*Econometrica Supplementary Material*

# SUPPLEMENT TO "GENERALIZED METHOD OF MOMENTS WITH MANY WEAK MOMENT CONDITIONS"

(*Econometrica*, Vol. 77, No. 3, May 2009, 687–719)

BY WHITNEY K. NEWEY AND FRANK WINDMEIJER

## APPENDIX: PROOFS

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–Schwarz, Markov, and triangle inequalities, respectively. Let S denote the Slutzky lemma and CMT denote the continuous mapping theorem. 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 1." The following standard matrix result is used repeatedly.

LEMMA A0: *If $A$ and $B$ are symmetric, positive semidefinite matrices, then*

$$|\xi_{\min}(A) - \xi_{\min}(B)| \le \|A - B\|, \quad |\xi_{\max}(A) - \xi_{\max}(B)| \le \|A - B\|.$$

*Also, if $\|\hat{A} - A\| \xrightarrow{p} 0$, $\xi_{\min}(A) \ge 1/C$, and $\xi_{\max}(A) \le C$, then w.p.a.1 $\xi_{\min}(\hat{A}) \ge 1/2C$ and $\xi_{\max}(\hat{A}) \le 2C$.*

### A.1. *Consistency Proofs for General CUE*

For Lemmas A1 and A10, let $Y_i, Z_i$ $(i = 1, \ldots, n)$ be i.i.d. $m \times 1$ random vectors that depend on $n$ and have fourth moments but where we suppress an $n$ subscript for notational convenience. Also, let

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

and let objects with $Z$ in place of $Y$ be defined in the corresponding way.

LEMMA A1: *If $(Y_i, Z_i)$ $(i = 1, \ldots, n)$ are i.i.d., $\xi_{\max}(AA') \le C$, $\xi_{\max}(A'A) \le C$, $\xi_{\max}(\Sigma_{YY}) \le C$, $\xi_{\max}(\Sigma_{ZZ}) \le C$, $m/a_n^2 \longrightarrow 0$, $a_n/n \le C$, $E[(Y_i'Y_i)^2]/na_n^2 \longrightarrow 0$, $E[(Z_i'Z_i)^2]/na_n^2 \longrightarrow 0$, $n\mu_Y'\mu_Y/a_n^2 \longrightarrow 0$, and $n\mu_Z'\mu_Z/a_n^2 \longrightarrow 0$, then*

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

PROOF: Let $W_i = AZ_i$. Then $A\Sigma_{YZ}' = \Sigma_{YW}'$, $A\mu_Z = \mu_W$,

$$\xi_{\max}(E[W_iW_i']) = \xi_{\max}(A\Sigma_{ZZ}A') \le C\xi_{\max}(AA') \le C,$$
$$E[(W_i'W_i)^2]/na_n^2 = E[(Z_i'A'AZ_i)^2]/na_n^2 \le CE[(Z_i'Z_i)^2]/na_n^2 \longrightarrow 0.$$

 DOI: 10.3982/ECTA6224

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

$$E[(Y_i'Z_i)^2] \le E[(Y_i'Y_i)^2] + E[(Z_i'Z_i)^2],$$
$$E[Y_i'Z_jZ_j'Y_i] = E[Y_i'\Sigma_{ZZ}Y_i] \le CE[Y_i'Y_i] = C\,\mathrm{tr}(\Sigma_{YY}) \le Cm,$$
$$|E[Y_i'Z_jY_j'Z_i]| \le C(E[Y_i'Z_jZ_j'Y_i] + E[Y_j'Z_iZ_i'Y_j]) \le Cm.$$

For the moment suppose $\mu_Y = \mu_Z = 0$. Let $W_n = n\bar{Y}'\bar{Z}/a_n$. Then $E[W_n] = E[Y_i'Z_i]/a_n = \mathrm{tr}(\Sigma_{YZ})/a_n$ and

$$\begin{aligned} E[W_n]^2/n &\le E[(Y_i'Z_i)^2]/na_n^2 \\ &\le \big\{E[(Y_i'Y_i)^2] + E[(Z_i'Z_i)^2]\big\}/na_n^2 \longrightarrow 0. \end{aligned}$$

We also have

$$\begin{aligned} E[W_n^2] &= E\Bigg[\sum_{i,j,k,\ell} Y_i'Z_jY_k'Z_\ell/n^2a_n^2\Bigg] \\ &= E[(Y_i'Z_i)^2]/na_n^2 \\ &\quad + (1-1/n)\big\{E[W_n]^2 + E[Y_i'Z_jY_j'Z_i]/a_n^2 + E[Y_i'Z_jZ_j'Y_i]/a_n^2\big\} \\ &= E[W_n]^2 + o(1), \end{aligned}$$

so that by M,

$$W_n = \mathrm{tr}(\Sigma_{YZ}')/a_n + o_p(1).$$

In general, when $\mu_Y$ or $\mu_Z$ is nonzero, note that $E[\{(Y_i - \mu_Y)'(Y_i - \mu_Y)\}^2] \le CE[(Y_i'Y_i)^2]$ and $\xi_{\max}(\mathrm{Var}(Y_i)) \le \xi_{\max}(\Sigma_{YY})$, 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,

$$\begin{aligned} W_n = n\bar{Y}'\bar{Z}/a_n &= n(\bar{Y} - \mu_Y)'(\bar{Z} - \mu_Z)/a_n + n\mu_Y'(\bar{Z} - \mu_Z)/a_n \\ &\quad + n(\bar{Y} - \mu_Y)'\mu_Z/a_n + n\mu_Y'\mu_Z/a_n. \end{aligned} \tag{A.1}$$

Note that

$$\begin{aligned} E\big[\{n\mu_Y'(\bar{Z} - \mu_Z)/a_n\}^2\big] &= n\mu_Y'(\Sigma_{ZZ} - \mu_Z\mu_Z')\mu_Y/a_n^2 \le n\mu_Y'\Sigma_{ZZ}\mu_Y/a_n^2 \\ &\le Cn\mu_Y'\mu_Y/a_n^2 \longrightarrow 0, \end{aligned}$$

so by M, the second and third terms in eq. (A.1) (with $Y$ and $Z$ interchanged) are $o_p(1)$. Also, $\mathrm{tr}(\mu_Z\mu_Y')/a_n = a_nn^{-1}(n\mu_Y'\mu_Z/a_n^2) \longrightarrow 0$. Applying the result for the zero mean case then gives

$$W_n = \mathrm{tr}(\Sigma'_{YZ} - \mu_Z\mu'_Y)/a_n + n\mu'_Y\mu_Z/a_n + o_p(1)$$
$$= \mathrm{tr}(\Sigma'_{YZ})/m + n\mu'_Y\mu_Z/m + o_p(1). \qquad \textit{Q.E.D.}$$

It is useful to work with a reparameterization

$$\delta = S'_n(\beta - \beta_0)/\mu_n.$$

For notational simplicity we simply change the argument to denote the reparameterized functions, for example, $\hat{Q}(\delta)$ will denote $\hat{Q}(\beta_0 + \mu_n S_n^{-1\prime}\delta)$. Let $\hat{Q}^*(\delta) = \hat{g}(\delta)'\hat{\Omega}(\delta)^{-1}\hat{g}(\delta)/2$ be the objective function for quadratic $\rho(v)$, let $\tilde{Q}(\delta) = \hat{g}(\delta)'\Omega(\delta)^{-1}\hat{g}(\delta)/2$, and let $Q(\delta) = \bar{g}(\delta)'\Omega(\delta)^{-1}\bar{g}(\delta)/2 + m/2n$.

LEMMA A2: *If Assumption* 3 *is satisfied, then for any* $C > 0$, $\sup_{\beta\in B,\|\delta\|\le C} \mu_n^{-2} \times n|\hat{Q}^*(\delta) - Q(\delta)| \xrightarrow{p} 0$.

PROOF: Note that by Assumption 3(ii), $\mu_n^{-2} nE[\|\hat{g}(0)\|^2] = \mu_n^{-2}\,\mathrm{tr}(\Omega(\beta_0)) \le C$, so by Assumption 3(v) and T,

$$\sup_{\|\delta\|\le C} \|\hat{g}(\delta)\| \le \|\hat{g}(0)\| + \sup_{\|\delta\|\le C} \|\hat{g}(\delta) - \hat{g}(0)\| = O_p(\mu_n/\sqrt{n}).$$

Let $\hat{a}(\delta) = \mu_n^{-1}\sqrt{n}\Omega(\delta)^{-1}\hat{g}(\delta)$. By Assumption 3(ii),

$$\|\hat{a}(\delta)\|^2 = \mu_n^{-2} n\hat{g}(\delta)'\Omega(\delta)^{-1/2}\Omega(\delta)^{-1}\Omega(\delta)^{-1/2}\hat{g}(\delta) \le C\mu_n^{-2} n\|\hat{g}(\delta)\|^2,$$

so that $\sup_{\|\delta\|\le C} \|\hat{a}(\delta)\| = O_p(1)$. Also, by Assumption 3(iii) we have

$$\left|\xi_{\min}(\hat{\Omega}(\delta)) - \xi_{\min}(\Omega(\delta))\right| \le \sup_{\|\delta\|\le C} \|\hat{\Omega}(\delta) - \Omega(\delta)\| \xrightarrow{p} 0,$$

so that $\xi_{\min}(\hat{\Omega}(\delta)) \ge C$, and hence $\xi_{\max}(\hat{\Omega}(\delta)^{-1}) \le C$ for all $\|\delta\| \le C$, w.p.a.1. Therefore,

$$\begin{aligned}
&\mu_n^{-2} n|\hat{Q}^*(\delta) - \tilde{Q}(\delta)| \\
&\quad \le \left|\hat{a}(\delta)'[\hat{\Omega}(\delta) - \Omega(\delta)]\hat{a}(\delta)\right| \\
&\qquad + \left|\hat{a}(\delta)'[\hat{\Omega}(\delta) - \Omega(\delta)]\hat{\Omega}(\delta)^{-1}[\hat{\Omega}(\delta) - \Omega(\delta)]\hat{a}(\delta)\right| \\
&\quad \le \|\hat{a}(\delta)\|^2\big(\|\hat{\Omega}(\delta) - \Omega(\delta)\| + C\|\hat{\Omega}(\delta) - \Omega(\delta)\|^2\big) \xrightarrow{p} 0.
\end{aligned}$$

Next, let $a(\tilde{\delta},\delta)=\mu_n^{-1}\sqrt{n}\Omega(\delta)^{-1}\bar{g}(\tilde{\delta})$ and $Q(\tilde{\delta},\delta)=\bar{g}(\tilde{\delta})'\Omega(\delta)^{-1}\bar{g}(\tilde{\delta})/2+m/2n$. By Assumption 3, $\sup_{\|\delta\|\le C,\|\tilde{\delta}\|\le C}\|a(\delta,\tilde{\delta})\|\le C$. Then by Assumption 3(iv), for $\|\delta\|\le C$ and $\|\tilde{\delta}\|\le C$, it follows by $\mu_n S_n^{-1}$ bounded that

$$\begin{aligned}\mu_n^{-2}n|Q(\tilde{\delta},\tilde{\delta})-Q(\tilde{\delta},\delta)| &= \left|a(\tilde{\delta},\tilde{\delta})'[\Omega(\tilde{\delta})-\Omega(\delta)]a(\tilde{\delta},\delta)\right| \\ &\le C\|\mu_n S_n^{-1\prime}(\tilde{\delta}-\delta)\|\le C\|\tilde{\delta}-\delta\|.\end{aligned}$$

Also, by T and Assumption 3, for $\|\delta\|\le C$ and $\|\tilde{\delta}\|\le C$,

$$\begin{aligned}&\mu_n^{-2}n|Q(\tilde{\delta},\delta)-Q(\delta,\delta)| \\ &\quad\le C\mu_n^{-2}n\big(\|\bar{g}(\tilde{\delta})-\bar{g}(\delta)\|^2+\|\bar{g}(\delta)\|\|\bar{g}(\tilde{\delta})-\bar{g}(\delta)\|\big) \\ &\quad\le C\|\tilde{\delta}-\delta\|.\end{aligned}$$

Then by T it follows that $\mu_n^{-2}n|Q(\tilde{\delta})-Q(\delta)|=\mu_n^{-2}n|Q(\tilde{\delta},\tilde{\delta})-Q(\delta,\delta)|\le C\|\tilde{\delta}-\delta\|$. Therefore, $\mu_n^{-2}nQ(\delta)$ is equicontinuous on $\|\tilde{\delta}\|\le C$ and $\|\delta\|\le C$. An analogous argument with $\hat{a}(\tilde{\delta},\delta)=\mu_n^{-1}\sqrt{n}\Omega(\delta)^{-1}\hat{g}(\tilde{\delta})$ and $\tilde{Q}(\tilde{\delta},\delta)=\hat{g}(\tilde{\delta})'\Omega(\delta)^{-1}\hat{g}(\tilde{\delta})$ replacing $a(\tilde{\delta},\delta)$ and $Q(\tilde{\delta},\delta)$, respectively, implies that $\mu_n^{-2}n|\tilde{Q}(\tilde{\delta})-\tilde{Q}(\delta)|=\mu_n^{-2}n|\tilde{Q}(\tilde{\delta},\tilde{\delta})-\tilde{Q}(\delta,\delta)|\le\hat{M}\|\tilde{\delta}-\delta\|$ on $\|\tilde{\delta}\|\le C$ and $\|\delta\|\le C$, with $\hat{M}=O_p(1)$, giving stochastic equicontinuity of $\mu_n^{-2}n\tilde{Q}(\delta)$.

Since $\mu_n^{-2}n\tilde{Q}(\delta)$ and $\mu_n^{-2}nQ(\delta)$ are stochastically equicontinuous, it suffices by Newey (1991, Theorem 2.1) to show that $\mu_n^{-2}n\tilde{Q}(\delta)=\mu_n^{-2}nQ(\delta)+o_p(1)$ for each $\delta$. Apply Lemma A1 with $Y_i=Z_i=g_i(\delta)$, $A=\Omega(\delta)^{-1}$, and $a_n=\mu_n^2$. By Assumption 3, $\xi_{\max}(A'A)=\xi_{\max}(AA')=\xi_{\max}(\Omega(\delta)^{-2})\le C$, $\xi_{\max}(\Sigma_{YY})=\xi_{\max}(\Omega(\delta))\le C$, $E[(Y_i'Y_i)^2]/na_n^2=E[\{g_i(\delta)'g_i(\delta)\}^2]/n\mu_n^4\longrightarrow 0$, and $n\mu_Y'\mu_Y/a_n^2\le Cn\bar{g}(\delta)'\Omega(\delta)^{-1}\bar{g}(\delta)/\mu_n^4=C(nQ(\delta)/\mu_n^2-m/\mu_n^2)\,/\mu_n^2\longrightarrow 0$, where the last expression follows by equicontinuity of $\mu_n^{-2}nQ(\delta)$. 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,

$$\begin{aligned}\mu_n^{-2}n\tilde{Q}(\delta) &= \mathrm{tr}(I_m)/\mu_n^2+\mu_n^{-2}n\bar{g}(\delta)'\Omega(\delta)^{-1}\bar{g}(\delta)+o_p(1) \\ &= \mu_n^{-2}nQ(\delta)+o_p(1). \qquad\qquad \textit{Q.E.D.}\end{aligned}$$

Let $\hat{P}(\beta,\lambda)=\sum_{i=1}^n\rho(\lambda'g_i(\beta))/n$.

LEMMA A3: *If Assumptions* 3 *and* 4 *are satisfied, then w.p.a.*1 $\hat{\beta}=\arg\min_{\beta\in B}\hat{Q}(\beta)$, $\hat{\lambda}=\arg\max_{\lambda\in\hat{L}_n(\hat{\beta})}\hat{P}(\hat{\beta},\lambda)$, *and* $\tilde{\lambda}=\arg\max_{\lambda\in\hat{L}(\beta_0)}\hat{P}(\beta_0,\lambda)$ *exist*, $\|\tilde{\lambda}\|=O_p(\sqrt{m/n})$, $\|\hat{\lambda}\|=O_p(\sqrt{m/n})$, $\|\hat{g}(\hat{\beta})\|=O_p(\sqrt{m/n})$, *and* $\hat{Q}^*(\hat{\beta})\le\hat{Q}^*(\beta_0)+o_p(m/n)$.

