Many Weak Moment Asymptotics for the Continuously Updated GMM Estimator*

Whitney K. Newey
Department of Economics
M.I.T.

September 2004

Abstract

Many instruments can result in improved inference in weakly identified models. Estimators with low bias for many instruments and standard errors that adjust for the presence of many instruments are useful for this purpose. This paper gives standard errors for the continuously updated GMM estimator that adjust for the number of overidentifying restrictions. This adjustment is based on many weak instrument asymptotics of Chao and Swanson (2002) and Han and Phillips (2003).

JEL Classification:
Keywords: GMM, Continuous Updating, Many Moments, Variance Adjustment

*The NSF provided partial financial support for this paper.

1 Introduction

Many applications of generalized method of moments (GMM, Hansen, 1982) tend to be weakly identified. Examples include natural experiments (Angrist and Krueger, 1991), consumption asset pricing models (Hansen and Singleton, 1982), and dynamic panel models (Holzt-Eakin, Newey, and Rosen, 1990). In these settings the use of many instruments may improve accuracy. For example, Hansen, Hausman, and Newey (2004) have recently found that using all 180 instruments in the Angrist and Krueger (1991) application shrinks correct confidence intervals substantially. In such settings accuracy of asymptotic approximations may depend on accounting for many instruments. Many instruments can lead to bias in the usual two-step GMM methods (Newey and Smith, 2004). This problem can be dealt with by using alternative estimators with smaller bias, but then it may be important to account for the effect of many instruments on standard errors. For example, Hansen, Hausman, and Newey (2004) recently found that in the Angrist and Krueger (1991) application, asymptotic confidence intervals based on limited information maximum likelihood (LIML) with the usual standard errors did not have the right size, but using Bekker (1994) standard errors corrects the size problem, and still results in a substantial shortening of confidence intervals.

This paper proposes a solution to the inference problem with many weak moment conditions in GMM estimation. We consider the continuous updating estimator (CUE) of Hansen, Heaton, and Yaron (1996). Their Monte Carlo results and the theory of Donald and Newey (2000) and Newey and Smith (2004) suggest that the CUE may have low bias relative to other GMM estimators. Here we give standard errors for the CUE that account for many instruments. We find that under many weak moments the CUE is asymptotically normal with an asymptotic variance that is larger than the usual formula due to many instruments. We give a consistent estimator of the asymptotic variance that is straightforward to compute, so that Wald inference can be carried out in the usual way. We argue that these standard errors provide an analog for GMM of the Bekker (1994) standard errors for LIML. We show that the asymptotic variance reduces to the Stock and Yogo (2004) and Hansen, Hausman, and Newey (2004) version of Bekker (1994) in a homoskedastic linear model.

Our results can be related to the higher-order variance of the CUE derived in Donald and Newey (2003). The higher-order variance has two types of terms, one type corresponding to variability of the Jacobian of the moment function and the other type to estimation of the weighting matrix. As they show, in weakly identified models the Jacobian variability term will dominate. The many weak moment approximation given here can be considered as a limit of the higher order approximation as identification weakens and the number of instruments grow, with the nice feature of having a Gaussian limit.

One could also try to adapt the Bekker (1994) asymptotics to GMM by letting the number of instruments grow as fast as the sample size. A technical difficulty with doing so is that the weighting matrix (which has dimension equal to the square of the number of moments) could be unstable. We finesse this difficulty by allowing the number of moments to grow more slowly than the sample size, but restricting the GMM asymptotic variance to grow at the same rate as the number of moments, and hence slower than the sample size. This kind of asymptotics seems well suited as an approximation for applications like those mentioned above where identification is weak, there are many instruments, and the sample size is quite large relative to the number of instruments. However, it does "wash out" the estimation of the weighting matrix from the asymptotic approximation, which could lead to a poor approximation in applications that are more strongly identified.

The asymptotic sequence we consider here is a special case of those considered in Chao and Swanson (2002) and Han and Phillips (2003). It is included in the Han and Phillips (2003) cases where there is both a noise and a signal component to the limit of the GMM objective function. Here we give a general consistent variance estimator for the CUE under many weak moments and general conditions for the CUE that account for estimation of the weighting matrix, neither of which was done in Han and Phillips (2003). Our asymptotic variance approximation also differs from Chao and Swanson (2004) in accounting for the both the usual term and the additional, Jacobian variability term in the asymptotic variance.

A formal derivation of the asymptotic distribution results can be obtained using results of Stock and Wright (2002). If one takes the limiting distribution of the CUE under weak identification, and allows the identification and the number of moment conditions to grow at the same rate, one obtains the asymptotic variance given here, as we discuss below. This derivation corresponds to a sequential asymptotics, where one lets the number of observations go to infinity and then lets identification and the number of moments grow. We give here simultaneous asymptotics, where identification and the number of moments grows along with, but slower than, the sample size.

The variance adjustment that comes out of the many weak instrument asymptotics is different than that of Windmeijer (2004). He adjusts for the variability of the weight matrix while the many instrument asymptotics adjust for the variability of the moment derivative. Adjusting only for variability of the moment derivative is appropriate in weakly identified cases, as mentioned above.

In Section 2 we describe the model, the CUE estimator, the asymptotic variance estimator that accounts for many weak moment conditions, and discuss the asymptotic sequence we consider. Section 3 gives the consistency results. Section 4 gives the asymptotic normality. Section 5 offers some conclusions and some possible directions for future work.

2 The Model and Estimators

The model we consider is for stationary data where there is a countable number of moment restrictions. We allow the moment generating process to depend on the sample size to model weak identification. To describe the model, let w_i, $(i = 1, ..., n)$, be strictly stationary observations on a data vector w. Also, let β be a $p \times 1$ parameter vector and $g(w, \beta) = (g_1^m(w, \beta), ..., g_m^m(w, \beta))'$ be an $m \times 1$ vector of functions of the data observation w and the parameter, where $m \geq p$. For notational convenience we suppress an m superscript on $g(w, \beta)$. The model has a true parameter β_0 satisfying the moment condition

$$E[g(w,\beta_0)]=0,$$

where $E[.]$ denotes expectation taken with respect to the distribution of w_i for sample size n, and we suppress the dependence on n for notational convenience.

To describe the CUE estimator, we need to describe the weighting matrix. To do so let $g_i(\beta)\equiv g(w_i,\beta)$, $\bar{g}(\beta)=E[g_i(\beta)]$, $\hat{g}(\beta)\equiv n^{-1}\sum_{i=1}^n g_i(\beta)$, and

$$\Omega(\delta,\beta)=n\{E[\hat{g}(\delta)\hat{g}(\beta)']-\bar{g}(\delta)\bar{g}(\beta)'\}=Cov(\sqrt{n}\hat{g}(\delta),\sqrt{n}\hat{g}(\beta)).$$

This matrix is the covariance between $\sqrt{n}\hat{g}(\delta)$ and $\sqrt{n}\hat{g}(\beta)$ in a sample of size n. In the i.i.d. case it will not depend on n but will with dependent observations, although for notational convenience we do not index $\Omega(\delta,\beta)$ by n. Let $\hat{\Omega}(\delta,\beta)$ be an estimator of $\Omega(\delta,\beta)$ and $\hat{\Omega}(\beta)=\hat{\Omega}(\beta,\beta)$. For i.i.d. data we will let this estimator be a standard one,

$$\hat{\Omega}(\delta,\beta)=\sum_{i=1}^n g_i(\delta)g_i(\beta)'/n-\hat{g}(\delta)\hat{g}(\beta). \tag{2.1}$$

For dependent data we will assume that $\hat{\Omega}(\delta,\beta)$ is a specification robust, autocorrelation consistent variance estimator, i.e. that it estimates $\Omega(\delta,\beta)$ for all values of δ and β. The specification robustness feature of $\hat{\Omega}(\delta,\beta)$ is important for the consistency of the CUE under many instrument asymptotics, as further discussed below.

The CUE is given by

$$\begin{aligned}\hat{\beta} &= \arg\min_{\beta\varepsilon\mathcal{B}}\hat{Q}(\beta),\\ \hat{Q}(\beta) &= \hat{g}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{g}(\beta)n/2m.\end{aligned}$$

where B is the parameter set. This estimator minimizes over β in the moments and the weighting matrix simultaneously. It is due to Hansen, Heaton and Yaron (1996), although it differs from the estimator they consider in the choice of $\hat{\Omega}(\beta)$.[1] Our requirement that $\hat{\Omega}(\beta)$ be a specification robust estimator means that we cannot exclude from $\hat{\Omega}\,(\beta)$ autocovariances that are zero only at the truth.

[1] In the i.i.d. case the CUE will be the same as an estimator obtained by minimizing $\tilde{Q}(\beta)=\hat{g}(\beta)'[\sum_{i=1}^n g_i(\beta)g_i(\beta)']^{-1}\hat{g}(\beta)$, as discussed in Newey and Smith (2004).

To describe the estimator of the asymptotic variance let

$$\begin{aligned}\hat{D}(\beta) &= [\hat{D}_1(\beta), ..., \hat{D}_p(\beta)], \\ \hat{D}_j(\beta) &= \sqrt{n/m}\left[\partial\hat{g}(\beta)/\partial\beta_j - \partial\hat{\Omega}(\delta,\beta)/\partial\delta_j|_{\delta=\beta}\hat{\Omega}(\beta)^{-1}\hat{g}(\beta)\right].\end{aligned}$$

An estimator of the asymptotic variance is given by

$$\hat{V} = \hat{H}^{-1}\hat{D}(\hat{\beta})'\hat{\Omega}^{-1}\hat{D}(\hat{\beta})\hat{H}^{-1}, \hat{H} \overset{def}{=} \partial^2\hat{Q}(\hat{\beta})/\partial\beta\partial\beta', \hat{\Omega} \overset{def}{=} \hat{\Omega}(\hat{\beta}).$$

The "sandwich" form of the asymptotic variance estimator is important under this asymptotics. Unlike the usual asymptotics, the middle matrix estimates a different, larger object than the Hessian. Also, the use of the Hessian is important. Here we cannot replace $\hat{H}$ by the more common formula $\hat{G}'\hat{\Omega}^{-1}\hat{G}n/m$, where $\hat{G} = \partial\hat{g}(\hat{\beta})/\partial\beta$, because $\hat{G}'\hat{\Omega}^{-1}\hat{G}n/m$ has extra random terms that are eliminated in the Hessian. A similar phenomena occurs in Bekker (1994), where it is important to use the Hessian of the LIML objective function rather than the usual instrumental variables formula.

The Hessian term on the outside of $\hat{V}$ is familiar from other estimation environments. The middle term $\hat{D}(\hat{\beta})'\hat{\Omega}^{-1}\hat{D}(\hat{\beta})$ is an estimator of the asymptotic variance of $\partial\hat{Q}(\beta_0)/\partial\beta$ that is due to Kleibergen (2004). He shows that this estimator is consistent under weak identification with fixed m. We give conditions for consistency when m is allowed to grow with the sample size.

It will be shown below that, under certain conditions, there will be a matrix V such that

$$\sqrt{m}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V), \hat{V} \xrightarrow{p} V, \tag{2.2}$$

Therefore, standard (Wald) confidence intervals and test statistics that treat $\hat{\beta}$ as if it were normally distributed with mean β_0 and variance $\hat{V}/m$ will be asymptotically correct. We use m as the growth rate for the asymptotics because it is a convenient, scalar quantity. We could also have used the degree of identification, but that is more cumbersome. We will give the form of V below.

The many weak moment condition asymptotics has the individual components of $E[\partial g_i(\beta_0)/\partial\beta]$ shrink in magnitude and the number of moments grow with the sample size. Specifically, we will impose the following condition, for $G_n = \sqrt{n/m}\ E[\partial g_i(\beta_0)/\partial\beta]$

Assumption 1: $G_n'\Omega^{-1}G_n \longrightarrow H$ and H is nonsingular.

