller samples especially for the unweighted GMM.

Figure 3 shows the simulated densities of the t-ratios computed according to the conventional formula "$n^{-1}(D'\Omega^{-1}D)^{-1}$" of two step GMM (on the left) and CUE (on the right). In both cases, the conventional variance estimates look as if they understate the true variances.

Table 3 summarizes the simulated means and standard deviations of the unweighted GMM, the usual two step GMM and the CUE. ■

A slight generalization is possible to the case where $\Omega_n(\theta)$ is a linear combination of a finite number of known functions, viz.,

$$\Omega_n(\theta) = \sum_{j=1}^{k} \psi_{jn}(\theta) Q_{jn}(\theta), \quad k \text{ finite, } Q_{jn}(\theta) \text{ positive semidefinite,}$$

where $\psi_{jn}(\theta)$ is a scalar function bounded away from zero and containing estimable unknown parameters and $Q_{jn}(\theta)$ is a known function of an increasing dimension. Let $\hat{\Omega}_n(\theta) = \sum_{j=1}^{k} \hat{\psi}_{jn}(\theta) Q_{jn}(\theta)$. Then (34) can be shown to hold from (suppressing the θ argument)

$$\begin{aligned} \frac{1}{q_n}\zeta_n'[\Omega_n^{-1} - \hat{\Omega}_n^{-1}]\zeta_n &= \frac{1}{q_n}\zeta_n'\Omega_n^{-1}[\hat{\Omega}_n - \Omega_n]\hat{\Omega}_n^{-1}\zeta_n \\ &= \sum_{j=1}^{k}(\hat{\psi}_{jn} - \psi_{jn})\frac{1}{q_n}\zeta_n'\Omega_n^{-1}Q_{jn}\hat{\Omega}_n^{-1}\zeta_n, \end{aligned}$$

whose absolute value is bounded by

$$\sum_{j=1}^{k} |\hat{\psi}_{jn} - \psi_{jn}| \frac{1}{q_n} \zeta_n' \Omega_n^{-1} Q_{jn} \hat{\Omega}_n^{-1} \zeta_n \leq \sum_{j=1}^{k} \left| \frac{\hat{\psi}_{jn} - \psi_{jn}}{\hat{\psi}_{jn}} \right| \frac{1}{q_n} \zeta_n' \Omega_n^{-1} \zeta_n \to_p 0.$$

The above construction of feasible weighting for consistent CUE estimation seems to allow for some further generalization to nonlinear models as long as there is a result similar to (34). However, it seems more productive to work with specific applications rather than attempt to develop a general theorem, and hence we do not pursue that direction further here.

6 Conclusion

The primary contribution of this paper is to provide a limit theory for GMM estimation that allows for cases where the number of moment conditions may grow with the sample size and where the moment conditions may individually be only weakly identifying. Under such circumstances, GMM estimation can be consistent but the rate of convergence depends on both the sample size and the number and the quality of the moment conditions. In addition to the usual source of signal from the moment conditions, the variability across moment conditions also plays a role in the asymptotic theory and can influence the point of consistency for the GMM estimator. Interestingly, consistent estimation is possible even in the case of apparently irrelevant instruments, provided there are "enough" of these instruments asymptotically.

The limit distribution theory presents more difficulties than usual. Our approach has been to employ some high level conditions and epiconvergence arguments to provide a general theory. Our analytical framework is best seen as a start in designing appropriate methods of inference in such general models, including the analysis of statistical tests, and further research on these problems will be needed to obtain useful inferential procedures for practical work at a reasonable level of generality. The theory does make it clear that the form of the limit distribution need not be normal. And some special cases where normality does apply are considered and, of course, the usual GMM limit theory is one such case. Additionally, the interesting case of linear structural equation estimation with many weak instruments falls within our framework. Here, we see that consistent estimation is possible even with apparently irrelevant instruments and the new limit theory highlights the interacting role in estimation between the quality of the instruments in their totality (as their numbers increase) on the one hand and the degree of endogeneity in the system on the other.

As discussed in the Introduction, there are now many instances in empirical research where large numbers of instruments are employed and where there is evidence of weak instrumentation. Asymptotic methods of the type given here seem likely to be useful in such contexts and show that rates of convergence different from the usual $\sqrt{n}$ are to be expected and that the point of consistency is not always the true parameter.

The asymptotic framework provides the opportunity to consider other interesting issues, one being the effect of dependence. When the independence assumption on the component

random variables is relaxed, the analysis naturally becomes more complex but some intriguing things can occur. For example, in a linear structural model with $I(1)$ instruments, the unweighted and optimally weighted GMM estimators have different rates of convergence. The analysis of this case and others of interest that arise in practical work are left for future research.

7 Notation

$g_{nk}(w_{ni}, \theta)$	moment functions, $k = 1, \ldots, q_n$
$\bar{g}_{nk}(\theta)$	sample moment functions, $k = 1, \ldots, q_n$
$\xi_{nk}(w_{ni}, \theta)$	$g_{nk}(w_{ni}, \theta) - Eg_{nk}(w_{ni}, \theta)$
$\bar{m}_{nk}(\theta)$	$E\bar{g}_{nk}(\theta)$
$\zeta_{nk}(\theta)$	$\sqrt{n}[\bar{g}_{nk}(\theta) - E\bar{g}_{nk}(\theta)]$
$\gamma_n(\theta)$, $\gamma(\theta)$	$c_n^{-1}\bar{m}_n(\theta)'\bar{m}_n(\theta)$, and its limit
$\delta_n(\theta)$, $\delta(\theta)$	$q_n^{-1}E\zeta_n(\theta)'\zeta_n(\theta)$, and its limit
α_n, α	q_n/nc_n, and its limit
$f_n(\theta)$, $f_\infty(\theta)$	$c_n^{-1}\bar{g}_n(\theta)'\bar{g}_n(\theta)$, and its limit
$\bar{f}_n(\theta)$	$Ef_n(\theta) = \gamma_n(\theta) + \alpha_n\delta_n(\theta)$
$W_n(\theta)$	$f_n(\theta) - \bar{f}_n(\theta)$
n, q_n, p	sample size, # of moment conditions, and # of parameters
$\lvert \cdot \rvert$	Euclidean distance from the origin
$o_p(1)$	tends to zero in probability
$O_p(1)$	bounded in probability
$A \lesssim B$, $B \gtrsim A$	$A \leq B$ up to a universal constant
$\to_p$	convergence in probability
$\to_d$, $\to$	convergence in distribution, weak convergence
$\mathrm{tr}(A)$	trace of A
$\mathbb{R}$, $\bar{\mathbb{R}}$	$(-\infty, \infty)$, and $\mathbb{R} \cup \{-\infty, \infty\}$
$\mathrm{cov}(x, y)$	$E(x - Ex)(y - Ey)'$
$\mathrm{var}(x)$	$\mathrm{cov}(x, x)$
$\nabla x(\theta)$	$\frac{\partial}{\partial \theta'}x(\theta)$ if x is a vector function, $\frac{\partial}{\partial \theta}x(\theta)$ otherwise.
$\nabla x'$	$(\nabla x)'$
$A(\theta) \lesssim B(\theta)$	$A(\theta) \leq kB(\theta)$, for some finite universal constant k.

