GMM WITH MANY WEAK MOMENT CONDITIONS

Whitney K. Newey
Frank Windmeijer

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP18/05

GMM with Many Weak Moment Conditions*

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

Frank Windmeijer
Department of Economics
University of Bristol

September 2004
Revised, November 2005

Abstract

Using many moment conditions can improve efficiency but makes the usual GMM inferences inaccurate. Two step GMM is biased. Generalized empirical likelihood (GEL) has smaller bias but the usual standard errors are too small. In this paper we use alternative asymptotics, based on many weak moment conditions, that addresses this problem. This asymptotics leads to improved approximations in overidentified models where the variance of the derivative of the moment conditions is large relative to the squared expected value of the moment conditions and identification is not too weak. We obtain an asymptotic variance for GEL that is larger than the usual one and give a "sandwich" estimator of it. In Monte Carlo examples we find that this variance estimator leads to a better Gaussian approximation to t-ratios in a range of cases. We also show that Kleibergen (2005) K statistic is valid under these asymptotics. We also compare these results with a jackknife GMM estimator, finding that GEL is asymptotically more efficient under many weak moments.

JEL Classification: C12, C13, C23
Keywords: GMM, Continuous Updating, Many Moments, Variance Adjustment

*The NSF provided financial support for this paper under Grant No. 0136869. Helpful comments were provided by Tom Rothenberg and Jim Powell. This paper has been presented at Cal Tech, NYU, Ohio State, Stanford, UC Berkeley, University College London, UCLA, and USC.

1 Introduction

Many applications of generalized method of moments (GMM, Hansen, 1982) have low precision. Examples include some natural experiments (Angrist and Krueger, 1991), consumption asset pricing models (Hansen and Singleton, 1982), and dynamic panel models (Holtz-Eakin, Newey and Rosen, 1988). In these settings the use of many moments can improve estimator accuracy. For example, Hansen, Hausman and Newey (2005) have recently found that in an application from Angrist and Krueger (1991), using 180 instruments, rather than 3, shrinks correct confidence intervals substantially.

A problem with using many moments is that the usual Gaussian asymptotic approximation can be poor. The two-step GMM estimator can be very biased. Generalized empirical likelihood (GEL, Smith 1997) and other estimators have smaller bias but the usual standard errors are found to be too small in examples in Han and Phillips (2005) and here. In this paper we use alternative asymptotics that addresses this problem in overidentified models where the variance of the derivative of the moment conditions is large relative to the squared expected value of the moment conditions and identification is not too weak. Such environments seem quite common in econometric applications of instrumental variables (IV). Under the alternative asymptotics we find that GEL has a Gaussian limit distribution with asymptotic variance larger than the usual one. We give a consistent, "sandwich" estimator of the alternative asymptotic variance. We find in instrumental variable Monte Carlo examples that, in a range of cases where identification is not very weak, the new t-ratios have a better Gaussian approximation than the usual ones. We also show that the Kleibergen (2005) K statistic is valid under these asymptotics.

For comparison purposes we also consider a jackknife GMM estimator that generalizes jackknife IV estimators of Phillips and Hale (1977) and Angrist, Imbens and Krueger (1998). This estimator should also be less biased than the two-step GMM estimator. In the IV case Chao and Swanson (2004) derived its limiting distribution under the alternative asymptotics. Here we show that jackknife IV is asymptotically less efficient than GEL.

The alternative asymptotics is based on many weak moment sequences like those of Chao and Swanson (2004, 2005), Stock and Yogo (2004), and Han and Phillips (2005). This paper picks up where Han and Phillips (2005) leave off, by showing asymptotic normality when the convergence rate of the estimator is the square root of the number of moment conditions, deriving an explicit formula for the asymptotic variance that is larger than the usual one, and giving a consistent variance estimator. This paper also extends Han and Phillips (2005) by giving primitive conditions for consistency and a limiting distribution when a general weight matrix is used for the continuous updating estimator (CUE), by analyzing GEL estimators other than the CUE, and by treatment of jackknife GMM.

The standard errors we give can be thought of as an extension of the Bekker (1994) standard errors from homoskedasticity and the limited information maximum likelihood (LIML) estimator to heteroskedasticity and GEL. Under many weak moments, in a homoskedastic linear model, Hansen, Hausman and Newey (2005) show that the Bekker (1994) standard errors are consistent for LIML. In the same model we show that GEL has the same asymptotic variance as LIML, so that the standard errors here have the same limit as those of Bekker (1994). However, the standard errors we give are also consistent for GEL with heteroskedasticity.

In the asymptotics here the variability of the derivative of the moments affects the limiting distribution but the variability of the weight matrix has no effect. The difference between the usual asymptotic variance and the one given here corresponds to a GEL higher-order variance term from Donald and Newey (2003), that depends on sample variability of the Jacobian of the moment functions. In Donald and Newey (2003) there are also higher-order variance terms corresponding to variability of the weight matrix, but these are relatively small when the Jacobian variance is large relative to squared average Jacobian, as happens under many weak moment asymptotics. Thus, the asymptotic variance we give will approximately be the higher order variance of GEL when the Jacobian variance is relatively large. This kind of approximation seems appropriate for many IV settings, where the sample variability of the Jacobian can be relatively large. It would not lead to improvements in minimum distance settings where the Jacobian does not depend on data. In that case the asymptotic variance here will equal the usual one.

The limiting distribution for GEL can be derived by increasing the number of moments in the Stock and Wright (2002) limiting distribution of the continuous updating estimator (CUE). This derivation corresponds to sequential asymptotics, where one lets the number of observations go to infinity and then lets the number of moments grow. We give here simultaneous asymptotics, where the number of moments grows along with, but slower than, the sample size.

One might also consider asymptotics where the number of moments increases at the same rate as the sample size, as did Bekker (1994). Theory for this case would be difficult because the dimension of the weighting matrix would grow at the same rate as the sample size.

The variance adjustment that comes out of the many weak instrument asymptotics is different from that of Windmeijer (2005). He adjusts for the variability of the weight matrix while the many instrument asymptotics adjusts for the variability of the moment derivative.

In Section 2 we describe the model, the estimators, the new asymptotic variance estimator, and the alternative asymptotics we consider. Section 3 gives the consistency results and Section 4 gives the asymptotic normality results. There we give regularity conditions for the CUE and reserve to Appendix B the regularity conditions for GEL. Section 5 reports some Monte Carlo results. Section 6 offers some conclusions and some possible directions for future work. Appendix A gives proofs of Theorems in Sections 3 and 4.

2 The Model and Estimators

The model we consider is for i.i.d. data where there is a countable number of moment restrictions. In the asymptotics we allow the data generating process to depend on the sample size. To describe the model, let w_i, $(i = 1, ..., n)$, be i.i.d. 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 GMM estimators let $g_i(\beta) = g(w_i, \beta)$, $\hat{g}(\beta) = \sum_{i=1}^n g_i(\beta)$, and $\hat{\Omega}(\beta) = \sum_{i=1}^n g_i(\beta) g_i(\beta)'$. Also let $\bar{\beta}$ be a preliminary estimator and B be a compact set of parameter values. The usual two-step GMM estimator is given by

$$\tilde{\beta} = \arg\min_{\beta \in B} \tilde{Q}(\beta), \tilde{Q}(\beta) = \hat{g}(\beta)' \hat{W} \hat{g}(\beta)/2, \hat{W} = \hat{\Omega}(\bar{\beta})^{-1}.$$

where $\bar{\beta}$ is some preliminary estimator. The weighting matrix $\hat{W} = \hat{\Omega}(\bar{\beta})^{-1}$ is optimal in minimizing the asymptotic variance of $\tilde{\beta}$ under standard asymptotics.

The jackknife GMM estimator is obtained as

$$\breve{\beta} = \arg\min_{\beta \in B} \breve{Q}(\beta), \breve{Q}(\beta) = \tilde{Q}(\beta) - tr(\hat{W} \hat{\Omega}(\beta))/2 = \sum_{i \neq j} [g_i(\beta)' \hat{W} g_j(\beta)]/2.$$

This estimator equals the JIVE2 estimator of Angrist, Imbens, and Krueger (1998) in a linear model when $\hat{W}$ is the inverse of the second moment matrix of the instruments. The first-order conditions for this estimator are

$$0 = \frac{\partial \hat{g}(\beta)}{\partial \beta}' \hat{W} \hat{g}(\beta) - \sum_{i=1}^n \frac{\partial \hat{g}_i(\beta)}{\partial \beta}' \hat{W} g_i(\beta).$$

This can be interpreted as a bias corrected version of the two-step GMM first order condition. The first term $\partial \hat{g}(\beta)/\partial \beta' \hat{W} \hat{g}(\beta)$ is the derivative of the GMM objective function. When evaluated at β_0 this term is biased, in the sense of having nonzero expectation (for $\hat{W}$ fixed). The second term is an estimator of the expectation of the first term (for $\hat{W}$ fixed). Subtracting out the second term makes the expectation exactly zero (for fixed $\hat{W}$), i.e. makes the first order conditions unbiased at the true parameter.

To describe a GEL estimator let $\rho(v)$ be a function of a scalar v that is concave on an open interval $\mathcal{V}$ containing zero and let $\rho_j(0) = \partial^j \rho(0)/\partial v^j$. We normalize $\rho(v)$ so that $\rho_0(0) = 0$, $\rho_1(0) = 1$ and $\rho_2(0) = -1$. Let $\hat{\Lambda}_n(\beta) = \{\lambda : \lambda' g_i(\beta) \in \mathcal{V}, i = 1, ..., n\}$. A GEL estimator is given by

$$\hat{\beta} = \arg\min_{\beta \in B} \hat{Q}(\beta),\ \hat{Q}(\beta) = \sup_{\lambda \in \hat{\Lambda}_n(\beta)} \sum_{i=1}^{n} \rho(\lambda' g_i(\beta)),$$

as in Smith (1997). The empirical likelihood (EL; Qin and Lawless, 1994) estimator is obtained when $\rho(v) = \ln(1 - v)$ (and $\mathcal{V} = (-\infty, 1)$), and exponential tilting (ET, Kitamura and Stutzer, 1997) when $\rho(v) = -e^v$. When $\rho(v)$ is quadratic, $\hat{Q}(\beta)$ has an explicit form, given by

$$\hat{Q}(\beta) = \hat{g}(\beta)' \hat{\Omega}(\beta)^{-1} \hat{g}(\beta)/2.$$

Newey and Smith, 2004). Here the GEL estimator $\hat{\beta}$, that minimizes $\hat{Q}(\beta)$, is the continuous updating estimator (CUE, Hansen, Heaton and Yaron, 1996).

The estimator of the asymptotic variance makes use of weights associated with the GEL estimator. Let

$$\begin{aligned} \hat{\rho}_{1i}(\beta) &= \rho_1(\hat{\lambda}(\beta)' g_i(\beta)), (i = 1, ..., n), \hat{\lambda}(\beta) = \arg\max_{\lambda} \sum_{i=1}^{n} \rho(\lambda' g_i(\beta)), \\ \hat{H} &= \partial \hat{Q}(\hat{\beta})/\partial\beta\partial\beta', \hat{D}(\beta) = \sum_{i=1}^{n} \hat{\rho}_{1i}(\beta) \partial g_i(\beta)/\partial\beta. \end{aligned}$$

Here $\hat{D}(\hat{\beta})/\sum_{i=1}^{n} \hat{\rho}_{1i}(\hat{\beta})$ is an efficient estimator of $G = E[\partial g_i(\beta_0)/\partial\beta]$, like that considered by Brown and Newey (1998). Let $\hat{D} = \hat{D}(\hat{\beta})$ and $\hat{\Omega} = \hat{\Omega}(\hat{\beta})$. The estimator of the asymptotic variance is given by

$$\hat{V} = \hat{H}^{-1} \hat{D}' \hat{\Omega}^{-1} \hat{D} \hat{H}^{-1}.$$

The "sandwich" form of the asymptotic variance estimator is important under the alternative asymptotics. Unlike the usual asymptotics, the middle matrix $\hat{D}'\hat{\Omega}^{-1}\hat{D}$ 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 $\hat{G}'\hat{\Omega}^{-1}\hat{G}$, where $\hat{G} = \partial \hat{g}(\hat{\beta})/\partial\beta$, because $\hat{G}'\hat{\Omega}^{-1}\hat{G}$ has extra random terms that are eliminated in the Hessian in the alternative asymptotics.

The Hessian term on the outside of $\hat{V}$ is familiar from other estimation environments. The middle term $\hat{D}'\hat{\Omega}^{-1}\hat{D}$ is an estimator of the asymptotic variance of $\partial\hat{Q}(\beta_0)/\partial\beta$ due to Kleibergen (2005) for the CUE and Guggenberger and Smith (2005) for other GEL settings. They show that this estimator can be used to construct a test statistic under weak identification with fixed m. We give conditions for consistency when m is allowed to grow with the sample size.

The Kleibergen (2005) K statistic will also be valid under many weak moment conditions. For the null hypothesis $H_0 : \beta_0 = \bar{\beta}$, where $\bar{\beta}$ is known, the K statistic is

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

Under the null hypothesis and the alternative asymptotics this statistic will have a $\chi^2(p)$ under the alternative asymptotics. As a result we can form joint confidence intervals for the vector β_0 by inverting $\hat{T}(\beta)$. Specifically, for the $1-\alpha$ quantile q of a $\chi^2(p)$ distribution, an asymptotic $1-\alpha$ confidence interval is $\{\beta : \hat{T}(\beta) \leq q\}$. These confidence intervals are also correct in the weak identification setting of Stock and Wright (2000). In general though, these intervals are much more difficult to compute than Wald confidence intervals.

The alternative variance estimator and associated asymptotics should provide a better approximation than the usual one when, for $G = E[\partial g_i(\beta_0)/\partial\beta]$ and $\Omega = E[g_i(\beta_0)g_i(\beta_0)']$,

$$\begin{aligned} 1 &: m > p, \\ 2 &: Var(\Omega^{-1/2}\partial g_i(\beta_0)/\partial\beta) >> G'\Omega^{-1}G, \\ 3 &: nG'\Omega^{-1}G >> 0. \end{aligned}$$

That is, the approximation should be better in 1) overidentified models where 2) the variance of the Jacobian of the moment functions is large relative to its average and 3) the model is not too weakly identified. Condition 2) is often true in IV settings, tending to hold when reduced form R^2s are low. Condition 3) is also often true in IV settings, corresponding to a model not being "too weakly" identified (e.g. see the brief applications survey in Hansen, Hausman and Newey, 2005). Condition 2) would not to be satisfied in minimum distance settings, where $Var(\partial g_i(\beta_0)/\partial\beta) = 0$, and so we expect that $\hat{V}$ would not provide an improvement there.

Conditions 1), 2) and 3) are simultaneously imposed in the many weak moment condition asymptotics, where m grows, $G'\Omega^{-1}G$ goes to zero, and $nG'\Omega^{-1}G$ grows. For this asymptotics we will give conditions under which there is $\mu_n \longrightarrow \infty$ and a matrix V such that

$$\mu_n(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V), \mu_n^2 \hat{V} \xrightarrow{p} V,$$

Therefore, standard (Wald) confidence intervals and test statistics that treat $\hat{\beta}$ as if it were normally distributed with mean β_0 and variance $\hat{V}$ will be asymptotically correct. The convergence rate of the estimator will be $1/\mu_n$.

We impose conditions so that μ_n^2 might be considered a generalization of the concentration parameter, that plays such a central role in the asymptotic theory of instrumental variable estimators. Let

$$\bar{g}(\beta) = E[g_i(\beta)], \Omega(\beta) = E[g_i(\beta)g_i(\beta)'], \Omega = \Omega(\beta_0),$$

where we suppress m subscripts and/or superscripts for convenience. We require that μ_n^2 behave as follows:

Assumption 1: i) there is $\mu_n \longrightarrow \infty$ such $\mu_n^2/n \longrightarrow 0, m \leq \mu_n^2, m/\mu_n^2 \longrightarrow \kappa, 0 \leq \kappa \leq 1$; ii) $(n/\mu_n^2)G'\Omega^{-1}G \longrightarrow H$ and H is nonsingular; iii) For all β and m, $\Omega(\beta)$ is nonsingular and there is a continuous function $\Delta(a) > 0$ for all $a \neq 0$ and $(n/\mu_n^2)\bar{g}(\beta)'\Omega(\beta)^{-1}\bar{g}(\beta) \geq \Delta(\|\beta - \beta_0\|)$.

This assumption means that μ_n^2 characterizes the growth rate of $nG'\Omega^{-1}G$, similarly to the concentration parameter of the simultaneous equations literature. When $\kappa > 0$, so that the number of instruments grows as fast as the concentration parameter, the convergence rate will also be $1/\sqrt{m}$. This formulation is a GMM version of Chao and Swanson (2005) that is similar to Han and Phillips (2005).

A special case is the linear model, where

$$y_i = x_i'\beta_0 + \varepsilon_i, \ x_i = \Upsilon_i + \eta_i, \tag{2.1}$$

$$0 = E[\varepsilon_i | z_i, \Upsilon_i],\ 0 = E[\eta_i | z_i, \Upsilon_i].$$

Here z_i is an $m \times 1$ vector of instrumental variables, where we suppress the m argument for convenience, and we will impose the normalization $E[z_i z_i'] = I_m$. Also, Υ_i is a $p \times 1$ vector of reduced form values. The moment functions are

$$g(w_i, \beta) = z_i(y_i - x_i'\beta).$$

Here $G = -E[z_i x_i'] = -E[z_i \Upsilon_i']$ and $\Omega = E[z_i z_i' \varepsilon_i^2] = E[\sigma_i^2 z_i z_i']$, where $\sigma_i^2 = E[\varepsilon_i^2 | z_i, \Upsilon_i]$. Then Assumption 1 means that for $\Upsilon_i^* = \Upsilon_i/\sigma_i^2$ and $\Pi^* = \Omega^{-1} E[\sigma_i^2 z_i \Upsilon_i^{*\prime}]$

$$\begin{aligned}(n/\mu_n^2) G'\Omega^{-1} G = &\ (n/\mu_n^2) E[\sigma_i^2 \Upsilon_i^* z_i'] \Omega^{-1} E[\sigma_i^2 z_i \Upsilon_i^{*\prime}] \\ = &\ (n/\mu_n^2) \Pi^{*\prime} \Omega \Pi^* \longrightarrow H.\end{aligned}$$

Here Π^* can be thought of as the coefficients from a population weighted regression of optimal instruments Υ_i^* on the instrumental variables z_i, with weight σ_i^2. Assumption 1 specifies that μ_n^2 gives the growth rate of $n\Pi^{*\prime}\Omega\Pi^*$ that can be interpreted as a weighted sum of squares of reduced form predicted values.