An important example is the linear model where

$$\begin{aligned} y &= x'\beta_0 + \varepsilon, x' = z'\pi_{mn} + \eta, \\ 0 &= E[\varepsilon|z], 0 = E[\eta|z]. \end{aligned}$$

Here z is an $m \times 1$ vector of instrumental variables, where we suppress the m argument for convenience, and we will impose the normalization $E[zz'] = I_m$. Also, π_{mn} is $m \times p$ matrix of reduced form coefficients. The moment functions are

$$g(w, \beta) = z(y - x'\beta).$$

Here $\Omega = E[zz'\varepsilon^2]$ and $G_n = -\sqrt{n/m}E[z_i x_i'] = -\sqrt{n/m}\pi_{mn}$, so that Assumption 1 is equivalent to

$$G_n'\Omega^{-1}G_n = (n/m)\pi_{mn}'\Omega^{-1}\pi_{mn} \longrightarrow H. \qquad (2.3)$$

For instance, if x is a scalar and $\pi_{mn} = De/\sqrt{n}$, where e is an $m \times 1$ vector of ones, and ε is homoskedastic, with $E[\varepsilon^2|z] = \sigma_\varepsilon^2$, then

$$G_n'\Omega^{-1}G_n = (n/m)D^2e'(\sigma_\varepsilon^2 I_m)^{-1}e/n = D^2/\sigma_\varepsilon^2 = H.$$

This linear model example shows how the asymptotics here is like that of Chao and Swanson (2002) and Han and Phillips (2003). Each reduced form coefficient vanishes at rate $1/\sqrt{n}$, that is like the weak instruments case of Stock and Staiger (1997), but the number of instruments grows, which will lead to consistency and asymptotic normality. Thus, the asymptotics is described as "many weak instrument" asymptotics.

3 Consistency

The usual extremum estimator analysis can be used to show consistency of the CUE under many weak instrument asymptotics. The CUE objective function $\hat{Q}(\beta)$ will converge to a function that is minimized at the true parameter β_0. Under appropriate regularity conditions we can interchange the limiting and minimization steps to get consistency, i.e. the limit of the minimand will be the minimand of the limit.

Intuition about consistency of the CUE is provided by the limiting objective function. Under conditions given below $\hat{\Omega}(\beta) = \hat{\Omega}(\beta, \beta)$ will be close to $\Omega(\beta) = \Omega(\beta, \beta)$ in such a way that the limit of $\hat{Q}(\beta)$ will coincide with the limit of $\tilde{Q}(\beta) = (n/m)\hat{g}(\beta)'\Omega(\beta)^{-1}\hat{g}(\beta)$. Also, $\tilde{Q}(\beta)$ should be close to its expectation $\bar{Q}(\beta)$ in large samples. Let $S_n(\beta) = \bar{g}(\beta)'\Omega(\beta)^{-1}\bar{g}(\beta)n/2m$. Then

$$\begin{aligned} & \bar{Q}(\beta) = E[\tilde{Q}(\beta)] = (n/m)E[\{\hat{g}(\beta) - \bar{g}(\beta)\}'\Omega(\beta)^{-1}\{\hat{g}(\beta) - \bar{g}(\beta)\}] \\ & +(n/m)2E[\hat{g}(\beta)'\Omega(\beta)^{-1}\bar{g}(\beta)] - S_n(\beta) \\ = \; & tr(\Omega(\beta)^{-1}Var(\sqrt{n}\hat{g}(\beta))/m + S_n(\beta) = tr(I_m)/m + S_n(\beta) = 1 + S_n(\beta). \end{aligned}$$

Since $S_n(\beta)$ has a minimum (of zero) at β_0, $\bar{Q}(\beta)$ will be minimized at β_0, leading to consistency of the CUE.

The fact that $\Omega(\beta)$ is a specification robust variance matrix, i.e. that $\Omega(\beta) = Var(\sqrt{n}\hat{g}(\beta))$ for all β, is important for $\bar{Q}(\beta)$ to be minimized at β_0. If the middle, weighting matrix was not the inverse of a specification robust variance matrix, then $\bar{Q}(\beta)$ would also depend on β through the other term, and hence $\bar{Q}(\beta)$ need not be minimized at the truth. For example, if we replace $\Omega(\beta)^{-1}$ by a fixed weighting matrix W, the limiting objective function would be

$$\bar{Q}(\beta) = tr(WVar(\sqrt{n}\hat{g}(\beta)))/n + S_n(\beta).$$

This function need not be minimized at β_0, so that the GMM estimator may not be consistent under many moment asymptotics. Han and Phillips (2003) interpret $tr(WVar(\sqrt{n}\hat{g}(\beta)))/n$ as a "noise" term that contaminates the "signal" term $S_n(\beta)$. The CUE eliminates β from the noise term and so leads to consistency.

The importance for consistency of choosing $\Omega(\beta)$ to be a robust variance estimator may have implications for practice. For instance, in dependent data there is often structure to the autocovariances for $g_i(\beta_0)$, such as autocovariances being zero. Generally $g_i(\beta)$ does not have this structure for $\beta \neq \beta_0$, so that excluding autocovariances from $\hat{\Omega}(\beta)$ would be misspecification that could lead to inconsistency of the CUE under these asymptotics. For this reason it might be good to use an autocorrelation consistent variance matrix for the CUE, even when $g_i(\beta_0)$ is not autocorrelated. The theoretical potential of using autocorrelation consistent variances to reduce bias of the CUE was also noted by Donald and Newey (2003). It should also be noted that Heaton, and Yaron (1996) found large bias reductions in their Monte Carlo, with the CUE based on zero autocovariance, so it may be that the CUE has low bias in practice, even without $\hat{\Omega}(\beta)$ being an autocorrelation consistent variance estimator.

We will give general regularity conditions for i.i.d. data and primitive conditions for the linear model. The first condition specifies properties of the function $S_n(\beta)$, including an identifiable uniqueness condition for the minimizer β_0 of $S_n(\beta)$.

Assumption 2: There is a continuous function $\Delta(a)$ such that $\Delta(0) = 0$, $\Delta(a) > 0$ for all $a \neq 0$, and $S_n(\beta) \geq \Delta(\|\beta - \beta_0\|)$.

The next condition specifies the properties of the weighting matrix. For a matrix F let $\|F\| = trace(F'F)^{1/2}$ denote its Euclidean norm and for symmetric F let $\lambda_{\min}(F)$ and $\lambda_{\min}(F)$ denote its smallest and largest eigenvalues, respectively. Also, define stochastic equicontinuity of a sequence of random functions $\{\hat{A}_n(\beta)\}_{n=1}$ to mean that for any $\Delta_n \longrightarrow 0$, $\sup_{\|\tilde{\beta}-\beta\| \leq \Delta_n} |\hat{A}(\tilde{\beta}) - \hat{A}(\beta)| \xrightarrow{p} 0$.

Assumption 3: The data are i.i.d., $\beta_0 \in B$ with B compact, there is a constant C with $\lambda_{\min}(\Omega(\beta)) \geq C$, $E[\{g_i(\beta)'g_i(\beta)\}^2]/m^2n \longrightarrow 0$ for each β, $\lambda_{\max}(E[g_i(\beta)g_i(\beta)']) \leq C$, $\sup_{\beta \in B} \|\hat{\Omega}(\beta) - \Omega(\beta)\| \xrightarrow{p} 0$, $S_n(\beta)$ is equicontinuous, and $(n/m)\hat{g}(\beta)'\Omega(\beta)^{-1}\hat{g}(\beta)$ is stochastically equicontinuous.

The condition that $\sup_{\beta \in B} \|\hat{\Omega}(\beta) - \Omega(\beta)\| \xrightarrow{p} 0$ puts restrictions on the rate at which

m can grow with the sample size. If $E[g_{ij}(\beta)^2]$ is bounded uniformly in j, m, and β then a sufficient condition for pointwise convergence would be that $m^2/n \longrightarrow 0$. The uniformity condition may impose further restrictions.

The following is a consistency result for the general i.i.d. case.

THEOREM 1: *If Assumptions 2 and 3 are satisfied then* $\hat{\beta} \xrightarrow{p} \beta_0$.

We also give more primitive regularity conditions for consistency for the linear model example. Let $\Sigma(z_i) = E[(\varepsilon_i, \eta_i')'(\varepsilon_i, \eta_i')|z_i]$.

Assumption 4: The linear model holds, there is a constant C with $E[\varepsilon_i^4|z_i] \leq C$, $E[\|\eta_i\|^4|z_i] \leq C$, $\lambda_{\min}(\Sigma(z_i)) \geq 1/C$, and $\|z_i'\pi_{mn}\| \leq C$, and $E[(z_i'z_i)^2]/n \longrightarrow 0$.

The conditions puts restrictions on the rate at which m can grow with the sample size. If z_{ij} is bounded uniformly in j and m, then these conditions will hold if $m^2/n \longrightarrow 0$, for in that case,

$$\begin{aligned} \|z_i'\pi_{mn}\| &\leq \|z_i\|\,\|\pi_{mn}\| \leq \sqrt{m}O(\sqrt{m/n}) = O(\sqrt{m^2/n}), \\ E[(z_i'z_i)^2]/n &\leq O(m^2)/n = O(m^2/n). \end{aligned}$$

THEOREM 2: *If Assumptions 1 and 4 are satisfied then* $\hat{\beta} \xrightarrow{p} \beta_0$.

4 Asymptotic Normality

To explain the asymptotic normality result it is helpful to consider the first-order conditions to the CUE, given by

$$\begin{aligned} 0 &= \partial\hat{Q}(\hat{\beta})/\partial\beta_j, (j = 1, ..., p), \qquad (4.1) \\ \partial\hat{Q}(\beta)/\partial\hat{\beta}_j &= \{\partial\hat{g}(\beta)/\partial\beta_j'\hat{\Omega}(\beta)^{-1}\hat{g}(\beta) - \hat{g}(\beta)'\hat{A}_j(\beta)'\hat{\Omega}(\beta)^{-1}\hat{g}(\beta)\}n/m, \\ \hat{A}_j(\beta) &= \partial\hat{\Omega}(\delta,\beta)/\partial\delta_j|_{\delta=\beta}\hat{\Omega}(\beta)^{-1}, \end{aligned}$$

where the expression for the derivative holds by $\hat{\Omega}(\delta,\beta) = \hat{\Omega}(\beta,\delta)'$ (see Donald and Newey, 2000). As usual, a mean value expansion of the first order conditions give

$$\sqrt{m}(\hat{\beta} - \beta_0) = -\bar{H}^{-1}\sqrt{m}\partial\hat{Q}(\beta_0)/\partial\beta, \bar{H} = \partial^2\hat{Q}(\bar{\beta})/\partial\beta\partial\beta', \qquad (4.2)$$

where $\bar{\beta}$ is an intermediate value for β, being on the line joining $\hat{\beta}$ and β_0, that actually differs from row to row of $\bar{H}$. Under regularity conditions given below we will have $\bar{H} \xrightarrow{p} H$, for H from Assumption 1. The asymptotic distribution of $\hat{\beta}$ will then be determined by the asymptotic distribution of $\sqrt{m}\partial\hat{Q}(\beta_0)/\partial\beta$.