PROOF: Let $b_i = \sup_{\beta\in B}\|g_i(\beta)\|$. A standard result gives $\max_{i\le n} b_i = O_p(n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma})$. Also, by Assumption 4 there exists $\tau_n$ such that $\sqrt{m/n} = o(\tau_n)$ and $\tau_n = o(n^{-1/\gamma}(E[b_i^\gamma])^{-1/\gamma})$. Let $L_n = \{\lambda : \|\lambda\| \le \tau_n\}$. Note that

$$\sup_{\lambda\in L_n,\beta\in B,i\le n} |\lambda' g_i(\beta)| \le \tau_n \max_{i\le n} b_i = O_p\big(\tau_n n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma}\big) \longrightarrow 0.$$

Then there is $C$ such that w.p.a.1, for all $\beta\in B$, $\lambda\in L_n$, and $i\le n$, we have

$$L_n \subset \hat{L}(\beta), \quad -C \le \rho_2(\lambda' g_i(\beta)) \le -C^{-1}, \quad \big|\rho_3(\lambda' g_i(\beta))\big| \le C.$$

By a Taylor expansion around $\lambda = 0$ with Lagrange remainder, for all $\lambda\in L_n$,

$$\hat{P}(\beta,\lambda) = -\lambda'\hat{g}(\beta) + \lambda'\left[\sum_{i=1}^n \rho_2(\bar{\lambda}' g_i(\beta)) g_i(\beta) g_i(\beta)'/n\right]\lambda,$$

where $\bar{\lambda}$ lies on the line joining $\lambda$ and 0. Then by Lemma A0, w.p.a.1 for all $\beta\in B$ and $\lambda\in L_n$,

$$\begin{aligned}(\text{A.2}) \qquad -\lambda'\hat{g}(\beta) - C\|\lambda\|^2 \le \hat{P}(\beta,\lambda) &\le -\lambda'\hat{g}(\beta) - C^{-1}\|\lambda\|^2 \\ &\le \|\lambda\|\|\hat{g}(\beta)\| - C^{-1}\|\lambda\|^2.\end{aligned}$$

Let $\tilde{g} = \hat{g}(\beta_0)$ and $\tilde{\lambda} = \arg\max_{\lambda\in L_n} \hat{P}(\beta_0,\lambda)$. By $\xi_{\max}(\Omega(\beta_0)) \le C$ it follows that $E[\|\tilde{g}\|^2] = \operatorname{tr}(\Omega)/n \le Cm/n$, so by M, $\|\tilde{g}\| = O_p(\sqrt{m/n})$. By the right-hand side inequality in eq. (A.2),

$$0 = \hat{P}(\beta_0, 0) \le \hat{P}(\beta_0,\tilde{\lambda}) \le \|\tilde{\lambda}\|\|\tilde{g}\| - C^{-1}\|\tilde{\lambda}\|^2.$$

Subtracting $C^{-1}\|\tilde{\lambda}\|^2$ from both sides and dividing through by $C^{-1}\|\tilde{\lambda}\|$ gives

$$\|\tilde{\lambda}\| \le C\|\tilde{g}\| = O_p(\sqrt{m/n}).$$

Since $\sqrt{m/n} = o(\tau_n)$ it follows that, w.p.a.1, $\tilde{\lambda}\in \operatorname{int}(L_n)$ and is therefore a local maximum of $\hat{P}(\beta_0,\lambda)$ in $\hat{L}(\beta)$. By concavity of $P(\beta_0,\lambda)$ in $\lambda$, a local maximum is a global maximum, that is,

$$\hat{P}(\beta_0,\tilde{\lambda}) = \max_{\lambda\in\hat{L}(\beta_0)} \hat{P}(\beta_0,\lambda) = \hat{Q}(\beta_0).$$

Summarizing, w.p.a.1 $\tilde{\lambda} = \arg\max_{\lambda\in\hat{L}(\beta_0)} \hat{P}(\beta_0,\lambda)$ exists and $\|\tilde{\lambda}\| = O_p(\sqrt{m/n})$. Also, plugging $\tilde{\lambda}$ back into the previous inequality gives

$$\hat{Q}(\beta_0) = O_p(m/n).$$

Next, let $\hat{Q}_{\tau_n}(\beta) = \max_{\lambda \in L_n} \hat{P}(\beta, \lambda)$. By continuity of $g_i(\beta)$ and $\rho(v)$, and by the theorem of the maximum, $\hat{Q}_{\tau_n}(\beta)$ is continuous on $B$, so $\hat{\beta}_{\tau_n} = \arg\min_{\beta \in B} \hat{Q}_{\tau_n}(\beta)$ exists by compactness of $B$. Let $\hat{g}_{\tau_n} = \hat{g}(\hat{\beta}_{\tau_n})$. By the left-hand side inequality in eq. (A.2), for all $\lambda \in L_n$,

$$\text{(A.3)} \qquad -\lambda' \hat{g}_{\tau_n} - C\|\lambda\|^2 \le \hat{P}(\hat{\beta}_{\tau_n}, \lambda) \le \hat{Q}_{\tau_n}(\hat{\beta}_{\tau_n}) \le \hat{Q}_{\tau_n}(\beta_0) \le \hat{Q}(\beta_0) = O_p(m/n).$$

Consider $\lambda = -(\hat{g}_{\tau_n}/\|\hat{g}_{\tau_n}\|)\tau_n$. Plugging this into eq. (A.3) gives

$$\tau_n \|\hat{g}_{\tau_n}\| - c\tau_n^2 = O_p(m/n).$$

Note that for $n$ large enough, $m/n \le C\tau_n^2$, so that dividing by $\tau_n^2$ gives

$$\|\hat{g}_{\tau_n}\| \le O_p(\tau_n^{-1} m/n) + C\tau_n = O_p(\tau_n).$$

Consider any $\alpha_n \longrightarrow 0$ and let $\breve{\lambda} = -\alpha_n \hat{g}_{\tau_n}$. Then $\|\breve{\lambda}\| = o_p(\tau_n)$ so that $\breve{\lambda} \in L_n$ w.p.a.1. Substituting this $\breve{\lambda}$ in the above inequality gives

$$\alpha_n \|\hat{g}_{\tau_n}\|^2 - C\alpha_n^2 \|\hat{g}_{\tau_n}\|^2 = \alpha_n(1 - C\alpha_n)\|\hat{g}_{\tau_n}\|^2 = O_p\left(\frac{m}{n}\right).$$

Note that $1 - C\alpha_n \to 1$, so that this inequality implies that $\alpha_n \|\hat{g}_{\tau_n}\|^2 = O_p(m/n)$. Since $\alpha_n$ goes to zero as slowly as desired, it follows that

$$\|\hat{g}(\hat{\beta}_{\tau_n})\| = \|\hat{g}_{\tau_n}\| = O_p(\sqrt{m/n}).$$

Let $\hat{\lambda} = \arg\max_{\lambda \in L_n} \hat{P}(\hat{\beta}_{\tau_n}, \lambda)$. It follows exactly as for $\tilde{\lambda}$, with $\hat{\beta}_{\tau_n}$ replacing $\beta_0$, that $\|\hat{\lambda}\| = O_p(\sqrt{m/n})$ and, w.p.a.1, $\hat{\lambda} = \arg\max_{\lambda \in \hat{L}(\beta)} \hat{P}(\hat{\beta}_{\tau_n}, \lambda)$, so that

$$\hat{Q}_{\tau_n}(\hat{\beta}_{\tau_n}) = \hat{P}(\hat{\beta}_{\tau_n}, \hat{\lambda}) = \max_{\lambda \in \hat{L}(\beta)} \hat{P}(\hat{\beta}_{\tau_n}, \lambda) = \hat{Q}(\hat{\beta}_{\tau_n}).$$

Then w.p.a.1, by the definition of $\hat{Q}_{\tau_n}(\beta)$ and $\hat{\beta}_{\tau_n}$, for all $\beta \in B$,

$$\hat{Q}(\hat{\beta}_{\tau_n}) = \hat{Q}_{\tau_n}(\hat{\beta}_{\tau_n}) \le \hat{Q}_{\tau_n}(\beta) = \max_{\lambda \in L_n} \hat{P}(\beta, \lambda) \le \hat{Q}(\beta).$$

Thus, w.p.a.1 we can take $\hat{\beta} = \hat{\beta}_{\tau_n}$.

Now expand around $\lambda = 0$ to obtain, for $\hat{g}_i = g_i(\hat{\beta})$ and $\hat{\Omega} = \hat{\Omega}(\hat{\beta})$, w.p.a.1,

$$\hat{Q}(\hat{\beta}) = \hat{P}(\hat{\beta}, \hat{\lambda}) = -\hat{g}'\hat{\lambda} - \frac{1}{2}\hat{\lambda}'\hat{\Omega}\hat{\lambda} + \hat{r}, \quad \hat{r} = \frac{1}{6}\sum \rho_3(\bar{\lambda}'\hat{g}_i)(\hat{\lambda}'\hat{g}_i)^3/n,$$

where $\|\bar{\lambda}\| \leq \|\hat{\lambda}\|$ and $\hat{r} = 0$ for the CUE (where $\rho(v)$ is quadratic). When $\hat{\beta}$ is not the CUE, w.p.a.1,

$$|\hat{r}| \leq \|\hat{\lambda}\| \max_i b_i C \hat{\lambda}' \hat{\Omega}(\hat{\beta}) \hat{\lambda} \leq O_p\big(\sqrt{m/n} n^{1/\gamma} (E[b_i^\gamma])^{1/\gamma}\big) C \|\bar{\lambda}\|^2$$
$$= o_p(m/n).$$

Also, $\hat{\lambda}$ satisfies the first-order conditions $\sum_{i=1}^n \rho_1(\hat{\lambda}'\hat{g}_i)\hat{g}_i/n = 0$. By an expansion, $\rho_1(\hat{\lambda}'\hat{g}_i) = -1 - \hat{\lambda}'\hat{g}_i + \rho_3(\bar{v}_i)(\lambda'\hat{g}_i)^2/2$, where $\bar{v}_i$ lies in between 0 and $\lambda'\hat{g}_i$ and either $\rho_3(\bar{v}_i) = 0$ for the CUE or $\max_{i \leq n} |\bar{v}_i| \leq \max_{i \leq n} |\hat{\lambda}'\hat{g}_i| \leq \tau_n \longrightarrow 0$. Expanding around $\lambda = 0$ gives

$$0 = -\hat{g} - \hat{\Omega}\hat{\lambda} + \hat{R}, \quad \hat{R} = \frac{1}{2} \sum_{i=1}^n \rho_3(\bar{v}_i)(\hat{\lambda}'\hat{g}_i)^2 \hat{g}_i/n = 0.$$

Then either $\hat{R} = 0$ for the CUE or we have

$$\|\hat{R}\| \leq C \max_i b_i |\rho_3(\bar{v}_i)| \hat{\lambda}' \hat{\Omega} \hat{\lambda} = O_p\big(n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma} m/n\big) = o_p(\sqrt{m/n}).$$

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

$$\hat{Q}(\hat{\beta}) = -\hat{g}'\hat{\Omega}^{-1}(-\hat{g} + \hat{R}) - \frac{1}{2}(-\hat{g} + \hat{R})'\hat{\Omega}^{-1}(-\hat{g} + \hat{R}) + o_p(m/n)$$
$$= \hat{Q}^*(\hat{\beta}) - \hat{R}'\hat{\Omega}^{-1}\hat{R}/2 + o_p(m/n) = \hat{Q}^*(\hat{\beta}) + o_p(m/n).$$

An exactly analogous expansion, replacing $\hat{\beta}$ with $\beta_0$, gives

$$\hat{Q}(\hat{\beta}) = \hat{Q}^*(\beta_0) + o_p(m/n).$$

Then by the definition of $\hat{\beta}$,

$$\hat{Q}^*(\hat{\beta}) = \hat{Q}(\hat{\beta}) + o_p(m/n) \leq \hat{Q}(\beta_0) + o_p(m/n)$$
$$= \hat{Q}^*(\beta_0) + o_p(m/n).$$

*Q.E.D.*

LEMMA A4: *If Assumptions* 2–4 *are satisfied*, *then* $\|\hat{\delta}\| = O_p(1)$.

PROOF: By Lemma A3, w.p.a.1, $\|\hat{g}(\hat{\beta})\| = O_p(\sqrt{m/n})$, so that Assumption 2(iii) and $m/\mu_n^2 \leq C$ give

$$\|\hat{\delta}\| \leq C\mu_n^{-1}\sqrt{n}\|\hat{g}(\hat{\beta})\| + O_p(1) = O_p(\sqrt{m}/\mu_n) + O_p(1) = O_p(1).$$

*Q.E.D.*

PROOF OF THEOREM 1: By Lemma A3 and $m/\mu_n^2 \le C$ it follows that, parameterizing in terms of $\delta = S_n'(\beta - \beta_0)/\mu_n$ (where $\delta_0 = 0$),

$$\mu_n^{-2} n\hat{Q}^*(\hat{\delta}) \le \mu_n^{-2} n\hat{Q}^*(0) + o_p(1).$$

Consider any $\varepsilon, \gamma > 0$. By Lemma A4 there is $C$ such that $\Pr(\mathcal{A}_1) \ge 1 - \varepsilon/3$ for $\mathcal{A}_1 = \{\|\hat{\delta}\| \le C\}$. In the notation of Lemma A2 let $\mathcal{A}_2 = \{\sup_{\|\delta\| \le C} \mu_n^{-2} n|\hat{Q}^*(\delta) - Q(\delta)| < \gamma/3\}$ and $\mathcal{A}_3 = \{\mu_n^{-2} n\hat{Q}^*(\hat{\delta}) \le \mu_n^{-2} n\hat{Q}^*(0) + \gamma/3\}$. By Lemma A2, for all $n$ large enough, $\Pr(\mathcal{A}_2) \ge 1 - \varepsilon/3$ and by Lemma A3, $\Pr(\mathcal{A}_3) \ge 1 - \varepsilon/3$. Then $\Pr(\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3) \ge 1 - \varepsilon$ and on $\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3$,

$$\begin{aligned}\mu_n^{-2} nQ(\hat{\delta}) &\le \mu_n^{-2} n\hat{Q}^*(\hat{\delta}) + \gamma/3 \le \mu_n^{-2} n\hat{Q}^*(0) + 2\gamma/3 \\ &\le \mu_n^{-2} nQ(0) + \gamma = m/\mu_n^2 + \gamma,\end{aligned}$$

where the second inequality follows by $\hat{\delta} \in \mathcal{A}_3$. Subtracting $m/\mu_n^2$ from both sides it follows that $\mathcal{A}$ implies $\mu_n^{-2} n\bar{g}(\hat{\delta})'\Omega(\hat{\delta})^{-1}\bar{g}(\hat{\delta}) \le \gamma$. Since $\varepsilon, \gamma$ can be any positive constants, we have $\mu_n^{-2} n\bar{g}(\hat{\delta})'\Omega(\hat{\delta})^{-1}\bar{g}(\hat{\delta}) \xrightarrow{p} 0$. Then, by Assumptions 2(ii) and 3(ii),

$$\mu_n^{-2} n\bar{g}(\hat{\delta})'\Omega(\hat{\delta})^{-1}\bar{g}(\hat{\delta}) \ge C\mu_n^{-2} n\bar{g}(\bar{\beta})'\bar{g}(\bar{\beta}) \ge C\|\hat{\delta}\|^2,$$

so that $\|\hat{\delta}\| \xrightarrow{p} 0$. *Q.E.D.*

## A.2. *Conditions for the Linear Model*

LEMMA A5: *If Assumption 5 is satisfied, then* $\xi_{\min}(E[(y_i - x_i'\beta)^2 | Z_i, Y_i]) \ge C$. *Also, for* $X_i = (y_i, x_i')'$, $E[\|X_i\|^4 | Z_i, Y_i] \le C$.