One example has $\Upsilon_i = \pi_{mn}' z_i$, so that the reduced form is a linear combination of the instrumental variables. If ε_i is homoskedastic with $\sigma_i^2 = \sigma_\varepsilon^2$ constant, Assumption 1 follows from

$$\frac{n}{\mu_n^2} \pi_{mn}' \pi_{mn} / \sigma_\varepsilon^2 \longrightarrow H.$$

When $p = 1$ this equation would be satisfied when $\pi_{mn} = \sigma_\varepsilon H^{1/2} (\mu_n/\sqrt{nm}, ..., \mu_n/\sqrt{nm})'$. When μ_n^2 grows at the same rate as m, each reduced form coefficient follows the weak instrument assumption of Staiger and Stock (1997), but the number of instruments is growing, which makes the concentration parameter grow. This example is a special case of the many weak instrument asymptotics of Chao and Swanson (2005).

Another example is given by

$$\Upsilon_i = \mu_n f_0(Z_i)/\sqrt{n}, z_i' = p^m(Z_i)', p^m(Z) = (p_{1m}(Z), ..., p_{mm}(Z))',$$

where $f_0(Z)$ is an unknown function of fixed dimensional vector of exogenous variables Z and $p_{1m}(Z), ..., p_{mm}(Z)$ are approximating functions for f_0, such as power series or splines. Here Assumption 1 will be satisfied if $E[f_0(Z_i)f_0(Z_i)'/\sigma_i^2]$ is nonsingular and there is γ_m such that

$$\lim_{m \longrightarrow \infty} E[\sigma_i^2 \| f_0(Z_i)/\sigma_i^2 - \gamma_m p^m(Z_i)\|^2] = 0.$$

This example is like Newey (1990) where z_i are approximating functions for the optimal (asymptotic variance minimizing) instruments Υ_i^*, but with μ_n^2 growing more slowly than n.

3 Consistency

We first give a brief explanation of the consistency results. As usual, the crucial condition for consistency of an extremum estimator is that the limit of the objective function is minimized at the truth. Under many weak instruments the limit of the objective function will be the limit of its expectation with the weighting matrix $\hat{W}$ replaced by its limit W and the expectation divided by μ_n^2.

As in Han and Phillips (2005), for two step GMM,

$$E[\hat{g}(\beta)'W\hat{g}(\beta)/n\mu_n^2] = (n-1)\bar{g}(\beta)'W\bar{g}(\beta)/\mu_n^2 + tr(W\Omega(\beta))/\mu_n^2.$$

The term $(n-1)\bar{g}(\beta)'W\bar{g}(\beta)/\mu_n^2$ is a "signal" term that is minimized at β_0. The second term is a "noise" term that is not minimized at β_0, and is not dominated by the signal term when μ_n^2 grows at the same rate as m. Consequently, two-step GMM will not be consistent, when μ_n^2 grows at the same rate as m. The jackknife GMM estimator eliminates the noise term. We have

$$E[\sum_{i \neq j} g_i(\beta)'Wg_j(\beta)/n\mu_n^2] = (n-1)\bar{g}(\beta)'W\bar{g}(\beta)/\mu_n^2.$$

This function is minimized at the truth, leading to consistency of the jackknife GMM estimator.

The CUE estimator makes the noise term not depend on β. We have

$$\begin{aligned} E[\hat{g}(\beta)'\Omega(\beta)^{-1}\hat{g}(\beta)/n\mu_n^2] &= (n-1)\bar{g}(\beta)'\Omega(\beta)^{-1}\bar{g}(\beta)/\mu_n^2 + tr(\Omega(\beta)^{-1}\Omega(\beta))/\mu_n^2 \\ &= (n-1)\bar{g}(\beta)'\Omega(\beta)^{-1}\bar{g}(\beta)/\mu_n^2 + m/\mu_n^2. \end{aligned}$$

This function is minimized at the truth, leading to consistency of the CUE. Also, it turns out that under many weak moments the objective function of every GEL estimator behaves like that of the CUE, leading to their consistency as well. The reason for this is that that for all β the vector $\hat{\lambda}(\beta)$ converges to zero, and so the GEL objective function $\hat{Q}(\beta)$ is approximately quadratic, i.e. is approximately the CUE objective function.

Turning now to precise results, for a matrix F let $\|F\| = trace(F'F)^{1/2}$ denote its Euclidean norm and for symmetric F let $\lambda_{\min}(F)$ and $\lambda_{\max}(F)$ denote its smallest and largest eigenvalues, respectively. Also, define stochastic equicontinuity of a sequence of random functions $\{\hat{S}_n(\beta)\}_{n=1}$ to mean that for any $\Delta_n \longrightarrow 0$, $\sup_{\|\tilde{\beta}-\beta\|\leq\Delta_n} |\hat{S}_n(\tilde{\beta}) - \hat{S}_n(\beta)| \xrightarrow{p} 0$.

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

The condition that $\sup_{\beta\in B} \|\hat{\Omega}(\beta)/n - \Omega(\beta)\| \xrightarrow{p} 0$ puts restrictions on the rate at which m can grow with the sample size. If $E[g_{ij}(\beta)^4]$ 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 CUE.

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

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

Assumption 3: The linear model holds and there is a constant C with $E[\varepsilon_i^4|z_i, \Upsilon_i] \leq C$, $E[\|\eta_i\|^4|z_i, \Upsilon_i] \leq C$, $\lambda_{\min}(\Sigma_i) \geq 1/C$, $\|\Upsilon_i\| \leq C$, $E[(z_i'z_i)^2]/n \longrightarrow 0$, and $nE[(\Upsilon_i'\Upsilon_i)^2]/\mu_n^4 \leq C$.

The conditions put 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 for the CUE if $m^2/n \longrightarrow 0$, for in that case, $E[(z_i'z_i)^2]/n = O(m^2/n)$.

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

4 Asymptotic Normality

We first give an explanation of the asymptotic normality results. The usual Taylor expansion of the first-order condition $\partial\hat{Q}(\hat{\beta})/\partial\beta = 0$ gives

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

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 $\mu_n^{-1}\partial\hat{Q}(\beta_0)/\partial\beta$. This reasoning also holds for the jackknife GMM estimator.

To simplify notation we focus on the scalar β case. Also, as for consistency we can take the weighting matrix equal to its limit. Let $g_i = g_i(\beta_0)$, $G_i = \partial g_i(\beta_0)/\partial\beta$, $\hat{g} = \hat{g}(\beta_0)$, and $G_n = (\sqrt{n}/\mu_n)G$. Then differentiating the jackknife GMM objective function, with Ω^{-1}/n replacing $\hat{W}$ gives

$$\begin{aligned} \mu_n^{-1}\partial\breve{Q}(\beta_0)/\partial\beta &= \sum_{i\neq j} G_i'\Omega^{-1}g_j/n\mu_n = (1-n^{-1})G_n'\Omega^{-1}\hat{g}/n^{1/2} + \sum_{j<i}\psi_{ij}^J/n\mu_n, \\ \psi_{ij}^J &= (G_j - G)'\Omega^{-1}g_i + (G_i - G)'\Omega^{-1}g_j, \end{aligned}$$

where the second equality holds by adding and subtracting G to G_i. The $G_n'\Omega^{-1}\hat{g}/n^{1/2}$ term is the usual GMM one, having asymptotic variance H. The other term $\sum_{j<i}\psi_{ij}^J/n\mu_n$ is a martingale sum, as in Hall (1984). Specifically, it is a degenerate U-statistic that is asymptotically normal. Also,

$$\begin{aligned} E[(\psi_{ij}^J)^2]/2 &= E[\{(G_j - G)'\Omega^{-1}g_i\}^2] + E[\{(G_j - G)'\Omega^{-1}g_i\}\{(G_i - G)'\Omega^{-1}g_j\}] \\ &= E[(G_j - G)'\Omega^{-1}(G_j - G)] + E[G_j'\Omega^{-1}g_iG_i'\Omega^{-1}g_j] \\ &= E[G_j'\Omega^{-1}G_j] - G'\Omega^{-1}G + tr(\{\Omega^{-1}E[g_iG_i']\}^2) \end{aligned}$$

The asymptotic variance of $\sum_{j<i} \psi_{ij}^J / n\mu_n$ will be (using $G'\Omega^{-1}G/m \longrightarrow 0$),

$$\begin{aligned}\Lambda_J &= \lim_{m,n\longrightarrow\infty} E[(\sum_{j<i} \psi_{ij}^J/n\mu_n)^2] = \lim_{m,n\longrightarrow\infty}\{(m/\mu_n^2)[n(n-1)/2]E[(\psi_{ij}^J)^2]/mn^2\} \\ &= \kappa \lim_{m\longrightarrow\infty} E[(\psi_{ij}^J)^2]/2m = \kappa \lim_{m\longrightarrow\infty}\{E[G_i'\Omega^{-1}G_i] + tr(\{\Omega^{-1}E[g_iG_i']\}^2)\}/m.\end{aligned}$$

The U-statistic term is uncorrelated with the usual GMM term, so by the central limit theorem, $\mu_n^{-1}\partial\breve{Q}(\beta_0)/\partial\beta \xrightarrow{d} N(0, H + \Lambda_J)$. It then follows that

$$\mu_n(\breve{\beta} - \beta_0) \xrightarrow{d} N(0, V_J), V_J = H^{-1} + H^{-1}\Lambda_J H^{-1},$$

a result that was previously derived for linear IV by Chao and Swanson (2004).

For the CUE, let $B = \Omega^{-1}E[g_iG_i']$ be the coefficients from the population regression of G_i on g_i and $U_i = G_i - G - B'g_i$ be the corresponding residual. Assuming we can differentiate under the integral we have

$$\partial\Omega(\beta_0)^{-1}/\partial\beta = -\Omega^{-1}[\partial\Omega(\beta_0)/\partial\beta]\Omega^{-1} = -B\Omega^{-1} - \Omega^{-1}B'$$

Then differentiating the CUE objective function with $\Omega(\beta)^{-1}/n$ replacing $\hat{\Omega}(\beta)^{-1}$ we have

$$\begin{aligned}\mu_n^{-1}\partial\hat{Q}(\beta_0)/\partial\beta &= \mu_n^{-1}\{\partial[\hat{g}(\beta_0)'\Omega^{-1}\hat{g}(\beta_0)]/\partial\beta + \partial[\hat{g}'\Omega(\beta_0)^{-1}\hat{g}]/\partial\beta\}/2n \\ &= \mu_n^{-1}\{\frac{\partial\hat{g}(\beta_0)'}{\partial\beta}\Omega^{-1}\hat{g} - \hat{g}'B\Omega^{-1}\hat{g}\}/n \\ &= G_n'\Omega^{-1}\hat{g}/n^{1/2} + \sum_{j<i}\psi_{ij}^*/n\mu_n + \sum_{i=1}^n U_i'\Omega^{-1}g_i/n\mu_n, \\ \psi_{ij}^* &= U_j{}'\Omega^{-1}g_i + U_i'\Omega^{-1}g_j.\end{aligned}$$

By the law of large numbers, $\sum_{i=1}^n U_i'\Omega^{-1}g_i/n\mu_n \xrightarrow{p} 0$. Also note that $E[(\psi_{ij}^*)^2]/2 = E[U_i'\Omega^{-1}U_i]$. It then follows similarly to the jackknife GMM that $\mu_n^{-1/2}\partial\hat{Q}(\beta_0)/\partial\beta \xrightarrow{d} N(0, H + \Lambda^*)$, $\Lambda^* = \kappa \lim_{m\longrightarrow\infty} E[U_i'\Omega^{-1}U_i]/m$. Then it follows that

$$\mu_n(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V), V = H^{-1} + H^{-1}\Lambda^* H^{-1}.$$

We now show that the CUE is asymptotically efficient relative to the jackknife GMM, i.e. that $V \leq V_J$ in the positive semidefinite sense. Let $\Delta_{ij} = g_j'B\Omega^{-1}g_i + g_i'B\Omega^{-1}g_j$. Note that by $E[U_ig_i'] = E[U_jg_j'] = 0$ we have

$$\psi_{ij}^J = \psi_{ij}^* + \Delta_{ij}, E[\psi_{ij}^*\Delta_{ij}] = 0.$$

It follows that $E[(\psi_{ij}^J)^2] = E[(\psi_{ij}^*)^2] + E[(\Delta_{ij})^2] \geq E[(\psi_{ij}^*)^2]$, so that

$$\Lambda^* = \kappa \lim_{m \longrightarrow \infty} E[(\psi_{ij}^*)^2]/2m \leq \kappa \lim_{m \longrightarrow \infty} E[(\psi_{ij}^J)^2]/2m = \Lambda_J.$$

Thus we have

$$V = H^{-1} + H^{-1}\Lambda^* H^{-1} \leq H^{-1} + H^{-1}\Lambda_J H^{-1} = V_J,$$

showing the asymptotic relative efficiency of CUE.

The linear model provides an example of the asymptotic variance. Continuing to assume that β is a scalar, we have

$$\begin{aligned} B &= -\Omega^{-1}E[z_i z_i' x_i \varepsilon_i] = -\Omega^{-1}E[z_i z_i' \eta_{ij} \varepsilon_i], \\ U_i &= -z_i x_i + E[z_i x_i] - B' z_i \varepsilon_i = -z_i \Upsilon_i + E[z_i \Upsilon_i] + u_i,\ u_i = -z_i \eta_i - B' z_i \varepsilon_i. \end{aligned}$$

Then we have,

$$E[U_i'\Omega^{-1}U_i]/m = E[u_i'\Omega^{-1}u_i]/m + E[\{z_i\Upsilon_i - E[z_i\Upsilon_i]\}'\Omega^{-1}\{z_i\Upsilon_i - E[z_i\Upsilon_i]\}]/m.$$

Under many weak instruments Υ_i is small, so that

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

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

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

Then, assuming $\pi_{mn}'\pi_{mn}n/\mu_n^2 \longrightarrow A$ for a nonsingular matrix A, the asymptotic variance matrix for $\mu_n(\hat{\beta} - \beta_0)$ will be

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

This variance for the CUE is identical to the asymptotic variance of LIML under many weak instrument asymptotics calculated by Stock and Yogo (2005). Thus we find that in the linear homoskedastic model the CUE and LIML have the same asymptotic variance under many weak moment asymptotics. As shown by Hansen, Hausman, and Newey (2005), the Bekker (1994) standard errors are consistent under many weak instruments, so that $\mu_n^2 \hat{V}$ will have the same limit as the Bekker standard errors in a homoskedastic linear model. Since $\mu_n^2 \hat{V}$ will also be consistent with heteroskedasticity, one can think of $\hat{V}$ as an extension of the Bekker (1994) variance estimator to GEL with heteroskedasticity.

It is interesting to compare the asymptotic variance V of the CUE with the usual asymptotic variance formula H^{-1} for GMM. When $\kappa = \lim(m/\mu_n^2) = 0$ or $\partial g_i(\beta_0)/\partial\beta$ is constant $V = H^{-1}$, but otherwise the variance here is larger than the standard formula. For further comparison we consider a corresponding variance approximation V_n for $\hat{\beta}$ for a sample size of size n. Replacing H with $(n/\mu_n^2)G'\Omega^{-1}G$ and Λ^* by $\Lambda_n = (m/\mu_n^2)E[U_i'\Omega^{-1}U_i]/m$, and dividing by μ_n^2 (the square of the convergence rate) gives the variance approximation for sample size n of

$$\begin{aligned} V_n &= (G'\Omega^{-1}Gn/\mu_n^2)^{-1}/\mu_n^2 + (G'\Omega^{-1}Gn/\mu_n^2)^{-1}\Lambda_n(G'\Omega^{-1}Gn/\mu_n^2)^{-1}/\mu_n^2 \\ &= (G'\Omega^{-1}G)^{-1}/n + \frac{m}{n}(G'\Omega^{-1}G)^{-1}(E[U_i'\Omega^{-1}U_i]/m)(G'\Omega^{-1}G)^{-1}/n. \end{aligned}$$

The usual variance approximation for $\hat{\beta}$ is $(G'\Omega^{-1}G)^{-1}/n$. The approximate variance V_n includes an additional term which can be important in practice. When $Var(\Omega^{-1/2}\partial g_i(\beta_0)/\partial\beta) >> G'\Omega^{-1}G$, as seems descriptive of many IV settings, $E[U_i'\Omega^{-1}U_i]/m$ may be very large relative to $G'\Omega^{-1}G$, leading to the additional term being important, even when m/n is small.

It is interesting to note that the usual term is divided by n and the additional term by n^2. In asymptotic theory with fixed m this makes the additional term a "higher-order" variance term. Indeed, by inspection of Donald and Newey (2003), one can see that the additional term corresponds to a higher order variance term involving sample variability of the Jacobian. There are also additional higher order terms that come from the estimation of the weight matrix, but the Jacobian term dominates as identification becomes weak. For example, in the linear homoskedastic example suppose that $E[\varepsilon_i^3|z_i] = 0$ and $E[\varepsilon_i^4|z_i] = E[\varepsilon_i^4]$ and let $A_n = \pi_{mn}'\pi_{mn}$. The higher-order variance approximation for GEL from Donald and Newey (2003) is

$$\begin{aligned} V_n &= \sigma_\varepsilon^2 A_n^{-1}/n + (m/n)\sigma_\varepsilon^2 A_n^{-1}(\Sigma_\eta - \sigma_{\eta\varepsilon}\sigma_{\eta\varepsilon}'/\sigma_\varepsilon^2)A_n^{-1}/n \\ &\quad +[(5-\kappa)+\rho_3(0)(3-\kappa)]\sigma_\varepsilon^2 A_n^{-1}(\pi_{mn}' E[\|z_i\|^2 z_i z_i']\pi_{mn})A_n^{-1}/n^2. \end{aligned}$$

The last term corresponds to estimating of the weight matrix and will tend to be small when π_{mn} is small, as it is under the asymptotics we consider. In this sense the many weak moment asymptotics accounts well for variability of the derivative of the moment conditions but takes no account of variability of the weight matrix.