To describe this distribution, let $\Omega = \Omega(\beta_0)$ and

$$\begin{aligned} A_j &= \partial\Omega(\delta,\beta_0)/\partial\delta_j|_{\delta=\beta_0}\Omega^{-1} = nCov(\partial\hat{g}(\beta_0)/\partial\beta_j, \hat{g}(\beta_0))\Omega^{-1}, \\ \tilde{U}_j &= \sqrt{n}\{\partial\hat{g}(\beta_0)/\partial\beta_j - E[\partial\hat{g}(\beta_0)/\partial\beta_j] - A_j\hat{g}(\beta_0)\}, \tilde{U} = [\tilde{U}_1, ..., \tilde{U}_p]. \end{aligned}$$

Under conditions given below the $\hat{\Omega}(\beta_0)^{-1}$ and $\hat{A}_j(\beta_0)$ terms in $\sqrt{m}\partial\hat{Q}(\beta_0)/\partial\beta$ can be replaced by Ω^{-1} and A_j respectively such a way that

$$\begin{aligned} \sqrt{m}\partial\hat{Q}(\beta_0)/\partial\beta_j &= \{\partial\hat{g}(\beta_0)/\partial\beta_j'\Omega^{-1}\hat{g}(\beta_0) - \hat{g}(\beta_0)'A_j'\Omega^{-1}\hat{g}(\beta_0)\}n/\sqrt{m} + o_p(1) \\ &= \sqrt{n/m}E[\partial g_i(\beta_0)/\partial\beta_j]'\Omega^{-1}\sqrt{n}\hat{g}(\beta_0) + \tilde{U}^{j\prime}\Omega^{-1}\sqrt{n}\hat{g}(\beta_0)/\sqrt{m} + o_p(1). \end{aligned}$$

Stacking these equations gives

$$\sqrt{m}\partial\hat{Q}(\beta_0)/\partial\beta = G_n'\Omega^{-1}\sqrt{n}\hat{g}(\beta_0) + \tilde{U}'\Omega^{-1}\sqrt{n}\hat{g}(\beta_0)/\sqrt{m} + o_p(1).$$

The first term following the equality is a random vector with variance $G_n'\Omega^{-1}G_n$ that converges to H by Assumption 1. Thus, a central limit theorem should apply to give asymptotic normality, with variance H, of the first term. To understand the second term's behavior, consider for the moment the case where m is fixed. In that case, as $n \longrightarrow \infty$, a central limit theorem will imply that $\tilde{U}$ and $\sqrt{n}\hat{g}(\beta_0)$ converge (jointly) to Gaussian vectors. These Gaussian vectors will be uncorrelated, and hence independent, because $\tilde{U}$ is the matrix of residuals from the projection of the derivatives on the moment functions (see also Donald and Newey, 2000). Because of this independence, in the limit the second term will be asymptotically normal, conditional on the limit of $\tilde{U}$, with variance equal to the limit of $\tilde{U}'\Omega^{-1}\tilde{U}/m$. As m grows this conditional variance matrix will converge in probability to

$$\Lambda^* = \lim_{n\longrightarrow\infty}\Lambda_n, \Lambda_n = E[\tilde{U}'\Omega^{-1}\tilde{U}]/m,$$

leading to asymptotic normality. Furthermore, for fixed m the limit of the first and second terms will be uncorrelated, so that the asymptotic variance of $\sqrt{m}\partial\hat{Q}(\beta_0)/\partial\beta$ will be the sum of the variances of the two terms. This sum of variances will be $H+\Lambda^*$, so that

$$\sqrt{m}\partial\hat{Q}(\beta_0)/\partial\beta \xrightarrow{d} N(0,H+\Lambda^*).$$

Then by equation (4.2) $\sqrt{m}(\hat{\beta}-\beta_0)$ will be asymptotically normal with asymptotic variance

$$V \overset{def}{=} H^{-1}+H^{-1}\Lambda^*H^{-1}.$$

This interpretation is given only for the purpose of providing some intuition about the asymptotic distribution. In the actual theorems we allow m to grow with n, and asymptotic normality is based on a using a Martingale central limit similarly to Hall (1994).

For comparison purposes it is useful to consider a corresponding variance approximation V_n for $\hat{\beta}$ for a sample size of size n. We can compare the variance approximation here with the standard GMM variance approximation. Let $\bar{g}_\beta = E[\partial g(w_i,\beta_0)/\partial\beta]$ and note that H is the limit of $\bar{g}_\beta'\Omega^{-1}\bar{g}_\beta n/m$. Replacing H with this object, Λ^* by Λ_n, and dividing by m (the square of the convergence rate) gives the variance approximation for sample size n of

$$\begin{aligned} V_n &= (\bar{g}_\beta'\Omega^{-1}\bar{g}_\beta n/m)^{-1}/m+(\bar{g}_\beta'\Omega^{-1}\bar{g}_\beta n/m)^{-1}\Lambda_n(\bar{g}_\beta'\Omega^{-1}\bar{g}_\beta n/m)^{-1}/m \\ &= (\bar{g}_\beta'\Omega^{-1}\bar{g}_\beta)/n+\frac{m}{n}(\bar{g}_\beta'\Omega^{-1}\bar{g}_\beta)^{-1}\Lambda_n(\bar{g}_\beta'\Omega^{-1}\bar{g}_\beta)^{-1}/n \end{aligned}$$

The usual variance approximation for GMM is $(\bar{g}_\beta'\Omega^{-1}\bar{g}_\beta)^{-1}/n$. The approximate varianc V_n includes an additional term. In a strongly identified model with fixed m the additional term is order $1/n^2$ and is a higher-order variance term. Indeed, by inspection of Donald and Newey (2003), one can see that the additional term corresponds to one of the higher order variance terms, that is largest when identification is weak. It can be shown that this term would dominate the higher-order variance as identification becomes weak and the number of moments grows. The present result also suggests that the higher-order distributional approximation would be approximately normal.

The variance formula V_n also suggests that the correction may be important in practice when the model is weakly identified. Weak identification will mean that $\bar{g}_\beta$ close to zero relative to the variance of $\partial g(w_i, \beta_0)/\partial\beta$, which will lead to $\bar{g}_\beta'\Omega^{-1}\bar{g}_\beta$ being much smaller than Λ_n. In those cases, the additional term may be important even when m/n is small.

The linear model provides an example of the asymptotic variance. We have

$$\begin{aligned} A_j &= -E[z_i z_i' x_{ij}\varepsilon_i]\Omega^{-1} = -E[z_i z_i' \eta_{ij}\varepsilon_i]\Omega^{-1}, (j = 1, ..., p), \\ U_i^j &= (-z_i x_{ij} + E[z_i x_{ij}] - A_j z_i \varepsilon_i) \\ &= -(z_i z_i' - I)\pi_{mnj} + u_i^j, u_i^j = -z_i\eta_{ij} - A_j z_i \varepsilon_i. \end{aligned}$$

Then for $u_i = [u_i^1, ..., u_i^p]$ we have ,

$$\Lambda = E[u_i'\Omega^{-1}u_i]/m + E[\pi_{mnj}'(z_i z_i' - I)\Omega^{-1}(z_i z_i' - I)\pi_{mnj}]/m. .$$

Under the conditions below, the second term will go to zero, so that

$$\Lambda^* = \lim_{m \longrightarrow \infty} E[u_i'\Omega^{-1}u_i]/m.$$

For instance, in the homoskedastic case where $E[\varepsilon^2|z] = \sigma_\varepsilon^2$, $E[\eta\eta'|z] = \Sigma_\eta$, $E[\varepsilon\eta|z] = \sigma_{\eta\varepsilon}$, we have $u_i = -z_i(\eta_i' - \sigma_{\eta\varepsilon}'\varepsilon_i/\sigma_\varepsilon^2)$, so that

$$\begin{aligned} E[u_i'\Omega^{-1}u_i]/m &= E[(\eta_i - \sigma_{\eta\varepsilon}\varepsilon_i/\sigma_\varepsilon^2)(\eta_i - \sigma_{\eta\varepsilon}\varepsilon_i/\sigma_\varepsilon^2)' z_i'\Omega^{-1}z_i]/m \\ &= (\Sigma_\eta - \sigma_{\eta\varepsilon}\sigma_{\eta\varepsilon}'/\sigma_\varepsilon^2)E[z_i'(\sigma_\varepsilon^2 I)^{-1}z_i]/m \\ &= (\Sigma_\eta - \sigma_{\eta\varepsilon}\sigma_{\eta\varepsilon}'/\sigma_\varepsilon^2)/\sigma_\varepsilon^2 = \Lambda^*. \end{aligned}$$

Then, assuming $\pi_{mn}'\pi_{mn}n/m \longrightarrow B$ for a nonsingular matrix B, the asymptotic variance matrix for $\sqrt{m}(\hat{\beta} - \beta_0)$ will be

$$V = \sigma_\varepsilon^2 B^{-1} + \sigma_\varepsilon^2 B^{-1}(\Sigma_\eta - \sigma_{\eta\varepsilon}\sigma_{\eta\varepsilon}'/\sigma_\varepsilon^2)B^{-1}.$$

This variance for the CUE is identical to the asymptotic variance of LIML under many weak instrument asymptotics of Stock and Yogo (2003) and Hansen, Hausman, and Newey (2004). In this sense, CUE is an extension of LIML to the heteroskedastic case.

For asymptotic normality in the general i.i.d. case we make the following assumption:

Assumption 6: $g(z,\beta)$ is twice continuously differentiable in a neighborhood N of β_0, $E[\|g_i(\beta_0)\|^4](m/n + 1/m\sqrt{n}) \longrightarrow 0$, $E[\|\partial g_i(\beta_0)/\partial\beta\|^4](m/n + 1/m\sqrt{n}) \longrightarrow 0$, and for all $\beta \in N$ we have $\lambda_{\max}(E[g_i(\beta)g_i(\beta)']) \leq C$, $\lambda_{\max}(E[\partial g_i(\beta)/\partial\beta_j\{\partial g_i(\beta)/\partial\beta_j\}']) \leq C$, $\lambda_{\max}(E[\partial^2 g_i(\beta)/\partial\beta_j\partial\beta_k\{\partial^2 g_i(\beta)/\partial\beta_j\partial\beta_k\}']) \leq C$ for a constant C.

This condition imposes a stronger restriction on the growth rate of the number of moment conditions than was imposed for consistency. If $g_{ij}(\beta_0)$ were uniformly bounded a sufficient condition would be that $m^3/n \longrightarrow 0$.

Assumption 7: For all β on a neighborhood N of β_0 i) each $\sqrt{n/m}\sup_{\beta\in N}\|\hat{g}(\beta)\|$, $\sqrt{n/m}\sup_{\beta\in N}\|\partial\hat{g}(\beta)/\partial\beta_j\|$, and $\sqrt{n/m}\sup_{\beta\in N}\|\partial^2\hat{g}(\beta)/\partial\beta_j\partial\beta_k\|$ are bounded in probability; ii) each of $E[\|g_i(\beta)\|^4]/n$, $E[\|\partial g_i(\beta)/\partial\beta_j\|^4]/n$, $E[\|\partial^2 g_i(\beta)/\partial\beta_j\partial\beta_k\|^4]/n$ converge to zero; iii) $\sup_{\beta\in N}\|\hat{\Omega}(\beta) - \Omega(\beta)\| \xrightarrow{p} 0$, $\sup_{\beta\in N}\|\partial\hat{\Omega}(\beta)/\partial\beta_j - \partial\Omega(\beta)/\partial\beta_j\| \xrightarrow{p} 0$, $\sup_{\beta\in N}\|\partial^2\hat{\Omega}(\beta)/\partial\beta_j\partial\beta_k - \partial^2\Omega(\beta)/\partial\beta_j\partial\beta_k\| \xrightarrow{p} 0$.

Let $\tilde{D}_j(\beta) = \sqrt{n/m}\left[\partial\hat{g}(\beta)/\partial\beta_j - A_j(\beta)\hat{g}(\beta)\right]$, where $A_j(\beta) = \partial\Omega(\delta,\beta)/\partial\delta_j|_{\delta=\beta}\Omega(\beta)^{-1}$, and $\tilde{D}(\beta) = [\tilde{D}_1(\beta), ..., \tilde{D}_p(\beta)]$.

Assumption 8: $\partial^2\tilde{Q}(\beta)/\partial\beta\partial\beta'$ is stochastically equicontinuous and $\tilde{D}(\beta)'\Omega^{-1}\tilde{D}(\beta)$ is stochastically equicontinuous.

Under these and other regularity conditions we can show that $\hat{\beta}$ is asymptotically normal and that the variance.

THEOREM 3: *If Assumptions 1-3, and 6-8 are satisfied and* $\Lambda_n \longrightarrow \Lambda^*$ *then*

$$\sqrt{m}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0,V), \hat{V} \xrightarrow{p} V.$$

This result specializes to the linear model under previous conditions and a slight strengthening of the fourth moment condition for the instruments.

THEOREM 4: *If Assumptions 1 and 4 are satisfied and* $E[(z_i'z_i)^2](m/n+1/m\sqrt{n}) \longrightarrow 0$ *then*

$$\sqrt{m}(\hat{\beta}-\beta_0) \xrightarrow{d} N(0,V), \hat{V} \xrightarrow{p} V$$

This limiting distribution can also be derived by a sequential asymptotics calculation based on Stock and Wright (2002). If one takes their limiting distribution of the CUE under weak identification and lets the number of moment restrictions and the degree of identification grow at the same rate then one obtains the same limiting distribution as in Theorem 3. This asymptotic distribution is sequential in the sense that we first let the number of observations go to infinity to obtain the Stock and Wright (2002) limit and then take the limit as the number of moment restrictions grows.