PROOF: Let $\Delta = \beta_0 - \beta$ and let $\tilde{\Delta}$ be the elements of $\Delta$ corresponding to the vector $\tilde{\eta}_i$ of nonzero elements of $\eta_i$ from Assumption 5. Then $y_i - x_i'\beta = \varepsilon_i + \tilde{\eta}_i'\tilde{\Delta} + Y_i'\Delta$, so that

$$\begin{aligned}E[(y_i - x_i'\beta)^2 | Z_i, Y_i] &\ge E[(\varepsilon_i + \tilde{\eta}_i'\tilde{\Delta})^2 | Z_i, Y_i] \\ &= (1, \tilde{\Delta}')\Sigma_i(1, \tilde{\Delta}')' \ge \xi_{\min}(\Sigma_i)(1 + \tilde{\Delta}'\tilde{\Delta}) \ge C,\end{aligned}$$

giving the first conclusion. Also, $E[\|x_i\|^4 | Z_i, Y_i] \le CE[\|\eta_i\|^4 | Z_i, Y_i] + CE[\|Y_i\|^4 | Z_i, Y_i] \le C$ and $E[y_i^4 | Z_i, Y_i] \le CE[\|x_i\|^4 \|\beta_0\|^4 | Z_i, Y_i] + E[\varepsilon_i^4 | Z_i, Y_i] \le C$, giving the second conclusion. *Q.E.D.*

LEMMA A6: *If Assumption 5 is satisfied, then there is a constant C such that for every* $\beta \in B$ *and* $m$, $C^{-1} I_m \le \Omega(\beta) \le CI_m$.

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 A7: *If Assumption* 5 *is satisfied*, *then Assumption* 3(v) *is satisfied*, $\|n^{-1}\sum_i Z_i z_i' - E[Z_i z_i']\| \xrightarrow{p} 0$, *and* $\|n^{-1}\sum_i Z_i \eta_i'\| = O_p(\sqrt{m/n})$.

PROOF: For the last conclusion, by $E[\eta_i'\eta_i|Z_i] \leq C$ we have

$$E\left[\left\|n^{-1}\sum_i Z_i\eta_i'\right\|^2\right] = n^{-1}E[Z_i'Z_i\eta_i'\eta_i] \leq Cn^{-1}E[Z_i'Z_i] = Cm/n,$$

so the last conclusion follows by M. For the second to last conclusion, we have

$$\begin{aligned} E\left[\left\|n^{-1}\sum_i Z_i z_i' - E[Z_i z_i']\right\|^2\right] &\leq E[Z_i'Z_i z_i' z_i]/n \\ &\leq \sqrt{E[\|Z_i\|^4]/n}\sqrt{E[\|z_i\|^4]/n} \longrightarrow 0, \end{aligned}$$

so it also follows by M.

Next, by Assumption 5 and Lemma A6 we have

$$\|E[Z_i z_i']\|^2 = \mathrm{tr}\big\{E[z_i Z_i'](E[Z_i Z_i'])^{-1}E[Z_i z_i']\big\} \leq \mathrm{tr}(E[z_i z_i']) \leq C.$$

Then we have by CS, $Y_i = S_n z_i/\sqrt{n}$, $G = -E[Z_i z_i']S_n'/\sqrt{n}$, and

$$\begin{aligned} \mu_n^{-1}\sqrt{n}\|\bar{g}(\tilde{\beta}) - \bar{g}(\beta)\| &= \mu_n^{-1}\sqrt{n}\|G(\tilde{\beta} - \beta)\| = \|E[Z_i z_i'](\tilde{\delta} - \delta)\| \\ &\leq \|E[Z_i z_i']\|\|\tilde{\delta} - \delta\| \leq C\|\tilde{\delta} - \delta\|. \end{aligned}$$

Also, by $\hat{G} = \hat{G}(\beta)$ not depending on $\beta$, by $\|S_n^{-1\prime}\| \leq C/\mu_n$, and by T,

$$\begin{aligned} \|\hat{G}\sqrt{n}S_n^{-1\prime}\| &\leq \left\|\frac{1}{\sqrt{n}}\sum_i Z_i\eta_i' S_n^{-1\prime}\right\| \\ &\quad + \left\|\frac{1}{n}\sum_i Z_i z_i' - E[Z_i z_i']\right\| + \|E[Z_i z_i']\| \\ &= O_p\left(\frac{\sqrt{n}}{\mu_n}\sqrt{\frac{m}{n}}\right) + o_p(1) + O(1) = O_p(1), \end{aligned}$$

so that for $\hat{M} = \|\hat{G}\sqrt{n}S_n^{-1\prime}\| = O_p(1)$, by CS,

$$\begin{aligned} \mu_n^{-1}\sqrt{n}\|\hat{g}(\tilde{\beta}) - \hat{g}(\beta)\| &= \mu_n^{-1}\sqrt{n}\|\hat{G}(\tilde{\beta} - \beta)\| \\ &= \|\hat{G}\sqrt{n}S_n^{-1\prime}(\tilde{\delta} - \delta)\| \leq \hat{M}\|\tilde{\delta} - \delta\|. \end{aligned}$$

*Q.E.D.*

LEMMA A8: *If Assumption* 5 *is satisfied*, *then Assumption* 3(iii) *and Assumption* 8(i) *are satisfied*.

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

$$\hat{\Omega}(\beta) - \Omega(\beta) = \sum_{k,\ell=1}^{p+1} \hat{F}_{k\ell}\alpha_k\alpha_\ell, \hat{F}_{k\ell}$$

$$= \sum_{i=1}^{n} Z_iZ_i'X_{ik}X_{i\ell}/n - E[Z_iZ_i'X_{ik}X_{i\ell}].$$

Then $E[X_{ik}^2X_{i\ell}^2|Z_i] \leq C$ by Lemma A4 so that

$$E[\|\hat{F}_{k\ell}\|^2] \leq CE\big[(Z_i'Z_i)^2E[X_{ik}^2X_{i\ell}^2|Z_i]\big]/n \leq CE[(Z_i'Z_i)^2]/n \longrightarrow 0.$$

Then $\sup_{\beta\in B}\|\hat{\Omega}(\beta) - \Omega(\beta)\| \xrightarrow{p} 0$ follows by $B$ bounded. The other parts of Assumption 8(i) follow similarly upon noting that

$$\hat{\Omega}^k(\beta) - \Omega^k(\beta) = \sum_{\ell=1}^{p+1} \hat{F}_{k\ell}\alpha_\ell, \hat{\Omega}^{k,\ell}(\beta) - \Omega^{k,\ell}(\beta)$$

$$= \hat{F}_{k\ell}, \hat{\Omega}^{k\ell}(\beta) = \Omega^{k\ell}(\beta) = 0. \qquad \textit{Q.E.D.}$$

LEMMA A9: *If Assumption* 5 *is satisfied*, *then Assumption* 3(iv) *and Assumption* 8(ii) *are satisfied*.

PROOF: Let $\tilde{\Sigma}_i = E[X_iX_i'|Z_i]$, which is bounded by Lemma A5. Then by $\alpha = (1, -\beta)$ bounded on $B$ we have $|\tilde{\alpha}'\tilde{\Sigma}_i\tilde{\alpha} - \alpha'\tilde{\Sigma}_i\alpha| \leq C\|\tilde{\beta} - \beta\|$. Also, $E[(a'Z_i)^2] = a'E[Z_iZ_i']a = \|a\|^2$. Therefore,

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

We also have

$$\begin{aligned}|a'\Omega^k(\tilde{\beta})b - a'\Omega^k(\beta)b| &= \big|2E\big[(a'Z_i)(b'Z_i)E[x_{ik}X_i'(\tilde{\alpha} - \alpha)|Z_i]\big]\big| \\ &\leq CE\big[|a'Z_i||b'Z_i|E[|x_{ij}|\|X_i\||Z_i]\big]\|\tilde{\beta} - \beta\| \\ &\leq C\|a\|\|b\|\|\tilde{\beta} - \beta\|.\end{aligned}$$

The other parts of Assumption 8(ii) follow by $\Omega^{k,\ell}(\beta)$ and $\Omega^{k\ell}(\beta)$ not depending on $\beta$. *Q.E.D.*

PROOF OF THEOREM 2: The result will follow by Theorem 1 upon showing that Assumptions 2 and 3 are true. We now verify Assumption 2. Assumption 2(i) holds by hypothesis. For Assumption 2(ii), note that by $G = -E[Z_i z_i']S_n'/\sqrt{n}$,

$$\mu_n^{-1}\sqrt{n}\bar{g}(\beta) = \sqrt{n}G(\beta - \beta_0)/\mu_n = -\sqrt{n}GS_n^{-1\prime}\delta.$$

Then by $nS_n^{-1}G'GS_n^{-1\prime} \geq CnS_n^{-1}G'\Omega^{-1}GS_n^{-1\prime}$ and Assumption 1 we have

$$\mu_n^{-1}\sqrt{n}\|\bar{g}(\beta)\| = (\delta'[nS_n^{-1}G'GS_n^{-1\prime}]\delta)^{1/2} \geq C\|\delta\|.$$

Next, let $\hat{R} = \sum_i (Z_i z_i' - E[Z_i z_i'])/n$ and note that

$$\begin{aligned}\hat{g}(\beta) &= \hat{g}(\beta_0) - \frac{1}{n}\sum_i Z_i x_i'(\beta - \beta_0) \\ &= \hat{g}(\beta_0) - \frac{1}{n}\sum_i Z_i \eta_i'(\beta - \beta_0) + \mu_n n^{-1/2}(-\hat{R} + E[Z_i z_i'])\delta.\end{aligned}$$

By Lemma A7, $\|\hat{R}\| \xrightarrow{p} 0$, so that by T and CS, w.p.a.1,

$$\left\|(-\hat{R} + E[Z_i z_i'])\delta\right\| \geq \|E[Z_i z_i']\delta\| - \|\hat{R}\delta\| \geq (C - \|\hat{R}\|)\|\delta\| \geq C\|\delta\|.$$

Also, as previously discussed, $\mu_n^{-1}\sqrt{n}\|\hat{g}(\beta_0)\| = O_p(1)$ and by Lemma A7, $\mu_n^{-1}\sqrt{n}\|\sum_i Z_i\eta_i'/n\| = O_p(1)$, so that by $B$ compact,

$$\hat{M} = \mu_n^{-1}\sqrt{n}\sup_{\beta\in B}\left\|\hat{g}(\beta_0) - \frac{1}{n}\sum_i Z_i\eta_i'(\beta - \beta_0)\right\| = O_p(1).$$

Then by T it follows that w.p.a.1 for all $\beta \in B$,

$$\|\delta\| \leq C\left\|(-\hat{R} + E[Z_i z_i'])\delta\right\| \leq \mu_n^{-1}\sqrt{n}\|\hat{g}(\beta)\| + \hat{M},$$

giving Assumption 2(iii).

Next, Assumption 3(i) holds by Lemma A5 and $E[(Z_i'Z_i)^2]/n \longrightarrow 0$, (ii) holds by Lemma A6, (iii) holds by Lemma A9, (iv) holds by Lemma A8, and (v) holds by Lemma A7. *Q.E.D.*

### A.3. *Asymptotic Normality*

The next result is a general result on asymptotic normality of the sum of a linear and a quadratic form. Let $X_i$ denote a scalar random variable where we also suppress dependence on $n$, let $Z_i$ and $Y_i$ be $m \times 1$ random vectors as in Lemma A1, and let $\Psi = \Sigma_{ZZ}\Sigma_{YY} + \Sigma_{ZY}^2$, $\bar{\xi}_Z = \xi_{\max}(\Sigma_{ZZ})$, and $\bar{\xi}_Y = \xi_{\max}(\Sigma_{YY})$.

LEMMA A10: *If* $(X_i, Y_i, Z_i)$ $(i = 1, \ldots, n)$ *are i.i.d.*, $E[X_i] = 0$, $E[Z_i] = E[Y_i] = 0$, $\Sigma_{ZZ}$ *and* $\Sigma_{YY}$ *exist*, $nE[X_i^2] \to A$, $n^2 \operatorname{tr}(\Psi) \to \Lambda$, $nE[X_i^4] \longrightarrow 0$, $mn^4\bar{\xi}_Z^2\bar{\xi}_Y^2 \longrightarrow 0$, $n^3(\bar{\xi}_Z^2 E[\|Y_i\|^4] + \bar{\xi}_Y^2 E[\|Z_i\|^4]) \longrightarrow 0$, *and* $n^2 E[\|Y_i\|^4] \times E[\|Z_i\|^4] \longrightarrow 0$, *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 $j < i$, let $\psi_{ij} = Z_i'Y_j + Z_j'Y_i$. Note that

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

We have

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

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

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

Therefore,

$$s_n = \sum_{i=2}^n E[(X_i + B_{in})^2] = (n-1)E[X_i^2] + 2\sum_{i=2}^n (i-1)\operatorname{tr}(\Psi) \tag{A.4}$$
$$= \frac{n-1}{n} nE[X_i^2] + \left(\frac{n^2 - n}{n^2}\right) n^2 \operatorname{tr}(\Psi) \longrightarrow A + \Lambda.$$

Next, note that

$$E[B_{in}^2 | w_{i-1}, \ldots, w_1] = T_{1i} + T_{2i} + 2T_{3i},$$

$$T_{1i} = \left(\sum_{j<i} Z_j'\right)\Sigma_{YY}\left(\sum_{j<i} Z_j\right),$$

$$T_{2i} = \left(\sum_{j<i} Y_j'\right)\Sigma_{ZZ}\left(\sum_{j<i} Y_j\right), \quad T_{3i} = \left(\sum_{j<i} Y_j'\right)\Sigma_{ZY}\left(\sum_{j<i} Z_j\right).$$

We also have

$$T_{3i} - E[T_{3i}] = T_{31i} + T_{32i} + T_{33i},$$

$$T_{31i} = \sum_{j<i} R_j, \quad R_j = [Y_j'\Sigma_{ZY}Z_j - \operatorname{tr}(\Sigma_{ZY}^2)],$$

$$T_{32i} = \sum_{k<i} S_k, \quad S_k = \left(\sum_{j<k} Y_j\right)'\Sigma_{ZY}Z_k, \quad T_{33i} = \sum_{j<k<i} Y_k'\Sigma_{ZY}Z_j.$$

By $E[(Y_i', Z_i')'(Y_i', Z_i')']$ being positive semi-definite it follows that $|Y_j'\Sigma_{ZY}Z_j| \leq (Y_j'\Sigma_{ZZ}Y_j + Z_j'\Sigma_{YY}Z_j)/2$. Note that

$$\begin{aligned} E[(Y_j'\Sigma_{ZY}Z_j)^2] &\leq CE[(Y_j'\Sigma_{ZZ}Y_j)^2] + CE[(Z_j'\Sigma_{YY}Z_j)^2] \\ &\leq C\bar{\xi}_Z^2 E[\|Y_j\|^4] + C\bar{\xi}_Y^2 E[\|Z_j\|^4]. \end{aligned}$$

Note that $\sum_{i=2}^n T_{31i} = \sum_{i=2}^n (n-i+1)R_i$ so that

$$\begin{aligned} E\left[\left(\sum_{i=2}^n T_{31i}\right)^2\right] &\leq E[(Y_j'\Sigma_{ZY}Z_j)^2]\sum_{i=2}^n (n-i+1)^2 \\ &\leq Cn^3\left\{\bar{\xi}_Z^2 E[\|Y_j\|^4] + \bar{\xi}_Y^2 E[\|Z_j\|^4]\right\} \longrightarrow 0, \end{aligned}$$

so that $\sum_{i=2}^n T_{31i} \xrightarrow{p} 0$ by M. We also have

$$\begin{aligned} E[Y_i'\Sigma_{ZY}\Sigma_{ZZ}\Sigma_{YZ}Y_i] &\leq \bar{\xi}_Z E[Y_i'\Sigma_{ZY}\Sigma_{YZ}Y_i] = \bar{\xi}_Z \operatorname{tr}(\Sigma_{YZ}\Sigma_{YY}\Sigma_{ZY}) \\ &\leq \bar{\xi}_Z\bar{\xi}_Y \operatorname{tr}(\Sigma_{YZ}\Sigma_{ZY}) \leq \bar{\xi}_Z^2\bar{\xi}_Y \operatorname{tr}(\Sigma_{YZ}\Sigma_{ZZ}^{-1}\Sigma_{ZY}) \\ &\leq \bar{\xi}_Z^2\bar{\xi}_Y \operatorname{tr}(\Sigma_{YY}) \leq m\bar{\xi}_Z^2\bar{\xi}_Y^2, \end{aligned}$$

so that $E[S_i^2] \leq (i-1)m\bar{\xi}_Z^2\bar{\xi}_Y^2$. In addition $E[S_i|w_{i-1}, \ldots, w_1] = 0$, so that

$$\begin{aligned} E\left[\left(\sum_{i=3}^n T_{32i}\right)^2\right] &= E\left[\left\{\sum_{i=3}^n (n-i+1)S_i\right\}^2\right] \\ &= \sum_{i=3}^n (n-i+1)^2 E[S_i^2] \end{aligned}$$

$$\leq \sum_{i=3}^{n}(n-i+1)^2(i-1)m\bar{\xi}_Z^2\bar{\xi}_Y^2$$
$$\leq mn^4\bar{\xi}_Z^2\bar{\xi}_Y^2 \longrightarrow 0$$

and hence $\sum_{i=3}^{n} T_{32i} \xrightarrow{p} 0$. It follows analogously that $\sum_{i=3}^{n} T_{33i} \xrightarrow{p} 0$, so by T, $\sum_{i=3}^{n}\{T_{3i} - E[T_{3i}]\} \xrightarrow{p} 0$. By similar arguments we have $\sum_{i=2}^{n}\{T_{ri} - E[T_{ri}]\} \xrightarrow{p} 0$ $(r = 1, 2)$, so by T,