A Technical Appendix and Proofs

As usual, the first step in the proof of Theorem 5 is to establish the uniform convergence in probability of the properly scaled criterion function $f_n(\theta)$ to the limit $\gamma(\theta) + \alpha\delta(\theta)$. Noting that $\bar{f}_n(\theta) = Ef_n(\theta) \to f_\infty(\theta)$ uniformly by assumption, we only have to show that $W_n(\theta) = f_n(\theta) - \bar{f}_n(\theta)$ vanishes in probability uniformly as $n \to \infty$. In fact, we have

$$
\begin{aligned}
W_n(\theta) &= c_n^{-1}\bar{g}_n(\theta)'\bar{g}_n(\theta) - Ec_n^{-1}\bar{g}_n(\theta)'\bar{g}_n(\theta) \\
&= c_n^{-1}\bar{m}_n(\theta)'\bar{m}_n(\theta) + 2c_n^{-1}n^{-1/2}\bar{m}_n(\theta)'\zeta_n(\theta) + (nc_n)^{-1}\zeta_n(\theta)'\zeta_n(\theta) \\
&\quad - c_n^{-1}\bar{m}_n(\theta)'\bar{m}_n(\theta) - (nc_n)^{-1}E\zeta_n(\theta)'\zeta_n(\theta) \\
&= 2c_n^{-1}n^{-1/2}\bar{m}_n(\theta)'\zeta_n(\theta) + \alpha_n q_n^{-1}[\zeta_n(\theta)'\zeta_n(\theta) - E\zeta_n(\theta)'\zeta_n(\theta)] \\
&= 2\Psi_{1n}(\theta) + \alpha_n\Psi_{2n}(\theta), \quad \text{say.}
\end{aligned} \tag{42}
$$

The following lemmas show that (42) converges in probability to zero uniformly in θ.

Lemma 19 *Under Assumption 1 (iv), $q_n^{-1}\zeta_n(\theta)'\zeta_n(\theta)$ is tight.*

Proof. Obviously,

$$
q_n^{-1}\zeta_n(\theta)'\zeta_n(\theta) \le \max_{1\le k\le q_n}\zeta_{nk}(\theta)^2 = \left(\max_{1\le k\le q_n}|\zeta_{nk}(\theta)|\right)^2,
$$

and the right hand side is tight by Assumption 1 (*iv*). ■

Lemma 20 *Under Assumptions 1 (i), 1 (iv), 2 and 4, $\Psi_{1n}(\theta) \to_p 0$ uniformly over all $\theta \in \Theta$.*

Proof. First we note that $\Psi_{1n}(\theta) \to_p 0$ for all $\theta \in \Theta$ because $E\Psi_{1n}(\theta) \equiv 0$ and

$$
\begin{aligned}
E\Psi_{1n}(\theta)^2 &= \frac{1}{nc_n^2}\bar{m}_n(\theta)'E[\zeta_n(\theta)\zeta_n(\theta)']\bar{m}_n(\theta) \\
&\lesssim (nc_n)^{-1}c_n^{-1}\bar{m}_n(\theta)'\bar{m}_n(\theta) = (nc_n)^{-1}\gamma_n(\theta) \to 0,
\end{aligned} \tag{43}
$$

where the curly inequality comes from Assumption 1 (*i*), and the last convergence is obtained because $nc_n \to \infty$ and $\gamma_n(\theta)$ converges by Assumption 2.

Next we show that $\Psi_{1n}(\theta)$ is tight. By the triangle inequality and Hölder's inequality, we have

$$
\begin{aligned}
|\Psi_{1n}(\theta)| &\le \frac{q_n}{c_n\sqrt{n}}\cdot\frac{1}{q_n}\sum_{k=1}^{q_n}|\bar{m}_{nk}(\theta)\zeta_{nk}(\theta)| \\
&\le \frac{q_n}{c_n\sqrt{n}}\left(q_n^{-1}\sum_{k=1}^{q_n}\bar{m}_{nk}(\theta)^2\right)^{1/2}\left(q_n^{-1}\sum_{k=1}^{q_n}\zeta_{nk}(\theta)^2\right)^{1/2} \\
&= \alpha_n^{1/2}\gamma_n(\theta)^{1/2}\cdot\left(q_n^{-1}\zeta_n(\theta)'\zeta_n(\theta)\right)^{1/2}.
\end{aligned} \tag{44}
$$

Clearly, $\gamma_n(\theta)$ is asymptotically uniformly bounded because $\gamma_n(\theta)$ is continuous by Assumption 4 and uniformly convergent by Assumption 2. Also, α_n converges, and tightness of $q_n^{-1}\zeta_n(\theta)'\zeta_n(\theta)$ has been established in Lemma 19. ■

We next prove that $\alpha_n\Psi_{2n}(\theta)$ vanishes uniformly.