The Kleibergen (2004) test that is asymptotically correct under weak instruments is also asymptotically correct with many weak moment conditions. As a test of the null hypothesis that $\beta_0 = \bar{\beta}$ where $\bar{\beta}$ is known, this statistic is given by

$$\hat{T}(\bar{\beta}) = m\hat{g}(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1}\hat{D}(\bar{\beta})[\hat{D}(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1}\hat{D}(\bar{\beta})]^{-1}\hat{D}(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1}\hat{g}(\bar{\beta}).$$

The following results shows that this Kleibergen statistic (2004) has the usual chi-squared distribution:

THEOREM 5: *If Assumptions 1-3, 6-8 are satisfied,* $E[\tilde{U}'\Omega^{-1}\tilde{U}]/m \longrightarrow \Lambda^*$, *and* $\beta_0 = \bar{\beta}$ *then*

$$\hat{T}(\bar{\beta}) \xrightarrow{d} \chi^2(p).$$

Because of this one can form joint confidence intervals for the vector β_0 by inverting the Kleibergen (2004). However, since we have asymptotic normality and a consistent estimator of the asymptotic variance, it is simpler to just proceed with Wald type inference in the usual way.

5 Appendix: Proofs of Theorems.

Throughout the Appendix, let C denote a generic positive constant that may be different in different uses. Let CS, M, and T denote the Cauchy-Schwartz, Markov, and triangle inequalities respectively. Also, let CM denote the conditional Markov inequality that if $E[|A_n||B_n] = O_p(\varepsilon_n)$ then $A_n = O_p(\varepsilon_n)$ and let w.p.a.1 stand for "with probability approaching one."

For the next two results let $Y_i, Z_i, (i = 1, ..., n)$ be i.i.d. $m \times 1$ random vectors with 4th moments, that can depend on n, but where we suppress an n subscript for notational convenience. Also, let

$$\begin{aligned} \bar{Y} &= \sum_{i=1}^{n} Y_i/n, \bar{Z} = \sum_{i=1}^{n} Z_i/n, \mu_y = E[Y_i], \mu_z = E[Z_i], \\ \Sigma_{yy} &= \text{var}(Y_i), \Sigma_{zz} = \text{var}(Z_i), \Sigma_{yz} = E[Y_i Z_i'] - \mu_y \mu_z', \end{aligned}$$

and A be a symmetric matrix.

Lemma A1: *If* $\lambda_{\max}(A^2) \leq C$, $\lambda_{\max}(\Sigma_{yy}) \leq C$, $\lambda_{\max}(\Sigma_{zz}) \leq C$, $E[Y_i'Y_iZ_i'Z_i]/nm^2 \longrightarrow 0$, $n\mu_y'A\mu_z/m \leq C$, *then*

$$n\bar{Y}'A\bar{Z}/m = \text{tr}\left(A\Sigma_{yz}'\right)/m + n\mu_y'A\mu_z/m + o_p(1).$$

Proof: By the eigenvalue conditions and $n\mu_y'A\mu_z/m \leq C$ we have

$$\begin{aligned} \left|\text{tr}\left(A\Sigma_{yz}'\right)/m\right| &\leq C, \text{tr}\left(\left(A\Sigma_{yz}'\right)^2\right)/m \leq C, \text{tr}\left(A\Sigma_{yy}A\Sigma_{zz}\right)/m \leq C, \\ |n\mu_y'A\Sigma_{zz}A\mu_y/m| &\leq C, |n\mu_z'A\Sigma_{yy}A\mu_z/m| \leq C. \end{aligned}$$

We also have

$$E\left[(Y_i - \mu_y)'A(Z_i - \mu_z)^2\right]/nm^2 \leq CE[Y_i'Y_iZ_i'Z_i]/nm^2 \longrightarrow 0.$$

For the moment suppose $\mu_y = \mu_z = 0$. Let $W_n = n\bar{Y}'A\bar{Z}/m$. Then $E[W_n] = \text{tr}\left(A\Sigma_{yz}'\right)/m$ and

$$\begin{aligned} E[W_n^2] &= E\left[\sum_{i,j,k,\ell} Y_i'AZ_jY_k'AZ_\ell/n^2m^2\right] = E\left[(Y_i'AZ_i)^2\right]/nm^2 + (1 - 1/n)\{E[W_n]^2 \\ &\quad + \text{tr}(A\Sigma_{zz}A\Sigma_{yy})/m^2 + \text{tr}\left(\left(A\Sigma_{yz}'\right)^2\right)/m^2\} = E[W_n]^2 + o(1), \end{aligned}$$

so that by M,

$$W_n = \text{tr}\left(A\Sigma_{yz}'\right)/m + o_p(1). \qquad (5.3)$$

In general, when μ_y or μ_z are nonzero we have,

$$W_n = n\left(\bar{Y} - \mu_y\right)' A(\bar{Z} - \mu_z)/m + n\mu_y' A(\bar{Z} - \mu_z)/m + n(\bar{Y} - \mu_y)' A\mu_z/m + n\mu_y' A\mu_z/m.$$

Now $E[W_n] = tr(A\Sigma_{yz}')/m + n\mu_y' A\mu_z/m$. Note that

$$E\left[\left\{nu_y' A(\bar{Z} - \mu_z)/m\right\}^2\right] = \mu_y' A\Sigma_{zz} A\mu_y/m^2 \longrightarrow 0, E\left[\left\{nu_y' A(\bar{Z} - \mu_z)/m\right\}^2\right] \longrightarrow 0,$$

so by M,

$$W_n = n\left(\bar{Y} - \mu_y\right)' A(\bar{Z} - \mu_z)/m + n\mu_y' A\mu_z/m + o_p(1).$$

Applying eq. (5.3) to $n\left(\bar{Y} - \mu_y\right)' A(\bar{Z} - \mu_z)/m$ and using T gives the result. Q.E.D.

LEMMA A2: *For* $\Psi = E[Z_i Z_i']$, *if* $E[Z_i] = E[Y_i] = 0$, $E[Y_i Y_i'] = I_m$, $E[Y_i Z_i'] = 0$, $na_n' a_n \to H$, $n^2 \operatorname{tr}(\Psi) \to \Lambda^*$, $n^3 a_n' \Psi a_n \to 0$, $\operatorname{tr}(\Psi^2)/[\operatorname{tr}(\Psi)]^2 \to 0$, $nE\left[|a_n' Y_i|^4\right] \to 0$, $n^{-1}E\left[|Y_1' Z_2|^4\right]/\operatorname{tr}(\Psi)^2 \to 0$, *and* $nE[|Y_i' Z_i|^2] \longrightarrow 0$, *then*

$$\sum_{i=1}^n a_n' Y_i + \sum_{i,j=1}^n Z_i' Y_j \xrightarrow{d} N(0, H + \Lambda^*)$$

Proof: Let w denote all possible data for a single observation that includes all of the elements of Y and $H_n(w, \tilde{w}) = Z'\tilde{Y} + \tilde{Z}'Y$. Then we have

$$\sum_{i=1}^n a_n' Y_i + \sum_{i,j=1}^n Z_i' Y_j = \sum_{i=2}^n (A_{in} + B_{in}) + R_n$$

$$A_{in} = a_n' Y_i, \; B_{in} = \sum_{j<i} H_n(w_i, w_j), \; R_n = \sum_{i=1}^n Z_i' Y_i + a_n' Y_1.$$

We have, by $E[Z_i' Y_i] = \operatorname{tr} E[Y_i Z_i'] = 0$,

$$E[(\sum_{i=1}^n Z_i' Y_i)^2] = nE\left[(Z_i' Y_i)^2\right] \to 0.$$

Also, $E\left[(a_n' Y_1)^2\right] = a_n' a_n \to 0$. Therefore by M, we have $R_n \xrightarrow{p} 0$. Next, note that

$$\begin{aligned} E[A_{in}^2] &= a_n' E[Y_i Y_i'] a_n = a_n' a_n, \\ E[A_{in} B_{in}] &= E\left[(a_n' Y_i)\left(\sum_{j<i}\left\{Z_i' Y_j + Z_j' Y_i\right\}\right)\right] = 0, \\ E\left[B_{in}^2\right] &= E\left[\left(\sum_{j<i} Z_i' Y_j + Z_j' Y_i\right)^2\right] = E\left[\sum_{j,k<i}\left(Z_i' Y_j Y_k' Z_i + Z_j' Y_i Y_i' Z_k + 2Y_j' Z_i Y_i' Z_k\right)\right] \\ &= \sum_{j<i} E\left[Z_i' E\left[Y_j Y_j'\right] Z_i\right] + E\left[Z_j' E[Y_i Y_i']\, Z_j\right] = 2(i-1) \operatorname{tr} E[Z_i' Z_i] = 2(i-1) \operatorname{tr}(\Psi). \end{aligned}$$

Therefore

$$\begin{aligned} s_n &= \sum_{i=2}^{n} E[(A_{in}+B_{in})^2] = (n-1)a_n'a_n + 2\sum_{i=2}^{n}(i-1)\operatorname{tr}(\Psi) \\ &= \frac{n-1}{n} n a_n' a_n + \left(\frac{n^2-n}{n^2}\right) n^2 \operatorname{tr}(\Psi) \to H + \Lambda^*. \end{aligned}$$

Next, define

$$\begin{aligned} & G_n(w,\tilde{w}) \overset{\text{def}}{=} E\left[H_n(w_1,w)H_n(w_1,\tilde{w})\right] \\ = \; & E\left[\left(Z_1'Y + Z'Y_1\right)\left(Z_1'\tilde{Y} + \tilde{Z}'Y_1\right)\right] = Y'\Psi\tilde{Y} + Z'\tilde{Z} \end{aligned}$$

We have, by $E\left[H_n(w_1,w_2)^2\right] = 2\operatorname{tr}(\Psi)$,

$$\begin{aligned} & E\left[G_n(w_1,w_2)^2\right]/E\left[H_n(w_1,w_2)^2\right]^2 \\ = \; & \left(E\left[(Y_1'\Psi Y_2)^2\right] + E\left[(Z_1'Z_2)^2\right] + 2E\left[Y_1'\Psi Y_2 Z_2' Z_1\right]\right)/4\operatorname{tr}(\Psi)^2 \\ = \; & \left(E\left[Y_1'\Psi E\left[Y_2Y_2'\right]\Psi Y_1\right] + E\left[Z_1'E\left[Z_2Z_2'\right]Z_1\right] + 0\right)/4\operatorname{tr}(\Psi)^2 \\ = \; & \left(E\left[\operatorname{tr}(\Psi^2 Y_1Y_1')\right] + E\left[\operatorname{tr}(\Psi Z_1 Z_1')\right]\right)/4\operatorname{tr}(\Psi)^2 = 2\operatorname{tr}(\Psi^2)/4\operatorname{tr}(\Psi)^2 \to 0. \end{aligned}$$

It then follows as in the proof of Theorem 1 of Hall (1984) that

$$\sum_{i=1}^{n}\left(E\left[B_{in}^2 \mid w_{i-1},...,w_1\right] - E[B_{in}^2]\right) \xrightarrow{p} 0.$$

Note also that $E[A_{in}^2] = E[A_{in}^2 \mid w_{i-1},...,w_1]$ and that

$$E[A_{in}B_{in}|w_{i-1},...,w_1] = \sum_{j<i} E\left[(a_n'Y_i)\left(Z_i'Y_j + Z_j'Y_i\right) \mid w_{i-1},...,w_1\right] = a_n'\left(\sum_{j<i} Z_j\right).$$

Therefore

$$\begin{aligned} & E\left[\left(\sum_{i=2}^{n} E[A_{in}B_{in}|w_{i-1},...,w_1]\right)^2\right] = E\left[(a_n'\sum_{i=1}^{n-1}(n-i)Z_i)^2\right] \\ = \; & a_n'\Psi a_n \sum_{i=1}^{n-1}(n-i)^2 \le n^3 a_n'\Psi a_n \to 0. \end{aligned}$$

Then by M, we have

$$\sum_{i=2}^{n} E[A_{in}B_{in} \mid w_{i-1},...,w_1] \xrightarrow{p} 0.$$

By T it then follows that

$$\begin{aligned}
&\sum_{i=2}^{n} E\left[(A_{in}+B_{in})^2 \mid w_{i-1},...,w_1\right] - E\left[(A_{in}+B_{in})^2\right] \\
= \; &\sum_{i=2}^{n}\left(E\left[B_{in}^2 \mid w_{i-1},...,w_1\right] - E[B_{in}^2]\right) + 2\sum_{i=2}^{n} E[A_{in}B_{in} \mid w_{i-1},...,w_1] \xrightarrow{p} 0
\end{aligned}$$

Next, note that

$$n^{-1}E\left[H_n(w_1,w_2)^4\right]/E[H_n(w_1,w_2)^2]^2 \le Cn^{-1}E\left[|Z_1'Y_2|^4\right]/\operatorname{tr}(\Psi)^2 \to 0.$$

It then follows as in the proof of Theorem 1 of Hall (1984) that $\sum_{i=1}^{n} E[B_{in}^4] \to 0$. Therefore, by T,

$$\sum_{i=1}^{n} E\left[(A_{in}+B_{in})^4\right] \le CnE[|a_n'Y_i|^4] + C\sum_{i=1}^{n} E[B_{in}^4] \to 0,$$

so that, as in Hall (1984), for any $\varepsilon > 0$

$$\sum_{i=1}^{n} E\left[(A_{in}+B_{in})^2 1\left(|A_{in}+B_{in}| > \varepsilon s_n\right)\right] \to 0.$$

The conclusion then follows from Brown's (1971) Martingale central limit theorem, similarly to Hall (1984). Q.E.D.