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

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

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

Therefore,

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

Then by M, we have

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

By T it then follows that

$$\text{(A.5)}\qquad \sum_{i=2}^{n}\big\{E[(X_i+B_{in})^2 \mid w_{i-1},\dots,w_1]-E[(X_i+B_{in})^2]\big\}$$
$$=\sum_{i=2}^{n}(E[B_{in}^2 \mid w_{i-1},\dots,w_1]-E[B_{in}^2])$$
$$+2\sum_{i=2}^{n}E[X_iB_{in} \mid w_{i-1},\dots,w_1]\xrightarrow{p}0.$$

Next, note that

$$\sum_{i=2}^{n}E\Big[\Big(\sum_{j<i}Y_j'Z_i\Big)^4\Big]$$
$$=\sum_{i=2}^{n}\sum_{j,k,\ell,m<i}E[Y_j'Z_iY_k'Z_iY_\ell'Z_iY_m'Z_i]$$
$$=\sum_{i=2}^{n}\Big\{3\sum_{j\neq k<i}E[Z_i'Y_jY_j'Z_iZ_i'Y_kY_k'Z_i]+\sum_{j<i}E[(Z_i'Y_j)^4]\Big\}$$
$$=E[(Z_1'\Sigma_{YY}Z_1)^2]\sum_{i=2}^{n}3(i-1)(i-2)+E[(Z_1'Y_2)^4]\sum_{i=2}^{n}(i-1)$$
$$\le n^3\bar{\xi}_Y^2E[\|Z_i\|^4]+n^2E[\|Z_i\|^4]E[\|Y_i\|^4]\longrightarrow 0.$$

It follows similarly that $\sum_{i=2}^{n}E[(\sum_{j<i}Z_j'Y_i)^4]\longrightarrow 0$. Then by T,

$$\sum_{i=2}^{n}E[B_{in}^4]\le\sum_{i=2}^{n}\Big\{CE\Big[\Big(\sum_{j<i}Y_j'Z_i\Big)^4\Big]+CE\Big[\Big(\sum_{j<i}Z_j'Y_i\Big)^4\Big]\Big\}\longrightarrow 0.$$

Therefore,

$$\text{(A.6)}\qquad \sum_{i=2}^{n}E[(X_i+B_{in})^4]\le CnE[X_i^4]+C\sum_{i=1}^{n}E[B_{in}^4]\to 0.$$

The conclusion then follows from eqs. (A.4), (A.5), and (A.6) and the martingale central limit theorem applied to $\sum_{i=2}^{n}(X_i+B_{in})$. *Q.E.D.*

We again consider the parameterization where $\delta=S_n'(\beta-\beta_0)/\mu_n$ and $\beta=\beta_0+\mu_nS_n^{-1\prime}\delta$. We will let a $\delta$ subscript denote derivatives with respect to $\delta$, for example, so that $g_{i\delta_k} = \partial g_i(0)/\partial\delta_k = G_i S_n^{-1\prime} e_k \mu_n$, where $e_k$ is the $k$th unit vector. Also let $\tilde{\Omega} = \hat{\Omega}(\beta_0)$, $\tilde{\Omega}^k = \sum_{i=1}^n g_i g_{i\delta_k}'/n$, $\Omega^k = E[\tilde{\Omega}^k]$, $\tilde{B}^k = \tilde{\Omega}^{-1}\tilde{\Omega}^k$, and $B^k = \Omega^{-1}\Omega^k$.

LEMMA A11: *If Assumptions* 1–4 *and* 6–9 *are satisfied*, *then*

$$\sqrt{m}\|\tilde{\Omega} - \Omega\| \xrightarrow{p} 0, \quad \mu_n\sqrt{m}\|\tilde{\Omega}^k - \Omega^k\| \xrightarrow{p} 0,$$
$$\sqrt{m}\|\tilde{B}^k - B^k\| \xrightarrow{p} 0.$$

PROOF: Note that $\mu_n S_n^{-1}$ is bounded, so that $\|g_{i\delta_k}\| \le C\|G_i\|$. Then by standard arguments and Assumption 6,

$$E[m\|\tilde{\Omega} - \Omega\|^2] \le CmE[\|g_i\|^4]/n \longrightarrow 0,$$
$$E[m\|\tilde{\Omega}^k - \Omega^k\|^2] \le CmE\big[\big\|g_{i\delta_k}\big\|^2\|g_i\|^2\big]/n \longrightarrow 0,$$

so the first two conclusions hold by M. Also, note that $\Omega^{k\prime}\Omega^k \le C\Omega^{k\prime}\Omega^{-1}\Omega^k \le CE[g_{i\delta_k} g_{i\delta_k}']$, so that by Assumption 6, $\xi_{\max}(\Omega^{k\prime}\Omega^k) \le C$. Also, $B^{k\prime}B^k \le C\Omega^{k\prime}\Omega^k \le CE[g_{i\delta_k} g_{i\delta_k}']$. Then w.p.a.1,

$$\sqrt{m}\|\tilde{B}^k - B^k\| \le \sqrt{m}\|(\tilde{\Omega}^{k\prime} - \Omega^{k\prime})\tilde{\Omega}^{-1}\| + \sqrt{m}\|B^{k\prime}(\Omega - \tilde{\Omega})\tilde{\Omega}^{-1}\|$$
$$\le C\sqrt{m}\|\tilde{\Omega}^k - \Omega^k\| + C\sqrt{m}\|\tilde{\Omega} - \Omega\| \xrightarrow{p} 0. \qquad Q.E.D.$$

LEMMA A12: *If Assumptions* 1–4 *and* 6–9 *are satisfied*, *then*

$$nS_n^{-1}\frac{\partial\hat{Q}(\beta_0)}{\partial\beta} = \mu_n^{-1} n \frac{\partial\hat{Q}(0)}{\partial\delta} \xrightarrow{d} N(0, H + \Lambda) = N(0, HVH).$$

PROOF: Let $\tilde{g} = \hat{g}(\beta_0)$, $\tilde{g}_{\delta_k} = \partial\hat{g}(0)/\partial\delta_k = \sum_i G_i S_n^{-1\prime} e_k \mu_n/n$, $\bar{g}_{\delta_k} = E[\partial g_i(0)/\partial\delta_k] = GS_n^{-1\prime} e_k \mu_n$, $\hat{U}^k = \tilde{g}_{\delta_k} - \bar{g}_{\delta_k} - \tilde{B}^{k\prime}\tilde{g}$, and let $\tilde{\lambda}$ be as defined in Lemma A3. Consider an expansion $\rho_1(\tilde{\lambda}' g_i) = -1 - \tilde{\lambda}' g_i + \rho_3(\bar{v}_i)(\tilde{\lambda}' g_i)^2/2$, where $|\bar{v}_i| \le |\tilde{\lambda}' g_i|$. By the envelope theorem and by $\hat{Q}(\delta) = \hat{Q}(\beta_0 + \mu_n S_n^{-1\prime}\delta)$,

$$ne_k' S_n^{-1}\,\partial\hat{Q}(\beta_0)/\partial\beta = n[\partial\hat{Q}(\beta_0)/\partial\beta]' S_n^{-1\prime} e_k = \mu_n^{-1} n(\partial\hat{Q}/\partial\delta_k)(0)$$
$$= \mu_n^{-1}\sum_i \tilde{\lambda}' g_{i\delta_k}\rho_1(\tilde{\lambda}' g_i)$$
$$= -\mu_n^{-1} n\tilde{g}_{\delta_k}'\tilde{\lambda} - \mu_n^{-1} n\tilde{\lambda}'\tilde{\Omega}^k\tilde{\lambda} + \hat{r},$$
$$\hat{r} = \mu_n^{-1}\sum_i \tilde{\lambda}' g_{i\delta_k}\rho_3(\bar{v}_i)(\tilde{\lambda}' g_i)^2/2.$$

By Lemma A3, $\|\tilde{\lambda}\| = O_p(\sqrt{m/n})$. Note that either $\beta$ is the CUE or $\max_{i\le n}|\bar{v}_i| \le \|\tilde{\lambda}\|\hat{b}$ for $\hat{b} = \max_{i\le n}\|g_i\|$, and that $\hat{b} = O_p(n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma})$ by a standard result. Therefore, by Assumption 9, either $\hat{\beta}$ is the CUE or $\max_{i\le n}|\bar{v}_i| \le O_p(\sqrt{m/n})\hat{b} = O_p(n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma}\sqrt{m/n}) \xrightarrow{p} 0$. It follows that $\max_{i\le n}\rho_3(\bar{\xi}_i' g_i) \le C$ w.p.a.1 and, by $\xi_{\max}(\tilde{\Omega}) = O_p(1)$, $\sqrt{m}/\mu_n \le C$, and by Assumption 9 that either $\hat{r} = 0$ for the CUE or

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

As in Lemma A3, w.p.a.1 $\tilde{\lambda}$ satisfies the first-order conditions

$$\sum_i \rho_1(\tilde{\lambda}' g_i)g_i/n = 0.$$

Plugging in the expansion for $\rho_1(\tilde{\lambda}' g_i)$ and solving gives

$$\tilde{\lambda} = -\tilde{\Omega}^{-1}\tilde{g} + \hat{R}, \quad \hat{R} = \tilde{\Omega}^{-1}\sum_i \rho_3(\bar{v}_i)g_i(\tilde{\lambda}' g_i)^2/n.$$

Either $\hat{R} = 0$ for the CUE or by $\xi_{\max}(\tilde{\Omega}^{-1}) \le C$ and $\xi_{\max}(\tilde{\Omega}) \le C$ w.p.a.1,

$$\|\hat{R}\| \le C\max_{i\le n}\|g_i\|\tilde{\lambda}'\tilde{\Omega}\tilde{\lambda} \le C\hat{b}\|\tilde{\lambda}\|^2 = O_p\big(n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma}m/n\big).$$

Now, plug $\hat{\lambda}$ back into the expression for $\partial\hat{Q}(0)/\partial\delta_k$ to obtain

$$\mu_n^{-1}n\frac{\partial\hat{Q}}{\partial\delta_k}(0) = \mu_n^{-1}n\tilde{g}_{\delta_k}'\tilde{\Omega}^{-1}\tilde{g} - \mu_n^{-1}n\tilde{g}'\tilde{B}^k\tilde{\Omega}^{-1}\tilde{g} + \hat{r}$$
$$+ \mu_n^{-1}n\tilde{g}_{\delta_k}'\hat{R} - \mu_n^{-1}n\hat{R}'\tilde{\Omega}^k\hat{R} + \mu_n^{-1}n\hat{R}(\tilde{\Omega}^k + \tilde{\Omega}^{k\prime})\tilde{\Omega}^{-1}\tilde{g}.$$

Note that by Assumption 6 and $\mu_n S_n^{-1}$ bounded, $E[\|g_{i\delta_k}\|^2] = \mathrm{tr}(E[g_{i\delta_k}g_{i\delta_k}']) \le Cm\xi_{\max}(E[G_iG_i']) \le Cm$. Therefore, $\|\tilde{g}_{\delta_k} - \bar{g}_{\delta_k}\| = O_p(\sqrt{m/n})$. We also have $\|\mu_n^{-1}\sqrt{n}\bar{g}_{\delta_k}\| \le \|\sqrt{n}GS_n^{-1\prime}\| \le C$, so that $\|\bar{g}_{\delta_k}\| = O(\mu_n/\sqrt{n})$. Therefore, by $\sqrt{m}/\mu_n \le C$ and T, $\|\tilde{g}_{\delta_k}\| = O_p(\mu_n/\sqrt{n})$, so by CS,

$$\big|\mu_n^{-1}n\tilde{g}_{\delta_k}'\hat{R}\big| \le \mu_n^{-1}n\|\tilde{g}_{\delta_k}\|\|\hat{R}\| = O_p\big(\sqrt{n}n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma}m/n\big) \xrightarrow{p} 0.$$

Let $\tilde{\Omega}^{k,k} = \sum_i g_{i\delta_k}g_{i\delta_k}'/n$ and $\Omega^{k,k} = E[g_{i\delta_k}g_{i\delta_k}']$. By Assumption 6 and M we have $\|\tilde{\Omega}^{k,k} - \Omega^{k,k}\| \xrightarrow{p} 0$, so by Lemma A0, Assumption 6, and $\mu_n S_n^{-1}$ bounded, w.p.a.1

$$\xi_{\max}(\tilde{\Omega}^{k,k}) \le \xi_{\max}(\Omega^{k,k}) + 1 \le C\xi_{\max}(E[G_iG_i']) + 1.$$

Therefore, $\hat{M} = \sqrt{\xi_{\max}(\tilde{\Omega})\xi_{\max}(\tilde{\Omega}^{k,k})} = O_p(1)$, so that for any $a$, $b$, by CS,

$$|a'\tilde{\Omega}^k b| \le [a'\tilde{\Omega} a b'\tilde{\Omega}^{k,k} b]^{1/2} \le \hat{M}\|a\|\|b\|.$$

Then

$$|\mu_n^{-1} n\hat{R}'\tilde{\Omega}^k\hat{R}| \le \hat{M}\mu_n^{-1} n\|\hat{R}\|^2 = O_p\big(\mu_n^{-1}\{n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma} m/\sqrt{n}\}^2\big) \xrightarrow{p} 0.$$

We also have $\|\tilde{\Omega}^{-1}\tilde{g}\| = O_p(\sqrt{m/n})$, so that by $\sqrt{m}/\mu_n \le C$,

$$|\mu_n^{-1} n\hat{R}(\tilde{\Omega}^k + \tilde{\Omega}^{k\prime})\tilde{\Omega}^{-1}\tilde{g}| \le C\hat{M}\mu_n^{-1} n\|\hat{R}\|\|\tilde{\Omega}^{-1}\tilde{g}\| = O_p\big(n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma} m/\sqrt{n}\big) \xrightarrow{p} 0.$$

By T it now follows that

$$\mu_n^{-1} n\frac{\partial\hat{Q}}{\partial\delta_k}(0) = \mu_n^{-1} n\tilde{g}_{\delta_k}'\tilde{\Omega}^{-1}\tilde{g} - \mu_n^{-1} n\tilde{g}'\tilde{B}^k\tilde{\Omega}^{-1}\tilde{g} + o_p(1) = \bar{g}_{\delta_k}'\tilde{\Omega}^{-1}\tilde{g} + \hat{U}^{k\prime}\tilde{\Omega}^{-1}\tilde{g} + o_p(1),$$

where $\hat{U}^k = \tilde{g}_{\delta_k} - \bar{g}_{\delta_k} - \tilde{B}^{k\prime}\tilde{g}$. For $B^k$ defined preceding Lemma A11, let $\tilde{U}^k = \tilde{g}_{\delta_k} - \bar{g}_{\delta_k} - B^{k\prime}\tilde{g}$. Note that $n\|\tilde{g}\|^2 = O_p(m)$. By Lemma A11 and $m/\mu_n^2 \le C$ we have

$$n\mu_n^{-1}|(\hat{U}^{k\prime}\tilde{\Omega}^{-1} - \tilde{U}^{k\prime}\tilde{\Omega}^{-1})\tilde{g}| \le Cn\mu_n^{-1}|\tilde{g}'(\tilde{B}^k - B^k)\tilde{\Omega}^{-1}\tilde{g}| \le Cn\mu_n^{-1}\|\tilde{g}\|^2\|\tilde{B}^k - B^k\| \xrightarrow{p} 0.$$