Lemma 21 *Under Assumptions 1 and 2 (ii), $\alpha_n\Psi_{2n}(\theta) \to_p 0$ uniformly in $\theta \in \Theta$.*

Proof. We abbreviate by writing $\xi_{n,i} := \xi_n(w_{ni}, \theta)$ without specifying the argument θ (a notation that is used only in this proof). Note that

$$\begin{aligned}\Psi_{2n}(\theta) &= n^{-1}\sum_{i=1}^{n}[q_n^{-1}\xi'_{n,i}\xi_{n,i} - Eq_n^{-1}\xi'_{n,i}\xi_{n,i}] + q_n^{-1/2}n^{-1}\sum_{i\neq j} q_n^{-1/2}\xi'_{n,i}\xi_{n,j} \\ &= \Psi_{2n}^{(1)} + \Psi_{2n}^{(2)}, \text{ say.}\end{aligned} \tag{45}$$

We will first show that $\alpha_n\Psi_{2n}(\theta) \to_p 0$ for each θ by showing $\alpha_n\Psi_{2n}^{(1)} \to_p 0$ and $\alpha_n\Psi_{2n}^{(2)} \to_p 0$. The convergence of $\alpha_n\Psi_{2n}^{(1)}$ for each θ follows straightforwardly from the finiteness of the fourth moments of $\xi_{n,i}$ as assumed in Assumption 1(*ii*) and the Chebyshev inequality. The convergence of $\alpha_n\Psi_{2n}^{(2)}$ to zero for each θ holds because it has zero mean and its variance diminishes to zero, as is now shown. The variance of $\Psi_{2n}^{(2)}$ is

$$\text{var}\,(\Psi_{2n}^{(2)}) = q_n^{-1}n^{-2}\sum_{i\neq j} q_n^{-1}\text{tr}\left(\Omega_{ni}(\theta)\Omega_{nj}(\theta)\right),$$

where $\Omega_{ni}(\theta) := E\xi_{n,i}\xi'_{n,i}$. Because $\text{tr}\,(AB)^2 \leq \text{tr}\,(A^2)\text{tr}\,(B^2)$ for any symmetric A and B, we have

$$q_n^{-1}\text{tr}\left(\Omega_{ni}(\theta)\Omega_{nj}(\theta)\right) \leq \left(q_n^{-1}\text{tr}\left(\Omega_{ni}(\theta)^2\right)\cdot q_n^{-1}\text{tr}\left(\Omega_{nj}(\theta)^2\right)\right)^{1/2}.$$

But $\text{tr}\left(\Omega_{ni}(\theta)^2\right)$, which is the sum of the squared eigenvalues of $\Omega_{ni}(\theta)$, is $O(q_n)$. Hence, the above displayed expresssion is a bounded sequence, whose upper bound is universal for all i and j by Assumption 1 (*i*), and therefore we have $\text{var}\,(\Psi_{2n}^{(2)}) = O(q_n^{-1})$. Finally,

$$\text{var}\,(\alpha_n\Psi_{2n}^{(2)}) = \alpha_n^2 O(q_n^{-1}) = \alpha_n O(1/nc_n) \to 0 \text{ by Assumption 2 } (ii).$$

So far, we have shown that $\alpha_n\Psi_{2n}(\theta) \to_p 0$ for each θ. Again by standard arguments (e.g., Billingsley, 1968), it suffices to prove tightness for $\Psi_{2n}(\theta)$. This part is easy. Clearly,

$$|\Psi_{2n}(\theta)| = q_n^{-1}|\zeta_n(\theta)'\zeta_n(\theta) - E\zeta_n(\theta)'\zeta_n(\theta)| \leq q_n^{-1}\zeta_n(\theta)'\zeta_n(\theta) + \delta_n(\theta). \tag{46}$$

We have already shown that $q_n^{-1}\zeta_n(\theta)'\zeta_n(\theta)$ is tight in Lemma 19, and $\delta_n(\theta)$ is asymptotically bounded because of the uniform convergence condition 1 (*iii*). The result follows straightforwardly. ■

Proof of Theorem 5. Lemmas 20 and 21 imply that $f_n(\theta) - \bar{f}_n(\theta) \to_p 0$ uniformly in θ, which together with the uniform convergence $\bar{f}_n(\theta) \to \gamma(\theta) + \alpha\delta(\theta)$ imply that $f_n(\theta) \to_p$

$\gamma(\theta) + \alpha\delta(\theta)$ uniformly. The result follows from Theorem 9.3.1 of Davidson (2000). In its proof, (7) of our Assumption 3 can obviously be used instead of the condition that θ^* is an interior point of Θ. ■

We now move on to consider the rate of convergence and the asymptotic distribution. Instead of Proposition 11, we first prove a somewhat more general version of the result, which is similar to Theorem 5.52 of van der Vaart (1998).

Proposition 22 *Suppose that the assumptions of Proposition 11 are satisfied with (19) replaced by*

$$\bar{f}_n(\theta) - \bar{f}_n(\theta_n^*) \gtrsim |\theta - \theta_n^*|^b, \quad n > n_0, \tag{47}$$

and (20) by

$$E \sup_{|\theta - \theta_n^*| < \delta} \left| r_n W_n(\theta) - r_n W_n(\theta_n^*) \right| \lesssim \delta^a, \quad a < b. \tag{48}$$

Then

$$r_n^{1/(b-a)}(\hat{\theta} - \theta_n^*) = O_p(1).$$

Proof. The proof follows Theorem 5.52 of van der Vaart (1998) with minor changes to accommodate the moving center asymptotics and to simplify the exposition. Fix $M > 0$. Let $s_n = r_n^{1/(b-a)}$ for notational simplicity. Let $S_{j,n} = \{\theta : 2^{j-1} < s_n|\theta - \theta_n^*| \leq 2^j\}$. Then, for every $\eta > 0$,

$$\begin{aligned} P(s_n|\hat{\theta} - \theta_n^*| > 2^M) \leq & \sum_{j \geq M, 2^j \leq \eta s_n} P\big(\inf_{\theta \in S_{j,n}} [f_n(\theta) - f_n(\theta_n^*)] \leq 0 \big) \\ & + P(2|\hat{\theta} - \theta_n^*| \geq \eta). \end{aligned} \tag{49}$$

By Theorem 5, the second probability on the right diminishes to 0 for every $\eta > 0$. Because

$$\inf[f_n(\theta) - f_n(\theta_n^*)] \geq \inf[W_n(\theta) - W_n(\theta_n^*)] + \inf[\bar{f}_n(\theta) - \bar{f}_n(\theta_n^*)]$$

for any common set over which the 'inf' operator applies, we have for each j

$$\begin{aligned} & P\big(\inf_{\theta \in S_{j,n}} [f_n(\theta) - f_n(\theta_n^*)] \leq 0 \big) \\ & \quad \leq P\big(- \inf_{\theta \in S_{j,n}} [W_n(\theta) - W_n(\theta_n^*)] \geq \inf_{\theta \in S_{j,n}} [\bar{f}_n(\theta) - \bar{f}_n(\theta_n^*)] \big) \\ & \quad \leq P\big(\sup_{\theta \in S_{j,n}} |W_n(\theta) - W_n(\theta_n^*)| \geq \inf_{\theta \in S_{j,n}} [\bar{f}_n(\theta) - \bar{f}_n(\theta_n^*)] \big). \end{aligned}$$