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

Assumption 4: $g(z,\beta)$ is twice continuously differentiable in a neighborhood N of β_0, $\{E[\|g_i(\beta_0)\|^4] + E[\|\partial g_i(\beta_0)/\partial\beta\|^4]\}(m/n + 1/m\sqrt{n}) \longrightarrow 0$, and for all $\beta \in N$ we have $\lambda_{\max}(E[\partial g_i(\beta)/\partial\beta_j\{\partial g_i(\beta)/\partial\beta_j\}']) \leq C$, $\lambda_{\max}(E[\partial^2 g_i(\beta)/\partial\beta_j\partial\beta_k\{\partial^2 g_i(\beta)/\partial\beta_j\partial\beta_k\}']) \leq C$ 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 $m^3/n \longrightarrow 0$. Let $\tilde{A}^j(\beta) = \sum_{i=1}^n \{\partial g_i(\beta)/\partial\beta_j\}g_i(\beta)'$ and $A^j(\beta) = E[\{\partial g_i(\beta)/\partial\beta_j\}g_i(\beta)']$.

Assumption 5: For all β on a neighborhood N of β_0 i) each of $\sup_{\beta\in N}\|\hat{g}(\beta)\|/(\mu_n\sqrt{n})$, $\sup_{\beta\in N}\|\partial\hat{g}(\beta)/\partial\beta_j\|/(\mu_n\sqrt{n})$, and $\sup_{\beta\in N}\|\partial^2\hat{g}(\beta)/\partial\beta_j\partial\beta_k\|/(\mu_n\sqrt{n})$ 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) $\sup_{\beta\in N}\|n^{-1}\tilde{A}^j(\beta) - A^j(\beta)\| \xrightarrow{p} 0$, $\sup_{\beta\in N}\|n^{-1}\partial^2\hat{\Omega}(\beta)/\partial\beta_j\partial\beta_k - \partial^2\Omega(\beta)/\partial\beta_j\partial\beta_k\| \xrightarrow{p} 0$.

Let $\dot{Q}(\beta) = \hat{g}(\beta)'\Omega(\beta)^{-1}\hat{g}(\beta)/2n\mu_n^2$, $\tilde{D}^j(\beta) = [\partial\hat{g}(\beta)/\partial\beta_j - A^j(\beta)\Omega(\beta)^{-1}\hat{g}(\beta)]/\mu_n\sqrt{n}$, and $\tilde{D}(\beta) = [\tilde{D}^1(\beta), ..., \tilde{D}^p(\beta)]$.

Assumption 6: $\partial^2\dot{Q}(\beta)/\partial\beta\partial\beta'$ and $\tilde{D}(\beta)'\Omega(\beta)^{-1}\tilde{D}(\beta)$ are stochastically equicontinuous.

Under these and other regularity conditions we can show that $\hat{\beta}$ is asymptotically normal and that the variance estimator is consistent. Let $B^j = \Omega^{-1}E[g_i \partial g_i(\beta_0)/\partial \beta_j']$, $U_i^j = \partial g_i(\beta_0)/\partial \beta_j - E[\partial g_i(\beta_0)/\partial \beta_j] - B^{j\prime} g_i$, and $U_i = [U_i^1, ..., U_i^p]$.

Theorem 3: *If Assumptions 1, 2, and 4-6 are satisfied and* $E[U_i'\Omega^{-1}U_i]/\mu_n^2 \longrightarrow \Lambda^*$ *then for* $V = H^{-1} + H^{-1}\Lambda^* H^{-1}$

$$\mu_n(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V), \mu_n^2 \hat{V} \xrightarrow{p} V.$$

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

Theorem 4: *If Assumptions 1 i), ii), and 3 are satisfied,* $E[U_i'\Omega^{-1}U_i]/\mu_n^2 \longrightarrow \Lambda^*$, *and* $E[(z_i'z_i)^2]m/n \longrightarrow 0$ *then*

$$\mu_n(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V), \mu_n^2 \hat{V} \xrightarrow{p} V.$$

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.

The last result shows that the Kleibergen (2005) K-statistic has the usual chi-squared distribution:

Theorem 5: *If Assumptions 1, 2, and 4-6 are satisfied* $\Lambda_n \longrightarrow \Lambda^*$ *and* $\beta_0 = \bar{\beta}$ *then*

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

5 Monte Carlo Results

We first carry out a Monte Carlo for the linear IV model of equation (2.1) where the disturbances and instruments have a Gaussian distribution, $\Upsilon_i = z_i'\pi$. The parameters of this experiment are the correlation coefficient ρ between the structural and reduced form errors, the concentration parameter $E[\pi' Z' Z\pi]/Var(\eta)$, and the number of instruments m.

The data generating process is given by

$$
\begin{aligned}
y_i &= x_i\beta_0 + \varepsilon_i \\
x_i &= z_i'\pi + \eta_i \\
\varepsilon_i &= \rho\eta_i + \sqrt{1-\rho^2}v_i \\
\eta_i &\sim N(0,1)\,;\; v_i \sim N(0,1)\,;\; z_i \sim N(0, I_m) \\
\pi &= \sqrt{\frac{CP}{mn}}\iota_m,
\end{aligned}
$$

where ι_m is an m-vector of ones. The concentration parameter in this design is equal to CP. We generate samples of size $n = 200$, with values of CP equal to 10, 20 or 35; number of instruments m equal to 3, 10 or 15; values of ρ equal to 0.3 or 0.5; and $\beta_0 = 0$.

Table 1 presents the estimation results for $10,000$ Monte Carlo replications. We report median bias and interquartile range (IQR) of 2SLS, GMM, LIML and CUE. The results for 2SLS and GMM are as expected. They are upward biased, with the bias increasing with the number of instruments, the degree of endogeneity and a decreasing concentration parameter. LIML and CUE are close to being median unbiased, although they display some small biases, accompanied by large interquartile ranges, when $CP = 10$ and the number of instruments is larger than 3. There is a clear reduction in IQR for LIML and CUE when both the number of instruments and the concentration parameter increase, whereas the biases for 2SLS and GMM remain.

Table 2 presents rejection frequencies of Wald tests at 5% nominal level. The estimators and standard errors utilised in the Wald tests are the two-step GMM estimator with the usual standard errors (GMM2), with the Windmeijer (2005) standard errors (GMM2C), the continuous updating estimator with the usual standard errors (CUE) and with the standard errors presented here (CUEC). For purposes of comparison we also give results for 2SLS and LIML with Bekker (1994) standard errors (LIMLC), and the Kleibergen test statistic (KST).

Table 1. Simulation results for linear IV model

	$CP = 10$		$CP = 20$		$CP = 35$	
	Med Bias	IQR	Med Bias	IQR	Med Bias	IQR
$\rho = 0.3$						
$m = 3$						
2SLS	0.0474	0.3891	0.0258	0.2876	0.0145	0.2217
GMM	0.0466	0.3964	0.0248	0.2896	0.0151	0.2242
LIML	-0.0017	0.4839	-0.0049	0.3238	-0.0016	0.2356
CUE	-0.0055	0.4955	-0.0042	0.3245	-0.0012	0.2392
$m = 10$						
2SLS	0.1438	0.3009	0.0972	0.2449	0.0615	0.1991
GMM2	0.1431	0.3140	0.0990	0.2499	0.0586	0.2066
LIML	0.0076	0.6060	0.0046	0.3725	-0.0034	0.2558
CUE	0.0140	0.6481	0.0041	0.4020	-0.0064	0.2771
$m = 15$						
2SLS	0.1792	0.2661	0.1262	0.2267	0.0847	0.1910
GMM2	0.1800	0.2791	0.1249	0.2364	0.0878	0.1986
LIML	0.0207	0.6572	0.0021	0.4111	-0.0021	0.2801
CUE	0.0339	0.7183	0.0044	0.4552	-0.0033	0.3159
$\rho = 0.5$						
$m = 3$						
2SLS	0.0970	0.3764	0.0494	0.2793	0.0297	0.2177
GMM	0.0970	0.3786	0.0502	0.2845	0.0308	0.2216
LIML	0.0099	0.4696	0.0011	0.3153	0.0020	0.2365
CUE	0.0092	0.4786	0.0031	0.3238	0.0022	0.2383
$m = 10$						
2SLS	0.2384	0.2786	0.1575	0.2364	0.1062	0.1908
GMM2	0.2386	0.2940	0.1580	0.2446	0.1060	0.1987
LIML	0.0122	0.5680	-0.0001	0.3599	0.0019	0.2518
CUE	0.0226	0.6052	-0.0015	0.3862	0.0039	0.2692
$m = 15$						
2SLS	0.2985	0.2475	0.2122	0.2154	0.1458	0.1789
GMM2	0.2994	0.2590	0.2093	0.2222	0.1460	0.1895
LIML	0.0297	0.6335	0.0040	0.3980	-0.0025	0.2759
CUE	0.0384	0.7096	0.0030	0.4348	-0.0029	0.3091

Notes: $n = 200$; $\beta_0 = 0$; $10,000$ replications

Table 2. Rejection frequencies of Wald tests for linear IV model

	$\rho = 0.3$			$\rho = 0.5$		
	$CP = 10$	$CP = 20$	$CP = 35$	$CP = 10$	$CP = 20$	$CP = 35$
$m = 3$						
2SLS	0.0448	0.0441	0.0507	0.0836	0.0707	0.0633
GMM	0.0477	0.0472	0.0539	0.0862	0.0761	0.0664
GMMC	0.0471	0.0452	0.0510	0.0805	0.0715	0.0626
LIML	0.0380	0.0388	0.0448	0.0609	0.0521	0.0516
LIMLC	0.0304	0.0334	0.0407	0.0490	0.0457	0.0480
CUE	0.0749	0.0605	0.0620	0.0932	0.0710	0.0639
CUEC	0.0338	0.0359	0.0442	0.0527	0.0475	0.0457
KST	0.0476	0.0448	0.0465	0.0461	0.0479	0.0448
$m = 10$						
2SLS	0.1088	0.0923	0.0739	0.2546	0.1838	0.1393
GMM	0.1357	0.1155	0.0973	0.2806	0.2113	0.1674
GMMC	0.1091	0.0922	0.0757	0.2333	0.1727	0.1315
LIML	0.0770	0.0675	0.0595	0.0998	0.0749	0.0597
LIMLC	0.0344	0.0369	0.0391	0.0536	0.0465	0.0437
CUE	0.3384	0.2293	0.1606	0.3073	0.2104	0.1447
CUEC	0.0542	0.0496	0.0452	0.0773	0.0568	0.0477
KST	0.0371	0.0334	0.0344	0.0375	0.0375	0.0339
$m = 15$						
2SLS	0.1654	0.1296	0.1127	0.3993	0.3079	0.2231
GMM	0.2083	0.1732	0.1440	0.4391	0.3473	0.2649
GMMC	0.1565	0.1242	0.1012	0.3608	0.2730	0.1964
LIML	0.1054	0.0865	0.0813	0.1300	0.0894	0.0736
LIMLC	0.0381	0.0391	0.0438	0.0602	0.0495	0.0460
CUE	0.4741	0.3408	0.2516	0.4534	0.3176	0.2322
CUEC	0.0733	0.0621	0.0531	0.0963	0.0697	0.0558
KST	0.0346	0.0330	0.0315	0.0316	0.0328	0.0304

Notes: $n = 200$; $H_0 : \beta_0 = 0$; $10,000$ replications, 5% nominal size

The LIML Wald test using the Bekker standard errors has rejection frequencies very close to the nominal size, correcting the usual asymptotic Wald test which is oversized. Kleibergen's K-statistic shows a tendency to be undersized with an increasing number of instruments. The results for the rejection frequencies of the Wald test show that even with low numbers of instruments the corrected standard errors for the continuous updating estimator produce large improvements in the accuracy of the approximation. When the instruments are not too weak, i.e. when $CP = 20$ and larger, the observed rejection frequencies are very close to the nominal size for all values of m, whereas those based on the usual asymptotic standard errors are much larger than the nominal size. When we consider the "diagonal" elements, i.e. increasing the number of instruments and the concentration parameter at the same time, we see that the CUEC Wald test performs very well in terms of size.

We next analyze the properties of the CUE using the many weak instrument asymptotics for the estimation of the parameters in a panel data process, generated as in Windmeijer (2005):

$$\begin{aligned} y_{it} &= \beta_0 x_{it} + u_{it},\ u_{it} = \eta_i + v_{it} \\ x_{it} &= \gamma x_{it-1} + \eta_i + 0.5 v_{it-1} + \varepsilon_{it}, \eta_i \sim N(0,1), \varepsilon_{it} \sim N(0,1) \\ v_{it} &= \delta_i \tau_t \omega_{it}, \omega_{it} \sim \left(\chi_1^2 - 1\right), \delta_i \sim U[0.5, 1.5], \tau_t = 0.5 + 0.1(t-1). \end{aligned}$$

Fifty time periods are generated, with $\tau_t = 0.5$ for $t = -49, ..., 0$ and $x_{i,-49} \sim N\left(\frac{\eta_i}{1-\gamma}, \frac{1}{1-\gamma^2}\right)$, before the estimation sample is drawn. $n = 250$, $T = 6$, $\beta_0 = 1$ and $10,000$ replications are drawn. For this data generating process the regressor x_{it} is correlated with the unobserved constant heterogeneity term η_i and is predetermined due to its correlation with v_{it-1}. The idiosyncratic shocks v_{it} are heteroskedastic over time and at the individual level, and have a skewed chi-squared distribution. The model parameter β_0 is estimated by first-differenced GMM (see Arellano and Bond (1991)). As x_{it} is predetermined the sequential moment conditions used are

$$g_i(\beta) = Z_i' \Delta u_i(\beta),$$

where

$$Z_i = \begin{bmatrix} x_{i1} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & x_{i1} & x_{i2} & \cdots & 0 & \cdots & 0 \\ . & . & . & \cdots & . & \cdots & . \\ 0 & 0 & 0 & \cdots & x_{i1} & \cdots & x_{iT-1} \end{bmatrix},$$

$$\Delta u_i(\beta) = \begin{bmatrix} \Delta u_{i2}(\beta) \\ \Delta u_{i3}(\beta) \\ \vdots \\ \Delta u_{iT}(\beta) \end{bmatrix} = \begin{bmatrix} \Delta y_{i2} - \beta \Delta x_{i2} \\ \Delta y_{i3} - \beta \Delta x_{i3} \\ \vdots \\ \Delta y_{iT} - \beta \Delta x_{iT} \end{bmatrix}.$$

This results in a total of 15 moment conditions in this case, but only a maximum of 5 instruments for the cross section in the last time period.

The first two sets of results in Table 3 are the estimation results for values of $\gamma = 0.40$ and $\gamma = 0.85$ respectively. When $\gamma = 0.40$ the instruments are relatively strong, but they are weaker for $\gamma = 0.85$. The reported empirical concentration parameter is an object corresponding to the reduced form of this panel data model and is equal to 261 when $\gamma = 0.4$ and 35 when $\gamma = 0.85$. This is estimated simply from the linear reduced form estimated by OLS and ignores serial correlation and heteroskedasticity over time. This CP is therefore only indicative and does not play the same role as in the linear homoskedastic IV model. Median bias and interquartile range (IQR) are reported for the standard linear one-step and two-step GMM estimators and the CUE. When $\gamma = 0.40$, median biases are negligible for both GMM and CUE, with comparable interquartile ranges. When $\gamma = 0.85$ and the instruments are weaker, the linear GMM estimators are downward biased, whereas the CUE is median unbiased but exhibits a larger interquartile range than the linear GMM estimators.

Table 3. Simulation results for panel data model, $N = 250$, $T = 6$

	$\gamma = 0.40$ $(CP = 261)$		$\gamma = 0.85$ $(CP = 35)$		$\gamma = 0.85$ $(CP = 54)$	
	Med Bias	IQR	Med Bias	IQR	Med Bias	IQR
GMM1	-0.0087	0.0784	-0.0689	0.2059	-0.0842	0.1780
GMM2	-0.0056	0.0714	-0.0508	0.1896	-0.0565	0.1617
CUE	-0.0001	0.0740	0.0000	0.2557	0.0000	0.2186
Instr:	$x_{it-1}, ..., x_{i1}$		$x_{it-1}, ..., x_{i1}$		$x_{it-1}, ..., x_{i1}; y_{it-2}, ..y_{i1}$	

Figures 1 and 2 present p-value plots for the Wald tests for the hypothesis $H_0 : \beta_0 = 1$, based on one-step GMM estimates (W_{GMM1}), on two-step GMM estimates (W_{GMM2}), on the Windmeijer (2005) corrected two-step Wald (W_{GMM2C}), on the continuously updated Wald test using the conventional asymptotic variance (W_{CUE}) and on the continuously updated Wald test using the variance estimate $\hat{V}$ described in Section 2, W_{CUEC}. Further displayed is the p-value plot for Kleibergen's (2005) K statistic. It is clear that the usual asymptotic variance estimate for the CUE is too small, especially when $\gamma = 0.85$. This problem is similar to that of the linear two-step GMM estimator, leading to rejection frequencies that are much larger than the nominal size. In contrast, use of the variance estimator under many weak instrument asymptotics leads to rejection frequencies that are very close to the nominal size.

The third set of results presented in Table 3 is for the design with $\gamma = 0.85$, but with lags of the dependent variable y_{it} included as sequential instruments $(y_{i,t-2}, ..., y_{i1})$ additional to the sequential lags of x_{it}. As there is feedback from y_{it-1} to x_{it} and x_{it} is correlated with η_i the lagged values of y_{it} could improve the strength of the instrument set. The total number of instruments increases to 25, with a maximum of 11 for the cross section in the final period. The empirical concentration parameter increases from 35 to 54. The GMM estimators are slightly more downward biased, especially GMM1, when the extra instruments are included. The CUE is still median unbiased and its IQR has decreased by 15%. As the p-value plot in Figure 3 shows, use of the proposed variance estimator results in rejection frequencies that are virtually equal to the nominal size. Although W_{GMM2C} had good size properties when using the smaller instrument set, use of the additional instruments leads to rejection frequencies that are larger than the nominal size.