Note also that by the usual properties of projections and Assumption 6, $nE[\|\tilde{U}^k\|^2] \le CE[\|g_{i\delta_k}\|^2] \le Cm$, so that $n\mu_n^{-1}|\tilde{U}^{k\prime}(\Omega^{-1} - \tilde{\Omega}^{-1})\tilde{g}| \xrightarrow{p} 0$. Similarly we have $\mu_n^{-1}\bar{g}_{\delta_k}'(\tilde{\Omega}^{-1} - \Omega^{-1})\tilde{g} \xrightarrow{p} 0$, so that by T

$$n\mu_n^{-1}\frac{\partial\hat{Q}}{\partial\delta_k}(0) = n\mu_n^{-1}\big(\bar{g}_{\delta_k} + \tilde{U}^k\big)'\Omega^{-1}\tilde{g} + o_p(1).$$

It is straightforward to check that for $U_i$ defined in Section 2 we have

$$\tilde{U}^k = n^{-1}\sum_{i=1}^n U_i S_n^{-1\prime} e_k\mu_n, \quad \bar{g}_{\delta_k} = GS_n^{-1\prime} e_k\mu_n.$$

Then stacking over $k$ gives

$$\text{(A.7)} \qquad n\mu_n^{-1}\frac{\partial\hat{Q}}{\partial\delta}(0) = nS_n^{-1}\left[G'\Omega^{-1}\tilde{g} + n^{-1}\sum_{i=1}^{n} U_i'\Omega^{-1}\tilde{g}\right] + o_p(1).$$

For any vector $\lambda$ with $\|\lambda\| = 1$, let $X_i = \lambda' S_n^{-1} G'\Omega^{-1} g_i$, $Y_i = \Omega^{-1} g_i$, $Z_i = U_i S_n^{-1\prime}\lambda/n$, and $A = \lambda' H\lambda$. Then from the previous equation we have

$$n\mu_n^{-1}\lambda'\frac{\partial\hat{Q}}{\partial\delta}(0) = \sum_{i=1}^{n} X_i + \sum_{i,j=1}^{n} Y_i' Z_i + o_p(1).$$

Note that $E[Z_i'Y_i] = 0$ by each component of $U_i$ being uncorrelated with every component of $g_i$. Also, by $\|S_n^{-1}\| \le C/\mu_n$,

$$\begin{aligned} nE[|Y_i'Z_i|^2] &\le CE[\|g_i'\Omega^{-1}U_i\|^2]/n\mu_n^2 \\ &\le C\big(E[\|g_i\|^4] + E[\|G_i\|^4]\big)/n\mu_n^2 \longrightarrow 0. \end{aligned}$$

Then $\sum_{i=1}^{n} Z_i'Y_i \xrightarrow{p} 0$ by M. Then by eq. (A.7),

$$n\mu_n^{-1}\lambda'\frac{\partial\hat{Q}(0)}{\partial\delta} = \sum_{i=1}^{n} X_i + \sum_{i\neq j} Z_i'Y_j + o_p(1).$$

Now apply Lemma A10. Note that $\Sigma_{YY} = \Omega^{-1}$ and $\Sigma_{ZY} = 0$, so that $\Psi = \Sigma_{ZZ}\Sigma_{YY} = n^{-2}E[U_i S_n^{-1\prime}\lambda\lambda' S_n^{-1} U_i']\Omega^{-1}$. By Assumption 1 and the hypothesis of Theorem 3, we have

$$\begin{aligned} &nE[X_i^2] = n\lambda' S_n^{-1} G'\Omega^{-1} G S_n^{-1\prime}\lambda \longrightarrow \lambda' H\lambda = A, \\ &n^2\,\mathrm{tr}(\Psi) = \lambda' S_n^{-1} E[U_i'\Omega^{-1}U_i] S_n^{-1\prime}\lambda \longrightarrow \lambda'\Lambda\lambda. \end{aligned}$$

Also, note that $\xi_{\max}(S_n^{-1\prime}\lambda\lambda' S_n^{-1}) \le C/\mu_n^2$, so that $\bar{\xi}_Z \le C/\mu_n^2 n^2$. We also have $\|\sqrt{n}S_n^{-1}G'\Omega^{-1}\| \le C$ by Assumption 1 and $\xi_{\max}(\Omega^{-1}) \le C$. Then

$$\begin{aligned} &nE[|X_i|^4] \le nE[\|\lambda'\sqrt{n}S_n^{-1}G'\Omega^{-1}g_i\|^4]/n^2 \le CE[\|g_i\|^4]/n \longrightarrow 0, \\ &mn^4\bar{\xi}_Y^2\bar{\xi}_Z^2 \le Cmn^4/(\mu_n^2 n^2)^2 \le Cm/\mu_n^4 \longrightarrow 0, \\ &n^3\big(\bar{\xi}_Z^2 E[\|Y_i\|^4] + \bar{\xi}_Y^2 E[\|Z_i\|^4]\big) \\ &\quad \le n^3 C\big(E[\|g_i\|^4] + E[\|G_i\|^4]\big)/\mu_n^4 n^4 \longrightarrow 0, \\ &n^2 E[\|Y_i\|^4]E[\|Z_i\|^4] \\ &\quad \le n^2 CE[\|g_i\|^4]\big(E[\|g_i\|^4] + E[\|G_i\|^4]\big)/\mu_n^4 n^4 \longrightarrow 0. \end{aligned}$$

The conclusion then follows by the conclusion of Lemma A10 and the Cramer–Wold device. *Q.E.D.*

LEMMA A13: *If Assumptions* 1–4 *and* 6–9 *are satisfied*, *then there is an open convex set* $N_n$ *such that* $0 \in N_n$ *and w.p.a.1* $\hat{\delta} \in N_n$, $\hat{Q}(\delta)$ *is twice continuous differentiable on* $N_n$, *and for any* $\bar{\delta}$ *that is an element of* $N_n$ *w.p.a.*1,

$$nS_n^{-1}[\partial^2 \hat{Q}(\bar{\delta})/\partial\beta\,\partial\beta']S_n^{-1\prime} = \mu_n^{-2} n\,\partial^2\hat{Q}(\bar{\delta})/\partial\delta\,\partial\delta' \xrightarrow{p} H.$$

PROOF: By Theorem 1, $\hat{\delta} \xrightarrow{p} 0$. Then there is $\zeta_n \longrightarrow 0$ such that w.p.a.1 $\bar{\delta} \in N_n = \{\delta : \|\delta\| < \zeta_n\}$. By Assumption 3, for all $\delta \in N_n$,

$$\mu_n^{-1}\sqrt{n}\|\hat{g}(\delta) - \hat{g}(0)\| \leq \hat{M}\|\delta\| \leq \hat{M}\zeta_n \xrightarrow{p} 0.$$

As previously shown, $\mu_n^{-1}\sqrt{n}\|\hat{g}(0)\| = O_p(\mu_n^{-1}\sqrt{n}\sqrt{m/n}) = O_p(1)$, so $\sup_{\delta\in N_n}\mu_n^{-1}\sqrt{n}\|\hat{g}(\delta)\| = O_p(1)$ by T. Now let $\tau_n$ go to zero slower than $\mu_n/\sqrt{n}$ but faster than $n^{-1/\gamma}E[\sup_{\beta\in B}\|g_i(\beta)\|^\gamma]^{-1/\gamma}$, which is possible by Assumption 9, and let $L_n = \{\lambda : \|\lambda\| \leq \tau_n\}$. Then $\max_{i\leq n}\sup_{\beta\in B,\lambda\in L_n}|\lambda' g_i(\beta)| \xrightarrow{p} 0$ similarly to the proof of Lemma A3. For all $\delta \in N_n$, let $\hat{\lambda}(\delta) = \arg\max_{\lambda\in L_n}\hat{P}(\delta,\lambda)$. By an argument similar to the proof of Lemma A3, an expansion of $S(\delta, \hat{\lambda}(\delta))$ around $\lambda = 0$ gives

$$\begin{aligned}
0 &= \hat{P}(\delta, 0) \leq \hat{P}(\delta, \hat{\lambda}(\delta)) \\
&= \hat{g}(\delta)'\hat{\lambda}(\delta) + \frac{1}{2}\hat{\lambda}(\delta)'\left[\sum_{i=1}^{n}\rho_2(\lambda' g_i(\delta))g_i(\delta)g_i(\delta)'/n\right]\hat{\lambda}(\delta) \\
&\leq \|\hat{g}(\delta)\|\|\hat{\lambda}(\delta)\| - C\|\hat{\lambda}(\delta)\|^2.
\end{aligned}$$

Adding $C\|\hat{\lambda}(\delta)\|^2$ and dividing through by $C\|\hat{\lambda}(\delta)\|$ gives

$$\text{(A.8)} \qquad \|\hat{\lambda}(\delta)\| \leq C\|\hat{g}(\delta)\| \leq C\sup_{\delta\in N_n}\|\hat{g}(\delta)\| = O_p(\mu_n/\sqrt{n}).$$

It follows that w.p.a.1 $\hat{\lambda}(\delta) \in \operatorname{int} L_n$ for all $\delta \in N_n$. Since a local maximum of a concave function is a global maximum, w.p.a.1 for all $\delta \in N_n$,

$$\hat{Q}(\delta) = \hat{P}(\delta, \hat{\lambda}(\delta)).$$

Furthermore w.p.a.1 the first-order conditions

$$\sum_{i=1}^{n}\rho_1(\hat{\lambda}(\delta)' g_i(\delta))g_i(\delta)/n = 0$$

will be satisfied for all $\delta$, so that by the implicit function theorem, $\hat{\lambda}(\delta)$ is twice continuously differentiable in $\delta \in N_n$ and hence so is $\hat{Q}(\delta)$.

Here let $\hat{g}_i = g_i(\bar{\delta})$, $\hat{g} = \hat{g}(\bar{\delta})$, $\hat{\lambda} = \hat{\lambda}(\bar{\delta})$, $\hat{\Omega} = -\sum_{i=1}^n \rho_2(\hat{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_i'/n$, $\hat{g}_{i\delta_k} = \partial g_i(\bar{\delta})/\partial\delta_k$, $\hat{g}_{\delta_k} = \partial\hat{g}(\bar{\delta})/\partial\delta_k$, and $\hat{\Omega}^k = -\sum_i \rho_2(\hat{\lambda}'\hat{g}_i)\hat{g}_i\hat{g}_{i\delta_k}'/n$. Then expanding $\rho_1(\hat{\lambda}'\hat{g}_i) = -1 + \rho_2(\bar{v}_i)\hat{\lambda}'\hat{g}_i$ for $|\bar{v}_i| \le |\hat{\lambda}'\hat{g}_i|$ and letting $\bar{\Omega}^k = -\sum_i \rho_2(\bar{v}_i)\hat{g}_i \times \hat{g}_{i\delta_k}'/n$, the implicit function theorem gives

$$\hat{\lambda}_{\delta_k} = \frac{\partial\hat{\lambda}}{\partial\delta_k}(\bar{\delta}) = \hat{\Omega}^{-1}\left[\sum_i \rho_1(\hat{\lambda}'\hat{g}_i)\frac{\hat{g}_{i\delta_k}}{n} - \frac{\hat{\Omega}^k\hat{\lambda}}{n}\right]$$
$$= -\hat{\Omega}^{-1}\big[\hat{g}_{\delta_k} + (\bar{\Omega}^{k\prime} + \hat{\Omega}^k)\hat{\lambda}\big].$$

Also, for $\bar{\Omega} = -\sum_i \rho_2(\bar{v}_i)\hat{g}_i\hat{g}_i'/n$, the first-order conditions $0 = \sum_i \rho_1(\hat{\lambda}'\hat{g}_i) \times \hat{g}_i/n = -\hat{g} - \bar{\Omega}\hat{\lambda}$ imply that

$$\hat{\lambda} = -\bar{\Omega}^{-1}\hat{g}.$$

Next, by the envelope theorem it follows that

$$\hat{Q}_{\delta_k}(\bar{\delta}) = \sum_i \rho_1(\hat{\lambda}'\hat{g}_i)\hat{\lambda}'\hat{g}_{i\delta_k}/n.$$

Let $\hat{g}_{i\delta_k\delta_\ell} = \partial^2 g_i(\hat{\delta})/\partial\delta_k\,\partial\delta_\ell$, $\hat{g}_{\delta_k\delta_\ell} = \partial^2\hat{g}(\hat{\delta})/\partial\delta_k\,\partial\delta_\ell$, $\hat{\Omega}^{k,\ell} = -\sum_i \rho_2(\hat{\lambda}'\hat{g}_i)\hat{g}_{i\delta_k} \times \hat{g}_{i\delta_\ell}'/n$, and $\bar{\Omega}^{k\ell} = -\sum_i \rho_2(\bar{v}_i)\hat{g}_i\hat{g}_{i\delta_k\delta_\ell}'/n$. Differentiating again yields

$$\hat{Q}_{\delta_k\delta_\ell}(\bar{\delta}) = \sum_i \big[\rho_1(\hat{\lambda}'g_i)\big(\hat{\lambda}_{\delta_k}'\hat{g}_{i\delta_\ell} + \hat{\lambda}'\hat{g}_{i\delta_k\delta_\ell}\big)$$
$$+ \rho_2(\hat{\lambda}'\hat{g}_i)\big(\hat{\lambda}_{\delta_k}'\hat{g}_i + \hat{\lambda}'\hat{g}_{i\delta_k}\big)\hat{\lambda}'\hat{g}_{i\delta_\ell}\big]/n$$
$$= n^{-1}\sum_i \big[(-1 + \rho_2(\bar{v}_i)\hat{\lambda}'\hat{g}_i)\big(\hat{\lambda}_{\delta_k}'\hat{g}_{i\delta_\ell} + \hat{\lambda}'\hat{g}_{i\delta_k\delta_\ell}\big)\big]$$
$$- \hat{\lambda}_{\delta_k}'\hat{\Omega}^\ell\hat{\lambda} - \hat{\lambda}'\hat{\Omega}^{k,\ell}\hat{\lambda}$$
$$= -\hat{\lambda}_{\delta_k}'\hat{g}_{\delta_\ell} - \hat{\lambda}'\hat{g}_{\delta_k\delta_\ell} - \hat{\lambda}'(\bar{\Omega}^\ell + \hat{\Omega}^{\ell\prime})\hat{\lambda}_{\delta_k} - \hat{\lambda}'(\bar{\Omega}^{k\ell} + \hat{\Omega}^{k,\ell})\hat{\lambda}.$$

Substituting in the formula for $\hat{\lambda}_{\delta_k}$ and then $\hat{\lambda}$ we obtain

$$\text{(A.9)}\qquad \hat{Q}_{\delta_k\delta_\ell}(\bar{\delta}) = \hat{g}_{\delta_k}'\hat{\Omega}^{-1}\hat{g}_{\delta_\ell} + \hat{\lambda}'(\bar{\Omega}^k + \hat{\Omega}^{k\prime})\hat{\Omega}^{-1}\hat{g}_{\delta_\ell} - \hat{\lambda}'\hat{g}_{\delta_k\delta_\ell}$$
$$+ \hat{\lambda}'(\bar{\Omega}^\ell + \hat{\Omega}^{\ell\prime})\hat{\Omega}^{-1}\hat{g}_{\delta_k}$$
$$+ \hat{\lambda}'(\bar{\Omega}^\ell + \hat{\Omega}^{\ell\prime})\hat{\Omega}^{-1}(\bar{\Omega}^{k\prime} + \hat{\Omega}^k)\hat{\lambda} - \hat{\lambda}'(\bar{\Omega}^{k\ell} + \hat{\Omega}^{k,\ell})\hat{\lambda}$$

$$
\begin{aligned}
&= \hat{g}_{\delta_k}' \hat{\Omega}^{-1} \hat{g}_{\delta_\ell} + \hat{g}' \bar{\Omega}^{-1} \hat{g}_{\delta_k \delta_\ell} - \hat{g}' \bar{\Omega}^{-1} (\bar{\Omega}^k + \hat{\Omega}^{k\prime}) \hat{\Omega}^{-1} \hat{g}_{\delta_\ell} \\
&\quad - \hat{g}' \bar{\Omega}^{-1} (\bar{\Omega}^\ell + \hat{\Omega}^{\ell\prime}) \hat{\Omega}^{-1} \hat{g}_{\delta_k} \\
&\quad + \hat{g}' \bar{\Omega}^{-1} (\bar{\Omega}^\ell + \hat{\Omega}^{\ell\prime}) \hat{\Omega}^{-1} (\bar{\Omega}^{k\prime} + \hat{\Omega}^k) \bar{\Omega}^{-1} \hat{g} \\
&\quad - \hat{g}' \bar{\Omega}^{-1} (\bar{\Omega}^{k\ell} + \hat{\Omega}^{k,\ell}) \bar{\Omega}^{-1} \hat{g}.
\end{aligned}
$$