Choose η small enough that the first condition of the theorem holds for all θ such that $|\theta - \theta_n^*| < \eta$ and the second for every $\delta \leq \eta$. Then, for each j, the above probability is bounded by

$$P\big(\sup_{\theta \in S_{j,n}} \left| r_n W_n(\theta) - r_n W_n(\theta_n^*) \right| \geq m r_n 2^{(j-1)b} s_n^{-a} \big), \tag{50}$$

for some finite constant m by (47). By Chebyshev's inequality and (48), this last probability is again bounded by $m^{-1}2^b 2^{-(b-a)j}$. Therefore, the first term on the right of (49) is eventually bounded by $m^{-1}2^b \sum_{j\geq M} 2^{-(b-a)j}$, which converges to 0 as $M \to \infty$ when $a < b$. ■

Proof of Proposition 11. Use $a = 1$ and $b = 2$ for the above proposition. ■

The proof of Theorem 13 follows Geyer (1994)'s approach. Denote $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and let $r_n = (nc_n)^{1/2}$. The next lemma is a variant of Geyer (1994)'s LEMMA 4.1.

Lemma 23 *Define a random function $H_n(\cdot)$ from $\mathbb{R}^p$ to $\bar{\mathbb{R}}$ by*

$$H_n(t) = \begin{cases} r_n^2\Big[f_n(\theta_n^* + t/r_n) - f_n(\theta_n^*)\Big], & t \in r_n(\Theta - \theta_n^*), \\ +\infty, & \textit{otherwise} \end{cases} \tag{51}$$

and another random function $H(\cdot)$ from $\mathbb{R}^p$ to $\bar{\mathbb{R}}$ by

$$H(t) = \begin{cases} Z't + \frac{1}{2}t'Vt, & t \in T(\theta^*) \\ +\infty, & \textit{otherwise}, \end{cases} \tag{52}$$

where Z is defined in (24). Under the assumptions of Theorem 13, H_n epiconverges in distribution to H.

The proof below follows Geyer (1994)'s proof of his LEMMA 4.1 with some minor changes for the moving center asymptotics and nonstandard convergence rate in the present case.

Proof. Define

$$G_n(t) = \begin{cases} r_n\Delta_n(\theta_n^*)'t + \frac{1}{2}t'Vt, & t \in r_n(\Theta - \theta_n^*) \\ +\infty, & \text{otherwise.} \end{cases}$$

Then for $t \in r_n(\Theta - \theta_n^*)$,

$$\begin{aligned} H_n(t) - G_n(t) = {} & \tfrac{1}{2}t'[V_n(\theta_n^*) - V]t \\ & + r_n^2[\bar{R}_{2n}(\theta_n^* + t/r_n) + R_{1n}(\theta_n^* + t/r_n)]. \end{aligned}$$

Let

$$\|H_n - G_n\|_\rho = \sup_{t \in \rho B_p \cap r_n(\Theta - \theta_n^*)} |H_n(t) - G_n(t)|,$$

where B_p is the closed unit ball in $\mathbb{R}^p$. Then because $t'[V_n(\theta_n^*) - V]t = o(1) \cdot |t|^2$ and

$$r_n^2\bar{R}_{2n}(\theta_n^* + t/r_n) = \frac{\bar{R}_{2n}(\theta_n^* + t/r_n)}{|t/r_n|^2} \cdot |t|^2 = o(1) \cdot |t|^2,$$

we have

$$\|H_n - G_n\|_\rho \leq o(1)\rho^2 + \rho \sup_{\theta \in (\theta_n^* + r_n^{-1}\rho B_p) \cap \Theta} \frac{r_n|\mathbb{R}_{1n}(\theta)|}{|\theta - \theta_n^*|},$$

and the right hand side converges to 0 by (25).

All that remains is to show that $H_n(t)$ epiconverges in distribution to $H(t)$, which can be done with a little modification of the proof of Geyer (1994, Lemma 4.1). Indeed, following Geyer (1994, Lemma 4.1), set w_n to be the function that is 0 on $(nc_n)^{1/2}(\Theta - \theta_n^*)$ and $+\infty$ elsewhere, and let w be the same kind of indicator function for $T(\theta^*)$. Because of Assumption 12 (*iii*), w_n epiconverges to w. The rest of proof is identical to that of Geyer (1994, Lemma 4.1). ■

Lemma 24 *Under Assumptions 1–4 and 12,* $(nc_n)^{1/2}(\hat{\theta} - \theta_n^*) = O_p(1)$.

Proof. We first prove that Assumptions 1–4 and 12 (*i*) imply (19). Let λ_{min}^n and λ_{min}^* denote the minimal eigenvalues of $V_n(\theta_n^*)$ and V, respectively. Then, (21) implies that

$$\begin{aligned} \bar{f}_n(\theta) - \bar{f}_n(\theta_n^*) &= \tfrac{1}{2}(\theta - \theta_n^*)' V_n(\theta_n^*)(\theta - \theta_n^*) + \bar{R}_{2n}(\theta) \\ &\geq \tfrac{1}{2}\lambda_{min}^n (\theta - \theta_n^*)'(\theta - \theta_n^*) + \bar{R}_{2n}(\theta) \\ &\geq \tfrac{1}{2}\lambda_{min}^n (\theta - \theta_n^*)'(\theta - \theta_n^*) - |\bar{R}_{2n}(\theta)| \\ &= \left(\tfrac{1}{2}\lambda_{min}^n - \frac{|\bar{R}_{2n}(\theta)|}{|\theta - \theta_n^*|^2} \right) \cdot |\theta - \theta_n^*|^2. \end{aligned} \tag{53}$$

Since $V_n(\theta_n^*) \to V$, we have $\lambda_{min}^n \to \lambda_{min}^*$ as $n \to \infty$, where $\lambda_{min}^* > 0$ because V is positive definite, and we can choose an $n_1 < \infty$ such that

$$\lambda_{min}^n \geq \lambda_{min}^*/2, \quad n \geq n_1. \tag{54}$$

Moreover, (22) implies that there exist an $n_2 < \infty$ and a neighborhood B of θ^* such that

$$\sup_{\theta \in B} \frac{|\bar{R}_{2n}(\theta)|}{|\theta - \theta_n^*|^2} \leq \tfrac{1}{8}\lambda_{min}^*, \quad n \geq n_2. \tag{55}$$

Now (53), (54) and (55) imply that

$$\bar{f}_n(\theta) - \bar{f}_n(\theta_n^*) \geq \tfrac{1}{8}\lambda_{min}^* \, |\theta - \theta_n^*|^2, \quad \theta \in B$$

for $n \geq \max(n_1, n_2)$, and the first part is complete.