Recall that $\tilde{Q}(\beta) = \hat{g}(\beta)'\Omega(\beta)^{-1}\hat{g}(\beta)n/2m$ and $\bar{Q}(\beta) = (1+S_n(\beta))/2$.

LEMMA A2A: *If Assumption 3 is satisfied then $\bar{\alpha} \xrightarrow{p} 1$ and there is a g-inverse M^- with $\left\|\hat{M} - \hat{M} - \hat{M}\right\|^2/K \xrightarrow{p} 0$.*

Proof: Note that $\bar{\alpha}$ is invariant to nonsingular transformation of Z. Let $z = z(E[z_t z_t'])^{-1/2}$, so now Assumption 3 is $E[(z_t'z_t)^2]/KT \to 0$. Let $\hat{M} = B\Lambda B'$ where B is an orthogonal matrix, Λ is a diagonal matrix of eigenvalues of $\hat{M}$, and let $M^- = B\Lambda^- B'$, where Λ is the diagonal g-inverse. Then $\hat{A} = \hat{M}^- M$ is symmetric and idem potent so that

$$\bar{\alpha} = \operatorname{tr}(P_z)/K = \operatorname{tr}(A)/K = \left\|\hat{M}^-\hat{M}\right\|^2/K.$$

By Newey (1997) we have

$$\left\|\hat{M} - I_K\right\|^2/K = O_p\left(E[Z_t'Z_t)^2]/TK\right) \xrightarrow{p} 0.$$

We also have, for $\hat{A} = \hat{M}^{-}\hat{M}$

$$\begin{aligned} \left\|\hat{A} - \hat{M}\right\|^2 /K &= \left\|\hat{A} - \hat{A}\hat{M}\right\|^2 /K \leq \left\|\hat{A}(I - \hat{M})\right\|^2 /K \\ &\leq \operatorname{tr}((I - \hat{M})\hat{A}^2(I - \hat{M}))/K \leq \left\|I_K - \hat{M}\right\|^2 /K \xrightarrow{p} 0, \end{aligned}$$

giving second conclusion. It then follows by T that $\left\|\hat{A} - I_K\right\|^2 /K \xrightarrow{p} 0$, so that by $1 = \|I_K\|^2 /K$,

$$|\bar{\alpha} - 1| = \left|\left\|\hat{A}\right\|^2 - \|I_K\|^2\right| /K \leq \left\|\hat{A} - I_K\right\|^2 /K + 2\|I_K\| \left\|\hat{A} - I_K\right\| /K \xrightarrow{p} 0.$$

LEMMA A3: *If Assumptions 2 and 3 are satisfied then* $\sup_{\beta \in B} |\tilde{Q}(\beta) - \bar{Q}(\beta)| \xrightarrow{p} 0$.

Proof: Since $\tilde{Q}(\beta)$ and $\bar{Q}(\beta)$ are stochastically equicontinuous by Assumption 3, it suffices by Newey (1991, Theorem 2.1) to show that $\tilde{Q}(\beta) \xrightarrow{p} \bar{Q}(\beta)$ for each β. Apply Lemma A1 with $Y_i = Z_i = g_i(\beta)$ and $A = \Omega(\beta)^{-1}/2$. Note that $A\Sigma'_{yz} = A\Sigma_{zz} = A\Sigma_{yy} = I_m/2$ and that the hypotheses of Lemma A1 are satisfied by Assumption 3. Then by the conclusion of Lemma A1

$$\tilde{Q}(\beta) = tr(I_m/2)/m + S_n(\beta)/2 + o_p(1) = \bar{Q}(\beta) + o_p(1).$$

Q.E.D.

Proof of Theorem 1: By T, Lemma A3, and $\bar{Q}(\beta)$ bounded on B uniformly in m, we have $\sup_{\beta \in B} |\tilde{Q}(\beta)| = O_p(1)$. Let $\hat{a}(\beta) = \Omega(\beta)^{-1}\hat{g}(\beta)$. By Assumption 2,

$$\|\hat{a}(\beta)\|^2 = \hat{g}(\beta)'\Omega(\beta)^{-\frac{1}{2}}\Omega(\beta)^{-1}\Omega(\beta)^{-\frac{1}{2}}\hat{g}(\beta) \leq C\tilde{Q}(\beta),$$

so that $\sup_{\beta \in B} \|\hat{a}(\beta)\| = O_p(1)$. We also have

$$\left|\lambda_{\min}(\hat{\Omega}(\beta)) - \lambda_{\min}(\Omega(\beta))\right| \leq \left\|\hat{\Omega}(\beta) - \Omega(\beta)\right\|,$$

so that w.p.a.1 (with probability approaching 1) $\lambda_{\min}(\hat{\Omega}(\beta)) \geq C$, and hence $\lambda_{\max}(\hat{\Omega}(\beta)^{-1}) \leq C$ uniformly in $\beta \in B$. Therefore,

$$\begin{aligned} \left|\hat{Q}(\beta) - \tilde{Q}(\beta)\right| &\leq \left|\hat{a}(\beta)'\left[\hat{\Omega}(\beta) - \Omega(\beta)\right]\hat{a}(\beta)\right| \\ &\quad + \left|\hat{a}(\beta)'\left[\hat{\Omega}(\beta) - \Omega(\beta)\right]\hat{\Omega}(\beta)^{-1}\left[\hat{\Omega}(\beta) - \Omega(\beta)\right]\hat{a}(\beta)\right| \\ &\leq \|\hat{a}(\beta)\|^2\left(\left\|\hat{\Omega}(\beta) - \Omega(\beta)\right\| + C\left\|\hat{\Omega}(\beta) - \Omega(\beta)\right\|^2\right) \end{aligned}$$

It then follows by Assumption 2 that $\sup_{\beta\in B}\left|\hat{Q}(\beta)-\tilde{Q}(\beta)\right| \xrightarrow{p} 0$. Then $\sup\limits_{\beta\in B}\left|\hat{Q}(\beta)-\bar{Q}(\beta)\right| \xrightarrow{p} 0$ by T. Therefore, for any $\zeta>0$, w.p.a.1,

$$\bar{Q}(\hat{\beta}) \le \hat{Q}(\hat{\beta})+\zeta \le \hat{Q}(\beta_0)+2\zeta \le \bar{Q}(\beta_0)+3\zeta$$

Since $\bar{Q}(\beta)=1+S_n(\beta)$, it follows that w.p.a.1,

$$\Delta\left(\left\|\hat{\beta}-\beta_0\right\|\right) \le S_n(\beta) \le S(\beta_0)+3\zeta = 3\zeta.$$

Since ζ is anything positive, it follows that $\Delta\left(\left\|\hat{\beta}-\beta_0\right\|\right) \xrightarrow{p} 0$, so the conclusion follows by Assumption 2. Q.E.D.

LEMMA A4: *If Assumption 5 is satisfied then* $E[(y_i-x_i'\beta)^2|z_i] \ge C$. *Also, for* $X_i=(y_i,x_i')'$, $E[\|X_i\|^4|z_i] \le C$.

Proof: Note that for $\delta=\beta_0-\beta$ we have $y_i-x_i'\beta=\varepsilon_i+\eta_i'\delta+z_i'\pi_{mn}\delta$, so that

$$E[(y_i-x_i'\beta)^2|z_i] \ge E[(\varepsilon_i+\eta_i'\delta)^2|z_i] = (1,\delta')\Sigma(z_i)(1,\delta')' \ge \lambda_{\min}(\Sigma(z_i))(1+\delta'\delta) \ge C,$$

giving the first conclusion. Also, $E[\|x_i\|^4] \le CE[\|\eta_i\|^4|z_i]+CE[\|z_i'\pi_{mn}\|^4|z_i] \le C$ and $E[y_i^4] \le CE[\|x_i\|^4\|\beta_0\|^4|z_i]+E[\varepsilon_i^4|z_i] \le C$, giving the second conclusion. Q.E.D.

LEMMA A5: *If Assumption 4 is satisfied then* $C^{-1}I_m \le \Omega(\beta) \le CI_m$.

Proof: By Lemma A4 $C^{-1} \le E[(y_i-x_i'\beta)^2|z_i] \le C$, so that the conclusion follows by $I_m=E[z_iz_i']$ and $\Omega(\beta)=E[z_iz_i'E[(y_i-x_i'\beta)^2|z_i]]$. Q.E.D.

LEMMA A6: *If Assumptions 1 and 4 are satisfied then* $C^{-1}I_p \le \pi_{mn}'\pi_{mn}n/m \le CI_p$.

Proof: By Lemma A5, $\pi_{mn}'\Omega^{-1}\pi_{mn}n/m \le C\pi_{mn}'\pi_{mn}n/m$ and $\pi_{mn}'\Omega^{-1}\pi_{mn}n/m \ge C\pi_{mn}'\pi_{mn}n/m$, so the conclusion follows by Assumption 1. Q.E.D.

LEMMA A7: *If Assumptions 1 and 4 are satisfied then there is* $\hat{M}=O_p(1)$ *with* $i)\sqrt{n/m}\,\|\partial\bar{g}(\beta)/\partial\beta\|=O(1)$, $ii)\sqrt{n/m}\,\|\partial\hat{g}(\beta)/\partial\beta-\partial\bar{g}(\beta)/\partial\beta\|=O_p(1)$, $iii)\sup_{\beta\in B}\sqrt{n/m}\,\|\bar{g}(\beta)\|=O(1)$, $iv)\sup_{\beta\in B}\sqrt{n/m}\|\hat{g}(\beta)\|=O_p(1)$, $v)\sqrt{n/m}\|\bar{g}(\tilde{\beta})-\bar{g}(\beta)\| \le C\|\tilde{\beta}-\beta\|$, $vi)\sqrt{n/m}\|\hat{g}(\tilde{\beta})-\hat{g}(\beta)\| \le \hat{M}\|\tilde{\beta}-\beta\|$.

Proof: Note first that that $\partial\bar{g}(\beta)/\partial\beta = -\pi_{mn}$, so i) follows by Lemma A6. Also,

$$\begin{aligned} E[\left\|\sum_{i=1}^{n} z_i\eta_i'/n\right\|^2] &= trE[z_iz_i'\eta_i'\eta_i]/n \le CtrE[z_iz_i']/n = Cm/n, \\ E[\left\|\sum_{i=1}^{n} z_iz_i'/n - I_m\right\|^2] &\le E[(z_i'z_i)^2]/n \longrightarrow 0. \end{aligned}$$

Therefore by M and T we have

$$\|\partial\hat{g}(\beta)/\partial\beta - \partial\bar{g}(\beta)/\partial\beta\| \le \left\|\sum_{i=1}^{n} z_i\eta_i'/n\right\| + \left\|\sum_{i=1}^{n} z_iz_i'/n - I_m\right\| \|\pi_{mn}\| = O_p(\sqrt{m/n}),$$

giving ii). Then by M and T $v)$ holds. FIX Then $i)$ follows by $\hat{g}(\tilde{\beta})-\hat{g}(\beta) = [\partial\hat{g}(\beta)/\partial\beta](\tilde{\beta}-\beta)$ (with $\hat{M} = \sup_{\beta\in B}\sqrt{n/m}\|\partial\hat{g}(\beta)/\partial\beta\|$), and $ii)$ similarly. By $ii)$, $\|\bar{g}(\beta)\| = \|\bar{g}(\beta) - \bar{g}(\beta_0)\| \le C\sqrt{m/n}\|\beta-\beta_0\| \le C\sqrt{m/n}$, giving iv). Also, similarly to the proof of $v)$, we have $\|\hat{g}(\beta_0)\| = O_p(\sqrt{m/n})$ so by i) $\|\hat{g}(\beta)\| \le \|\hat{g}(\beta) - \hat{g}(\beta_0)\| + O_p(\sqrt{m/n}) \le \hat{M}(\sqrt{m/n})\|\beta-\beta_0\| + O_p(\sqrt{m/n}) \le O_p(\sqrt{m/n})$. Q.E.D.