Next, let $\breve{\Omega}^k = \sum_i \hat{g}_i \hat{g}_{i\delta_k}'/n$. Note that $|1 + \rho_2(\bar{v}_i)| \le C|\bar{v}_i| \le C|\hat{\lambda}'\hat{g}_i|$, so that by CS and M,

$$
\begin{aligned}
\|\bar{\Omega}^k - \breve{\Omega}^k\| &\le C \sum_i |\bar{v}_i| \|\hat{g}_i\| \|\hat{g}_{i\delta_k}\|/n \\
&\le \Bigl(C \sum_i \bar{v}_i^2/n\Bigr)^{1/2} \Bigl(\sum_i \|\hat{g}_i\|^2 \|\hat{g}_{i\delta_k}\|^2/n\Bigr)^{1/2} \\
&\le C(\hat{\lambda}'\hat{\Omega}\hat{\lambda})^{1/2} \Bigl[\sum_i \bigl(\|\hat{g}_i\|^4 + \|\hat{g}_{i\delta_k}\|^4\bigr)/n\Bigr]^{1/2} \\
&= O_p\bigl(\{\mu_n^2 E[d_i^4]/n\}^{1/2}\bigr) \xrightarrow{p} 0.
\end{aligned}
$$

Also, for $\Omega^k(\delta) = E[g_i(\delta) g_{i\delta_k}(\delta)']$, by Assumption 8(i) and $S_n^{-1}\mu_n$ bounded we have $\|\breve{\Omega}^k - \Omega^k(\bar{\delta})\| \xrightarrow{p} 0$. Then by T,

$$
\|\bar{\Omega}^k - \Omega^k(\bar{\delta})\| \xrightarrow{p} 0.
$$

Let $\Omega^{k,\ell}(\delta) = E[g_{i\delta_k}(\delta) g_{i\delta_\ell}(\delta)']$ and $\Omega^{k\ell}(\delta) = E[g_i(\delta) g_{i\delta_k\delta_\ell}(\delta)']$. Then it follows by arguments exactly analogous to those just given that

$$
\begin{aligned}
&\|\hat{\Omega} - \Omega(\bar{\delta})\| \xrightarrow{p} 0, \quad \|\bar{\Omega} - \Omega(\bar{\delta})\| \xrightarrow{p} 0, \quad \|\hat{\Omega}^k - \Omega^k(\bar{\delta})\| \xrightarrow{p} 0, \\
&\|\hat{\Omega}^{k,\ell} - \Omega^{k,\ell}(\bar{\delta})\| \xrightarrow{p} 0, \quad \|\bar{\Omega}^{k\ell} - \Omega^{k\ell}(\bar{\delta})\| \xrightarrow{p} 0.
\end{aligned}
$$

Next, as previously shown, $\mu_n^{-1}\sqrt{n}\|\hat{g}(\bar{\delta})\| = O_p(1)$. It follows similarly from Assumption 7 that

$$
\mu_n^{-1}\sqrt{n}\|\partial\hat{g}(\bar{\delta})/\partial\delta\| = \sqrt{n}\|\hat{G}(\bar{\beta})S_n^{-1\prime}\| = \sqrt{n}\|\hat{G}(\beta_0)S_n^{-1\prime}\| + o_p(1).
$$

Then by Assumption 6, $E[\|G_i\|^2] \le Cm$, so by M,

$$
\bigl(\sqrt{n}\bigl\|[\hat{G}(\beta_0) - G]S_n^{-1\prime}\bigr\|\bigr)^2 = O_p\bigl(E[\|G_i\|^2]\bigr)/\mu_n^2 = O_p(1).
$$

Also by Assumptions 1 and 3 we have $\sqrt{n}\|GS_n^{-1\prime}\| \le C$. Then by T and Assumption 1,

$$
\sqrt{n}\|\hat{G}(\beta_0)S_n^{-1}\| \le \sqrt{n}\bigl\|[\hat{G}(\beta_0) - G]S_n^{-1\prime}\bigr\| + \sqrt{n}\|GS_n^{-1\prime}\| = O_p(1).
$$

Then by T it follows that

$$\mu_n^{-1}\sqrt{n}\|\partial\hat{g}(\bar{\delta})/\partial\delta\| = O_p(1).$$

By similar arguments it follows by Assumption 6 that

$$\mu_n^{-1}\sqrt{n}\|\partial^2\hat{g}(\bar{\delta})/\partial\delta\,\partial\delta_k\| = O_p(1).$$

Next, for notational convenience let $\tilde{\Omega} = \Omega(\bar{\delta})$ and $\tilde{\Omega}^k = \Omega^k(\bar{\delta})$. By Assumption 2, $\xi_{\max}(\tilde{\Omega}^{-1}) \le C$, so that $\xi_{\max}(\tilde{\Omega}^{-2}) \le C$. It follows as previously that $\xi_{\max}(\bar{\Omega}^{-2}) \le C$ and $\xi_{\max}(\hat{\Omega}^{k\prime}\bar{\Omega}^{-2}\hat{\Omega}^k) \le C$ w.p.a.1, so that

$$\begin{aligned}
&\|\bar{\Omega}^{-1}\hat{\Omega}^k\bar{\Omega}^{-1} - \tilde{\Omega}^{-1}\tilde{\Omega}^k\tilde{\Omega}^{-1}\| \\
&\quad \le \|\bar{\Omega}^{-1}\hat{\Omega}^k(\bar{\Omega}^{-1} - \tilde{\Omega}^{-1})\| + \|\hat{\Omega}^{-1}(\hat{\Omega}^k - \tilde{\Omega}^k)\tilde{\Omega}^{-1}\| \\
&\qquad + \|(\hat{\Omega}^{-1} - \tilde{\Omega}^{-1})\tilde{\Omega}^k\tilde{\Omega}^{-1}\| \xrightarrow{p} 0.
\end{aligned}$$

Then by Assumption 8 it follows that

$$\begin{aligned}
&\mu_n^{-2}n\left|\hat{g}'\bar{\Omega}^{-1}\bar{\Omega}^k\hat{\Omega}^{-1}\hat{g}_{\delta_\ell} - \hat{g}'\tilde{\Omega}^{-1}\tilde{\Omega}^k\tilde{\Omega}^{-1}\hat{g}_{\delta_\ell}\right| \\
&\quad \le O_p(1)\|\bar{\Omega}^{-1}\bar{\Omega}^k\hat{\Omega}^{-1} - \tilde{\Omega}^{-1}\tilde{\Omega}^k\tilde{\Omega}^{-1}\| \xrightarrow{p} 0.
\end{aligned}$$

Therefore, we can replace $\bar{\Omega}$ and $\hat{\Omega}$ by $\tilde{\Omega}$ in the third term in eq. (A.9) without affecting its probability limit. Let $\tilde{Q}_{k,\ell}(\delta)$ denote the expression following the second equality in eq. (A.9), with $\tilde{\Omega}$ replacing $\bar{\Omega}$ and $\hat{\Omega}$ throughout. Then applying an argument similar to the one just given to each of the six terms following the second equality in eq. (A.9), it follows by T that

$$\mu_n^{-2}n|\hat{Q}_{\delta_k\delta_\ell}(\bar{\delta}) - \tilde{Q}_{k,\ell}(\bar{\delta})| \xrightarrow{p} 0.$$

Next, we will show that

$$\mu_n^{-2}n|\tilde{Q}_{k,\ell}(\bar{\delta}) - \tilde{Q}_{k,\ell}(0)| \xrightarrow{p} 0.$$

Working again with the third term, let $F(\delta) = \Omega(\delta)^{-1}\Omega^k(\delta)\Omega(\delta)^{-1}$. It follows from Assumptions 3 and 8 similarly to the previous argument that for any $a$ and $b$, $|a'[F(\bar{\delta}) - F(0)]b| \le C\|a\|\|b\|\|\bar{\delta}\|$. Also, by Assumptions 3 and 7 we have $\mu_n^{-1}\sqrt{n}\|\hat{g}(\bar{\delta}) - \hat{g}(0)\| \xrightarrow{p} 0$ and $\mu_n^{-1}\sqrt{n}\|\hat{g}_{\delta_k}(\bar{\delta}) - \hat{g}_{\delta_k}(0)\| \xrightarrow{p} 0$. It then follows by CS and T that

$$\begin{aligned}
&\mu_n^{-2}n\left|\hat{g}(\bar{\delta})'F(\bar{\delta})\hat{g}_{\delta_k}(\bar{\delta}) - \hat{g}(0)'F(0)\hat{g}_{\delta_k}(0)\right| \\
&\quad \le \mu_n^{-2}nC\big(\|\hat{g}(\bar{\delta})\|\left\|\hat{g}_{\delta_k}(\bar{\delta})\right\|\|\bar{\delta}\| + \|\hat{g}(\bar{\delta}) - \hat{g}(0)\|\left\|\hat{g}_{\delta_k}(\bar{\delta})\right\| \\
&\qquad + \|\hat{g}(0)\|\left\|\hat{g}_{\delta_k}(\bar{\delta}) - \hat{g}_{\delta_k}(0)\right\|\big) \xrightarrow{p} 0.
\end{aligned}$$

Applying a similar argument for each of the other six terms and using T gives $\mu_n^{-2}n|\tilde{Q}_{k,\ell}(\bar{\delta})-\tilde{Q}_{k,\ell}(0)|\xrightarrow{p}0$. It therefore suffices to show that $\mu_n^{-2}n\times\tilde{Q}_{k,\ell}(0)\xrightarrow{p}H_{k\ell}$.

Next, let $\Omega^k=\Omega^k(\beta_0)$, $\Omega^{k\ell}=\Omega^{k\ell}(\beta_0)$, $\Omega^{k,\ell}=\Omega^{k,\ell}(\beta_0)$, $\tilde{g}=\hat{g}(\beta_0)$, $\tilde{g}_{\delta_k}=\partial\hat{g}(0)/\partial\delta_k$, and $\tilde{g}_{\delta_k\delta_\ell}=\partial^2\hat{g}(0)/\partial\delta_\ell\,\partial\delta_k$. Note that

$$\begin{aligned}\tilde{Q}_{k,\ell}(0)&=\tilde{g}'_{\delta_k}\Omega^{-1}\tilde{g}_{\delta_\ell}+\tilde{g}'\Omega^{-1}\tilde{g}_{\delta_k\delta_\ell}-\tilde{g}'\Omega^{-1}(\Omega^k+\Omega^{k\prime})\Omega^{-1}\tilde{g}_{\delta_\ell}\\&\quad-\tilde{g}'\Omega^{-1}(\Omega^\ell+\Omega^{\ell\prime})\Omega^{-1}\tilde{g}_{\delta_k}\\&\quad+\tilde{g}'\Omega^{-1}(\Omega^\ell+\Omega^{\ell\prime})\Omega^{-1}(\Omega^{k\prime}+\Omega^k)\Omega^{-1}\tilde{g}\\&\quad-\tilde{g}'\Omega^{-1}(\Omega^{k\ell}+\Omega^{k,\ell})\Omega^{-1}\tilde{g}.\end{aligned}$$

Consider once again the third term in $\tilde{Q}_{k,\ell}(0)$, that is, $\tilde{g}'A\tilde{g}_{\delta_\ell}$, where $A=-\Omega^{-1}(\Omega^k+\Omega^{k\prime})\Omega^{-1}$. Now apply Lemma A1 with $Y_i=g_i$, $Z_i=G_iS_n^{-1\prime}\mu_ne_k$, and $a_n=\mu_n^2$ to obtain

$$\mu_n^{-2}n\tilde{g}'A\tilde{g}_{\delta_\ell}=-\operatorname{tr}(\Omega^{-1}(\Omega^k+\Omega^{k\prime})\Omega^{-1}\Omega^{\ell\prime})/\mu_n^2+o_p(1).$$

Let $H_n=nS_n^{-1}G'\Omega^{-1}GS_n^{-1\prime}$. Then applying a similar argument to each term in $\tilde{Q}_{k,\ell}(0)$ gives

$$\begin{aligned}\mu_n^{-2}n\tilde{Q}_{k,\ell}(0)&=H_{nk,\ell}+\mu_n^{-2}\operatorname{tr}\big[\Omega^{-1}\Omega^{k,\ell\prime}+\Omega^{-1}\Omega^{k\ell\prime}\\&\qquad-\Omega^{-1}(\Omega^k+\Omega^{k\prime})\Omega^{-1}\Omega^{\ell\prime}-\Omega^{-1}(\Omega^\ell+\Omega^{\ell\prime})\Omega^{-1}\Omega^{k\prime}\\&\qquad+\Omega^{-1}(\Omega^\ell+\Omega^{\ell\prime})\Omega^{-1}(\Omega^{k\prime}+\Omega^k)-\Omega^{-1}(\Omega^{k\ell}+\Omega^{k,\ell})\big]\\&\qquad+o_p(1)\\&=H_{nk,\ell}+\mu_n^{-2}\operatorname{tr}\big[\Omega^{-1}(\Omega^{k,\ell\prime}-\Omega^{k,\ell})+\Omega^{-1}(\Omega^{k\ell\prime}-\Omega^{k\ell})\\&\qquad-\Omega^{-1}(\Omega^k+\Omega^{k\prime})\Omega^{-1}\Omega^{\ell\prime}+\Omega^{-1}(\Omega^\ell+\Omega^{\ell\prime})\Omega^{-1}\Omega^k\big]\\&\qquad+o_p(1).\end{aligned}$$

By $\operatorname{tr}(AB)=\operatorname{tr}(BA)$ for any conformable matrices $A$ and $B$, we have

$$\operatorname{tr}[(\Omega^{-1}\Omega^{\ell\prime})(\Omega^{-1}\Omega^k)]=\operatorname{tr}(\Omega^{-1}\Omega^k\Omega^{-1}\Omega^{\ell\prime}).$$

Also, for a symmetric matrix $A$, $\operatorname{tr}(AB)=\operatorname{tr}(B'A)=\operatorname{tr}(AB')$, so that

$$\begin{aligned}&\operatorname{tr}(\Omega^{-1}\Omega^{k,\ell\prime})=\operatorname{tr}(\Omega^{-1}\Omega^{k,\ell}),\quad\operatorname{tr}(\Omega^{-1}\Omega^{k\ell\prime})=\operatorname{tr}(\Omega^{-1}\Omega^{k\ell}),\\&\operatorname{tr}[\Omega^{-1}(\Omega^\ell\Omega^{-1}\Omega^k)]=\operatorname{tr}[\Omega^{-1}(\Omega^{k\prime}\Omega^{-1}\Omega^{\ell\prime})].\end{aligned}$$

Then we have $\mu_n^{-2}n\tilde{Q}_{k,\ell}(0)=H_{nk,\ell}+o_p(1)$, so that the conclusion follows by T. *Q.E.D.*

LEMMA A14: *If Assumptions* 1–4 *and* 6–9 *are satisfied, then* $nS_n^{-1}\hat{D}(\hat{\beta})'\hat{\Omega}^{-1}\times \hat{D}(\hat{\beta})S_n^{-1\prime} \xrightarrow{p} H+\Lambda = HVH$.