Next, the uniform integrability condition (26) implies that the expectation on the left of (20) exists and is uniformly bounded with $r_n = (nc_n)^{1/2}$. Then, given this integrability, the linear approximation (23), the weak convergence (24) and the stochastic equicontinuity (25) of the residuals together imply the stated modulus of continuity (20). The conclusion now follows from Proposition 11. ■

Proof of Theorem 13. The proof that $H(t)$ has an almost surely unique minimizer over $T(\theta^*)$ is identical to Geyer's (1994) Proposition 4.2 and Theorem 4.4. >From Lemma 24, we have that $\tilde{t}_n \equiv (nc_n)^{1/2}(\hat{\theta} - \theta_n^*)$ is $O_p(1)$, and obviously $\tilde{t}_n$ minimizes $H_n(t)$ over $(nc_n)^{-1}(\Theta - \theta_n^*)$. Finally, we may invoke Proposition 10 to show that $\tilde{t}_n$ has the limit distribution of the minimizer of $H(t)$ over $T(\theta^*)$. ■

Proof of Corollary 14. Since $\alpha_n \to 0$, we have $Z \sim N(0, A)$ and the first result follows from Theorem 13. ■

Proof of Corollary 15. We will show that $r_n(\theta_n^* - \theta_0) \to 0$, where r_n is still $(nc_n)^{1/2}$. By the definition of θ_n^*, we have $\nabla \bar{f}_n(\theta_n^*) = 0$ and hence

$$r_n \nabla \bar{f}_n(\theta_n^*) = 0. \tag{56}$$

We expand $r_n \nabla \bar{f}_n(\theta_n^*)$ by (30) to get

$$r_n \nabla \bar{f}_n(\theta_n^*) = r_n \nabla \bar{f}_n(\theta_0) + V_n(\tilde{\theta}_n) r_n(\theta_n^* - \theta_0) = 0,$$

where $\tilde{\theta}_n$ lies on the line segment between θ_n^* and θ_0, and the last equality comes from (56). Noting that $\bar{f}_n(\theta) = \gamma_n(\theta) + \alpha_n \delta_n(\theta)$, we observe that

$$r_n \nabla \bar{f}_n(\theta_0) = r_n \gamma_n(\theta_0) + r_n \alpha_n \delta(\theta_0) = r_n \alpha_n \delta_n(\theta_0) \to 0,$$

because $r_n \alpha_n = q_n/(nc_n)^{1/2} \to 0$ by assumption, and $V_n(\tilde{\theta}_n) \to V(\theta_0)$, which is nonsingular. Therefore, $r_n(\theta_n^* - \theta_0) \to 0$, so that the limit of $r_n(\hat{\theta} - \theta_0)$ equals that of $r_n(\hat{\theta} - \theta_n^*)$. ■

B Asymptotic Variance Matrix Formula

This section calculates the asymptotic variance matrix of $(nc_n)^{1/2}(\hat{\theta} - \theta_n^*)$ under the assumption that $f_n(\cdot)$ permits a second order Taylor series expansion, i.e., for the case $\Delta_n(\theta) = \nabla W_n(\theta)$ and $V_n(\theta) = \nabla^2 \bar{f}_n(\theta) = \nabla^2[\gamma_n(\theta) + \alpha_n \delta_n(\theta)]$. Let $V = \lim_{n\to\infty} V_n(\theta_n^*)$. We usually have $V = \lim_{n\to\infty} \nabla^2[\gamma_n(\theta^*) + \alpha_n \delta_n(\theta^*)]$.

The limit variance of $(nc_n)^{1/2}(\hat{\theta} - \theta_n^*)$ is

$$V^{-1} \lim_n E(Z_n^* Z_n^{*\prime}) V^{-1}, \quad Z_n^* = (nc_n)^{1/2} \nabla W_n(\theta_n^*),$$

where

$$(nc_n)^{1/2} W_n(\theta) = 2c_n^{-1/2} \bar{m}_n(\theta)' \zeta_n(\theta) + \alpha_n^{1/2} q_n^{-1/2} \left[\zeta_n(\theta)' \zeta_n(\theta) - E\zeta_n(\theta)' \zeta_n(\theta)\right].$$

Suppressing the θ argument, we write

$$\tfrac{1}{2}(nc_n)^{1/2} \nabla W_n = c_n^{-1/2} D_n' \zeta_n + c_n^{-1/2} \nabla \zeta_n' \bar{m}_n + \alpha_n^{1/2} q_n^{-1/2} (\nabla \zeta_n' \zeta_n - E \nabla \zeta_n' \zeta_n).$$

Suppressing the argument again, but this time evaluating the functions at θ_n^*, we obtain the covariance matrix of Z_n^* as follows. (It is sometimes useful, e.g., in Example 16, to assume *iid* components across i and express the elements of the above formula in terms of the moments of more primitive variables.) Let $\xi_{n,i} = \xi_n(w_{ni}, \theta^*)$ and $\xi_{n,i}^{(k)} = \frac{\partial}{\partial \theta_k} \xi_n(w_{ni}, \theta^*)$. This short-cut notation applies only to (57) below. Let $D_{n,k}$ be the kth column of D_n, and

$$\begin{aligned}
&\Omega_n = E\xi_{n,i}\xi_{n,i}', \quad \Omega_{n,k}^{\dagger} = E\xi_{n,i}\xi_{n,i}^{(k)\prime}, \quad \Omega_{n,kl}^{\ddagger} = E\xi_{n,i}^{(k)}\xi_{n,i}^{(l)\prime}, \\
&\omega_{n,kl}^{\sharp} = q_n^{-2} \mathrm{cov}(\xi_{n,i}^{(k)\prime}\xi_{n,i},\, \xi_{n,i}^{(l)\prime}\xi_{n,i}), \\
&\omega_{n,kl}^{\flat} = q_n^{-1} E\left(\xi_{n,i}^{(k)\prime}\xi_{n,j}\xi_{n,i}'\xi_{n,j}^{(l)} + \xi_{n,i}^{(k)\prime}\xi_{n,j}\xi_{n,j}'\xi_{n,i}^{(l)}\right), \\
&\Omega_{n,k}^{\triangleright} = q_n^{-1} E\xi_{n,i}\xi_{n,i}'\xi_{n,i}^{(k)}, \text{ and } \Omega_{n,kl}^{\triangleleft} = q_n^{-1} E\xi_{n,i}^{(k)}\xi_{n,i}^{(l)\prime}\xi_{n,i}.
\end{aligned} \tag{57}$$