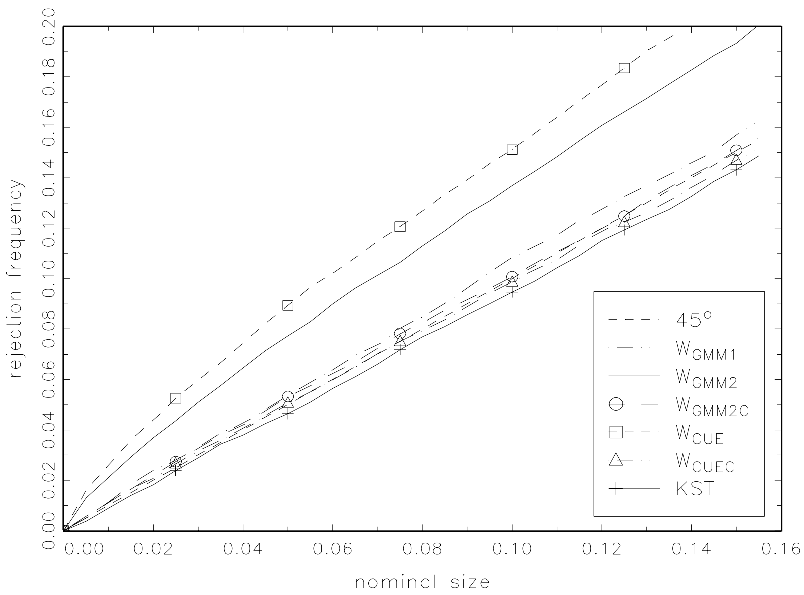

Fig. 1. P-value plot, $\gamma = 0.4$, $H_0 : \beta_0 = 1$, Panel data model

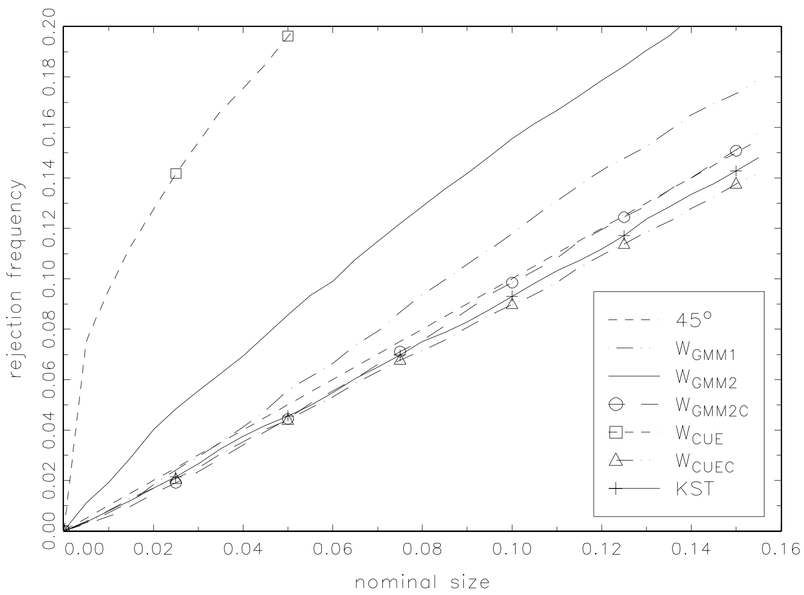

Fig. 2. P-value plot, $\gamma = 0.85$, $H_0 : \beta_0 = 1$, Panel data model

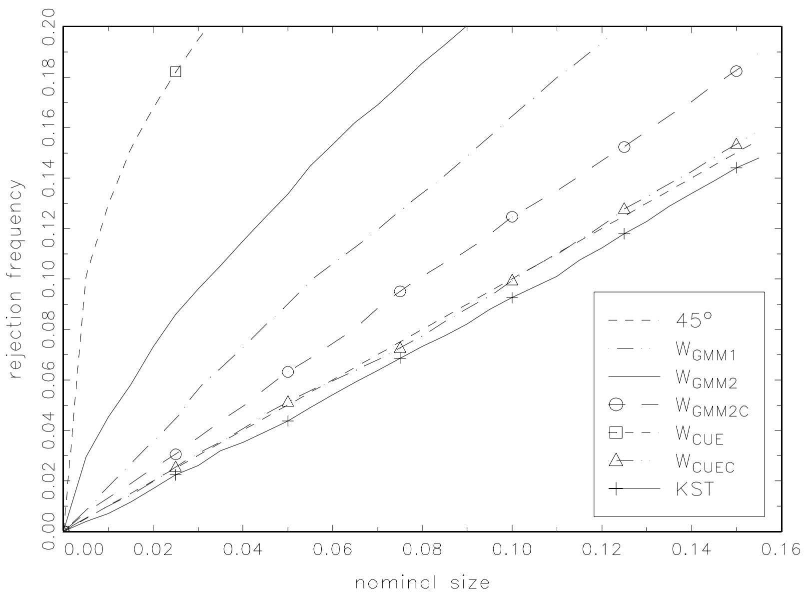

Fig. 3. P-value plot, $\gamma = 0.85$, $H_0 : \beta_0 = 1$, Panel data model, additional instr.

6 Conclusion

We have given an asymptotic approximation for generalized empirical likelihood estimators that accounts for many weak moment conditions by adding a term to the variance, and have suggested an estimator for that variance. This approximation is shown to perform well in a simple linear IV and panel data Monte Carlo.

There are several topics that could be considered in future research. One topic would be more refined asymptotics where the number of moment conditions m grows slower than the concentration parameter μ_n^2, i.e. where $\kappa = \lim_{n\longrightarrow\infty}(m/\mu_n^2) = 0$. Here we have focused on the case where $\kappa > 0$, leading to an asymptotic variance that is larger than the usual one. When $\kappa = 0$ the asymptotic variance is the same as the usual one, but the standard errors given here may provide an improvement over the usual standard errors. Intuitively, if m grows slower than, but close to μ_n^2, the standard errors may still help account for the extra term. Hansen, Hausman, and Newey (2005) have shown that Bekker (1994) standard errors in a homoskedastic linear model give an improved approximation if m^2 grows faster than μ_n^2. We expect that this result will also hold here.

Another interesting topic is the choice of moment conditions under many weak moment conditions. Donald, Imbens, and Newey (2003) give a criteria for moment choice for GMM and GEL that is quite complicated. Under many weak moment conditions this criteria should simplify. It would be useful in practice to have a simple criteria for choosing the moment conditions.

A third topic for future research is the extension of these results to dependent observations. It appears that the variance estimator for the CUE would be the same except that $\hat{\Omega}$ would include autocorrelation terms. It should also be possible to obtain similar results for GEL estimators based on time smoothed moment conditions, like those considered in Kitamura and Stutzer (1997).

7 Appendix A: Proofs of Theorems 1 - 5.

Throughout the Appendices, 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

$$\bar{Y} = \sum_{i=1}^{n} Y_i/n, \mu_Y = E[Y_i], \Sigma_{YY} = E[Y_i Y_i'], \Sigma_{YZ} = E[Y_i Z_i']$$

and define the corresponding object with Z in place of Y.

Lemma A1: *If* $\lambda_{\max}(AA') \leq C$, $\lambda_{\max}(A'A) \leq C$, $\lambda_{\max}(\Sigma_{YY}) \leq C$, $\lambda_{\max}(\Sigma_{ZZ}) \leq C$, $E[(Y_i'Y_i)^2]/nm \leq C$, $E[(Z_i'Z_i)^2]/nm \leq C$, $n\mu_Y'\mu_Y/m \leq C$, $n\mu_Z'\mu_Z/m \leq C$, *then*

$$n\bar{Y}'A\bar{Z}/m = \operatorname{tr}(A\Sigma_{YZ}')/m + n\mu_Y'A\mu_Z/m + O_p(1/\sqrt{m} + 1/n).$$

Proof: Let $W_i = AZ_i$. Then $A\Sigma'_{YZ} = \Sigma'_{YW}$, $A\mu_Z = \mu_W$,

$$\begin{aligned} \lambda_{\max}(E[W_iW_i']) &= \lambda_{\max}(A\Sigma_{ZZ}A') \leq C\lambda_{\max}(AA') \leq C \\ E[(W_i'W_i)^2]/nm &= E[(Z_i'A'AZ_i)^2]/nm \leq C. \end{aligned}$$

Thus the hypotheses and conclusion are satisfied with W in place of Z and $A = I$. Therefore, it suffices to show the result with $A = I$.

Note that

$$\begin{aligned} E[(Y_i'Z_i)^2] &\leq CE[(Y_i'Y_i)^2] + CE[(Z_i'Z_i)^2] \leq Cmn, \\ E[Y_i'Z_jZ_j'Y_i] &= E[Y_i'\Sigma_{ZZ}Y_i] \leq CE[Y_i'Y_i] = Ctr(\Sigma_{YY}) \leq Cm, \\ |E[Y_i'Z_jY_j'Z_i]| &\leq C(E[Y_i'Z_jZ_j'Y_i] + E[Y_j'Z_iZ_i'Y_j]) \leq Cm. \end{aligned}$$

For the moment suppose $\mu_y = \mu_z = 0$. Let $W_n = n\bar{Y}'\bar{Z}/m$. Then $E[W_n] = \mathrm{tr}\,(\Sigma'_{YZ})\,/m$ and

$$\begin{aligned} E[W_n^2] &= E\left[\sum_{i,j,k,\ell} Y_i'Z_jY_k'Z_\ell/n^2m^2\right] = E\left[(Y_i'Z_i)^2\right]/nm^2 + (1-1/n)\{E[W_n]^2 \\ &\quad + E[Y_i'Z_jY_j'Z_i]/m^2 + E[Y_i'Z_jZ_j'Y_i]/m^2\} = E[W_n]^2 + O(1/m), \end{aligned}$$

so that by M,

$$W_n = \mathrm{tr}\,(\Sigma'_{YZ})\,/m + O_p(1/\sqrt{m}).$$

In general, when μ_Y or μ_Z are nonzero, note that $E[\{(Y_i-\mu_Y)'(Y_i-\mu_Y)\}^2] \leq CE[(Y_i'Y_i)^2]$ and $\lambda_{\max}(Var(Y_i)) \leq \lambda_{\max}(\Sigma_{YY})$, so the hypotheses are satisfied with $Y_i - \mu_Y$ replacing Y_i and $Z_i - \mu_Z$ replacing Y_i and Z_i respectively. Also,

$$W_n = n\left(\bar{Y}-\mu_Y\right)'(\bar{Z}-\mu_Z)/m + n\mu_Y'(\bar{Z}-\mu_Z)/m + n(\bar{Y}-\mu_Y)'\mu_Z/m + n\mu_Y'\mu_Z/m. \tag{7.1}$$

Note that

$$E\left[\left\{n\mu_Y'(\bar{Z}-\mu_Z)/m\right\}^2\right] = n\mu_Y'(\Sigma_{ZZ} - \mu_Z\mu_Z')\mu_Y/m^2 \leq n\mu_Y'\Sigma_{ZZ}\mu_Y/m^2 \leq C/m.$$

so by M, the second and third terms in eq. (7.1) (with Y and Z interchanged) are $O_p(1/\sqrt{m})$. Also, $tr(\mu_Z\mu_Y')/m = n^{-1}(n\mu_Y'\mu_Z/m) = O(1/n)$. Applying the result for the zero mean case then gives

$$
\begin{aligned}
W_n &= tr(\Sigma'_{YZ} - \mu_Z \mu'_Y)/m + n\mu'_Y \mu_Z/m + O_p(1/\sqrt{m}) \\
&= tr(\Sigma'_{YZ})/m + n\mu'_Y \mu_Z/m + O_p(1/\sqrt{m} + 1/n), Q.E.D..
\end{aligned}
$$

For the next result, let X_i denote a scalar random variable where we also suppress dependence on n, let $\Psi = \Sigma_{ZZ}\Sigma_{YY} + \Sigma_{ZY}^2$, and let $\bar{\lambda}_Z = \lambda_{\max}(\Sigma_{ZZ})$ and $\bar{\lambda}_Y = \lambda_{\max}(\Sigma_{YY})$.

LEMMA A2: *If* $E[X_i] = 0$, $E[Z_i] = E[Y_i] = 0$, Σ_{ZZ} *and* Σ_{YY} *exist,* $nE[X_i^2] \to A$, $nE[X_i^4] \longrightarrow 0$, $n^2 \operatorname{tr}(\Psi) \to \Lambda$, $mn^4 \bar{\lambda}_Z^2 \bar{\lambda}_Y^2 \longrightarrow 0$, and $n^3 E\left[|Y_1' Z_2|^4\right] \longrightarrow 0$, *then*

$$
\sum_{i=1}^n X_i + \sum_{i \neq j} Z_i' Y_j \xrightarrow{d} N(0, A + \Lambda)
$$

Proof: Let $w_i = (X_i, Y_i, Z_i)$ and for any $j < i$, $\psi_{ij} = Z_i' Y_j + Z_j' Y_i$. Note that

$$
\begin{aligned}
E[\psi_{ij} | w_{i-1}, ..., w_1] &= 0, \\
E[\psi_{ij}^2] &= E[(Z_i' Y_j)^2 + (Z_j' Y_i)^2 + 2Z_i' Y_j Z_j' Y_i] = 2tr(\Psi).
\end{aligned}
$$

We have

$$
\sum_{i=1}^n X_i + \sum_{i \neq j} Z_i' Y_j = \sum_{i=2}^n (X_i + B_{in}) + X_1, B_{in} = \sum_{j<i} \psi_{ij}.
$$

Note that $E[X_1^2] = (nE[X_i^2])/n \longrightarrow 0$, so $X_1 \xrightarrow{p} 0$ by M. Also, $E[X_i B_{in}] = E\left[X_i \sum_{j<i} \psi_{ij}\right] = 0$ and

$$
E\left[B_{in}^2\right] = E\left[\sum_{j,k<i} \psi_{ij}\psi_{ik}\right] = (i-1)E[\psi_{ij}^2] = 2(i-1)\operatorname{tr}(\Psi).
$$

Therefore

$$
\begin{aligned}
s_n &= \sum_{i=2}^n E[(X_i + B_{in})^2] = (n-1)E[X_i^2] + 2\sum_{i=2}^n (i-1)\operatorname{tr}(\Psi) \\
&= \frac{n-1}{n} nE[X_i^2] + \left(\frac{n^2 - n}{n^2}\right) n^2 \operatorname{tr}(\Psi) \longrightarrow A + \Lambda^*.
\end{aligned}
$$

Next, for $k \neq i$ and $k \neq j$ define

$$
\varphi_{ij} = E[\psi_{ki}\psi_{kj} | w_i, w_j] = Y_i' \Sigma_{ZZ} Y_j + Z_i' \Sigma_{YY} Z_j + Z_i' \Sigma_{YZ} Y_j + Z_j' \Sigma_{YZ} Y_i.
$$

Let $\bar{\lambda}_Z = \lambda_{\max}(\Sigma_{ZZ})$ and $\bar{\lambda}_Y = \lambda_{\max}(\Sigma_{YY})$. Note that

$$\begin{aligned} E[(Y_i'\Sigma_{ZZ}Y_j)^2] &= E[Y_i'\Sigma_{ZZ}\Sigma_{YY}\Sigma_{ZZ}Y_i] \leq \bar{\lambda}_Y E[Y_i'\Sigma_{ZZ}^{1/2}\Sigma_{ZZ}\Sigma_{ZZ}^{1/2}Y_i] \leq \bar{\lambda}_Y\bar{\lambda}_Z E[Y_i'\Sigma_{ZZ}Y_i] \\ &\leq \bar{\lambda}_Y\bar{\lambda}_Z^2 E[Y_i'Y_i] \leq m\bar{\lambda}_Y^2\bar{\lambda}_Z^2. \end{aligned}$$

Similarly, $E[(Z_i'\Sigma_{YY}Z_j)^2] \leq m\bar{\lambda}_Y^2\bar{\lambda}_Z^2$. We also have, by $I_m \leq \bar{\lambda}_Z\Sigma_{ZZ}^{-1}$ and $\Sigma_{YZ}\Sigma_{ZZ}^{-1}\Sigma_{ZY} \leq \Sigma_{YY}$,

$$\begin{aligned} E[(Z_i'\Sigma_{YZ}Y_j)^2] &= E[Z_i'\Sigma_{YZ}\Sigma_{YY}\Sigma_{ZY}Z_i] \leq \bar{\lambda}_Y E[Z_i'\Sigma_{YZ}\Sigma_{ZY}Z_i] \leq \bar{\lambda}_Y\bar{\lambda}_Z E[Z_i'\Sigma_{YZ}\Sigma_{ZZ}^{-1}\Sigma_{ZY}Z_i] \\ &\leq \bar{\lambda}_Y\bar{\lambda}_Z E[Z_i'\Sigma_{YY}Z_i] \leq m\bar{\lambda}_Y^2\bar{\lambda}_Z^2. \end{aligned}$$

Therefore, it follows that $E[\varphi_{ij}^2] \leq Cm\bar{\lambda}_Y^2\bar{\lambda}_Z^2$, so that

$$E[\varphi_{ij}^2]/E[\psi_{ij}^2]^2 = E[\varphi_{ij}^2]/[4tr(\Psi)^2] \leq Cmn^4\bar{\lambda}_Y^2\bar{\lambda}_Z^2 \longrightarrow 0.$$

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

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

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

$$\begin{aligned} \sum_{i=2}^{n} E[X_iB_{in}|w_{i-1},...,w_1] &= \sum_{i=2}^{n}\sum_{j<i} E\left[X_i\left(Z_i'Y_j + Z_j'Y_i\right) \mid w_{i-1},...,w_1\right] \\ &= \sum_{i=2}^{n}\{E[X_iZ_i']\left(\sum_{j<i}Y_j\right) + E[X_iY_i']\left(\sum_{j<i}Z_j\right)\} \\ &= E[X_iZ_i']\sum_{i=1}^{n-1}(n-i)Y_i + E[X_iY_i']\sum_{i=1}^{n-1}(n-i)Z_i \end{aligned}$$

Therefore

$$\begin{aligned} & E\left[\left(\sum_{i=2}^{n} E[X_iB_{in}|w_{i-1},...,w_1]\right)^2\right] \\ \leq\; & C(E[X_iY_i']\Sigma_{ZZ}E[Y_iX_i] + E[X_iZ_i']\Sigma_{YY}E[Z_iX_i])\sum_{i=1}^{n-1}(n-i)^2 \\ \leq\; & Cn^3\bar{\lambda}_Y\bar{\lambda}_Z E[X_i^2] \leq C\bar{\lambda}_Y\bar{\lambda}_Z n^2 = C(mn^4\bar{\lambda}_Y^2\bar{\lambda}_Z^2)^{1/2}/m^{1/2} \longrightarrow 0. \end{aligned}$$

Then by M, we have

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

By T it then follows that

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

Next, note that

$$n^{-1}E\left[\psi_{ij}^4\right]/E[\psi_{ij}^2]^2 \le Cn^3E\left[|Z_1'Y_2|^4\right]/\left[n^2\operatorname{tr}(\Psi)\right]^2 \longrightarrow 0.$$

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

$$\sum_{i=2}^{n}E\left[(X_i+B_{in})^4\right] \le CnE[X_i^4]+C\sum_{i=1}^{n}E[B_{in}^4] \rightarrow 0,$$

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

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

The conclusion then follows from the martingale central limit theorem applied to $\sum_{i=2}^{n}(X_i+B_{in})$. Q.E.D.