PROOF: For $\hat{g}_i = g_i(\hat{\beta})$, an expansion like those above gives $\rho_1(\hat{\lambda}'\hat{g}_i) = -1 - \hat{\lambda}'\hat{g}_i + \rho_3(\bar{v}_i)(\hat{\lambda}'\hat{g}_i)^2$, so that w.p.a.1,

$$\frac{1}{n}\sum_i \rho_1(\hat{\lambda}'\hat{g}_i) = -1 - \hat{\lambda}'\hat{g} + r,$$

$$|r| \le C\max_i|\rho_3(\bar{v}_i)|\hat{\lambda}'\hat{\Omega}(\hat{\beta})\hat{\lambda} \le C\|\hat{\lambda}\|^2.$$

By $\|\hat{\lambda}\| = O_p(\sqrt{m/n})$ and $\|\hat{g}\| = O_p(\sqrt{m/n})$ we have $|\hat{\lambda}'\hat{g}| = O_p(m/n) \xrightarrow{p} 0$. Also, $|r| = O_p(m/n) \xrightarrow{p} 0$, so that by T,

$$\frac{1}{n}\sum_i \rho_1(\hat{\lambda}'\hat{g}_i) \xrightarrow{p} -1. \tag{A.10}$$

Next, consider the expansion $\rho_1(\hat{\lambda}'\hat{g}_i) = -1 + \rho_2(\bar{v}_i)\hat{\lambda}'\hat{g}_i$ as in the proof of Lemma A13. As discussed there, $\hat{\lambda}$ satisfies the first-order condition $0 = \sum_i \rho_1(\hat{\lambda}'\hat{g}_i)\hat{g}_i/n = -\hat{g} - \bar{\Omega}\hat{\lambda}$ for $\bar{\Omega} = -\sum_i \rho_2(\bar{v}_i)\hat{g}_i\hat{g}_i'/n$, so that for $\hat{g}_{i\delta_k} = \partial\hat{g}_i(\hat{\delta})/\partial\delta_k$, $\hat{g}_{\delta_k} = \partial\hat{g}(\hat{\delta})/\partial\delta_k$, and $\bar{\Omega}^k = -\sum_i \rho_2(\bar{v}_i)\hat{g}_i'\hat{g}_{i\delta_k}/n$ we have

$$\hat{\lambda} = -\bar{\Omega}^{-1}\hat{g}, \quad \sum_i \rho_1(\hat{\lambda}'\hat{g}_i)\hat{g}_{i\delta_k}/n = -\hat{g}_{\delta_k} - \bar{\Omega}^{k\prime}\hat{\lambda} = -\hat{g}_{\delta_k} + \bar{\Omega}^{k\prime}\bar{\Omega}^{-1}\hat{g}.$$

Also, note that for $\bar{U} = \sum_{i=1}^n U_i/n$, we have $\bar{U}S_n^{-1\prime}e_k\mu_n = \tilde{g}_{\delta_k} - \bar{g}_{\delta_k} - \Omega^{k\prime}\Omega^{-1}\tilde{g}$. Then, in terms of the notation of Lemma A13, it follows similarly to the arguments given there that

$$\begin{aligned}
&\left[\frac{1}{n}\sum_i \rho_1(\hat{\lambda}'\hat{g}_i)\right]^2 e_k' nS_n^{-1}\hat{D}(\hat{\beta})'\hat{\Omega}^{-1}\hat{D}(\hat{\beta})S_n^{-1\prime}e_\ell \\
&\quad = \mu_n^{-2}n\big(\hat{g}_{\delta_k}'\hat{\Omega}^{-1}\hat{g}_{\delta_\ell} - \hat{g}_{\delta_k}\hat{\Omega}^{-1}\bar{\Omega}^{\ell\prime}\bar{\Omega}^{-1}\hat{g} - \hat{g}'\bar{\Omega}^{-1}\bar{\Omega}^{k}\hat{\Omega}^{-1}\hat{g}_{\delta_\ell} \\
&\qquad + \hat{g}'\bar{\Omega}^{-1}\bar{\Omega}^{k}\hat{\Omega}^{-1}\bar{\Omega}^{\ell\prime}\bar{\Omega}^{-1}\hat{g}\big) \\
&\quad = \mu_n^{-2}n\big(\tilde{g}_{\delta_k}'\Omega^{-1}\tilde{g}_{\delta_\ell} - \tilde{g}_{\delta_k}\Omega^{-1}\Omega^{\ell\prime}\Omega^{-1}\tilde{g} - \tilde{g}'\Omega^{-1}\Omega^{k}\Omega^{-1}\tilde{g}_{\delta_\ell} \\
&\qquad + \tilde{g}'\Omega^{-1}\Omega^{k}\Omega^{-1}\Omega^{\ell\prime}\Omega^{-1}\tilde{g}\big) + o_p(1) \\
&\quad = \mu_n^{-2}n\big(\tilde{g}_{\delta_k} - \Omega^{k\prime}\Omega^{-1}\tilde{g}\big)'\Omega^{-1}\big(\tilde{g}_{\delta_\ell} - \Omega^{\ell\prime}\Omega^{-1}\tilde{g}\big) + o_p(1) \\
&\quad = ne_k'S_n^{-1}(G+\bar{U})'\Omega^{-1}(G+\bar{U})S_n^{-1\prime}e_\ell + o_p(1).
\end{aligned}$$

Note that by Assumption 1, $nS_n^{-1}G'\Omega^{-1}GS_n^{-1\prime} \longrightarrow H$. Also, $\xi_{\max}(E[U_iS_n^{-1\prime}e_\ell e_\ell' \times S_n^{-1}U_i']) \le C/\mu_n^2$, so that

$$\begin{aligned}&E[(ne_k'S_n^{-1}G'\Omega^{-1}\bar{U}S_n^{-1\prime}e_\ell)^2]\\&\quad = ne_k'S_n^{-1}G'\Omega^{-1}E[U_iS_n^{-1\prime}e_\ell e_\ell' S_n^{-1}U_i']\Omega^{-1}GS_n^{-1\prime}e_k\\&\quad \le Cne_k'S_n^{-1}G'\Omega^{-2}GS_n^{-1\prime}e_k/\mu_n^2 \le CH_{nkk}/\mu_n^2 \longrightarrow 0.\end{aligned}$$

Now apply Lemma A1 to $ne_k'S_n^{-1}\bar{U}'\Omega^{-1}\bar{U}S_n^{-1\prime}e_\ell$ for $A = \Omega^{-1}$, $Y_i = U_iS_n^{-1\prime}e_k\mu_n$, $Z_i = U_iS_n^{-1\prime}e_\ell\mu_n$, and $\mu_n^2 = a_n$. Note that $\xi_{\max}(A'A) = \xi_{\max}(AA') = \xi_{\max}(\Omega^{-2}) \le C$. Also, by $S_n^{-1}\mu_n$ bounded, $\xi_{\max}(\Sigma_{YY}) \le \xi_{\max}(E[U_iU_i']) \le C$ and $\xi_{\max}(\Sigma_{ZZ}) \le C$. Furthermore, $m/a_n^2 = m/\mu_n^4 \longrightarrow 0$, $a_n/n = \mu_n^2/n \le C$, $\mu_Y = \mu_Z = 0$, and

$$E[(Y_i'Y_i)^2]/na_n^2 \le CE[\|U_i\|^4]/na_n^2 \le CE[\|g_i\|^4 + \|G_i\|^4]/na_n^2 \longrightarrow 0.$$

Then by the conclusion of Lemma A1,

$$\begin{aligned}ne_k'S_n^{-1}\bar{U}'\Omega^{-1}\bar{U}S_n^{-1\prime}e_\ell &= n\bar{Y}'A\bar{Z}/a_n = \operatorname{tr}(A\Sigma_{YZ}')/a_n + o_p(1)\\&= \operatorname{tr}(\Omega^{-1}E[U_iS_n^{-1\prime}e_\ell e_k'S_n^{-1}U_i']) + o_p(1)\\&= e_k'S_n^{-1}E[U_i'\Omega^{-1}U_i]S_n^{-1\prime}e_\ell + o_p(1) \xrightarrow{p} \Lambda_{k\ell}.\end{aligned}$$

Then by T,

$$e_k'nS_n^{-1}\hat{D}(\hat{\beta})'\hat{\Omega}^{-1}\hat{D}(\hat{\beta})S_n^{-1\prime}e_\ell \xrightarrow{p} H_{k\ell} + \Lambda_{k\ell}.$$

The conclusion then follows by applying this result for each $k$ and $\ell$. *Q.E.D.*

PROOF OF THEOREM 3: Let $Y_n = n\mu_n^{-1}\,\partial\hat{Q}(0)/\partial\delta$. Then expanding the first-order conditions as outlined in Section 5 gives

$$0 = n\mu_n^{-1}\frac{\partial\hat{Q}(\hat{\delta})}{\partial\delta} = n\mu_n^{-1}\frac{\partial\hat{Q}(0)}{\partial\delta} + n\mu_n^{-2}\frac{\partial^2\hat{Q}(\bar{\delta})}{\partial\delta\,\partial\delta'}\mu_n\hat{\delta}.$$

By Lemma 13, $n\mu_n^{-2}\,\partial^2\hat{Q}(\bar{\delta})/\partial\delta\,\partial\delta'$ is nonsingular w.p.a.1. Then by CMT, Lemmas A12 and A13, and S,

$$\mu_n\hat{\delta} = S_n'(\hat{\beta} - \beta_0) = \left[n\mu_n^{-2}\frac{\partial^2\hat{Q}(\bar{\delta})}{\partial\delta\,\partial\delta'}\right]^{-1} n\mu_n^{-1}\frac{\partial\hat{Q}(0)}{\partial\delta} = H^{-1}Y_n + o_p(1).$$

Then by Lemma A12 and S,

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

Also, by Lemmas A13 and A14,

$$nS_n^{-1}\hat{H}S_n^{-1\prime} = \mu_n^{-2} n \frac{\partial^2 \hat{Q}(\hat{\delta})}{\partial\delta\,\partial\delta'} \xrightarrow{p} H, \quad nS_n^{-1}\hat{D}'\hat{\Omega}^{-1}\hat{D}S_n^{-1\prime} \xrightarrow{p} HVH.$$

Also, $\hat{H}$ is nonsingular w.p.a.1, so that

$$\begin{aligned} S_n' V S_n/n &= (nS_n^{-1}\hat{H}S_n^{-1\prime})^{-1} nS_n^{-1}\hat{D}'\hat{\Omega}^{-1}\hat{D}S_n^{-1\prime}(nS_n^{-1}\hat{H}S_n^{-1\prime})^{-1} \\ &\xrightarrow{p} H^{-1}HVHH^{-1} = V. \end{aligned}$$

To prove the last conclusion, note that $r_n S_n^{-1} c \longrightarrow c^*$ and S imply that

$$r_n c'(\hat{\beta} - \beta_0) = r_n c' S_n^{-1\prime} S_n'(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, c^{*\prime} V c^*),$$

$$r_n^2 c' \hat{V} c/n = r_n c' S_n^{-1\prime}(S_n' \hat{V} S_n/n) S_n^{-1} c r_n \xrightarrow{p} c^{*\prime} V c^*.$$

Therefore, by CMT and S,

$$\frac{c'(\hat{\beta} - \beta_0)}{\sqrt{c'\hat{V}c/n}} = \frac{r_n c' S_n^{-1\prime} S_n'(\hat{\beta} - \beta_0)}{\sqrt{r_n^2 c' S_n^{-1\prime}(S_n'\hat{V}S_n/n)S_n^{-1}c}} \xrightarrow{d} \frac{N(0, c^{*\prime}Vc^*)}{\sqrt{c^{*\prime}Vc^*}} = N(0, 1).$$

For the linear model we proceed by verifying all of the hypotheses of the general case. Note that $g_i(\beta) = Z_i(y_i - x_i'\beta)$ is twice continuously differentiable and that its first derivative does not depend on $\beta$, so Assumption 7 is satisfied. Also, by Lemma A5,

$$\big(E[\|g_i\|^4] + E[\|G_i\|^4]\big) m/n \le CE[\|Z_i\|^4] m/n \longrightarrow 0,$$

$$\xi_{\max}(E[G_i G_i']) \le \sum_{j=1}^{p} \xi_{\max}(E[Z_i Z_i' x_{ij}^2]) \le C\xi_{\max}(CI_m) \le C,$$

so that Assumption 6 is satisfied. Assumption 8 is satisfied by Lemmas A8 and A9. Assumptions 2–4 were shown to hold in the proof of Theorem 2. Assumption 9 can be shown to be satisfied similarly to the proof of Theorem 2. *Q.E.D.*

### A.4. *Large Sample Inference Proofs*

The following result improves upon Theorem 6.2 of Donald, Imbens, and Newey (2003). Let $\tilde{g} = \hat{g}(\beta_0)$ by only requiring that $m/n \longrightarrow 0$ in the case where the elements of $g_i$ are uniformly bounded.

LEMMA A15: *If* $E[(g_i'\Omega^{-1}g_i)^2]/mn \longrightarrow 0$, *then*

$$\frac{n\tilde{g}'\Omega^{-1}\tilde{g}-m}{\sqrt{2m}} \xrightarrow{d} N(0,1).$$

PROOF: Note that $E[g_i'\Omega^{-1}g_i]=m$ so that by M,

$$\frac{\sum_{i=1}^n g_i'\Omega^{-1}g_i/n-m}{\sqrt{2m}} = O_p\big(\big\{E[\{g_i'\Omega^{-1}g_i\}^2]/nm\big\}^{1/2}\big) \xrightarrow{p} 0.$$

Now apply Lemma A9 with $Y_i = Z_i = \Omega^{-1/2}g_i/\sqrt{n}(2m)^{1/4}$, so that $\bar{\xi}_Z = \bar{\xi}_Y = n^{-1}(2m)^{-1/2}$. Note that $\Psi = \Sigma_{YY}\Sigma_{ZZ} + \Sigma_{YZ}^2 = 2I_m/n^2 2m = I_m/n^2 m$, so that $n^2\operatorname{tr}(\Psi) = n^2\operatorname{tr}(I_m/n^2m) = 1$. Also note that

$$\begin{aligned}
&mn^4\bar{\xi}_Z^2\bar{\xi}_Y^2 = m/4m^2 \longrightarrow 0,\\
&n^3\big(\bar{\xi}_Z^2 E[\|Y_i\|^4] + \bar{\xi}_Y^2 E[\|Z_i\|^4]\big)\\
&\quad \le n^3 2\big\{n^{-2}(2m)^{-1}E[\{g_i'\Omega^{-1}g_i\}^2/n^2 2m]\big\} \longrightarrow 0,\\
&n^2E[\|Y_i\|^4]E[\|Z_i\|^4] = n^2\big\{E[\{g_i'\Omega^{-1}g_i\}^2]n^{-2}(2m)^{-1}\big\}^2 \longrightarrow 0.
\end{aligned}$$

It then follows by Lemma A10 that $\sum_{i\neq j} g_i'\Omega^{-1}g_j/\sqrt{2m} \xrightarrow{d} N(0,1)$, so the conclusion follows by T. *Q.E.D.*

PROOF OF THEOREM 4: By an expansion in $\lambda$ around $\lambda = 0$ we have

$$\hat{Q}(\beta_0) = -\tilde{\lambda}'\tilde{g} - \tilde{\lambda}'\bar{\Omega}\tilde{\lambda}/2,$$

where $\bar{\Omega} = -\sum_i \rho_2(\bar{v}_i)g_ig_i'/n$, $\bar{v}_i = \bar{\xi}'g_i$, and $\|\bar{\xi}\| \le \|\tilde{\lambda}\|$. Also, by an expansion around 0 we have $\rho_1(\tilde{\lambda}'g_i) = -1 + \rho_2(\breve{v}_i)\tilde{\lambda}'g_i$ with $|\breve{v}_i| \le |\tilde{\lambda}'g_i|$, so that for $\breve{\Omega} = -\sum_i \rho_2(\breve{v}_i)g_ig_i'/n$ the first-order conditions for $\tilde{\lambda}$ give $0 = -\tilde{g} - \breve{\Omega}\tilde{\lambda}$. Note that for $\Delta_n = n^{1/\gamma}(E[b_i^\gamma])^{1/\gamma}\sqrt{m/n}$ we have

$$\max_{i\le n}|1+\rho_2(\breve{v}_i)| \le C\|\tilde{\lambda}\|\max_{i\le n} g_i = O_p(\Delta_n).$$

Let $\tilde{\Omega} = \sum_i g_ig_i'/n$. By Lemma A0, $\xi_{\max}(\tilde{\Omega}) \le C$ w.p.a.1, so that for any $a, b$,

$$\begin{aligned}
|a'(\breve{\Omega}-\tilde{\Omega})b| &\le \sum_i |1+\rho(\breve{v}_i)||a'g_i||b'g_i|/n\\
&\le O_p(\Delta_n)\sqrt{a'\tilde{\Omega}ab'\tilde{\Omega}b} = O_p(\Delta_n)\|a\|\|b\|.
\end{aligned}$$

It follows similarly that

$$|a'(\breve{\Omega}-\tilde{\Omega})b| \le O_p(\Delta_n)\|a\|\|b\|.$$

It then follows from $\Delta_n \longrightarrow 0$, similarly to Lemma A0, that $\xi_{\min}(\breve{\Omega}) \ge C$ w.p.a.1, so $\tilde{\lambda} = -\breve{\Omega}^{-1}\tilde{g}$. Plugging into the above expansion gives

$$\hat{Q}(\beta_0) = \tilde{g}'\breve{\Omega}^{-1}\tilde{g} - \tilde{g}'\breve{\Omega}^{-1}\bar{\Omega}\breve{\Omega}^{-1}\tilde{g}/2.$$

