Supplemental Appendix for: Minimum Distance Approach to Inference with Many Instruments

Michal Kolesár*

This version 3.1.2, January 10, 2018[†] First circulated: November 2012.

SA.1 Additional details for calculations in main text

Let $e_1 = (1,0)'$ and let $e_2 = (0,1)'$.

SA.1.1 Additional details for Section 2.2

This section shows that the block of the inverse information matrix based on the limited information likelihood corresponding to β is given by $n^{-1}b'\Omega b \cdot a'\Omega^{-1}a/\lambda_n$.

The distribution of the statistics $\hat{\Pi}$ and *S* is given by

$$\operatorname{vec}(\widehat{\Pi}) \sim \mathcal{N}_{2k_n} \left((a' \Omega^{-1} a/n)^{-1/2} a \otimes \eta_n, \Omega \otimes I_{k_n} \right), \tag{SA-1}$$

$$(n-k_n-\ell_n)S \sim \mathcal{W}_2(n-k_n-\ell_n,\Omega),$$
 (SA-2)

with $\hat{\Pi}$ independent of *S*, where $W_2(n - k_n - \ell_n, \Omega)$ denotes a Wishart distribution with $n - k_n - \ell_n$ degrees of freedom, and scale matrix Ω . Their densities are therefore given by

$$f_{\hat{\Pi}}(\hat{\Pi};\beta,\eta_n,\Omega) = \frac{|\Omega|^{-k_n/2}}{(2\pi)^{k_n}} \exp\left(-\frac{n}{2}\left(\operatorname{tr}(\Omega^{-1}T) + \eta'_n\eta_n - 2\frac{\eta'_n\hat{\Pi}\Omega^{-1}a}{(na'\Omega^{-1}a)^{1/2}}\right)\right), \quad (SA-3)$$
$$f_S(S;\Omega) = C_{\nu}|S|^{(\nu-3)/2}|\Omega|^{-\nu/2}e^{-\frac{\nu}{2}\operatorname{tr}(\Omega^{-1}S)},$$

where $\nu = n - k_n - \ell_n$, and $C_{\nu}^{-1} = (2/\nu)^{\nu} \pi^{1/2} \Gamma(\nu/2) \Gamma((\nu-1)/2)$, with Γ denoting the gamma function.

It follows that the limited information likelihood is given by

$$\mathcal{L}_{\text{LI},n}(\beta,\eta_n,\Omega) = \frac{C_{\nu}|S|^{(\nu-3)/2}}{(2\pi)^{k_n}} \cdot |\Omega|^{-(n-\ell_n)/2} e^{-\frac{1}{2}\left(\operatorname{tr}(\Omega^{-1}\tilde{S}) + n\eta'_n\eta_n - 2n\frac{\eta'_n\Pi\Omega^{-1}a}{(na'\Omega^{-1}a)^{1/2}}\right)},$$
(SA-4)

^{*}Electronic correspondence: mkolesar@princeton.edu.

[†]Compiled January 23, 2019

where $\tilde{S} = nT + \nu S$.

The score is given by

$$\begin{split} \mathcal{S}_{\beta}(\beta,\eta_{n},\Omega) &= n^{1/2} \frac{\eta_{n}' \hat{\Pi} \Omega^{-1} d(\beta,\Omega)}{(a\Omega^{-1}a)^{1/2}}, \\ \mathcal{S}_{\eta_{n}}(\beta,\eta_{n},\Omega) &= n \left(\frac{n^{-1/2} \hat{\Pi} \Omega^{-1}a}{(a'\Omega^{-1}a)^{1/2}} - \eta_{n} \right), \\ \mathcal{S}_{\text{vech}(\Omega)}(\beta,\eta_{n},\Omega) &= \frac{1}{2} \tilde{D}' \left[\text{vec} \left(\tilde{S} - (n-\ell_{n})\Omega \right) - \frac{2n^{1/2}}{(a'\Omega^{-1}a)^{1/2}} \hat{\Pi}' \eta_{n} \otimes a + \frac{n^{1/2} \eta_{n}' \hat{\Pi} \Omega^{-1}a}{(a'\Omega^{-1}a)^{3/2}} a \otimes a \right], \end{split}$$