Let $\dot{Q}(\beta)=\hat{g}(\beta)'\Omega(\beta)^{-1}\hat{g}(\beta)/n\mu_n^2$ and $\bar{Q}(\beta)=n\bar{g}(\beta)'\Omega(\beta)^{-1}\bar{g}(\beta)/\mu_n^2+m/\mu_n^2$.

Lemma A3: *If Assumption 2 is satisfied then* $\sup_{\beta\in B}|\dot{Q}(\beta)-\bar{Q}(\beta)| \xrightarrow{p} 0$.

Proof: Since $\dot{Q}(\beta)$ and $\bar{Q}(\beta)$ are stochastically equicontinuous by Assumption 2, it suffices by Newey (1991, Theorem 2.1) to show that $\dot{Q}(\beta) \xrightarrow{p} \bar{Q}(\beta)$ for each β. Apply Lemma A1 with $Y_i=Z_i=m^{1/2}g_i(\beta)/\mu_n$ and $A=\Omega(\beta)^{-1}$. By Assumption 2, $\lambda_{\max}(A'A)=\lambda_{\max}(AA')=\lambda_{\max}(\Omega(\beta)^{-2})\le C$, $\lambda_{\max}(\Sigma_{YY})=(m/\mu_n^2)\lambda_{\max}(\Omega(\beta))\le C$, $E[(Y_i'Y_i)^2]/nm=mE[\{g_i(\beta)'g_i(\beta)\}^2]/n\mu_n^4\le C$, and $n\mu_Y'\mu_Y/m\le Cn\bar{g}(\beta)'\Omega(\beta)^{-1}\bar{g}(\beta)/\mu_n^2=C\bar{Q}(\beta)\le C$ where the last inequality follows by equicontinuity of $\bar{Q}(\beta)$ (which implies $\bar{Q}(\beta)$ is uniformly bounded on the compact set B). Thus, the hypotheses of Lemma A1 are satisfied. Note that $A\Sigma_{YZ}'=A\Sigma_{ZZ}=A\Sigma_{YY}=mI_m/\mu_n^2$, so by the conclusion of Lemma A1

$$\dot{Q}(\beta)=tr(I_m)/\mu_n^2+n\bar{g}(\beta)'\Omega(\beta)^{-1}\bar{g}(\beta)/\mu_n^2+o_p(1)=\bar{Q}(\beta)+o_p(1).$$

Q.E.D.

Proof of Theorem 1: We first will show that $\sup_{\beta \in B} \left| 2\hat{Q}(\beta)/\mu_n^2 - \bar{Q}(\beta) \right| \xrightarrow{p} 0$. By T, Lemma A3, equicontinuity of $\bar{Q}(\beta)$ and B compact, we have $\sup_{\beta \in B} |\dot{Q}(\beta)| = O_p(1)$. Let $\hat{a}(\beta) = \Omega(\beta)^{-1}\hat{g}(\beta)/\mu_n\sqrt{n}$. By Assumption 2,

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

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

$$\left| \lambda_{\min}(\hat{\Omega}(\beta)/n) - \lambda_{\min}(\Omega(\beta)) \right| \leq \sup_{\beta \in B} \left\| \hat{\Omega}(\beta)/n - \Omega(\beta) \right\| \xrightarrow{p} 0,$$

so that $\lambda_{\min}(\hat{\Omega}(\beta)/n) \geq C$, and hence $\lambda_{\max}((\hat{\Omega}(\beta)/n)^{-1}) \leq C$ for all $\beta \in B$, w.p.a.1. Therefore,

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

It then follows by Assumption 2 that $\sup_{\beta \in B} \left| 2\hat{Q}(\beta)/\mu_n^2 - \dot{Q}(\beta) \right| \xrightarrow{p} 0$. Then $\sup_{\beta \in B} \left| 2\hat{Q}(\beta)/\mu_n^2 - \bar{Q}(\beta) \right| \xrightarrow{p} 0$ by T and Lemma A3. The conclusion then follows by standard results. Q.E.D.

Lemma A4: *If Assumption 3 is satisfied then* $E[(y_i - x_i'\beta)^2 | z_i, \Upsilon_i] \geq C$. *Also, for* $X_i = (y_i, x_i')'$, $E[\|X_i\|^4 | z_i, \Upsilon_i] \leq C$.

Proof: Note that for $\delta = \beta_0 - \beta$ we have $y_i - x_i'\beta = \varepsilon_i + \eta_i'\delta + \Upsilon_i'\delta$, so that

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

giving the first conclusion. Also, $E[\|x_i\|^4 | z_i, \Upsilon_i] \leq CE[\|\eta_i\|^4 | z_i, \Upsilon_i] + CE[\|\Upsilon_i\|^4 | z_i, \Upsilon_i] \leq C$ and $E[y_i^4 | z_i, \Upsilon_i] \leq CE[\|x_i\|^4 \|\beta_0\|^4 | z_i, \Upsilon_i] + E[\varepsilon_i^4 | z_i, \Upsilon_i] \leq C$, giving the second conclusion. Q.E.D.

Lemma A5: *If Assumption 3 is satisfied then there is a constant* C *such that for every* β *and* m, $C^{-1}I_m \leq \Omega(\beta) \leq CI_m$.

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

LEMMA A6: *If Assumptions 1 i), ii), and 3 are satisfied then Assumption 1 iii) is satisfied.*

Proof: By Assumption 3 and Lemma A5, $\lambda_{\min}(G'G) \geq C\lambda_{\min}(G'\Omega^{-1}G)$ for large enough n. Then 1by Lemma A5 and Assumption 1 ii),

$$\begin{aligned} n\bar{g}(\beta)'\Omega(\beta)^{-1}\bar{g}(\beta)/\mu_n^2 &= (\beta - \beta_0)'[nG'\Omega(\beta)^{-1}G/\mu_n^2](\beta - \beta_0) \\ &\geq C(\beta - \beta_0)'(nG'G/\mu_n^2)(\beta - \beta_0) \geq C\left\|\beta - \beta_0\right\|^2 .Q.E.D. \end{aligned}$$

LEMMA A7: *If Assumptions 1 and 3 are satisfied then there is* $\hat{M} = O_p(1)$ *with* *i)* $\|\partial\bar{g}(\beta)/\partial\beta\| = O(\mu_n/\sqrt{n})$, *ii)* $\|n^{-1}\partial\hat{g}(\beta)/\partial\beta - \partial\bar{g}(\beta)/\partial\beta\| = O_p(\mu_n/\sqrt{n})$, *iii)* $\sup_{\beta\in B}\|\bar{g}(\beta)\| = O(\mu_n/\sqrt{n})$, *iv)* $\sup_{\beta\in B}\|\hat{g}(\beta)/n\| = O_p(\mu_n/\sqrt{n})$, *v)*$\mu_n^{-1}\sqrt{n}\|\bar{g}(\tilde{\beta}) - \bar{g}(\beta)\| \leq C\|\tilde{\beta} - \beta\|$, *vi)*$\mu_n^{-1}n^{-1/2}\|\hat{g}(\tilde{\beta}) - \hat{g}(\beta)\| \leq \hat{M}\|\tilde{\beta} - \beta\|$.

Proof: Note first that $\partial\bar{g}(\beta)/\partial\beta = -E[z_i\Upsilon_i'] = G$, so i) follows by $G'G \leq CG'\Omega^{-1}G = O(\mu_n^2/n)$. Also, we have

$$\begin{aligned} (n/\mu_n^2)E[\left\|\sum_{i=1}^n z_i\eta_i'/n\right\|^2] &= E[z_i'z_i\eta_i'\eta_i]/\mu_n^2 \leq CE[z_i'z_i]/\mu_n^2 = Cm/\mu_n^2, \\ (n/\mu_n^2)E[\left\|\sum_{i=1}^n z_i\Upsilon_i'/n - E[z_i\Upsilon_i']\right\|^2] &\leq E[z_i'z_i\Upsilon_i'\Upsilon_i]/\mu_n^2 \leq (E[(z_i'z_i)^2]/n)^{1/2}\{nE[(\Upsilon_i'\Upsilon_i)^2]/\mu_n^4\}^{1/2}. \end{aligned}$$

Therefore by M and T we have

$$\left\|n^{-1}\partial\hat{g}(\beta)/\partial\beta - \partial\bar{g}(\beta)/\partial\beta\right\| \leq \left\|\sum_{i=1}^n z_i\eta_i'/n\right\| + \left\|\sum_{i=1}^n z_i\Upsilon_i'/n - E[z_i\Upsilon_i']\right\| = O_p(\mu_n/\sqrt{n}),$$

giving ii). For iii), note that by β in a compact set,

$$\|\bar{g}(\beta)\| = \|\partial\bar{g}(\beta)/\partial\beta(\beta - \beta_0)\| \leq C(\mu_n/\sqrt{n})\|\beta - \beta_0\| \leq C\mu_n/\sqrt{n}.$$

For iv), note that by T, i), and ii), and $\|\hat{g}(\beta_0)/n\| = O_p(\mu_n/\sqrt{n})$ we have

$$\sup_{\beta\in B}\|\hat{g}(\beta)/n\| = \sup_{\beta\in B}\|n^{-1}[\partial\hat{g}(\beta)/\partial\beta](\beta - \beta_0)\| + \|\hat{g}(\beta_0)/n\| = O_p(\mu_n/\sqrt{n}).$$

Finally, v) follows by i) and CS and vi) by i), ii), CS, and T. Q.E.D.

LEMMA A8: *If Assumption 3 is satisfied, then* $\sup_{\beta\in B}\|\hat{\Omega}(\beta)/n-\Omega(\beta)\|\xrightarrow{p}0$,

$$\sup_{\beta\in B}\|n^{-1}\partial\hat{\Omega}(\beta)/\partial\beta_j-\partial\Omega(\beta)/\partial\beta_j\|\xrightarrow{p}0,\sup_{\beta\in B}\|n^{-1}\partial^2\hat{\Omega}(\beta)/\partial\beta_j\partial\beta_k-\partial^2\Omega(\beta)/\partial\beta_j\partial\beta_k\|\xrightarrow{p}0.$$

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

$$\hat{\Omega}(\beta)-\Omega(\beta)=\sum_{j,k=1}^{p+1}\hat{F}_{jk}\alpha_j\alpha_k,\hat{F}_{jk}=\sum_{i=1}^{n}z_iz_i'X_{ij}X_{ik}/n-E[z_iz_i'X_{ij}X_{ik}].$$

Then $E[X_{ij}^2X_{ik}^2|z_i]\le C$ by Lemma A4 so that

$$E[\left\|\hat{F}_{jk}\right\|^2]\le CE[(z_i'z_i)^2E[X_{ij}^2X_{ik}^2|z_i]]/n\le CE[(z_i'z_i)^2]/n\longrightarrow 0.$$

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

LEMMA A9: *If Assumption 3 is satisfied, then*

$$\begin{aligned}|a'\Omega(\tilde{\beta})b-a'\Omega(\beta)b| &\le C\|a\|\|b\|\|\tilde{\beta}-\beta\|,\\ |a'[\partial\Omega(\tilde{\beta})/\partial\beta_j]b-a'[\partial\Omega(\beta)/\partial\beta_j]b| &\le C\|a\|\|b\|\|\tilde{\beta}-\beta\|.\end{aligned}$$

Proof: Let $\tilde{\Sigma}_i=E[X_iX_i'|z_i,\Upsilon_i]$, which is bounded. Then by $\alpha=(1,-\beta)$ bounded on B we have $|\tilde{\alpha}'\tilde{\Sigma}_i\tilde{\alpha}-\alpha'\tilde{\Sigma}_i\alpha|\le C\left\|\tilde{\beta}-\beta\right\|$. Also, $E[(a'z_i)^2]=a'E[z_iz_i']a=\|a\|^2$. Therefore,

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

We also have

$$\begin{aligned}&|a'\partial\Omega(\tilde{\beta})/\partial\beta_jb-a'\partial\Omega(\beta)/\partial\beta_jb|=|2E[(a'z_i)(b'z_i)E[x_{ij}X_i'(\tilde{\alpha}-\alpha)|z_i]]|\\ \le\ & C\|a\|\|b\|\|\tilde{\beta}-\beta\|\le CE[|a'z_i|\,|b'z_i|\,E[|x_{ij}|\|X_i\||z_i]]\|\tilde{\beta}-\beta\|\le C\|a\|\|b\|\|\tilde{\beta}-\beta\|.Q.E.D.\end{aligned}$$

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 $\sup_{\beta\in B}\|\hat{\Omega}(\beta)/n-\Omega(\beta)\| \stackrel{p}{\longrightarrow} 0$. Let $a(\beta,\tilde{\beta})=\Omega(\beta)^{-1}\bar{g}(\tilde{\beta})\sqrt{n}/\mu_n$ and $\bar{Q}(\tilde{\beta},\beta)=(n/\mu_n^2)\bar{g}(\tilde{\beta})'\Omega(\beta)^{-1}\bar{g}(\tilde{\beta})$. By Lemmas A5 and A7, $\sup_{\beta,\tilde{\beta}\in B}\left\|a(\beta,\tilde{\beta})\right\|\leq C$. Let Then by Lemma A9,

$$\left|\bar{Q}(\tilde{\beta},\tilde{\beta})-\bar{Q}(\tilde{\beta},\beta)\right|=\left|a(\tilde{\beta},\tilde{\beta})'\left[\Omega(\beta)-\Omega(\tilde{\beta})\right]a(\beta,\tilde{\beta})/2\right|\leq C\left\|\tilde{\beta}-\beta\right\|.$$

Also, by T and Lemma A7,

$$\begin{aligned}\left|\bar{Q}(\tilde{\beta},\beta)-\bar{Q}(\beta)\right| &\leq C(n/\mu_n^2)(\left\|\bar{g}(\tilde{\beta})-\bar{g}(\beta)\right\|^2+\|\bar{g}(\beta)\|\left\|\bar{g}(\tilde{\beta})-\bar{g}(\beta)\right\|)\\ &\leq C\left\|\tilde{\beta}-\beta\right\|.\end{aligned}$$

Then by T it follows that $\left|\bar{Q}(\tilde{\beta})-\bar{Q}(\beta)\right|\leq C\left\|\tilde{\beta}-\beta\right\|$, implying equicontinuity of $\bar{Q}(\beta)$. An analogous argument with $\hat{a}(\beta,\tilde{\beta})=\Omega(\beta)^{-1}\hat{g}(\tilde{\beta})/\sqrt{n}\mu_n$ and $\check{Q}(\tilde{\beta},\beta)=\hat{g}(\tilde{\beta})'\Omega(\beta)^{-1}\hat{g}(\tilde{\beta})/n\mu_n^2$ replacing $a(\beta,\tilde{\beta})$ and $\bar{Q}(\tilde{\beta},\beta)$ respectively implies that $\left|\check{Q}(\tilde{\beta})-\check{Q}(\beta)\right|\leq\hat{M}\left\|\tilde{\beta}-\beta\right\|$, with $\hat{M}=O_p(1)$, giving stochastic equicontinuity of $\check{Q}(\beta)$. Thus, all the hypotheses of Assumption 2 are satisfied. Assumption 1 iii) follows by Lemma A6. Thus, all the hypotheses of Theorem 1 are satisfied. The conclusion then follows by Theorem 1. Q.E.D.

For the next results $g_i=g_i(\beta_0)$, $g_i^k=\partial g_i(\beta_0)/\partial\beta_k$, $\tilde{\Omega}=\hat{\Omega}(\beta_0)/n$, $\tilde{A}^k=\sum_{i=1}^n g_i g_i^{k\prime}/n$, $A^k=E[\tilde{A}^k]$, $\tilde{B}^k=\tilde{\Omega}^{-1}\tilde{A}^k$, and $B^k=\Omega^{-1}A^k$.