As above $\xi_{\min}(\tilde{\Omega}) \ge C$ w.p.a.1, so that $\|\tilde{\Omega}^{-1}\tilde{g}\| \le C\|\tilde{g}\| = O_p(\sqrt{m/n})$ and $\|\breve{\Omega}^{-1}\tilde{g}\| = O_p(\sqrt{m/n})$. Therefore, by $\Delta_n\sqrt{m} \longrightarrow 0$,

$$\begin{aligned}|\tilde{g}'(\breve{\Omega}^{-1}-\tilde{\Omega}^{-1})\tilde{g}| &= \left|\tilde{g}'\breve{\Omega}^{-1}(\tilde{\Omega}-\breve{\Omega})\tilde{\Omega}^{-1}\tilde{g}\right| \le O_p(\Delta_n)O_p(m/n)\\ &= o_p(\sqrt{m}/n).\end{aligned}$$

It follows similarly that $|\tilde{g}'(\breve{\Omega}^{-1}\bar{\Omega}\breve{\Omega}^{-1}-\tilde{\Omega}^{-1})\tilde{g}| = o_p(\sqrt{m}/n)$, so that by T,

$$\hat{Q}(\beta_0) = \tilde{g}'\tilde{\Omega}^{-1}\tilde{g}/2 + o_p(\sqrt{m}/n).$$

It follows by $mE[\|g_i\|^4]/n \longrightarrow 0$ that $\|\tilde{\Omega}-\Omega\| = o_p(1/\sqrt{m})$, so that $\tilde{g}'\tilde{\Omega}^{-1}\tilde{g} = \tilde{g}'\Omega^{-1}\tilde{g} + o_p(\sqrt{m}/n)$ and, by T,

$$\hat{Q}(\beta_0) = \tilde{g}'\Omega^{-1}\tilde{g}/2 + o_p(\sqrt{m}/n).$$

It then follows that

$$\frac{2n\hat{Q}(\beta_0)-m}{\sqrt{m}} - \frac{n\tilde{g}'\Omega^{-1}\tilde{g}-m}{\sqrt{m}} = \frac{2n}{\sqrt{m}}\left[\hat{Q}(\beta_0) - \frac{\tilde{g}'\Omega^{-1}\tilde{g}}{2}\right] = o_p(1).$$

Then by Lemma A15 and S we have

$$\frac{2n\hat{Q}(\beta_0)-m}{\sqrt{m}} \xrightarrow{d} N(0,1).$$

Also, by standard results for the chi-squared distribution, as $m \longrightarrow \infty$ the $(1-\alpha)$th quantile $q_\alpha^m$ of a $\chi^2(m)$ distribution has the property that $[q_\alpha^m - m]/\sqrt{2m}$ converges to the $(1-\alpha)$th quantile $q_\alpha$ of $N(0,1)$. Hence we have

$$\Pr(2n\hat{Q}(\beta_0) \ge q_\alpha^m) = \Pr\left(\frac{2n\hat{Q}(\beta_0)-m}{\sqrt{2m}} \ge \frac{q_\alpha^m - m}{\sqrt{2m}}\right) \longrightarrow \alpha. \qquad Q.E.D.$$

PROOF OF THEOREM 5: Let $\hat{B} = nS_n^{-1}\hat{D}(\beta_0)'\hat{\Omega}(\beta_0)^{-1}\hat{D}(\beta_0)S_n^{-1\prime}$ and $B = HVH$. It follows from Lemma A14, replacing $\hat{\beta}$ with $\beta_0$, that $\hat{B} \xrightarrow{p} B$. By the proof of Theorem 3, S, and CM we have

$$\hat{T} = (\hat{\beta} - \beta_0)' S_n (S_n' \hat{V} S_n / n)^{-1} S_n' (\hat{\beta} - \beta_0)$$
$$= Y_n' B^{-1} Y_n + o_p(1) \xrightarrow{d} \chi^2(p).$$

Then by Lemma A12,

$$\mathrm{LM}(\beta_0) = n \frac{\partial \hat{Q}(\beta_0)'}{\partial \beta} S_n^{-1\prime} \hat{B}^{-1} n S_n^{-1} \frac{\partial \hat{Q}(\beta_0)}{\partial \beta} = Y_n' (B + o_p(1))^{-1} Y_n$$
$$= Y_n' B^{-1} Y_n + o_p(1).$$

Therefore, we have $\mathrm{LM}(\beta_0) = \hat{T} + o_p(1)$.

Next, by an expansion, for $\bar{H} = n S_n^{-1} \partial^2 \hat{Q}(\bar{\beta}) / \partial \beta \, \partial \beta' S_n^{-1\prime}$,

$$2n[\hat{Q}(\beta_0) - \hat{Q}(\hat{\beta})] = n(\hat{\beta} - \beta_0)' [\partial^2 \hat{Q}(\bar{\beta}) / \partial \beta \, \partial \beta'] (\hat{\beta} - \beta_0)$$
$$= (\hat{\beta} - \beta_0)' S_n \bar{H} S_n' (\hat{\beta} - \beta_0),$$

where $\bar{\beta}$ lies on the line joining $\hat{\beta}$ and $\beta_0$ and $\bar{H} \xrightarrow{p} H$ by Lemma A13. Then by the proof of Theorem 3 and the CMT,

$$2n[\hat{Q}(\beta_0) - \hat{Q}(\hat{\beta})] = \{Y_n' H^{-1} + o_p(1)\}\{H + o_p(1)\}\{H^{-1} Y_n + o_p(1)\}$$
$$= Y_n' H^{-1} Y_n + o_p(1).$$

It follows that $2n[\hat{Q}(\beta_0) - \hat{Q}(\hat{\beta})] = O_p(1)$, so that

$$2n[\hat{Q}(\beta_0) - \hat{Q}(\hat{\beta})] / \sqrt{m - p} \xrightarrow{p} 0.$$

Therefore, it follows as in the proof of Theorem 4 that

$$\frac{2n\hat{Q}(\hat{\beta}) - (m - p)}{\sqrt{m - p}} = \frac{2n\hat{Q}(\beta_0) - (m - p)}{\sqrt{m - p}} + o_p(1)$$
$$= \sqrt{\frac{m}{m - p}} \frac{2n\hat{Q}(\beta_0) - m}{\sqrt{m}} + \frac{p}{\sqrt{m - p}} + o_p(1)$$
$$\xrightarrow{d} N(0, 1).$$

Next, note that $H^{-1} \leq V$ in the p.s.d. sense so that $V^{-1} \leq H$. It follows that

$$Y_n' H^{-1} Y_n \geq Y_n' B^{-1} Y_n \xrightarrow{d} \chi^2(p).$$

Then $\Pr(2n[\hat{Q}(\beta_0) - \hat{Q}(\hat{\beta})] > q_\alpha^p) = \Pr(Y_n' H^{-1} Y_n > q_\alpha^p) + o(1) \geq \alpha$.

Next, in considering the CLR test, for notational convenience evaluate at $\beta_0$ and drop the $\beta$ argument, for example, so that $\hat{R} = \hat{R}(\beta_0)$. By have $\hat{B} \xrightarrow{p} B$, it follows that $\hat{B} \geq (1-\varepsilon)B$ w.p.a.1 for all for $\varepsilon > 0$. Also by $m/\mu_n^2$ bounded, for any $C$ there is $\varepsilon$ small enough so that $(1-\varepsilon)C - \varepsilon m/\mu_n^2$ is positive and bounded away from zero, that is, so that $(1-\varepsilon)C - \varepsilon m/\mu_n^2 \geq C$ (the $C$s are different). Then by hypothesis, multiplying through by $1-\varepsilon$, and subtracting $\varepsilon m/\mu_n^2$ from both sides it will be the case that

$$\xi_{\min}(\mu_n^{-2} S_n (1-\varepsilon) B S_n') - (m/\mu_n^2) \geq (1-\varepsilon)C - \varepsilon m/\mu_n^2 \geq C.$$

Then w.p.a.1,

$$\begin{aligned}\hat{F} &= \frac{\hat{R}-m}{\mu_n^2} = \xi_{\min}(\mu_n^{-2} S_n \hat{B} S_n') - \frac{m}{\mu_n^2} \\ &\geq \xi_{\min}(\mu_n^{-2} S_n (1-\varepsilon) B S_n') - (m/\mu_n^2) \geq C.\end{aligned}$$

Also, by the proof of Theorem 4,

$$\frac{\mathrm{AR}-m}{\mu_n^2} = \frac{\sqrt{m}}{\mu_n^2}\frac{\mathrm{AR}-m}{\sqrt{m}} \xrightarrow{p} 0.$$

Therefore, we have w.p.a.1,

$$\frac{\mathrm{AR}-\hat{R}}{\mu_n^2} = \frac{\mathrm{AR}-m}{\mu_n^2} - \hat{F} \leq -C.$$

It follows that w.p.a.1,

$$\frac{\mathrm{AR}}{\hat{R}} = \frac{(\mathrm{AR}-m)/\mu_n^2 + m/\mu_n^2}{\hat{F} + m/\mu_n^2} \leq \frac{C/2 + m/\mu_n^2}{C + m/\mu_n^2} \leq 1 - C.$$

Therefore, by $\hat{R} \geq C\mu_n^2 + m \longrightarrow \infty$, w.p.a.1,

$$\frac{\hat{R}}{(\mathrm{AR}-\hat{R})^2} = \frac{1}{\hat{R}}\frac{1}{(1-\mathrm{AR}/\hat{R})^2} \xrightarrow{p} 0.$$

Note that $\mathrm{AR}-\hat{R} < 0$ w.p.a.1, so that $|\mathrm{AR}-\hat{R}| = \hat{R}-\mathrm{AR}$. Also, similarly to Andrews and Stock (2006), by a mean value expansion $\sqrt{1+x} = 1 + (1/2)(x + o(1))$, so that

$$\widehat{\mathrm{CLR}} = \frac{1}{2}\{\mathrm{AR}-\hat{R} + [(\mathrm{AR}-\hat{R})^2 + 4\,\mathrm{LM}\cdot\hat{R}]^{1/2}\}$$

$$= \frac{1}{2}\left\{\mathrm{AR} - \hat{R} + |\mathrm{AR} - \hat{R}|\left[1 + \frac{4\,\mathrm{LM}\cdot\hat{R}}{(\mathrm{AR} - \hat{R})^2}\right]^{1/2}\right\}$$
$$= \frac{1}{2}\left\{\mathrm{AR} - \hat{R} + |\mathrm{AR} - \hat{R}|\left[1 + 2\,\mathrm{LM}\frac{\hat{R}}{(\mathrm{AR} - \hat{R})^2}(1 + o_p(1))\right]\right\}$$
$$= \widehat{\mathrm{LM}}\frac{\hat{R}}{\hat{R} - \mathrm{AR}}(1 + o_p(1)).$$

Let $r_n = \xi_{\min}(S_n B S_n'/\mu_n^2)$. Then $r_n - m/\mu_n^2 \geq C$ by hypothesis. Then $\hat{R}/\mu_n^2 = r_n + o_p(1)$ as shown below. It then follows that

$$\frac{\hat{R}}{\hat{R} - \mathrm{AR}} = \frac{\hat{R}/\mu_n^2}{(\hat{R} - m)/\mu_n^2 - (\mathrm{AR} - m)/\mu_n^2}$$
$$= \frac{r_n + o_p(1)}{r_n - m/\mu_n^2 + o_p(1)} = \frac{r_n}{r_n - m/\mu_n^2} + o_p(1).$$

It then follows that

$$\mathrm{CLR} = \left(\frac{r_n}{r_n - m/\mu_n^2}\right)\mathrm{LM} + o_p(1).$$

Carrying out these same arguments with $q_s^{m-p} + q_s^p$ replacing AR it follows that

$$\hat{q}_s = \frac{1}{2}\left\{q_s^{m-p} + q_s^p - \hat{R} + [(q_s^{m-p} + q_s^p - \hat{R})^2 + 4q_s^p\cdot\hat{R}]^{1/2}\right\}$$
$$= \left(\frac{r_n}{r_n - m/\mu_n^2}\right)q_s^p + o_p(1),$$

giving the conclusion with $c_n = r_n/(r_n - m/\mu_n^2)$.

It now remains to show that $\hat{R}/\mu_n^2 = r_n + o_p(1)$. Note that for $\bar{S}_n = S_n/\mu_n$,

$$\hat{R}/\mu_n^2 = \min_{\|x\|=1} x'\bar{S}_n\hat{B}\bar{S}_n'x, \quad r_n = \min_{\|x\|=1} x'\bar{S}_n B\bar{S}_n'x.$$

By Assumption 1 we can assume without loss of generality that $\mu_n = \mu_{1n}$ and

$$\bar{S}_n = \tilde{S}_n\,\mathrm{diag}(1, \mu_{2n}/\mu_n, \ldots, \mu_{pn}/\mu_n).$$

Let $e_j$ denote the $j$th unit vector and consider $x_n$ such that $x_n'S_n e_j = 0$ ($j = 2, \ldots, p$) and $\|x_n\| = 1$. Then by $\tilde{S}_n$ bounded and CS,

$$\|x_n'\bar{S}_n\| = \left\|x_n'\tilde{S}_n\left[e_1 + \sum_{j=2}^{p}(\mu_{jn}/\mu_n)e_j\right]\right\| = \|x_n'\tilde{S}_n e_1\| \leq \|\tilde{S}_n\| \leq C.$$

Also, by $\hat{B} \xrightarrow{p} B$ there is $C$ such $\|\hat{B}\| \leq C$ and $\xi_{\min}(\hat{B}) \geq 1/C$. w.p.a.1. Let $\hat{x} = \arg\min_{\|x\|=1} x'\bar{S}_n\hat{B}\bar{S}_n'x$ and $x_n^* = \arg\min_{\|x\|=1} x'\bar{S}_nB\bar{S}_n'x$. Then w.p.a.1,

$$C^{-1}\|\hat{x}'\bar{S}_n\|^2 \leq \hat{R}/\mu_n^2 \leq x_n'\bar{S}_n\hat{B}\bar{S}_n'x_n \leq C,$$
$$C^{-1}\|x_n^{*\prime}\bar{S}_n\|^2 \leq r_n \leq x_n'\bar{S}_nB\bar{S}_n'x_n \leq C,$$

so that there is $\bar{C}$ such that w.p.a.1,

$$\|\hat{x}'\bar{S}_n\| \leq \bar{C}, \quad \|x_n^{*\prime}\bar{S}_n\| \leq \bar{C}.$$

Consider any $\varepsilon > 0$. By $\hat{B} \xrightarrow{p} B$, w.p.a.1 $\|\hat{B} - B\| \leq \varepsilon/\bar{C}^2$. Then, w.p.a.1,

$$\begin{aligned}
\hat{R}/\mu_n^2 &\leq x_n^{*\prime}\bar{S}_n\hat{B}\bar{S}_nx_n^* \\
&= r_n + x_n^{*\prime}\bar{S}_n(\hat{B} - B)\bar{S}_nx_n^* \leq r_n + |x_n^{*\prime}\bar{S}_n(\hat{B} - B)\bar{S}_nx_n^*| \\
&\leq r_n + \|x_n^{*\prime}\bar{S}_n\|^2\|\hat{B} - B\| \leq r_n + \bar{C}^2(\varepsilon/\bar{C}^2) = r_n + \varepsilon, \\
r_n &\leq \hat{x}'\bar{S}_nB\bar{S}_n\hat{x} = \hat{R}/\mu_n^2 + \hat{x}'\bar{S}_n(B - \hat{B})\bar{S}_n\hat{x} \leq \hat{R}/\mu_n^2 + \varepsilon.
\end{aligned}$$

Thus, w.p.a.1, $r_n - \hat{R}/\mu_n^2 \leq \varepsilon$ and $\hat{R}/\mu_n^2 - r_n \leq \varepsilon$, implying $|\hat{R}/\mu_n^2 - r_n| \leq \varepsilon$, showing $|\hat{R}/\mu_n^2 - r_n| \xrightarrow{p} 0$. *Q.E.D.*

## REFERENCES

ANDREWS, D. W. K., AND J. H. STOCK (2006): "Inference With Weak Instruments," in *Advances in Economics and Econometrics*, Vol. 3, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge, U.K.: Cambridge University Press.

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.

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

*Dept. of Economics, Massachusetts Institute of Technology, Cambridge, MA 02142-1347, U.S.A.; wnewey@mit.edu*

*and*

*Dept. of Economics, University of Bristol, 2 Prior Road, Bristol BS8 1TN, U.K.; F.Windmeijer@bristol.ac.uk.*

*Manuscript received August, 2006; final revision received January, 2008.*