Also let

$$\begin{aligned}\tfrac{1}{4}A_{n,kl} &= c_n^{-1}\left(D'_{n,k}\Omega_n D_{n,l} + D'_{n,k}\Omega^{\dagger}_{n,l}\bar{m}_n + D'_{n,l}\Omega^{\dagger}_{n,k}\bar{m}_n + \bar{m}'_n\Omega^{\ddagger}_{n,kl}\bar{m}_n\right)\\ &\quad + \alpha_n\left(\tfrac{q_n}{n}\omega^{\sharp}_{n,kl} + \omega^{\flat}_{n,kl}\right) - \alpha_n n^{-1}\omega^{\flat}_{n,kl}\\ &\quad + \alpha_n\left(D'_{n,k}\Omega^{\triangleright}_{n,l} + D'_{n,l}\Omega^{\triangleright}_{n,k} + \bar{m}'_n\Omega^{\triangleleft}_{n,kl} + \bar{m}'_n\Omega^{\triangleleft}_{n,lk}\right),\end{aligned}$$

and let A_n be the $p\times p$ matrix of $A_{n,kl}$, $k,l=1,\ldots,p$. Then, in this case the limit covariance of $(nc_n)^{1/2}(\hat{\theta}-\theta^*)$ is $V^{-1}AV^{-1}$ where $A=\lim_{n\to\infty}A_n$.

When $\alpha_n\to 0$, we have $\hat{\theta}\to\theta_0$ and the limit variance matrix of $\frac{1}{2}Z_n^*$ dramatically simplifies to $\lim_n c_n^{-1}D'_n\Omega_n D_n$, where

$$D_n=\nabla\bar{m}_n(\theta_0) \text{ and } \Omega_n := n^{-1}\sum_{i=1}^{n} E\xi_n(y_{ni},\theta_0)\xi_n(y_{ni},\theta_0)'.$$

The component $\frac{1}{2}V$ simplifies to $\lim c_n^{-1}D'_nD_n$, and then the limit variance of $(nc_n)^{1/2}(\hat{\theta}-\theta_0)$ is

$$\lim_{n\to\infty}(c_n^{-1}D'_nD_n)^{-1}c_n^{-1}D'_n\Omega_nD_n(c_n^{-1}D'_nD_n)^{-1},$$

which is comparable to that obtained in conventional large sample asymptotics except that the estimator itself is scaled by the additional factor of $c_n^{1/2}$ with corresponding changes in the scaling of the variance matrix.

Derivation of (32). Define $\varepsilon_i=y_i-\theta_0$ and let z_i (without the superscript n for simplicity) be the $q_n\times 1$ vector of the z^n_{ki} for $k=1,\ldots,q_n$. Let 1_q denote the $q_n\times 1$ vector with unity in each position. Let $v_i=z_i-Ez_i=z_i-n^{-1/2}r1_q$. Then, from $g_n(\cdot,\theta)=z_i\varepsilon_i-z_i(\theta-\theta_0)$**,** we have

$$\xi_n(\cdot,\theta)=z_i\varepsilon_i-v_i(\theta-\theta_0), \text{ and } \nabla\xi_n(\cdot,\theta)=-v_i.$$

(Of course, we have $E\varepsilon_i^2=\text{var}\,(y_i)=\sigma^2$ and $Ev_iv_i'=\sigma_z^2I$.) Therefore, using the notation in (57), we have

$$\xi_{n,i}=z_i\varepsilon_i, \text{ and } \xi^{(k)}_{n,i}=-v_i,$$

(because $\theta^*=\theta_0$). The explicit values for the elements defined in (57) are

$$\begin{aligned}&\Omega_n=\sigma^2(\sigma_z^2I_q+n^{-1}r^2 1_q1_q'),\quad \Omega_n^{\dagger}=0,\quad \Omega_n^{\ddagger}=\sigma_z^2I_q,\\ &\omega_n^{\sharp}=\sigma^2\sigma_z^4+q_n^{-1}\sigma^2(Ev_{ki}^{n\,4}-\sigma_z^4)+(nq_n)^{-1}r^2\sigma^2\sigma_z^2,\quad \omega_n^{\flat}=\sigma^2\sigma_z^4+n^{-1}r^2\sigma^2\sigma_z^2,\\ &\Omega_n^{\triangleright}=-n^{-1/2}r\sigma^2\sigma_z^2(1-q_n^{-1})1_q,\quad \text{and } \Omega_n^{\triangleleft}=0,\end{aligned}$$

where $v^n_{ki}=z^n_{ki}-Ez^n_{ki}$. In addition, $\bar{m}_n(\theta)=-n^{-1/2}r1_q(\theta-\theta_0)$, and therefore from the fact that $\theta^*=\theta_0$,

$$m_n(\theta^*)=0, \text{ and } D_n=-n^{-1/2}r1_q.$$

For the denominator, from (10), we have $\frac{1}{2}V = r^2 + \sigma_z^2$. The asymptotic variance (32) is then straightforwardly obtained by direct calculation and by taking limits. ■

Asymptotic normality when $r = 0$. Consider the above model again and develop the limit distribution theory when $r = 0$. Let z denote the $n \times q_n$ observation matrix of z_{ij}. Then the (unweighted) GMM estimator is

$$\hat{\theta} = \theta_0 + (1'zz'1)^{-1}1'zz'\varepsilon,$$

where 1 is the n-vector of ones and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$. We show that $q_n^{1/2}(\hat{\theta} - \theta_0) \to_d N[0, (1+c)\sigma^2]$ as $n \to \infty$ and $q_n \to \infty$.