Lemma A10: *If Assumption 5 is satisfied then*

$$\sqrt{m}\|\tilde{\Omega}-\Omega\|\stackrel{p}{\longrightarrow}0,\sqrt{m}\|\tilde{A}^k-A^k\|\stackrel{p}{\longrightarrow}0,\sqrt{m}\|\tilde{B}^k-B^k\|\stackrel{p}{\longrightarrow}0.$$

Proof: By standard arguments and Assumption 5,

$$E[m\|\tilde{\Omega}-\Omega\|^2]\leq CmE[\|g_i\|^4]/n\longrightarrow 0,E[m\|\tilde{A}^k-A^k\|^2]\leq CmE[\|g_i^k\|^2\|g_i\|^2]/n\longrightarrow 0,$$

so the first conclusion holds by M. Also, note that $A^kA^{k\prime}\leq CA^k\Omega^{-1}A^{k\prime}\leq CE[g_i^kg_i^{k\prime}]$ and $\lambda_{\min}(\tilde{\Omega})\geq C$ w.p.a.1. Also, $B^kB^{k\prime}\leq CA^kA^{k\prime}\leq CE[g_i^kg_i^{k\prime}]$. Then

$$\begin{aligned}\sqrt{m}\|\tilde{B}^k-B^k\| &\leq \sqrt{m}\|(\tilde{A}^k-A^k)\tilde{\Omega}^{-1}\|+\sqrt{m}\|B^k(\Omega-\tilde{\Omega})\tilde{\Omega}^{-1}\|\\ &\leq C\sqrt{m}\|\tilde{A}^k-A^k\|+C\sqrt{m}\|\tilde{\Omega}-\Omega\|\stackrel{p}{\longrightarrow}0.Q.E.D.\end{aligned}$$

Lemma A11: *If Assumption 5 is satisfied then,*

$$\mu_n^{-1}\partial\hat{Q}(\beta_0)/\partial\beta\stackrel{d}{\longrightarrow}N(0,H+\Lambda^*).$$

Proof: Let $\hat{g} = \hat{g}(\beta_0)$, $\tilde{G}^k = \partial\hat{g}(\beta_0)/\partial\beta_k$, $G^k = E[\partial g_i(\beta_0)/\partial\beta_k]$, and $\hat{U}^k = \tilde{G}^k - nG^k - \tilde{B}^{k\prime}\hat{g}$. Note that $n\partial\hat{\Omega}(\beta_0)^{-1}/\partial\beta_k = -\tilde{B}^k\tilde{\Omega}^{-1} - \tilde{\Omega}^{-1}\tilde{B}^{k\prime}$. Therefore,

$$\begin{aligned}\mu_n^{-1}\frac{\partial\hat{Q}}{\partial\beta_k}(\beta_0) &= n^{-1}\mu_n^{-1}(\tilde{G}^{k\prime}\tilde{\Omega}^{-1}\hat{g} - \hat{g}'\tilde{B}^k\tilde{\Omega}^{-1}\hat{g}) = \mu_n^{-1}(G^{k\prime}\tilde{\Omega}^{-1}\hat{g} + \hat{U}^{k\prime}\tilde{\Omega}^{-1}\hat{g}/n) \\ &= \mu_n^{-1}G^{k\prime}\tilde{\Omega}^{-1}\hat{g} + \sqrt{\kappa_n}\hat{U}^{k\prime}\tilde{\Omega}^{-1}\hat{g}/n\sqrt{m}.\end{aligned}$$

Let $\tilde{U}^k = \tilde{G}^k - nG^k - B^{k\prime}\hat{g}$. Note that $\|\hat{g}\|^2/n = O_p(m)$ by M and that $\lambda_{\max}(B^k\Omega^{-1}\Omega^{-1}B^{k\prime}) \leq C$. By Lemma A10 we have

$$\begin{aligned}&\sqrt{\kappa_n}|(\hat{U}^{k\prime}\tilde{\Omega}^{-1} - \tilde{U}^{k\prime}\Omega^{-1})\hat{g}/n\sqrt{m}| \\ \leq\ & C|\hat{g}'(\tilde{B}^k - B^k)\tilde{\Omega}^{-1}\hat{g}/n\sqrt{m}| + |\hat{g}'B^k\Omega^{-1}(\tilde{\Omega} - \Omega)\tilde{\Omega}^{-1}\hat{g}/n\sqrt{m}| \\ \leq\ & Cn^{-1}\|\hat{g}\|^2\|\tilde{B}^k - B^k\|/\sqrt{m} + Cn^{-1}\|\hat{g}'B^k\Omega^{-1}\|\|\hat{g}\|\|\tilde{\Omega} - \Omega\|/\sqrt{m} \leq O_p(m)o_p(1/\sqrt{m})/\sqrt{m} \xrightarrow{p} 0.\end{aligned}$$

Similarly we have $\mu_n^{-1}G^{k\prime}\tilde{\Omega}^{-1}\hat{g} - \mu_n^{-1}G^{k\prime}\Omega^{-1}\hat{g} \xrightarrow{p} 0$. Therefore, we have

$$\mu_n^{-1}\frac{\partial\hat{Q}}{\partial\beta_k}(\beta_0) = \mu_n^{-1}G^{k\prime}\Omega^{-1}\hat{g} + \sqrt{\kappa_n}\tilde{U}^{k\prime}\Omega^{-1}\hat{g}/n\sqrt{m} + o_p(1).$$

Let $\tilde{U} = [\tilde{U}_1, ..., \tilde{U}_p]$. Then stacking over k gives

$$\mu_n^{-1}\frac{\partial\hat{Q}}{\partial\beta}(\beta_0) = \mu_n^{-1}G'\Omega^{-1}\hat{g} + \sqrt{\kappa_n}\tilde{U}'\Omega^{-1}\hat{g}/n\sqrt{m} + o_p(1). \tag{7.2}$$

Next, let $U_i^k = g_i^k - G^k - B^{k\prime}g_i$ and $U_i = [U_i^1, ..., U_i^p]$, so that $\tilde{U} = \sum_{i=1}^n U_i$. For any vector λ with $\|\lambda\| = 1$ let $X_i = \mu_n^{-1}\lambda'G'\Omega^{-1}g_i$, $Y_i = \Omega^{-1/2}g_i$, $Z_i = \sqrt{\kappa_n}\Omega^{-1/2}U_i\lambda/n\sqrt{m}$, and $A = \lambda'H\lambda$. Note that $E[Z_i'Y_i] = 0$ and

$$nE[|Y_i'Z_i|^2] \leq CE[\left\|g_1'\Omega^{-1}U_1\right\|^2]/nm \leq (E[\|g_1\|^4] + E[\|g_{\beta 1}\|^4])/mn \longrightarrow 0.$$

Then $\sum_{i=1}^n Z_i'Y_i \xrightarrow{p} 0$ by M. Then by eq. (7.2),

$$\mu_n^{-1}\lambda'\frac{\partial\hat{Q}(\beta_0)}{\partial\beta} = \sum_{i=1}^n X_i + \sum_{i,j=1}^n Z_i'Y_j + o_p(1) = \sum_{i=1}^n X_i + \sum_{i\neq j} Z_i'Y_j + o_p(1).$$

Now apply Lemma A2. Note that $\Sigma_{YY} = I_m$ and $\Sigma_{ZY} = 0$, so that $\Psi = \Sigma_{ZZ} = \kappa_n\Omega^{-1/2}E[U_i\lambda\lambda'U_i]\Omega^{-1/2}/n^2m$. By Assumption 1 and the hypothesis of Theorem 3, we have

$$\begin{aligned}nE[X_i^2] &= (n/\mu_n^2)\lambda'G'\Omega^{-1}G\lambda \longrightarrow \lambda'H\lambda = A, \\ n^2tr(\Psi) &= n^2E[Z_i'Z_i] = \kappa_n\lambda'E[U_i'\Omega^{-1}U_i]\lambda/m \longrightarrow \kappa\lambda'\Lambda^*\lambda.\end{aligned}$$

Also, note that $E[U_iU_i'] \leq C\sum_{j=1}^{p} E[g_i^j g_i^{j\prime}]$ so that

$$\lambda_{\max}(\Sigma_{ZZ}) \leq C\lambda_{\max}(\Omega^{-1})/mn^2 \leq C/mn^2.$$

It then follows that

$$mn^4\lambda_{\max}(\Sigma_{YY})^2\lambda_{\max}(\Sigma_{ZZ})^2 \leq Cmn^4/(mn^2)^2 \leq C/m \longrightarrow 0.$$

In addition, by Assumption 5 and $\|\mu_n^{-1}\sqrt{n}G'\Omega^{-1}\| \leq C$ we have for $g_{\beta i} = \partial g_i(\beta_0)/\partial\beta$,

$$\begin{aligned} nE[|X_i|^4] &\leq nE[\left\|\mu_n^{-1}\sqrt{n}G'\Omega^{-1}g_i\right\|^4]/n^2 \leq CE[\|g_i\|^4]/n \longrightarrow 0, \\ n^3E\left[|Y_1'Z_2|^4\right] &\leq CE[\left\|g_1'\Omega^{-1}U_2\right\|^4]/nm^2 \leq CE[\|g_1\|^4]E[\|U_1\|^4]/nm^2 \\ &\leq C(E[\|g_1\|^4]/m\sqrt{n})(E[\|g_1\|^4] + E[\|g_{\beta 1}\|^4])/(m\sqrt{n}) \longrightarrow 0. \end{aligned}$$

The conclusion then follows by the conclusion of Lemma A2 and the Cramer-Wold device. Q.E.D.

Lemma A12: *If Assumptions 2 and 4-6 are satisfied then for any* $\bar{\beta} \xrightarrow{p} \beta_0$, $\mu_n^{-2}\partial^2\hat{Q}(\bar{\beta})/\partial\beta\partial\beta' \xrightarrow{p} H$.

Proof: For notational convenience, let $\tilde{g}(\beta) = \hat{g}(\beta)/\mu_n\sqrt{n}$, drop the β argument, replace $\hat{Q}$ by $\mu_n^{-2}\hat{Q}$, $\hat{\Omega}$ by $\hat{\Omega}/n$, 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 we have

$$\begin{aligned} \hat{Q}_k &= \tilde{g}_k'\hat{\Omega}^{-1}\tilde{g} - \frac{1}{2}\tilde{g}'\hat{\Omega}^{-1}\hat{\Omega}_k\hat{\Omega}^{-1}\tilde{g} \qquad (7.3) \\ \hat{Q}_{k,\ell} &= \tilde{g}_{k,\ell}'\hat{\Omega}^{-1}\tilde{g} + \tilde{g}_k'\hat{\Omega}^{-1}\tilde{g}_\ell' - \tilde{g}_k'\hat{\Omega}^{-1}\hat{\Omega}_\ell\hat{\Omega}^{-1}\tilde{g} - \tilde{g}_\ell'\hat{\Omega}^{-1}\hat{\Omega}_k\hat{\Omega}^{-1}\tilde{g} \\ &\quad + \tilde{g}'\hat{\Omega}^{-1}\hat{\Omega}_\ell\hat{\Omega}^{-1}\hat{\Omega}_k\hat{\Omega}^{-1}\tilde{g} - \frac{1}{2}\tilde{g}'\hat{\Omega}^{-1}\hat{\Omega}_{k,\ell}\hat{\Omega}^{-1}\tilde{g}. \end{aligned}$$

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 5 each of Ω^{-2}, Ω_k^2, and $\Omega_{k\ell}^2$ have largest eigenvalue bounded above by a constant. Then by Assumption 7 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 term in $\hat{Q}_{k,\ell}$, from eq. (7.3), 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 Assumption 8, $\tilde{Q}_{k,\ell}(\beta)$ is stochastically equicontinuous, so by $\bar{\beta} \xrightarrow{p} \beta_0$, the previous equation, and T,

$$\left\| \hat{Q}_{k,\ell}(\bar{\beta}) - \tilde{Q}_{k,\ell}(\beta_0) \right\| \leq \left\| \hat{Q}_{k,\ell}(\bar{\beta}) - \tilde{Q}_{k,\ell}(\bar{\beta}) \right\| + \left\| \tilde{Q}_{k,\ell}(\bar{\beta}) - \tilde{Q}_{k,\ell}(\beta_0) \right\| \xrightarrow{p} 0. \tag{7.4}$$

It therefore suffices to show that $\tilde{Q}_{k,\ell} \xrightarrow{p} H_{k\ell}$, where we now evaluate at β_0, i.e. $\tilde{Q}_{k,\ell} = \partial^2 \tilde{Q}(\beta_0)/\partial\beta_k\partial\beta_\ell$. Next, note that $\Omega_k = \Omega^k + \Omega^{k\prime}$ for $\Omega^k = E[g_i g_i^{k\prime}]$. Then by standard properties of the trace of a matrix,

$$\begin{aligned}
tr(\Omega^{-1}\Omega_k\Omega^{-1}\Omega^\ell) &= tr(\Omega^{-1}\Omega^\ell\Omega^{-1}\Omega^{k\prime}) + tr(\Omega^{-1}\Omega^\ell\Omega^{-1}\Omega^k) \\
&= tr(\Omega^{-1}\Omega^\ell\Omega^{-1}\Omega^{k\prime}) + tr(\Omega^{k\prime}\Omega^{-1}\Omega^{\ell\prime}\Omega^{-1}) \\
&= tr(\Omega^{-1}\Omega^\ell\Omega^{-1}\Omega^{k\prime}) + tr(\Omega^{\ell\prime}\Omega^{-1}\Omega^{k\prime}\Omega^{-1}) = tr(\Omega^{-1}\Omega_\ell\Omega^{-1}\Omega^{k\prime}).
\end{aligned}$$

Therefore, it follows by Lemma A1 that

$$\begin{aligned}
& -\tilde{g}_k'\Omega^{-1}\Omega_\ell\Omega^{-1}\tilde{g} - \tilde{g}_\ell'\Omega^{-1}\Omega_k\Omega^{-1}\tilde{g} + \tilde{g}'\Omega^{-1}\Omega_\ell\Omega^{-1}\Omega_k\Omega^{-1}\tilde{g} \\
= & -tr(\Omega^{-1}\Omega_\ell\Omega^{-1}\Omega^k)/\mu_n^2 - tr(\Omega^{-1}\Omega_k\Omega^{-1}\Omega^\ell)/\mu_n^2 \\
& + tr(\Omega^{-1}\Omega_\ell\Omega^{-1}\Omega_k\Omega^{-1}\Omega)/\mu_n^2 + o_p(1) \xrightarrow{p} 0.
\end{aligned}$$

It also follows similarly that for $\Omega^{k\ell} = E[g_i g_i^{k\ell\prime}]$ and $\Omega^{k,\ell} = E[g_i^k g_i^{\ell\prime}]$,

$$tr(\Omega^{-1}(\Omega^{k\ell} + \Omega^{k,\ell})) = \frac{1}{2} tr(\Omega^{-1}\Omega_{k,\ell}).$$

Then by Lemma A1, for $\bar{g} = \mu_n^{-1}\sqrt{n}E[g_i(\beta)]$,

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

By Assumption 1, $\bar{g}_k'\Omega^{-1}\bar{g}_\ell \longrightarrow H_{k\ell}$. It then follows by T that $\tilde{Q}_{k\ell} \xrightarrow{p} H_{k\ell}$. The conclusion for the CUE then follows by T and eq. (7.4). Q.E.D.

LEMMA A13: *If Assumptions 4-6 are satisfied then* $\hat{D}(\hat{\beta})'\hat{\Omega}^{-1}\hat{D}(\hat{\beta})/\mu_n^2 \xrightarrow{p} H + \Lambda^*$.

Proof: For the CUE $\rho(v) = v - v^2/2$ so that $\hat{\lambda}(\beta) = \hat{\Omega}(\beta)^{-1}\hat{g}(\beta)$ and $\hat{\rho}_{1i}(\beta) = 1 - \hat{g}(\beta)'\hat{\Omega}(\beta)^{-1}g_i(\beta)$. Let $\tilde{A}^j(\beta)$, $A^j(\beta)$, $\tilde{D}^j(\beta)$, and $\tilde{D}(\beta)$ be as defined in connection with Assumptions 5 and 6. For the j^{th} unit vector e_j we have

$$\hat{D}(\beta)e_j = \partial\hat{g}(\beta)/\partial\beta_j - \hat{A}^j(\beta)\hat{\Omega}(\beta)^{-1}\hat{g}(\beta).$$

By Assumption 5, $\sup_{\beta\in B}\left\|\hat{A}^j(\beta)/n - A^j(\beta)\right\| \xrightarrow{p} 0$. Then it follows similarly to the proof of Lemma A10 that $\sup_{\beta\in B}\left\|\hat{A}^j(\beta)\hat{\Omega}(\beta)^{-1} - A^j(\beta)\Omega(\beta)^{-1}\right\| \xrightarrow{p} 0$, so that by Assumption 6 i) and CS,

$$\sup_{\beta\in N}\left\|\hat{D}^j(\beta)/(\mu_n\sqrt{n}) - \tilde{D}^j(\beta)\right\| \le \sup_{\beta\in B}\left\|\hat{A}^j(\beta)\hat{\Omega}(\beta)^{-1} - A^j(\beta)\Omega(\beta)^{-1}\right\| \sup_{\beta\in N}\left\|\hat{g}(\beta)/\mu_n\sqrt{n}\right\| \xrightarrow{p} 0.$$

By Assumption 6 we also have $\sup_{\beta\in N}\left\|\tilde{D}(\beta)\right\| = O_p(1)$ so that by T and CS,

$$\begin{aligned} & \left\|\hat{D}(\hat{\beta})'\hat{\Omega}^{-1}\hat{D}(\hat{\beta})/n\mu_n^2 - \tilde{D}(\hat{\beta})'\Omega(\hat{\beta})^{-1}\tilde{D}(\hat{\beta})\right\| \qquad (7.5)\\ \le & \left\|\hat{D}(\hat{\beta})'\hat{\Omega}^{-1}\hat{D}(\hat{\beta})/n\mu_n^2 - \tilde{D}(\hat{\beta})'\hat{\Omega}^{-1}\tilde{D}(\hat{\beta})\right\| + \left\|\tilde{D}(\hat{\beta})'(\hat{\Omega}^{-1} - \Omega(\hat{\beta})^{-1})\tilde{D}(\hat{\beta})\right\| \xrightarrow{p} 0. \end{aligned}$$

Also, by Assumption 6, $\tilde{D}(\hat{\beta})'\Omega(\hat{\beta})^{-1}\tilde{D}(\hat{\beta}) - \tilde{D}(\beta_0)'\Omega^{-1}\tilde{D}(\beta_0) \xrightarrow{p} 0$. Note that $\tilde{D}(\beta_0) = \sum_{i=1}^n U_i/\mu_n\sqrt{n}$ so that, in the notation of Lemma A1,

$$\tilde{D}^j(\beta_0)'\Omega^{-1}\tilde{D}^k(\beta_0)e_k = n\bar{Y}'A\bar{Z}/m$$

for $A = \Omega^{-1}$, $Y_i = \sqrt{m}U_i^j/\mu_n$, and $Z_i = \sqrt{m}U_i^k/\mu_n$. Also note that

$$tr(A\Sigma_{YZ'})/m = E[U_i^{j\prime}\Omega^{-1}U_i^k]/\mu_n^2 \longrightarrow \Lambda_{jk}^*, n\mu_Y'A\mu_Z/m = nG^{j\prime}\Omega^{-1}G^k/\mu_n^2 \longrightarrow H_{jk}.$$

Therefore, it follows from the conclusion of Lemma A1 that

$$\tilde{D}^j(\beta_0)'\Omega^{-1}\tilde{D}^k(\beta_0) = \Lambda_{jk}^* + H_{jk} + o_p(1).$$

The conclusion for the CUE 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 4. Note that $g(w,\beta) = z(y - x'\beta)$ is twice continuously differentiable by inspection. Also, by Lemma A4 and the specified rate condition,

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

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 that Assumption 4 is satisfied.