LEMMA A8: *If Assumption 5 is satisfied, then* $\sup_{\beta\in B}\|\hat{\Omega}(\beta)-\Omega(\beta)\| \xrightarrow{p} 0$, $\sup_{\beta\in B}\|\partial\hat{\Omega}(\delta,\beta)/\partial\delta_j - \partial\Omega(\delta,\beta)/\partial\delta_j\||_{\delta=\beta} \xrightarrow{p} 0$, *and* $\|\partial^2\hat{\Omega}(\delta,\beta)/\partial\delta_j\partial\beta_k - \partial^2\Omega(\delta,\beta)/\partial\delta_j\partial\beta_k\||_{\delta=\beta} \xrightarrow{p} 0$.

Proof: Let $X_i = (y_i, x_i')'$ and $\alpha = (1, -\beta)$, so that $y_i - x_i'\beta = X_i'\delta$. Note that

$$\begin{aligned} \hat{\Omega}(\beta,\tilde{\beta}) - \Omega(\beta,\tilde{\beta}) &= \sum_{j,k=1}^{p+1}(\hat{D}_{jk}+\hat{F}_{jk})\alpha_j\tilde{\alpha}_k, \hat{D}_{jk} = \sum_{i=1}^{n} z_iz_i'X_{ij}X_{ik}/n - E[z_iz_i'X_{ij}X_{ik}]. \\ \hat{F}_{jk} &= (\sum_{i=1}^{n} z_iX_{ij}/n)(\sum_{i=1}^{n} z_i'X_{ik}/n) - E[z_iX_{ij}]E[z_i'X_{ik}]. \end{aligned}$$

Then for $\hat{\Delta}_j = \sum_{i=1}^{n} z_iX_{ij}/n - E[z_iX_{ij}]$

$$\begin{aligned} E[\left\|\hat{D}_{jk}\right\|^2] &\le CE[(z_i'z_i)^2E[X_{ij}^2X_{ik}^2|z_i]]/n \le CE[(z_i'z_i)^2]/n \longrightarrow 0, \\ E[\left\|\hat{\Delta}_j\right\|^2] &\le E[z_i'z_iE[X_{ij}^2|z_i]]/n \le Cm/n, \\ \|E[z_iX_{ij}]\| &\le (E[z_i'z_iE[X_{ij}^2|z_i]])^{1/2} \le C\sqrt{m}. \end{aligned}$$

From the last two lines, M, and T it follows that

$$\begin{aligned} \left\|\hat{F}_{jk}\right\| &\le \left\|\hat{\Delta}_j\right\|\left\|\hat{\Delta}_k\right\| + \|E[z_iX_{ij}]\|\left\|\hat{\Delta}_k\right\| + \|E[z_iX_{ik}]\|\left\|\hat{\Delta}_j\right\| \\ &= O_p(m/n) + O(\sqrt{m})O_p(\sqrt{m/n}) = O_p(\sqrt{m^2/n}) \xrightarrow{p} 0. \end{aligned}$$

The conclusion then follows by B bounded and by the fact that $\hat{\Omega}(\beta,\tilde{\beta}) - \Omega(\beta,\tilde{\beta})$ is a quadratic function of β and $\tilde{\beta}$. Q.E.D.

LEMMA A9: *If Assumption 5 is satisfied, then*

$$\begin{aligned} |a'\Omega(\tilde{\beta})b - a'\Omega(\beta)b| &\le C\|a\|\|b\|\|\tilde{\beta}-\beta\|, \\ |a'\partial\Omega(\delta,\tilde{\beta})/\partial\delta_j|_{\delta=\tilde{\beta}}b - a'\partial\Omega(\delta,\beta)/\partial\delta_j|_{\delta=\beta}b| &\le C\|a\|\|b\|\|\tilde{\beta}-\beta\|. \end{aligned}$$

Proof: Let $\tilde{\Sigma}_i = E[X_iX_i'|z_i]$, which is bounded. Then by $\delta = (1,-\beta)$ bounded on B, $|\tilde{\delta}'\tilde{\Sigma}_i\tilde{\delta} - \delta'\tilde{\Sigma}_i\delta| \le C\left\|\tilde{\beta}-\beta\right\|$, so that

$$\begin{aligned} &|a'\Omega(\tilde{\beta})b - a'\Omega(\beta)b| = |E[(a'z_i)(b'z_i)E[(X_i'\tilde{\delta})^2 - (X_i'\delta)^2|z_i]]| \\ \le\ & E[|a'z_i|\,|b'z_i|\,|\tilde{\delta}'\tilde{\Sigma}_i\tilde{\delta} - \delta'\tilde{\Sigma}_i\delta|] \le CE[(a'z_i)^2]^{1/2}E[(b'z_i)^2]^{1/2}\|\tilde{\beta}-\beta\| \le C\|a\|\|b\|\|\tilde{\beta}-\beta\|. \end{aligned}$$

We also have

$$\begin{aligned} &|a'\partial\Omega(\delta,\tilde{\beta})/\partial\delta_j|_{\delta=\tilde{\beta}}b - a'\partial\Omega(\delta,\beta)/\partial\delta_j|_{\delta=\beta}b| \le E[|a'z_i|\,|b'z_i|\,E[\|x_{ij}x_i'\|\,|z_i]]\|\tilde{\beta}-\beta\| \\ \le\ & C\|a\|\|b\|\|\tilde{\beta}-\beta\|.\ Q.E.D. \end{aligned}$$

Proof of Theorem 2: By Lemma A5, $\lambda_{\min}(\Omega(\beta)) \ge C$. Also, by Lemma A4,

$$E[\{g_i(\beta)'g_i(\beta)\}^2]/n = E[(z_i'z_i)^2E[(y_i - x_i'\beta)^4|z_i]]/n \le CE[(z_i'z_i)^2]/n \longrightarrow 0.$$

Lemma A8 gives $\sup_{\beta\in B}\|\hat{\Omega}(\beta) - \Omega(\beta)\| \xrightarrow{p} 0$. Let $a(\beta,\tilde{\beta}) = \sqrt{n/m}\Omega(\beta)^{-1}\bar{g}(\tilde{\beta})$. By Lemma A7, $\sup_{\beta,\tilde{\beta}\in B}\left\|a(\beta,\tilde{\beta})\right\| \le C$. Then by Lemma A9,

$$\left|S_n(\tilde{\beta}) - (n/m)\bar{g}(\tilde{\beta})'\Omega(\beta)^{-1}\bar{g}(\tilde{\beta})/2\right| = \left|a(\tilde{\beta},\tilde{\beta})'\left[\Omega(\beta) - \Omega(\tilde{\beta})\right]a(\beta,\tilde{\beta})/2\right| \le C\left\|\tilde{\beta}-\beta\right\|.$$

Also, by T and Lemma A7,

$$\begin{aligned} \left|(n/m)\bar{g}(\tilde{\beta})'\Omega(\beta)^{-1}\bar{g}(\tilde{\beta})/2 - S_n(\beta)\right| &\le C(n/m)\left\|\bar{g}(\tilde{\beta}) - \bar{g}(\beta)\right\|^2 + C(n/m)\left\|\bar{g}(\beta)\right\|\left\|\bar{g}(\tilde{\beta}) - \bar{g}(\beta)\right\| \\ &\le C\left\|\tilde{\beta}-\beta\right\|. \end{aligned}$$

Then by T it follows that $\left|S_n(\tilde{\beta}) - S_n(\beta)\right| \le C\left\|\tilde{\beta}-\beta\right\|$, implying equicontinuity of $S_n(\beta)$, and hence $\bar{Q}(\beta)$. An analogous argument with $\hat{a}(\beta,\tilde{\beta}) = \Omega(\beta)^{-1}\hat{g}(\beta)$ replacing $a(\beta,\tilde{\beta})$ implies that $\left|\hat{Q}(\tilde{\beta}) - \hat{Q}(\beta)\right| \le \hat{M}\left\|\tilde{\beta}-\beta\right\|$, with $\hat{M} = O_p(1)$, giving stochastic equicontinuity of $\tilde{Q}(\beta)$. Thus, all the hypotheses of Assumption 3 are satisfied. Finally, by Lemmas A5 and A6,

$$\begin{aligned} S_n(\beta) &= (n/m)\bar{g}(\beta)'\Omega(\beta)^{-1}\bar{g}(\beta) = (n/m)(\beta-\beta_0)'\pi_{mn}'\Omega(\beta)^{-1}\pi_{mn}(\beta-\beta_0) \\ &\geq C(\beta-\beta_0)'[(n/m)\pi_{mn}'\pi_{mn}](\beta-\beta_0) \geq C(\beta-\beta_0)'(\beta-\beta_0). \end{aligned}$$

so that Assumption 2 is satisfied. Hence, all the hypotheses of Theorem 1 are satisfied, so the conclusion to Theorem 2 follows by Theorem 1. Q.E.D.

For the next results let $\hat{\Omega} = \hat{\Omega}(\beta_0)$, $\hat{B}_j = \partial\hat{\Omega}(\delta,\beta_0)/\partial\delta_j|_{\delta=\beta_0}$, $\hat{A}_j = \hat{B}_j\hat{\Omega}^{-1}$.

Lemma A10: *If Assumption 6 is satisfied then*

$$\sqrt{m}\|\hat{\Omega}-\Omega\| \xrightarrow{p} 0, \sqrt{m}\|\hat{A}_j - A_j\| \xrightarrow{p} 0.$$

Proof: By standard arguments and Assumption 6,

$$\begin{aligned} E[m\|\hat{\Omega}-\Omega\|^2] &\leq CmE[\|g_i(\beta_0)\|^4]/n \longrightarrow 0, \\ E[m\|\hat{B}_j - B\|^2] &\leq CmE[\|\partial g_i(\beta_0)/\partial\beta_j\|^2\|g_i(\beta_0)\|^2]/n \longrightarrow 0, \end{aligned}$$

so the first conclusion holds by M. Also, note that $\lambda_{\max}(A_jA_j') \leq C$ by $A_jA_j' \leq CB_j\Omega^{-1}B_j' \leq CE[\partial g_i(\beta_0)/\partial\beta_j\{\partial g_i(\beta_0)/\partial\beta_j\}']$. Then by $\lambda_{\max}(\hat{\Omega}^{-1}) \leq C$ w.p.a.1we have

$$\begin{aligned} \sqrt{m}\|\hat{A}_j - A_j\| &\leq \sqrt{m}\|\hat{\Omega}^{-1}(\hat{\Omega}-\Omega)A_j\| + \sqrt{m}\|\Omega^{-1}(\hat{B}_j - B_j)\| \\ &\leq C\sqrt{m}\|(\hat{\Omega}-\Omega)A_j\| + C\sqrt{m}\|\hat{B}_j - B_j\| \\ &\leq C\sqrt{m}\|\hat{\Omega}-\Omega\| + C\sqrt{m}\|\hat{B}_j - B_j\| \xrightarrow{p} 0. \ Q.E.D. \end{aligned}$$

For the next result $g_i = g(w_i,\beta_0)$, $U_i^j = \partial g_i(\beta_0)/\partial\beta - E[\partial g_i(\beta_0)/\partial\beta_j] - A_jg_i$, and $U_i = [U_i^1, ..., U_i^p]$.