First, we can show that $(nq_n)^{-1}1'zz'1 \to_p \sigma_z^2$ as $n \to \infty$ by showing that its mean is σ_z^2 and its variance shrinks to zero. So it remains to show that

$$n^{-1}q_n^{-1/2}1'zz'\varepsilon \to_d N\left[0, (1+c)\sigma_z^4\sigma^2\right]. \quad (58)$$

Rewrite $n^{-1}q_n^{-1/2}1'zz'\varepsilon = n^{-1/2}\sum_{i=1}^n (nq_n)^{-1/2}1'zz_i\varepsilon_i$. Note that when z_{ij} has finite fourth moment,

$$(nq_n)^{-1}E[(1'zz_i)^2] = \omega_n^2 = \left[n^{-1}E(z_{ij}^4) + (1 + q_n/n - 2/n)\,\sigma_z^4\right].$$

Let $\eta_{ni} = \omega_n^{-1}(nq_n)^{-1/2}1'zz_i$ and $v_i = \sigma_\varepsilon^{-1}\varepsilon_i$ so that $E\eta_{ni}^2 = 1$ and $Ev_i^2 = 1$. Let $U_n = n^{-1/2}\sum_{i=1}^n \eta_{ni}v_i$. We will show that if $Ez_{ij}^8 < \infty$ and $E\varepsilon_i^2 < \infty$ then $U_n \to_d N(0,1)$ by showing that for all t,

$$Ee^{itU_n} \to e^{-t^2/2}, \quad (59)$$

which implies (58) because $\omega_n^2 \to (1+c)\sigma_z^4$.

As the first step, we can show that if $Ez_{ij}^8 < \infty$ then

$$\frac{1}{n}\sum_{i=1}^n \eta_{ni}^2 \to_p 1, \quad (60)$$

$$\max_{i \le n} n^{-1/2}|\eta_{ni}| \to_p 0, \quad (61)$$

where (60) holds because the variance shrinks to zero, and (61) holds because

$$P\left(\max_{i\le n} n^{-1/2}|\eta_{ni}| > c\right) \le \sum_{i=1}^n P\left(n^{-1/2}|\eta_{ni}| > c\right) \le n^{-2}c^{-2}\omega_n^4 \sum_{i=1}^n E\eta_{ni}^4 \to 0.$$

Let $\varphi_{nj}(t) = E(e^{itn^{-1/2}\eta_{nj}\varepsilon_j}|z)$ and $\psi_{nj} = 1 - t^2\eta_{nj}^2/2n$. For any sequence of (random) sets A_n, we have the inequality

$$\left|E^{itU_n} - e^{-t^2/2}\right| \le E\left|\prod_{i=1}^n \varphi_{ni}(t) - e^{-t^2/2}\right| \le E\left|\prod_{i=1}^n \varphi_{ni}(t) - \psi_{ni}(t)\right| 1_{A_n} + E\left|\prod_{i=1}^n \psi_{ni}(t) - e^{-t^2/2}\right| 1_{A_n} + E\left|\prod_{i=1}^n \varphi_{ni}(t) - e^{-t^2/2}\right| 1_{A_n^c}. \quad (62)$$

Now fix a $\delta > 0$. Let $c^* = \delta/t^3 > 0$, $d^* = \min\{t^{-2}, \delta/(3a_0t^2)\} > 0$, and b^* such that $(b^*)^2 \leq \min\{1, t^{-2}, 2/[a_1t^4(1+d^*)], 2\delta/[3a_0a_1t^4(1+d^*)]\}$ and $Ev_i^2\{|v_i| > c^*/b^*\} \leq \delta/(6t^2)$. Then, because $Ev_i^2 < \infty$, we can choose a $b^* > 0$.

Let $A_n = \{\max_{i\leq n} n^{-1/2}|\eta_{ni}| \leq b^*\} \cap \{|n^{-1}\sum_{i=1}^n \eta_{ni}^2 - 1| \leq d^*\}$. Then, because of (60) and (61), we can pick an $n^* < \infty$ such that $P(A_n) \geq 1 - \delta/6$ for all $n > n^*$. And for those chosen constants and A_n, we can show each of the three terms on the right of (62) to be bounded by $\delta/3$ for $n > n^*$, and hence[2] the sum is bounded by δ. ■

CUE Weighting. The CUE with $\Omega_n(\theta)^{-1}$ as weight function discussed in section 5 is equivalent to unweighted GMM using $\Omega_n(\theta)^{-1/2}g_n(\cdot,\theta)$, so the associated "$\delta(\theta)$" function is constant, $\theta_n^* \equiv \theta_0$, and the associated covariance matrix of the rescaled moment conditions (i.e, the new "$\Omega_n(\theta)$" matrix) is the identity. The $\frac{1}{2}V_n(\theta_0)$ matrix reduces to $c_n^{-1}D_n'\Omega_n^{-1}D_n$ where D_n and Ω_n are evaluated at θ_0. So we have

$$\tfrac{1}{2}V = \lim_{n\to\infty} \frac{1}{c_n} D_n'\Omega_n^{-1}D_n.$$

The "center" part in the sandwich form is more complicated. The CUE using $\Omega_n(\theta)^{-1}$ as weight has

$$\begin{aligned}\tfrac{1}{2}W_n(\theta) &= c_n^{-1}n^{-1/2}\bar{m}_n(\theta)'\Omega_n(\theta)^{-1}\zeta_n(\theta) + \alpha_n\left[q_n^{-1}\zeta_n(\theta)'\Omega_n(\theta)^{-1}\zeta_n(\theta) - 1\right] \\ &= c_n^{-1}n^{-1/2}\bar{m}_n(\theta)'\Omega_n(\theta)^{-1/2}\tilde{\zeta}_n(\theta) + \alpha_n\left[q_n^{-1}\tilde{\zeta}_n(\theta)'\tilde{\zeta}_n(\theta) - 1\right],\end{aligned}$$

where $\tilde{\zeta}_n(\theta) = \Omega_n(\theta)^{-1/2}\zeta_n(\theta)$ with $E\tilde{\zeta}_n(\theta)\tilde{\zeta}_n(\theta)' = I$, and $\Omega_n(\theta) = E\zeta_n(\theta)\zeta_n(\theta)'$ as before. Because $\bar{m}_n(\theta_0) \equiv 0$, we have

$$\begin{aligned}\tfrac{1}{2}Z_n &= \tfrac{1}{2}(nc_n)^{1/2}\nabla W_n(\theta_0) \\ &= c_n^{-1/2}D_n(\theta_0)'\Omega_n(\theta_0)^{-1/2}\tilde{\zeta}_n(\theta_0) + \alpha_n^{1/2}q_n^{-1/2}\nabla\tilde{\zeta}_n(\theta_0)'\tilde{\zeta}_n(\theta_0).\end{aligned}$$