It follows by Lemma A7 that Assumption 5 i) is satisfied. Assumption 5 ii) holds by $E[(z_i'z_i)^2]/n \longrightarrow 0$. Assumption 5 iii) holds by Lemma A8.

The proof of Assumption 6 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 $\hat{D}(\bar{\beta})'\hat{\Omega}(\bar{\beta})^{-1}\hat{D}(\bar{\beta}) \xrightarrow{p} H + \kappa\Lambda^*$. Also, Lemma A11 gives $\mu_n^{-1}\partial\hat{Q}(\bar{\beta})/\partial\beta \xrightarrow{d} N(0, H + \kappa\Lambda^*)$, so the conclusion follows in the usual way. Q.E.D.

8 Appendix B: Asymptotic Theory for GEL.

We give here results for GEL. For the consistency results we make use of the following condition:

Assumption 7: *i) $\rho(v)$ is three times continuously differentiable;ii) there is $\gamma > 2$ such that $n^{1/\gamma}(E[\sup_{\beta\in B}\|g_i(\beta)\|^\gamma])^{1/\gamma}\mu_n/\sqrt{n} \longrightarrow 0$.*

The first two results are consistency in the general case and in the linear model.

THEOREM 6: *If Assumptions 1, 2, and 7 are satisfied then* $\hat{\beta} \xrightarrow{p} \beta_0$ *for any GEL estimator* $\hat{\beta}$.

THEOREM 7: *If Assumptions 1 i), 1 ii), 3, and 7 i) are satisfied and for* $\gamma > 2$ *we have* $E[|\varepsilon_i|^\gamma|z_i] \le C$, $E[\|\eta_i\|^\gamma|z_i] \le C$, $n^{1/\gamma}(E[\|z_i\|^\gamma])^{1/\gamma}\mu_n/\sqrt{n} \longrightarrow 0$ *then* $\hat{\beta} \xrightarrow{p} \beta_0$ *for any GEL estimator* $\hat{\beta}$.

For the asymptotic normality results we make use of the following condition:

ASSUMPTION 8: *For* $b_i = \max_{\beta \in B} \max\{\|g_i(\beta)\|, \|\partial g_i(\beta)/\partial\beta\|\}$ *there is* $\gamma > 2$ *such that* $n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma}\mu_n/\sqrt{n} \longrightarrow 0$, *for* $d_i = \max_{\beta\in B}\max_j\{\|g_i(\beta)\|, \|\partial g_i(\beta)/\partial\beta\|, \|\partial^2 g_i(\beta)/\partial\beta\partial\beta_j\|\}$, $E[d_i^3]\sqrt{m/n} \longrightarrow 0$.

We now give the asymptotic normality results for the general case and the linear model, respectively.

THEOREM 8: *If Assumptions 1, 2, and 4-8 are satisfied, and* $E[U_i'\Omega^{-1}U_i]/\mu_n^2 \longrightarrow \Lambda^*$ *then for* $V = H^{-1} + H^{-1}\Lambda^* H^{-1}$,

$$\mu_n(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V), \mu_n^2\hat{V} \xrightarrow{p} V.$$

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

THEOREM 9: *If Assumptions 1 i), ii), 3, 7 i) are satisfied,and for* $\gamma > 2$ *we have* $E[|\varepsilon_i|^\gamma|z_i] \le C$, $E[\|\eta_i\|^\gamma|z_i] \le C$, $n^{1/\gamma}(E[\|z_i\|^\gamma])^{1/\gamma}\mu_n/\sqrt{n} \longrightarrow 0$, $E[U_i'\Omega^{-1}U_i]/\mu_n^2 \longrightarrow \Lambda^*$, *and* $E[(z_i'z_i)^2]m/n \longrightarrow 0$ *then*

$$\mu_n(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V), \mu_n^2\hat{V} \xrightarrow{p} V.$$

Before proving these results we first give two additional Lemmas.

LEMMA A14: *If Assumptions 2 and 7 are satisfied then there is* $\dot{C} > 0$ *such that w.p.a.1 for all* $\beta \in B$ $\hat{\lambda}(\beta) = \arg\max_{\lambda\in\hat{\Lambda}_n(\beta)} \sum_{i=1}^n \rho(\lambda' g_i(\beta))$ *exists,* $\left\|\hat{\lambda}(\beta)\right\| \le C\|\hat{g}(\beta)/n\|$, *and* $\sup_{\beta\in B}\left\|\hat{\lambda}(\beta)\right\| = O_p(\mu_n/\sqrt{n})$.

Proof: By $\lambda_{\max}(\Omega(\beta)) \leq C$ we have $\|\hat{g}(\beta)\|^2 / n\mu_n^2 \leq C\check{Q}(\beta)$ and by Lemma A3, $\sup_{\beta\in B} \check{Q}(\beta) = O_p(1)$, so by T, $\sup_{\beta\in B} \|\hat{g}(\beta)\|^2 = O_p(n\mu_n^2)$. Let $b_i = \sup_{\beta\in B} \|g_i(\beta)\|$. By a standard argument, $\max_{i\leq n} b_i = O_p(n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma})$. By Assumption 7 there exists δ_n with $\delta_n = o(n^{-\frac{1}{\gamma}}(E[b_i^\gamma])^{-1/\gamma})$ and $\mu_n/\sqrt{n} = o(\delta_n)$. Let $\Lambda_n = \{\lambda : \|\lambda\| \leq \delta_n\}$ and $\hat{S}(\beta,\lambda) = \sum_{i=1}^n \rho\left(\lambda' g_i(\beta)\right)$. Note that

$$\max_{\lambda\in\Lambda_n,\beta\in B,i\leq n} |\lambda' g_i(\beta)| \leq \delta_n \max_{i\leq n} b_i = O_p(\delta_n n^{\frac{1}{\gamma}} (E[b_i^\gamma])^{1/\gamma}) \xrightarrow{p} 0,$$

so that $\Lambda_n \subseteq \hat{\Lambda}_n(\beta)$ for all β w.p.a.1. Similarly, by continuity of $\rho_2(v)$ and $\rho_2(0) = -1$, $\sup_{\lambda\in\Lambda_n,\beta\in B,i\leq n} \rho_2(\lambda' g_i(\beta)) \leq -1/2$ w.p.a.1. Since $\hat{S}(\beta,\lambda)$ is concave in λ, $\tilde{\lambda}(\beta) = \arg\max_{\lambda\in\Lambda_n} \hat{S}(\beta,\lambda)$ exists (for all β). Also, w.p.a.1,

$$\lambda_{\min}\left(\hat{\Omega}(\beta)/n\right) \geq \lambda_{\min}(\Omega(\beta)) - \left\|\hat{\Omega}(\beta)/n - \Omega(\beta)\right\| \geq C - \sup_{\beta} \left\|\hat{\Omega}(\beta)/n - \Omega(\beta)\right\| \geq C.$$

Then by a Taylor expansion in λ around zero with Lagrange remainder, for $\rho_j(v) = \partial^j \rho(v)/\partial v^j$

$$\begin{aligned} 0 &= \hat{S}(\beta,0) \leq \hat{S}(\beta,\tilde{\lambda}(\beta)) = \tilde{\lambda}(\beta)'\hat{g}(\beta) + \tilde{\lambda}(\beta)'\left[\sum_{i=1}^n \rho_2(\bar{\lambda}(\beta)' g_i(\beta)) g_i(\beta) g_i(\beta)'\right]\tilde{\lambda}(\beta) \\ &\leq \tilde{\lambda}(\beta)'\hat{g}(\beta) - \tilde{\lambda}(\beta)'\hat{\Omega}(\beta)\tilde{\lambda}(\beta)/2 \leq \tilde{\lambda}(\beta)'\hat{g}(\beta) - Cn\left\|\tilde{\lambda}(\beta)\right\|^2, \end{aligned}$$

so that

$$Cn\left\|\tilde{\lambda}(\beta)\right\|^2 \leq \tilde{\lambda}(\beta)'\hat{g}(\beta) \leq |\tilde{\lambda}(\beta)'\hat{g}(\beta)| \leq \left\|\tilde{\lambda}(\beta)\right\| \|\hat{g}(\beta)\|.$$

Dividing through by $Cn\left\|\tilde{\lambda}(\beta)\right\|$ we find that w.p.a. 1 for all β

$$C\left\|\tilde{\lambda}(\beta)\right\| \leq \|\hat{g}(\beta)/n\| \leq \sup_{\beta\in B} \|\hat{g}(\beta)\|/n = O_p\left(\mu_n/\sqrt{n}\right).$$

By $\mu_n/\sqrt{n} = o(\delta_n)$ it follows that $\tilde{\lambda}(\beta) \in \operatorname{int}\Lambda_n$ for all β w.p.a.1. Then, since a local maximum of the convex function $\hat{S}(\beta,\lambda)$ over an open convex set $int(\Lambda_n)$ is a global maximum, it follows that $\tilde{\lambda}(\beta) = \hat{\lambda}(\beta)$ w.p.a.1, giving the conclusion. Q.E.D.

LEMMA A15: *If Assumptions 1, 2, and 7 are satisfied then for* $\hat{Q}^*(\beta) = \hat{g}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{g}(\beta)/2$, $\sup_{\beta\in B} |\hat{Q}(\beta) - \hat{Q}^*(\beta)| = o_p(\mu_n^2)$.

Proof: Expanding around $\lambda = 0$ gives

$$\begin{aligned}\hat{Q}(\beta) &= [\hat{g}(\beta)'\hat{\lambda}(\beta) - \frac{1}{2}\hat{\lambda}(\beta)'\hat{\Omega}(\beta)\hat{\lambda}(\beta) + \hat{r}(\beta)], \\ \hat{r}(\beta) &= \frac{1}{6}\sum_i \rho_3\left(\bar{\lambda}(\beta)'\hat{g}(\beta)\right)\left[g_i(\beta)'\hat{\lambda}(\beta)\right]^3,\end{aligned}$$

where $\bar{\lambda}(\beta)$ lies on the line joining $\hat{\lambda}(\beta)$ and 0. Similarly to previous arguments there is C with $\lambda_{\max}(\hat{\Omega}(\beta)/n) \leq C$ for all $\beta \in B$ w.p.a.1, $\sup_{\beta\in B, i\leq n}|\hat{\lambda}(\beta)'g_i(\beta)| \xrightarrow{p} 0$, and $\sup_{\beta\in B, i\leq n}|\bar{\lambda}(\beta)'g_i(\beta)| \xrightarrow{p} 0$. Therefore, by $\rho_3(v)$ bounded in a neighborhood of $v = 0$, It then follows that

$$|\hat{r}(\beta)| \leq C\sum_{i=1}^n \left|g_i(\beta)'\hat{\lambda}(\beta)\right|^3 \leq n \sup_{\beta\in B, i\leq n}|\hat{\lambda}(\beta)'g_i(\beta)|\hat{\lambda}(\beta)'[\hat{\Omega}(\beta)/n]\hat{\lambda}(\beta) = o_p(\mu_n^2).$$

Also, as shown above $\hat{\lambda}(\beta) \in \Lambda_n \subseteq \hat{\Lambda}(\beta)$, so that $\hat{\lambda}(\beta) \in \operatorname{int}\hat{\Lambda}(\beta)$ for all β w.p.a.1.. Thus, w.p.a.1 $\hat{\lambda}(\beta)$ satisfies the first-order conditions

$$0 = \sum_{i=1}^n \rho_1\left(\hat{\lambda}(\beta)'g_i(\beta)\right)g_i(\beta)$$

Note that $\max_{i\leq n} b_i\mu_n/\sqrt{n} \xrightarrow{p} 0$, i.e. $\max_{i\leq n} b_i = o_p(\sqrt{n}/\mu_n)$. Then expand around $\lambda = 0$ to obtain

$$\begin{aligned}0 &= \hat{g}(\beta) - \hat{\Omega}(\beta)\hat{\lambda}(\beta) + \hat{R}(\beta), \hat{R}(\beta) = \sum_{i=1}^n \rho_3(\bar{\lambda}(\beta)'g_i(\beta))\left[g_i(\beta)'\hat{\lambda}(\beta)\right]^2 g_i(\beta), \\ \left\|\hat{R}(\beta)\right\| &\leq C\max_{i\leq n} b_i\hat{\lambda}(\beta)'\hat{\Omega}(\beta)\hat{\lambda}(\beta) \leq o_p(\sqrt{n}/\mu_n)nO_p(\mu_n^2/n) = o_p(\sqrt{n}\mu_n).\end{aligned}$$

It follows from the last inequality and $\lambda_{\min}(\hat{\Omega}(\beta)/n) \geq C$ w.p.a.1 that

$$\hat{R}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{R}(\beta) \leq C\sup_{\beta\in B}\left\|\hat{R}(\beta)\right\|^2/n = o_p(\mu_n^2).$$

Then solving for $\hat{\lambda}(\beta) = \hat{\Omega}(\beta)^{-1}[\hat{g}(\beta) + \hat{R}(\beta)]$ and plugging into the expansion for $\hat{Q}(\beta)$ gives

$$\begin{aligned}\hat{Q}(\beta) &= \hat{g}(\beta)'\hat{\Omega}(\beta)^{-1}[\hat{g}(\beta) + \hat{R}(\beta)] - [\hat{g}(\beta) + \hat{R}(\beta)]'\hat{\Omega}(\beta)^{-1}[\hat{g}(\beta) + \hat{R}(\beta)]/2 + \hat{r}(\beta) \\ &= \hat{Q}^*(\beta) - \hat{R}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{R}(\beta)/2 + \hat{r}(\beta).\end{aligned}$$

It then follows by T that

$$\sup_{\beta\in B}|\hat{Q}(\beta)-\hat{Q}^*(\beta)|\leq\sup_{\beta\in B}\hat{R}(\beta)'\hat{\Omega}(\beta)^{-1}\hat{R}(\beta)/2+\sup_{\beta\in B}|\hat{r}(\beta)|=o_p(\mu_n^2).Q.E.D.$$

Proof of Theorem 6: The conclusion of Lemma A14 shows that the difference of the CUE and GEL objective functions, divided by μ_n^2, converges to zero uniformly in β. The remainder of the proof then follows from the proof for the CUE.

Proof of Theorem 7: All of the hypotheses of Theorem 6 are satisfied, so the conclusion follows by Theorem 6.

Proof of Theorem 8: First, we show that for GEL,

$$\mu_n^{-1}\partial\hat{Q}(\beta_0)/\partial\beta\xrightarrow{d}N(0,H+\kappa\Lambda^*).$$

Let $\tilde{G}_k=\sum g_i^k/n$ and $\tilde{A}^k=\sum_i g_i g_i^{k\prime}/n$, as before. Also, let $\hat{\lambda}=\hat{\lambda}(\beta_0)$. By the envelope theorem and an expansion,

$$\frac{\partial\hat{Q}}{\partial\beta_j}(\beta_0)=\sum_i\hat{\lambda}'g_i^j\rho_1(\hat{\lambda}'g_i)=n\tilde{G}^{j\prime}\hat{\lambda}-n\hat{\lambda}'\tilde{A}^j\hat{\lambda}+\hat{r},\hat{r}=\sum_i\hat{\lambda}'g_i^j\rho_3\left(\bar{\lambda}'g_i\right)\left(\bar{\lambda}'g_i\right)^2/2,$$

where $\left\|\bar{\lambda}\right\|\leq\left\|\hat{\lambda}\right\|$. By M, $\|\hat{g}\|=O_p\left(\sqrt{m}\right)$, so by Lemma A14, $\left\|\hat{\lambda}\right\|=O_p\left(\sqrt{m/n}\right)$. It follows as above that $\max_{i\leq n}\left|\hat{\lambda}'g_i\right|\xrightarrow{p}0$, so that $\max_{i\leq n}\left|\rho_3\left(\bar{\lambda}'g_i\right)\right|\leq C$ w.p.a. 1. Let $G_i=\partial g_i(\beta_0)/\partial\beta$ and $b_i=\max\{\|G_i\|,\|g_i\|\}$. As above, $\hat{b}=\max_{i\leq n}b_i=O_p(n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma})$. It then follows that, by $\lambda_{\max}\left(\tilde{\Omega}\right)=O_p(1)$ and Assumption 8,

$$\mu_n^{-1}|\hat{r}|\leq\mu_n^{-1}C\left\|\hat{\lambda}\right\|\hat{b}n\bar{\lambda}'\tilde{\Omega}\bar{\lambda}=O_p(\mu_n^{-1}m^{3/2}n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma}/\sqrt{n})\xrightarrow{p}0.$$

It also follows similarly to previous arguments that w.p.a.1 $\hat{\lambda}$ satisfies the first-order conditions

$$\sum_i\rho_1\left(\hat{\lambda}'g_i\right)g_i=0.$$

Expanding and solving give

$$\hat{\lambda}=\tilde{\Omega}^{-1}\hat{g}/\sqrt{n}+\hat{R},\hat{R}=\tilde{\Omega}^{-1}\textstyle\sum_i\rho_3\left(\bar{\lambda}'g_i\right)g_i\left(\bar{\lambda}'g_i\right)^2/n.$$

As previously,

$$\left\|\hat{R}\right\| \leq C\hat{b}\bar{\lambda}'\tilde{\Omega}\bar{\lambda} = O_p(n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma}m/n).$$

Now, plug this expansion for $\hat{\lambda}$ back in the expression for $\partial\hat{Q}(\beta_0)/\partial\beta_j$ to obtain

$$\begin{aligned}\mu_n^{-1}\frac{\partial\hat{Q}}{\partial\beta_j}(\beta_0) &= \mu_n^{-1}\sqrt{n}\tilde{G}^{j\prime}\tilde{\Omega}^{-1}\hat{g} - \mu_n^{-1}\hat{g}'\tilde{\Omega}^{-1}\tilde{A}^j\tilde{\Omega}^{-1}\hat{g} + \hat{r}/\mu_n \\ &+\mu_n^{-1}n\tilde{G}^{j\prime}\hat{R} - \mu_n^{-1}n\hat{R}'\tilde{A}^j\hat{R} - \mu_n^{-1}\sqrt{n}\hat{R}'\left(\tilde{A}^j + \tilde{A}^{j\prime}\right)\tilde{\Omega}^{-1}\hat{g}.\end{aligned}$$

Note that by Assumption 5, $E[\left\|g_i^j\right\|^2] = tr(E[g_i^j g_i^{j\prime}]) \leq m\lambda_{\max}(E[g_i^j g_i^{j\prime}]) \leq Cm$. Therefore, $\left\|\tilde{G}^j - G^j\right\| = O_p\left(\left(E\left[\left\|g_i^j\right\|^2\right]\right)^{1/2}/\sqrt{n}\right) = O_p(\sqrt{m/n}) = O_p(\mu_n/\sqrt{n})$. We also have $\|G^j\| = O\left(\mu_n/\sqrt{n}\right)$. Therefore,

$$\left|\mu_n^{-1}n\tilde{G}^{j\prime}\hat{R}\right| \leq \mu_n^{-1}n\left(\left\|\tilde{G}^j - G^j\right\| + \left\|G^j\right\|\right)\left\|\hat{R}\right\| = O_p(n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma}m/\sqrt{n}) \xrightarrow{p} 0.$$

It also follows that $\lambda_{\max}\left(\tilde{A}^j + \tilde{A}^{j\prime}\right) \leq C$ w.p.a.1, so that

$$\left|\mu_n^{-1}n\hat{R}'\tilde{A}^j\hat{R}\right| \leq \mu_n^{-1}n\left\|\hat{R}\right\|^2 = O_p\left(\{\mu_n^{-1/2}n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma}m/\sqrt{n}\}^2\right) \xrightarrow{p} 0.$$

We also have $\left\|\tilde{\Omega}^{-1}\hat{g}\right\| = O_p\left(\sqrt{m}\right)$, so that by $\sqrt{m} = O(\mu_n)$,

$$\left|\mu_n^{-1}\sqrt{n}\hat{R}\left(\tilde{A}_j + \tilde{A}_j'\right)\tilde{\Omega}^{-1}\hat{g}\right| = O_p\left(n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma}m/\sqrt{n}\right) \xrightarrow{p} 0.$$

By T it now follows that

$$\mu_n^{-1}\frac{\partial\hat{Q}}{\partial\beta}(\beta_0) = \mu_n^{-1}\sqrt{n}\tilde{G}^{j\prime}\tilde{\Omega}^{-1}\hat{g} - \mu_n^{-1}\hat{g}'\tilde{\Omega}^{-1}\tilde{A}^j\tilde{\Omega}^{-1}\hat{g} + o_p(1)$$

The expression between the equality sign and $o_p(1)$ is equal to $\mu_n^{-1}\partial\hat{Q}(\beta_0)/\partial\beta$ for the CUE, so the conclusion for a general GEL estimator follows by the conclusion for the CUE.