Lemma A11: *If Assumption 5 is satisfied then*

$$\sqrt{m}\frac{\partial\hat{Q}}{\partial\beta}(\beta_0) \xrightarrow{d} N(0, H+\Lambda^*).$$

Proof: Let $\hat{\Omega} = \hat{\Omega}(\beta_0)$, $\hat{B}_j = \partial\hat{\Omega}(\delta,\beta_0)/\partial\delta_j|_{\delta=\beta_0}$, $\hat{A}_j = \hat{B}_j\hat{\Omega}^{-1}$, $\hat{g} = \hat{g}(\beta_0)$ Stacking over j

from eq. (4.1) and evaluation at β_0 gives,

$$\begin{aligned}\sqrt{m}\frac{\partial\hat{Q}}{\partial\beta}(\beta_0) &= G_n'\hat{\Omega}^{-1}\sqrt{n}\hat{g}+\hat{U}'\hat{\Omega}^{-1}\sqrt{n}\hat{g}/\sqrt{m}, \\ \hat{U} &= [\hat{U}^1,...,\hat{U}^p],\hat{U}^j=\sum_{i=1}^{n}\hat{U}_i^j/\sqrt{n} \\ \hat{U}_i^j &= \partial g_i(\beta_0)\partial\beta_j-E[\partial g_i(\beta_0)\partial\beta_j]-\hat{A}_j g_i(\beta_0).\end{aligned}$$

By Lemma A10 we have

$$\begin{aligned}&|(\hat{U}^{j\prime}\hat{\Omega}^{-1}-U^{j\prime}\Omega^{-1})\sqrt{n}\hat{g}/\sqrt{m}| \\ \leq\ & |\sqrt{n}\hat{g}'(\hat{A}_j'-A_j')\hat{\Omega}^{-1}\sqrt{n}\hat{g}/\sqrt{m}|+|\sqrt{n}\hat{g}'A_j'\Omega^{-1}(\hat{\Omega}-\Omega)\hat{\Omega}^{-1}\sqrt{n}\hat{g}/\sqrt{m}| \\ \leq\ & C(n\|\hat{g}\|^2/m)\sqrt{m}\|\hat{A}_j-A_j\|+C(n/m)\|\hat{g}'A_j'\Omega^{-1}\|\|\hat{g}\|\sqrt{m}\|\hat{\Omega}-\Omega\| \\ \leq\ & o_p(1)+C(n/m)\|\hat{g}\|^2\sqrt{m}\|\hat{\Omega}-\Omega\|\xrightarrow{p}0.\end{aligned}$$

Similarly we have $G_n'\hat{\Omega}^{-1}\sqrt{n}\hat{g}-G_n'\Omega^{-1}\sqrt{n}\hat{g}\xrightarrow{p}0$. Therefore, by T,

$$\sqrt{m}\frac{\partial\hat{Q}}{\partial\beta}(\beta_0)=G_n'\Omega^{-1}\sqrt{n}\hat{g}+\tilde{U}'\Omega^{-1}\sqrt{n}\hat{g}/\sqrt{m}+o_p(1). \tag{5.4}$$

Next, for any nonzero vector λ consider $\sqrt{m}\lambda'\partial\hat{Q}(\beta_0)/\partial\beta$ and apply Lemma A2 with $a_n=\lambda'G_n'\Omega^{-1/2}/\sqrt{n}, Y_{in}=\Omega^{-1/2}g_i$, $Z_{in}=\Omega^{-1/2}U_i\lambda/n\sqrt{m}$, and $\bar{H}=\lambda'H\lambda$. By Assumption 1 and the hypothesis of Theorem 3, for $\Psi=E[Z_{in}Z_{in}']$ we have

$$na_n'a_n=\lambda'G_n'\Omega^{-1}G_n\lambda\longrightarrow\lambda'H\lambda=\bar{H},n^2tr(\Psi)=E[Z_{in}'Z_{in}]=\lambda'E[U_i'\Omega^{-1}U_i]\lambda/m\longrightarrow\lambda'\Lambda^*\lambda.$$

Also, note that $A=E[U_i\lambda\lambda'U_i']\leq CE[U_iU_i']\leq C\sum_{j=1}^{p}E[\{\partial g_i(\beta_0)/\partial\beta_j\}\{\partial g_i(\beta_0)/\partial\beta_j\}']$ so that $\lambda_{\max}(A)\leq C$ by Assumption 6. Therefore

$$n^3a_n'\Psi a_n=\lambda'G_n'\Omega^{-1}E[U_i\lambda\lambda'U_i']\Omega^{-1}G_n\lambda/m\leq C\lambda'G_n'\Omega^{-1}G_n\lambda/m\longrightarrow 0.$$

Also, it follows similarly that $\Psi\leq C\lambda_{\max}(A)/n^2m$, so that by $tr(\Psi)\geq C/n^2$

$$tr(\Psi^2)/tr(\Psi)^2\leq Cn^4tr(I_m^2)/n^4m^2=Ctr(I_m)/m^2\longrightarrow 0.$$

In addiiton, by Assumption 6 and $\|G_n'\Omega^{-1}\| \leq C$ we have for $g_{\beta i} = \partial g_i(\beta_0)/\partial\beta$,

$$\begin{aligned} nE[|a_n'Y_{in}|^4] &\leq CE[\left\|G_n'\Omega^{-1}g_i\right\|^4]/n \leq CE[\|g_i\|^4]/n \longrightarrow 0, \\ n^{-1}E\left[|Y_1'Z_2|^4\right]/\operatorname{tr}(\Psi)^2 &\leq CE[\left\|g_1'\Omega^{-1}U_2\right\|^4]/nm^2 \leq CE[\|g_1\|^4]E[\|U_1\|^4]/nm^2 \\ &\leq C(E[\|g_1\|^4]/m\sqrt{n})(E[\|g_1\|^4] + E[\|g_{\beta 1}\|^4]/m\sqrt{n}) \longrightarrow 0, \\ nE[|Y_i'Z_i|^2] &\leq CE[\left\|g_1'\Omega^{-1}U_1\right\|^2]/nm \leq (E[\|g_1\|^4] + E[\|g_{\beta 1}\|^4])/mn\backslash \longrightarrow 0. \end{aligned}$$

The conclusion then follows by the conclusion of Lemma A2, eq. (5.4), T, and the Cramer-Wold device. Q.E.D.

LEMMA A12: *If Assumptions 6-8 hold then for any $\bar{\beta} \xrightarrow{p} \beta_0$, $\partial^2\hat{Q}(\bar{\beta})/\partial\beta\partial\beta' \xrightarrow{p} H$.*

Proof: For notational convenience, drop the β argument and let k and ℓ denote derivatives with respect to β_k and β_ℓ, e.g. $\partial\hat{Q}(\beta)/\partial\beta_k = \hat{Q}_k$ and $\partial^2\hat{Q}(\beta)/\partial\beta_k\partial\beta_\ell = \hat{Q}_{k,\ell}$. Then differentiating twice for $\tilde{g}(\beta) = \sqrt{n/m}\hat{g}(\beta)$ we have

$$\begin{aligned} \hat{Q}_k &= \tilde{g}_k'\hat{\Omega}^{-1}\tilde{g} - \frac{1}{2}\tilde{g}'\hat{\Omega}^{-1}\hat{\Omega}_k\hat{\Omega}^{-1}\tilde{g} \qquad (5.5) \\ \hat{Q}_{k,\ell} &= \tilde{g}_{k,\ell}'\hat{\Omega}^{-1}\tilde{g} + \tilde{g}_k'\hat{\Omega}^{-1}\tilde{g}_\ell - \tilde{g}_k'\hat{\Omega}^{-1}\hat{\Omega}_\ell\hat{\Omega}^{-1}\tilde{g} - \tilde{g}_\ell'\hat{\Omega}^{-1}\hat{\Omega}_k\hat{\Omega}^{-1}\tilde{g} \\ &\quad +\tilde{g}'\hat{\Omega}^{-1}\hat{\Omega}_\ell\hat{\Omega}^{-1}\hat{\Omega}_k\hat{\Omega}^{-1}\tilde{g} - \frac{1}{2}\tilde{g}'\hat{\Omega}^{-1}\hat{\Omega}_{k,\ell}\hat{\Omega}^{-1}\tilde{g}. \end{aligned}$$

Note also that for $\tilde{Q} = \frac{1}{2}\tilde{g}'\Omega^{-1}\tilde{g}$, $\tilde{Q}_{k,\ell} = \partial^2\tilde{Q}(\beta)/\partial\beta_k\partial\beta_\ell$ has the same formula as $\hat{Q}_{k,\ell}$ with $\Omega = \Omega(\beta)$ replacing $\hat{\Omega}$. By Assumption 6 ii) each of Ω^{-2}, Ω_k^2, and $\Omega_{k\ell}^2$ have largest eigenvalue bounded above by a constant. Then by Assumption 7 i) and iii) and Lemma A12, it follows that

$$\begin{aligned} \sup_{\beta\in N}\left|\tilde{g}_k'\hat{\Omega}^{-1}\hat{\Omega}_\ell\hat{\Omega}^{-1}\tilde{g} - \tilde{g}_k'\Omega^{-1}\Omega_\ell\Omega^{-1}\tilde{g}\right| &\leq \sup_{\beta\in N}\|\tilde{g}_k\| \sup_{\beta\in N}\left\|\hat{\Omega}^{-1}\hat{\Omega}_\ell\hat{\Omega}^{-1} - \Omega^{-1}\Omega_\ell\Omega^{-1}\right\| \sup_{\beta\in N}\|\tilde{g}\| \\ &= O_p(1)o_p(1)O_p(1) \xrightarrow{p} 0. \end{aligned}$$

Therefore, we can replace $\hat{\Omega}$ by Ω in the third for $\hat{Q}_{k,\ell}$ without affecting its probability limit. Applying a similar argument to each of the six terms in the above expression for $\hat{Q}_{k,\ell}$, it follow that for $\tilde{Q} = \frac{1}{2}\tilde{g}'\Omega^{-1}\tilde{g}$, by T,

$$\sup_{\beta\in N}\left|\hat{Q}_{k,\ell} - \tilde{Q}_{k,\ell}\right| \xrightarrow{p} 0.$$

By hypothesis, $\tilde{Q}_{\beta\beta}(\beta)$ is stochastically equicontinuous, so by $\bar{\beta} \xrightarrow{p} \beta_0$, the previous equation, and T,

$$\left\|\hat{Q}_{\beta\beta'}(\bar{\beta}) - \tilde{Q}_{\beta\beta'}(\beta_0)\right\| \leq \left\|\hat{Q}_{\beta\beta'}(\bar{\beta}) - \tilde{Q}_{\beta\beta'}(\bar{\beta})\right\| + \left\|\tilde{Q}_{\beta\beta'}(\bar{\beta}) - \tilde{Q}_{\beta\beta'}(\beta_0)\right\| \xrightarrow{p} 0. \tag{5.6}$$

It therefore suffices to show that $\tilde{Q}_{k,\ell} \xrightarrow{p} H_{k\ell}$, where we now evaluate at β_0, i.e. $\tilde{Q}_{k,\ell} = \partial^2\tilde{Q}(\beta_0)/\partial\beta_k\partial\beta_\ell$. Let $\Upsilon_k = E\left[g_{ki}g_i'\right]$, $\Upsilon_{k,\ell} = E\left[(g_{ki} - \bar{g}_k)(g_{\ell i} - \bar{g}_\ell)'\right]$, $\Upsilon_{k\ell} = E\left[(g_{k\ell i} - \bar{g}_{k\ell})\, g_i'\right]$. Note that

$$\Omega_k = \Upsilon_k + \Upsilon_k', \Omega_{k,\ell} = \Upsilon_{k\ell} + \Upsilon_{k,\ell} + \Upsilon_{k,\ell}' + \Upsilon_{k\ell}'.$$

By Assumption 7 and Lemma A1 we have

$$\begin{aligned}
\tilde{g}_{k,\ell}'\Omega^{-1}\tilde{g} &= \operatorname{tr}\left(\Omega^{-1}\Upsilon_{k\ell}'\right) + o_p(1), \tilde{g}_k'\Omega^{-1}\tilde{g}_\ell = \bar{g}_k'\Omega^{-1}\bar{g}_\ell + \operatorname{tr}\left(\Omega^{-1}\Upsilon_{k,\ell}'\right) + o_p(1),\\
\tilde{g}_\ell'\Omega^{-1}\Omega_k\Omega^{-1}\tilde{g} &= \operatorname{tr}\left(\Omega^{-1}\Omega_k\Omega^{-1}\Upsilon_\ell'\right) + o_p(1), \tilde{g}'\Omega^{-1}\Omega_{k,\ell}\Omega^{-1} = \operatorname{tr}\left(\Omega^{-1}\Omega_{k,\ell}\right) + o_p(1),\\
\tilde{g}'\Omega^{-1}\Omega_\ell\Omega^{-1}\Omega_k\Omega^{-1}\tilde{g} &= \operatorname{tr}\left(\Omega^{-1}\Omega_\ell\Omega^{-1}\Omega_k\right) + o_p(1).
\end{aligned}$$