Clearly,

$$\begin{aligned}\tfrac{1}{4}EZ_nZ_n' &= c_n^{-1}D_n'\Omega_n^{-1}D_n + \alpha_nq_n^{-1}E\left[\nabla\tilde{\zeta}_n(\theta_0)'\tilde{\zeta}_n(\theta_0)\tilde{\zeta}_n(\theta_0)'\nabla\tilde{\zeta}_n(\theta_0)\right] \\ &\quad + c_n^{-1/2}\alpha_n^{1/2}q_n^{-1/2}D_n(\theta_0)'\Omega_n(\theta_0)^{-1/2}E\left[\tilde{\zeta}_n(\theta_0)\tilde{\zeta}_n(\theta_0)'\nabla\tilde{\zeta}_n(\theta_0)\right] \\ &\quad + c_n^{-1/2}\alpha_n^{1/2}q_n^{-1/2}E\left[\nabla\tilde{\zeta}_n(\theta_0)'\tilde{\zeta}_n(\theta_0)\tilde{\zeta}_n(\theta_0)'\right]\Omega_n(\theta_0)^{-1/2}D_n(\theta_0).\end{aligned}$$

Note that estimation of this variance form is, in practice, quite complicated even in the simplest model because it involves third and fourth moments and $\tilde{\zeta}_n(\theta)$ is a product of the matrix function $\Omega_n(\theta)^{-1/2}$ and the rescaled average $n^{-1/2}\sum_{i=1}^n \xi_n(w_{ni},\theta)$. Bekker (1994) provides an estimator for his linear model with Gaussian errors but it is unclear how this can be generalized to GMM.

[2] A complete demonstration is available on request.

References

Ahn, S. C. and P. Schmidt (1995). Efficient Estimation of Models for Dynamic Panel Data. *Journal of Econometrics*, 68, 5–27.

Anderson, T. W. (1977). Asymptotic Expansions of the Distributions of Estimates in Simultaneous Equations for Alternative Parameter Sequences. *Econometrica*, 45(2), 509–518.

Angrist, J. D. (1990). Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records. *American Economic Review*, 80(3), 313–336.

Angrist, J. D. and A. B. Krueger (1991). Does Compulsory School Attendance Affect Schooling and Earnings? Quarterly *Journal of Economics*, 106(4), 979–1014.

Bekker, P. A. (1994). Alternative Approximations to the Distributions of Instrumental Variable Estimators. *Econometrica*, 62(3), 657–681.

Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons, Inc.

Chamberlain, G. C. (1987). Asymptotic Efficiency in Estimation with Conditional Moment Restrictions. *Journal of Econometrics*, 34, 305–334.

Chao, J. C. and N. R. Swanson (2002). Consistent Estimation with a Large Number of Weak Instruments. Unpublished Manuscript.

Chao, J. C. and N. R. Swanson (2003). Asymptotic Normality of Single Equation Estimators for the Case with a Large Number of Weak Instruments. Unpublished Manuscript.

Davidson, J. (2000), *Econometric Theory*, Blackwell Publishers.

Donald, C. G, G. W. Imbens and W. K. Newey (2003). Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions. *Journal of Econometrics*, 117, 55–93.

Donald, C. G and W. K. Newey (2001). Choosing the Number of Instruments. *Econometrica*, 69(5), 1161–1191.

Geyer, C. J. (1994). On the Asymptotics of Constrained M-estimation. *Annals of Statistics*, 22(4), 1993–2010.

Han, C., L. Orea, and P. Schmidt (2003). A Panel Data Estimation with Parametric Temporal Variation in Individual Effects. Forthcoming in *Journal of Econometrics*.

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

Hillier, G. H. (2004). Yet More on the Exact Properties of Instrumental Variable Estimtors. Unpublished paper, Southampton University.

Kim, J. and D. Pollard (1990), Cube Root Asymptotics. *The Annals of Statistics*, 18, 191–219.

Knight, K. (2003), Epi-convergence in Distribution and Stochastic Equi-semicontinuity, Unpublished manuscript.

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

Lai, T. L. and C. Z. Wei (1982). Least Squares Estimates in Stochastic Regression Models with Applications to Identification and Control of Dynamic Systems. *Annals of Statistics*, 10, 154-166.

Morimune, K. (1983). Approximate Distributions of k-class Estimators When the Degree of Overidentification Is Large Compared with the Sample Size. *Econometrica*, 51(3), 821–842.

Phillips, P. C. B. (1980). "The exact finite sample density of instrumental variable estimators in an equation with $n+1$ endogenous variables," *Econometrica* 48:4, 861–878.

Phillips, P. C. B. (1983). "Exact small sample theory in the simultaneous equations model," Chapter 8 and pp. 449–516 in M. D. Intriligator and Z. Griliches (eds.), *Handbook of Econometrics*. Amsterdam: North–Holland.

Phillips, P. C. B. (1984). "The exact distribution of LIML: I". *International Economic Review*, 25, 249-261.

Phillips, P. C. B. (1989). Partially Identified Econometric Models. *Econometric Theory*, 5, 181–240.

Phillips, P.C.B. and H.R. Moon (1999) : Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica,* 67, 1057-1111.

Phillips, P. C. B and V. Solo (1992). Asymptotics for Linear Processes, *Annals of Statistics*, 20(2), 971–1001.

Rockafellar, R. T. and R. J.-B. Wets (1998). *Variational Analysis*, Springer.

Staiger, D. and J. H. Stock (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3), 557–586.

Stock, J.H., J.H. Wright, and M. Yogo, (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments", *Journal of Business and Economic Statistics, **20**, 518-529.*

Stock, J.H. and M. Yogo, (2004), "Asymptotic distributions of instrumental variables statistics with many weak instruments". in D. W. K. Andrews and J. H. Stock, eds., *Festschrift in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, forthcoming.

Stock, J. H. and J. H. Wright (2000). GMM with Weak Identification. *Econometrica*, 68(5), 1055–1096.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*, Cambridge University Press.

Van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*, Springer-Verlag.