Next we show that for any $\bar{\beta} \xrightarrow{p} \beta_0$, $\mu_n^{-2}\partial^2\hat{Q}(\bar{\beta})/\partial\beta\partial\beta' \xrightarrow{p} H$. It follows as in Lemma A14 of Donald, Imbens, and Newey (2003) that $\left\|\hat{g}\left(\hat{\beta}\right)/n\right\| = O_p\left(\sqrt{m/n}\right)$. Then by Lemma A14, we have $\left\|\hat{\lambda}\left(\hat{\beta}\right)\right\| = O_p\left(\sqrt{m/n}\right)$. Also, as in the proof of Lemma A14, w.p.a. 1 $\lambda\left(\beta\right) \in \text{int}\left(\hat{\Lambda}_n\left(\beta\right)\right)$ for all $\beta \in B$. Therefore w.p.a.1, for all $\beta \in B$, $\hat{\lambda}\left(\beta\right)$ solves

$$\sum_{i=1}^n \rho_1\left(\hat{\lambda}(\beta)'g_i(\beta)\right)g_i(\beta) = 0.$$

As above we have

$$\sup_{\beta\in B}\max_{i\leq n}\left|\hat{\lambda}(\beta)'g_i(\beta)\right| \xrightarrow{p} 0,$$

so that for all $\left\|\bar{\lambda}\right\| \leq \hat{\lambda}(\beta)$ and all β,

$$-\sum_i \rho_2\left(\bar{\lambda}'g_i(\beta)\right)g_i(\beta)g_i(\beta)'/n \geq C\sum_i g_i(\beta)g_i(\beta)'/n = C\hat{\Omega}(\beta),$$

and so w.p.a.1 the matrix preceding the inequality is nonsingular. It then follows by the implicit function theorem that $\hat{\lambda}(\beta)$ is differentiable and

$$\begin{aligned}\frac{\partial\hat{\lambda}}{\partial\beta_j}(\beta) &= \left[-\sum_i \rho_2\left(\hat{\lambda}(\beta)'g_i(\beta)\right)g_i(\beta)g_i(\beta)'\right]^{-1}\times \\ &\quad \sum_i \rho_1\left(\hat{\lambda}(\beta)'g_i(\beta)\right)\frac{\partial g_i(\beta)}{\partial\beta_j} + \sum_i \rho_2\left(\hat{\lambda}(\beta)'g_i(\beta)\right)g_i(\beta)\frac{\partial g_i(\beta)'}{\partial\beta_j}\hat{\lambda}\end{aligned}$$

To simplify derivations we will henceforth evaluate at $\hat{\beta}$, unless otherwise notified, and drop the β argument. Thus, in what follows $g_i = g_i\left(\hat{\beta}\right)$, $\hat{\lambda} = \hat{\lambda}\left(\hat{\beta}\right)$, etc. Also, let superscripts denote derivatives, e.g. so that $g_i^j = \partial g_i(\hat{\beta})/\partial\beta_j$. Then evaluating the previous equation at $\hat{\beta}$ and letting $\tilde{\Omega} = -\sum_{i=1}^n \rho_2\left(\hat{\lambda}'g_i\right)g_ig_i'/n$, we have

$$\hat{\lambda}^j = \tilde{\Omega}^{-1}\left[\sum_i\left\{\rho_1\left(\hat{\lambda}'g_i\right)g_i^j + \rho_2\left(\hat{\lambda}'g_i\right)g_ig_i^{j\prime}\hat{\lambda}\right\}/n\right]$$

Let $\tilde{\Omega}^j = -\sum_i \rho_2\left(\bar{\lambda}'g_i\right)g_ig_i^{j\prime}/n$, where $\bar{\lambda}$ is somewhere on the line between $\hat{\lambda}$ and zero, and where, for notational simplicity, we do not distinguish different such $\bar{\lambda}$. Then an expansion gives

$$\hat{\lambda}^j = \tilde{\Omega}^{-1}\left[g^j - \left(\tilde{\Omega}^j + \tilde{\Omega}^{j\prime}\right)\hat{\lambda}\right].$$

Next, by the envelope theorem it follows that

$$\frac{\mu_n^2}{n}\frac{\partial\hat{Q}}{\partial\beta_j}\left(\hat{\beta}\right) = \sum_i \rho_1\left(\hat{\lambda}'g_i\right)\hat{\lambda}'g_i^j/n.$$

Differentiating again and using the expansion $\rho_1\left(\hat{\lambda}'g_i\right) = 1 + \rho_2\left(\bar{\lambda}'g_i\right)\hat{\lambda}'g_i$, we obtain

$$\begin{aligned}\frac{\mu_n^2}{n}\hat{Q}_{jk} &= \sum_i[\rho_1\left(\hat{\lambda}'g_i\right)\left(\hat{\lambda}^{k\prime}g_i^j + \hat{\lambda}'g_i^{jk}\right) + \rho_2\left(\hat{\lambda}'g_i\right)\hat{\lambda}'g_i^j\left(\hat{\lambda}^{k\prime}g_i + \hat{\lambda}'g_i^k\right)]/n \\ &= n^{-1}\sum_i\left[\left(1+\rho_2\left(\bar{\lambda}'g_i\right)\hat{\lambda}'g_i\right)\left(\hat{\lambda}^{k\prime}g_i^j + \hat{\lambda}'g_i^{jk}\right)\right] - \hat{\lambda}^{k\prime}\hat{\Omega}^j\hat{\lambda} - \hat{\lambda}'\hat{\Omega}^{j,k}\hat{\lambda} \\ &= \hat{\lambda}^{k\prime}\tilde{g}^j + \hat{\lambda}'\tilde{g}^{jk} - \hat{\lambda}'\hat{\Omega}^j\hat{\lambda}^k - \hat{\lambda}^{k\prime}\hat{\Omega}^j\hat{\lambda} - \hat{\lambda}'\left(\hat{\Omega}^{jk} + \hat{\Omega}^{j,k}\right)\hat{\lambda}\end{aligned}$$

where $\hat{\Omega}^{jk} = -\sum_i \rho_2\left(\tilde{\lambda}' g_i\right) g_i g_i^{jk\prime}/n$, $\hat{\Omega}^{j,k} = -\sum_i \rho_2\left(\tilde{\lambda}' g_i\right) g_i^j g_i^{k\prime}/n$, $\hat{g}^j = \sum_i g_i^j/n$, and $\hat{g}^{jk} = \sum_i g^{jk}/n$. Substituting the formula for $\hat{\lambda}^k$ we obtain

$$\begin{aligned}\frac{\mu_n^2}{n}\hat{Q}_{jk} &= \hat{g}^{k\prime}\tilde{\Omega}^{-1}\hat{g}^j - \hat{\lambda}'\left(\tilde{\Omega}^k + \tilde{\Omega}^{j\prime}\right)\tilde{\Omega}^{-1}\hat{g}^j + \hat{\lambda}'\hat{g}^{jk} - \hat{\lambda}'\tilde{\Omega}^j\tilde{\Omega}^{-1}\hat{g}^k + \hat{\lambda}'\tilde{\Omega}^j\tilde{\Omega}^{-1}\left(\tilde{\Omega}^k + \tilde{\Omega}^{k\prime}\right)\hat{\lambda} \\ &\quad -\hat{g}^{k\prime}\tilde{\Omega}^{-1}\tilde{\Omega}^j\hat{\lambda} + \hat{\lambda}'\left(\tilde{\Omega}^k + \tilde{\Omega}^{k\prime}\right)\tilde{\Omega}^{-1}\tilde{\Omega}^j\hat{\lambda} - \hat{\lambda}'\left(\tilde{\Omega}^{jk} + \tilde{\Omega}^{j,k}\right)\hat{\lambda}.\end{aligned}$$

Next, a mean value expansion of the first-order conditions for $\hat{\lambda}$ gives

$$\hat{\lambda} = \tilde{\Omega}^{-1}\tilde{g}\mu_n/\sqrt{n}, \tilde{\Omega} = -\frac{1}{n}\sum_{i=1}^n \rho_2\left(\bar{\lambda} g_i\right) g_i g_i'.$$

where $\tilde{g}$ now comes from the result for the CUE. Noting also that $\hat{g}^j = \tilde{g}_j\mu_n/\sqrt{n}$

$$\begin{aligned}\hat{Q}_{jk} &= \tilde{g}_k\tilde{\Omega}^{-1}\tilde{g}_j + \tilde{g}'\tilde{\Omega}^{-1}\tilde{g}_{jk} \\ &\quad -\tilde{g}'\tilde{\Omega}^{-1}\left(\tilde{\Omega}^k + \tilde{\Omega}^{k\prime}\right)\tilde{\Omega}^{-1}\tilde{g}_j - \tilde{g}'\tilde{\Omega}^{-1}\left(\tilde{\Omega}^j + \tilde{\Omega}^{j\prime}\right)\tilde{\Omega}^{-1}\tilde{g}_k \\ &\quad +\tilde{g}'\tilde{\Omega}^{-1}\left(\tilde{\Omega}^j + \tilde{\Omega}^{j\prime}\right)\tilde{\Omega}^{-1}\left(\tilde{\Omega}^k + \tilde{\Omega}^{k\prime}\right)\tilde{\Omega}^{-1}\tilde{g} \\ &\quad -\frac{1}{2}\tilde{g}'\tilde{\Omega}^{-1}\left(\tilde{\Omega}^{jk} + \tilde{\Omega}^{j,k} + \tilde{\Omega}^{j,k\prime} + \tilde{\Omega}^{j,k\prime}\right)\tilde{\Omega}^{-1}\tilde{g}\end{aligned}$$

Comparing with equation (7.3) we can see that this expression is identical to that for the CUE with $\tilde{\Omega}^j + \tilde{\Omega}^{j\prime}$ replacing $\hat{\Omega}_j$ and $\tilde{\Omega}^{jk} + \tilde{\Omega}^{j,k} + \tilde{\Omega}^{j,k\prime} + \tilde{\Omega}^{jk\prime}$ replacing $\hat{\Omega}_{j,k}$. Evaluate the CUE expressions at the GEL estimator, and note that

$$\begin{aligned}\left\|\hat{\Omega}_j - \left(\tilde{\Omega}^j + \tilde{\Omega}^j\right)\right\| &\leq 2\sum_i \left|\rho_2\left(\bar{\lambda}' g_i\right) + 1\right| \|g_i\| \left\|g_i^j\right\|/n \\ &= O_p\left(E\left[\sup_{\beta\in B}(\|g_i\|^2\left\|g_i^j\right\|)\right]\sqrt{m}/\sqrt{n}\right) \xrightarrow{p} 0.\end{aligned}$$

by CS. It follows similarly that

$$\left\|\hat{\Omega} - \tilde{\Omega}\right\| \xrightarrow{p} 0, \quad \left\|\hat{\Omega}_{j,k} - \left(\tilde{\Omega}^{jk} + \tilde{\Omega}^{j,k} + \tilde{\Omega}^{j,k\prime} + \tilde{\Omega}^{jk\prime}\right)\right\| \xrightarrow{p} 0.$$

Then it follows similarly to the proof for the CUE that one can replace $\hat{\Omega}$ there by $\tilde{\Omega}$ that the difference of the expression for the CUE and for GEL converges in probability to zero.

Finally, we prove that $\hat{D}(\hat{\beta})'\hat{\Omega}^{-1}\hat{D}(\hat{\beta})/\mu_n^2 \xrightarrow{p} H + \kappa\Lambda^*$ for GEL. Like like those in previous proofs,

$$\begin{aligned}
\hat{D}_j &= \sum_i \rho_1(\hat{\lambda}'g_i)g_i^j = \hat{G}^j - \sum_i \rho_2(\hat{\lambda}'g)g_i^j g_i{}'\hat{\lambda} = \hat{G}^j - \bar{A}^j\bar{\Omega}^{-1}\hat{g}(\hat{\beta}), \\
\bar{A}^j &= -\sum_i \rho_2(\bar{\lambda}'g_i)g_i^j g_i'/n, \bar{\Omega} = \sum_i -\rho_2(\bar{\lambda}'g_i)g_i g_i'/n,
\end{aligned}$$

where the two $\bar{\lambda}$ may differ but each lies on the line joining $\hat{\lambda}(\hat{\beta})$ and zero. It follows as in previous proofs that

$$\left\|\bar{A}^j - \hat{A}^j\right\| \leq C\left\|\hat{\lambda}\right\| \sum_i \sup_{\beta\in B}(\left\|g_i^j(\beta)\right\| \|g_i(\beta)\|^2)/n = O_p(\sqrt{m/n}E[d_i^3]) \xrightarrow{p} 0.$$

It follows similarly that $\left\|\bar{\Omega} - \Omega\right\| \xrightarrow{p} 0$. Then it follows similarly to eq. (7.5) that $\sqrt{n}/\mu_n$ times the difference of the expressions of $\hat{D}_j$ for GEL and the CUE converges in probability to zero. It then follows by arguments similar to those already given that the μ_n^{-2} times the difference of $\hat{D}'\hat{\Omega}^{-1}\hat{D}$ for the GEL estimator and the CUE converges in probability to zero. The conclusion then follows by T. 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.

Arellano, M. and S.R. Bond (1991): "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies* 58, 277-297.

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

Bekker, P.A. and F. Kleibergen (2003): "Finite Sample Instrumental Variables Inference using an Asymptotically Pivotal Statistic," *Econometric Theory* 19, 744-753.

Brown, B.W. and W.K. Newey (1998): "Efficient Semiparametric Estimation of Expectations," *Econometrica* 66, 453-464.

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

Chao, J.C. and N.R. Swanson (2004): "Estimation and Testing Using Jackknife IV in Heteroskedastic Regression with Many Weak Instruments," working paper, Rutgers University.

Chao, J.C. and N.R. Swanson (2005): "Consistent Estimation With a Large Number of Weak Instruments," *Econometrica* 73, 1673-692.

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 (2003): "Empirical Likelihood Estimation and Consistent Tests With Conditional Moment Restrictions," *Journal of Econometrics* 117, 55-93.

Donald, S. G. and W. K. Newey (2003) "Choosing the Number of Moments in GMM and GEL Estimation," working paper, MIT.

Guggenberger, P. and R.J. Smith (2005): "Generalized Empirical Likelihood Estimators and Tests Under Partial, Weak, and Strong Identification," *Econometric Theory* 21, 667-709.

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

Han, C. and P.C.B. Phillips (2005): "GMM with Many Moment Conditions," *Econometrica,* forthcoming.

Hansen, C., J.A. Hausman, and W.K. Newey (2005): "Estimation with Many Instrumental Variables," working paper, MIT.

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

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

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.

Holtz-Eakin, D., W.K. Newey and H. Rosen (1988), "Estimating Vector Autoregressions With Panel Data," *Econometrica* 56, 1371-1396.

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

Kitamura, Y., and M. Stutzer (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation", *Econometrica* 65, 861-874.

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

Kleibergen, F. (2005): "Testing Parameters in GMM Without Assuming They are Identified," *Econometrica* 73, 1103-1123.

Newey, W.K. (1990): "Efficient Instrumental Variables Estimation of Nonlinear Models," *Econometrica* 58, 809-837.

Newey, W.K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica* 59, 1161-1167.

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

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

Phillips, G.D.A. and C. Hale (1977): "The Bias of Instrumental Variable Estimators of Simultaneous Equation Systems," *International Economic Review* 18, 219-228.

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

Smith, R. J. (1997): "Alternative Semi-Parametric Likelihood Approaches to Generalized Method of Moments Estimation," *Economic Journal* 107, 503-519.

Staiger, D. and J. Stock (1997): "Instrumental Variables Regression with Weak Instruments", *Econometrica* 65, 57-586.

Stock, J. and J. Wright (2000): "GMM With Weak Identification," *Econometrica* 68, 1055-1096.

Stock, J. and M. Yogo (2005): "Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments," in D.W.K. Andrews and J.H. Stock, eds., *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 109-120.

Windmeijer, F. (2005): "A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators," *Journal of Econometrics* 126, 25-51.