For a symmetric matrix A we have $tr(AB) = tr(B'A') = tr(A'B') = tr(AB')$. Then by T we have

$$\begin{aligned}
\tilde{Q}_{k,\ell} &= tr(\Omega^{-1}\Upsilon_{k\ell}') + \bar{g}_k'\Omega^{-1}\bar{g}_\ell + tr(\Omega^{-1}\Upsilon_{k,\ell}') - tr(\Omega^{-1}(\Upsilon_\ell + \Upsilon_\ell')\Omega^{-1}\Upsilon_k')\\
&\quad -tr(\Omega^{-1}(\Upsilon_k + \Upsilon_k')\Omega^{-1}\Upsilon_\ell') + tr(\Omega^{-1}(\Upsilon_k + \Upsilon_k')\Omega^{-1}(\Upsilon_\ell + \Upsilon_\ell'))\\
&\quad -(1/2)tr(\Omega^{-1}(\Upsilon_{k\ell} + \Upsilon_{k,\ell} + \Upsilon_{k,\ell}' + \Upsilon_{k\ell}') + o_p(1)\\
&= \bar{g}_k'\Omega^{-1}\bar{g}_\ell - tr(\Omega^{-1}(\Upsilon_\ell + \Upsilon_\ell')\Omega^{-1}\Upsilon_k') + tr(\Omega^{-1}(\Upsilon_k + \Upsilon_k')\Omega^{-1}\Upsilon_\ell) + o_p(1)\\
&= \bar{g}_k'\Omega^{-1}\bar{g}_\ell - tr(\Omega^{-1}\Upsilon_\ell\Omega^{-1}\Upsilon_k') + tr(\Omega^{-1}\Upsilon_k'\Omega^{-1}\Upsilon_\ell)\\
&\quad -tr(\Omega^{-1}\Upsilon_\ell'\Omega^{-1}\Upsilon_k') + tr(\Omega^{-1}\Upsilon_k\Omega^{-1}\Upsilon_\ell) + o_p(1)\\
&= \bar{g}_k'\Omega^{-1}\bar{g}_\ell - tr(\Upsilon_k\Omega^{-1}\Upsilon_\ell\Omega^{-1}) + tr(\Omega^{-1}\Upsilon_k\Omega^{-1}\Upsilon_\ell) + o_p(1) = \bar{g}_k'\Omega^{-1}\bar{g}_\ell + o_p(1).
\end{aligned}$$

Thus $\partial^2\tilde{Q}(\beta_0)/\partial\beta\partial\beta' = G_n'\Omega^{-1}G_n + o_p1)$. The conclusion then follows by the triangle inequality and eq. (??). Q.E.D.

LEMMA A13: *If Assumptions 6-8 are satisfied,* $\hat{D}(\hat{\beta})'\hat{\Omega}^{-1}\hat{D}(\hat{\beta}) \xrightarrow{p} H + \Lambda^*$.

Proof: Let $\tilde{D}_j(\beta) = \sqrt{n/m}\,[\partial\hat{g}(\beta)/\partial\beta_j - A_j(\beta)\hat{g}(\beta)]$, where $A_j(\beta) = \partial\Omega(\delta,\beta)/\partial\delta_j|_{\delta=\beta}\Omega(\beta)^{-1}$, and $\tilde{D}(\beta) = [\tilde{D}_1(\beta), ..., \tilde{D}_p(\beta)]$. It follows similarly to the proof of Lemma A10 that $\sup_{\beta\in B}\left\|\hat{A}(\beta) - A(\beta)\right\| \xrightarrow{p} 0$, so that

$$\sup_{\beta\in N}\left\|\hat{D}(\beta) - \tilde{D}(\beta)\right\| \leq \sup_{\beta\in B}\left\|\hat{A}(\beta) - A(\beta)\right\| \sup_{\beta\in N}\left\|\sqrt{n/m}\hat{g}(\beta)\right\| \xrightarrow{p} 0.$$

We also have $\sup_{\beta\in N}\left\|\tilde{D}(\beta)\right\| = O_p(1)$ so that by T and CS,

$$\left\|\hat{D}(\hat{\beta})'\hat{\Omega}^{-1}\hat{D}(\hat{\beta}) - \tilde{D}(\hat{\beta})'\Omega^{-1}\tilde{D}(\hat{\beta})\right\| \xrightarrow{p} 0.$$

Also, by Assumption 8, $\tilde{D}(\hat{\beta})'\Omega^{-1}\tilde{D}(\hat{\beta}) - \tilde{D}(\beta_0)'\Omega^{-1}\tilde{D}(\beta_0) \xrightarrow{p} 0$. Now apply Lemma A1 to $\tilde{D}_j(\beta_0)'\Omega^{-1}\tilde{D}_k(\beta_0)$ with $A = \Omega^{-1}$, $Y_i = \partial g(w_i,\beta_0)/\partial\beta_j - A_j g_i$, and $Z_i = \partial g(w_i,\beta_0)/\partial\beta_k - A_k g_i$. Note that for the j^{th} unit vector e_j,

$$n\mu_y' A\mu_z/m = e_j' G_n' \Omega^{-1} G_n e_k, tr(A\Sigma_{yz}')/m = tr(\Omega^{-1}E[U_i^k U_i^{j\prime}])/m = e_j'\Lambda_n e_k.$$

Therefore, it follows from the conclusion of Lemma A1 that

$$\tilde{D}_j(\beta_0)'\Omega^{-1}\tilde{D}_k(\beta_0) = e_j' G_n' \Omega^{-1} G_n e_k + e_j'\Lambda_n e_k + o_p(1) = H_{jk} + o_p(1).$$

The conclusion now follows by T. Q.E.D.

Proof of Theorem 3: The result follows from Lemmas A11, A12, and A13 in the usual way. Q.E.D.

Proof of Theorem 4: We proceed by verifying all of the hypotheses of Theorem 3. First consider Assumption 6. Note that $g(w,\beta) = z(y - x'\beta)$ is twice continuously differentiable by inspection. Also, by Lemma A4 and the specified rate condition,

$$E[\|g_i\|^4] + E[\|\partial g(w_i,\beta_0)/\partial\beta\|^4](m/n + 1/m\sqrt{n}) \leq CE[(z_i'z_i)^2](m/n + 1/m\sqrt{n}) \longrightarrow 0.$$

Also by Lemma A4,

$$\begin{aligned} \lambda_{\max}(E[\partial g_i(\beta)/\partial\beta_j\{\partial g_i(\beta)/\partial\beta_j\}']) &= \lambda_{\max}(E[z_i z_i' x_{ij}^2]) \leq \lambda_{\max}(CI_m) \leq C, \\ \lambda_{\max}(E[g_i(\beta)g_i(\beta)']) &\leq \lambda_{\max}(CE[g_i g_i'] + CE[\partial g_i(\beta)/\partial\beta_j\{\partial g_i(\beta)/\partial\beta_j\}']) \\ &\leq \lambda_{\max}(CI_m) + C \leq C. \end{aligned}$$

It follows Assumption 6 is satisfied.

It follows by Lemma A7 that Assumption 7 i) is satisfied. Assumption 7 ii) holds by $E[(z_i'z_i)^2]/n \longrightarrow 0$. Assumption 7 iii) holds by Lemma A8.

The proof of Assumption 8 follows similarly to the proof of stochastic equicontinuity in the proof $\tilde{Q}(\beta)$ in the proof of Theorem 2. Q.E.D..

Proof of Theorem 5: It follows from Lemma A13, replacing $\hat{\beta}$ with β_0, that $\hat{D}(\beta_0)'\hat{\Omega}(\beta_0)^{-1}\hat{D}(\beta_0) \xrightarrow{p} H + \Lambda^*$. Also, Lemma A11 gives $\sqrt{m}\partial\hat{Q}(\beta_0)/\partial\beta \xrightarrow{d} N(0, H + \Lambda^*)$, so the conclusion follows in the usual way. Q.E.D.

References

ANGRIST, J. AND A. KRUEGER (1991): "Does Compulsory School Attendance Affect Schooling and Earnings", *Quarterly Journal of Economics, 106*, 979–1014.

BEKKER, P.A. (1994): "Alternative Approximations to the Distributions of Instrumental Variables Estimators," *Econometrica,* 63, 657-681.

BEKKER, P. AND F. KLEIBEGEN (2003): "Finite Sample Instrumental Variables Inference using an Asymptotically Pivotal Statistic," *Econometric Theory 19*, 744-753.

Brown Martingale Central Limit Theorem

CHAMBERLAIN, G. (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," working paper, Rutgers University.

CHAO, J.C. AND N.R. SWANSON (2002): "Estimation and Testing Using Jackknife IV in Heteroskedastic Regressions With Many Weak Instruments," working paper, Rutgers University.

DONALD, S.G. AND W.K. NEWEY (2000): "A Jackknife Interpretation of the Continuous Updating Estimator," *Economics Letters* 67, 239-244.

DONALD, S.G., G.W. IMBENS, AND W.K. NEWEY (2002): "Choosing the Number of Instruments for GMM and GEL Estimators," Journal of Econometrics.

DONALD, S. G. AND W. K. NEWEY (2003) "Choosing the Number of Moments in GMM and GEL Estimation," working paper.

HAHN, J. AND A. INOUE (2002): "A Monte Carlo Comparison of Various Asymptotic Approximations to the Distribution of Instrumental Variables Estimators," *Econometric Reviews* 21, 309-336.

HAHN, J. AND J. HAUSMAN (2002): "A New Specification Test for the Validity of Instrumental Variables", *Econometrica 70,* 163-189.

HAHN, J., J.A. HAUSMAN, AND G.M. KUERSTEINER (2004): "Estimation with Weak Instruments: Accuracy of higher-order bias and MSE approximations," *Econometrics Journal,*

HALL, P. (1984): "Central Limit Theorem for Integrated Squared Error of Multivariate Nonparametric Density Estimators," J*ournal of Multivariate Analysis* 14, 1-16.

HAN, C. AND P.C.B. PHILLIPS (2003): "GMM with Many Moment Conditions," working paper.

HANSEN, C., J.,J.A.HAUSMAN, AND W.K. NEWEY, (2004): "Many Instruments, Weak Instruments, and Econometric Practice," working paper, MIT.

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

HANSEN, L.P., J. HEATON AND A. YARON (1996): "Finite-Sample Properties of Some Alternative GMM Estimators", *Journal of Business and Economic Statistics* 14, 262-280.

IMBENS, G.W. (1997): "One-Step Estimators for Over-Identified Generalized Method of Moments Models," *Review of Economic Studies* 64, 359-383.

KLEIBEGEN, F. (2002): "Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression," *Econometrica* 70, 1781-1803.

MORIMUNE, K. (1983): "Approximate Distributions of k-Class Estimators When the Degree of Overidentifiability is Large Compared with the Sample Size," *Econometrica* 51, 821-841.

NEWEY, W.K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics* 79, 147-168.

NEWEY, W.K. AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in Engle, R. and D. McFadden, eds., *Handbook of Econometrics, Vol. 4,* New York: North Holland.

NEWEY, W.K., AND R.J. SMITH (2004): "Higher-order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica* 72, 219-255.

QIN, J. AND LAWLESS, J. (1994): "Empirical Likelihood and General Estimating Equations", *Annals of Statistics* 22, 300-325.

STAIGER, D. AND J. STOCK (1997): "Instrumental Variables Regression with Weak Instruments", *Econometrica, 65*, 557-586.

STOCK, J. AND J. WRIGHT (2000): "GMM With Weak Identification," *Econometrica* 68, 1055-1096.

STOCK, J. AND M. YOGO (2004): "Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments," prepared for Festschrift in honor of Tom Rothenberg.

WINDMEIJER, F. (2004): "A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators, *Journal of Econometrics*, forthcoming.