where $\tilde{D} = (\Omega^{-1} \otimes \Omega^{-1})D_2$, D_2 is the duplication matrix, and $d(\Omega, \beta) = e_1 - a \frac{a'\Omega^{-1}e_1}{a'\Omega^{-1}a}$. Let $\mathcal{H}_{\beta\eta_n}(\beta, \eta_n, \Omega)$ denote the β - η_n block of the Hessian, and similarly for the other blocks. By taking derivatives of the score, we obtain that

$$\begin{aligned} \mathcal{H}_{\beta\beta}(\beta,\eta_n,\Omega) &= -2S_{\beta}(\beta,\eta_n,\Omega) \frac{a'\Omega^{-1}e_1}{a'\Omega^{-1}a} - \frac{n^{1/2}\eta'_n\hat{\Pi}\Omega^{-1}a}{(a'\Omega^{-1}a)^{1/2}} \frac{1}{b'\Omega b \cdot a'\Omega^{-1}a'} \\ \mathcal{H}_{\eta_n\beta}(\beta,\eta_n,\Omega) &= \frac{n^{1/2}\hat{\Pi}\Omega^{-1}d(\Omega,\beta)}{(a\Omega^{-1}a)^{1/2}}, \\ \mathcal{H}_{\text{vech}(\Omega)\beta}(\beta,\eta_n,\Omega) &= \tilde{D}' \left(\frac{S_{\beta}(\beta,\eta_n,\Omega)}{2a'\Omega^{-1}a}a \otimes a + \frac{n^{1/2}}{\sqrt{a'\Omega^{-1}a}} \left(\frac{\eta'_n\hat{\Pi}\Omega^{-1}a}{a'\Omega^{-1}a}a - 2\hat{\Pi}'\eta_n\right) \otimes d(\Omega,\beta)\right), \end{aligned}$$

where we use the identities $e'_1\Omega^{-1}e_1a'\Omega^{-1}a - (a'\Omega^{-1}e_1)^2 = |\Omega|^{-1}$, $D_2(v_1 \otimes v_2) = D_2(v_2 \otimes v_1)$ for any vectors v_1, v_2 , and $(\hat{\Pi}'\eta_n) \otimes a = \operatorname{vec}(a\eta'_n\hat{\Pi}) = (\hat{\Pi}' \otimes a)\eta_n$. Since $a'\Omega^{-1}d(\Omega,\beta) = 0$, it follows that $\mathbb{E}[\mathcal{H}_{\operatorname{vech}(\Omega)\beta}(\beta,\eta_n,\Omega)] = 0$ and $\mathbb{E}[\mathcal{H}_{\eta_n\beta}(\beta,\eta_n,\Omega)] = 0$. Thus, the (1,1) element block of the inverse information matrix is given by

$$\mathcal{I}_{\text{LL},n}^{11}(\beta,\eta_n,\Omega) = -\frac{1}{E[\mathcal{H}_{\beta\beta}(\beta,\eta_n,\Omega)]} = \frac{a'\Omega^{-1}ab'\Omega b}{n\eta'_n\eta_n} = \frac{a'\Omega^{-1}ab'\Omega b}{n\lambda_n},$$

as stated in the main text.

SA.1.2 Additional details for Section 4.2

Consider the groups example, so that $z_{ij}^* = 1$ if individual *i* belongs to group *j* and zero otherwise, and let $W = \iota_n$, where ι_n denotes to an *n*-vector of ones, so that $(W'(W'W)W')_{ij} = 1/n$. Let ν denote a k_n -vector with elements $\nu_j = n_j$, and let $\operatorname{diag}(\nu)$ denote a diagonal matrix with elements $\nu_j = n_j$ on the diagonal. It then follows that $\tilde{Z} = Z^* - W'(W'W)W'Z^* = Z^* - \iota_n \nu'/n$, $\tilde{Z}'\tilde{Z} = \operatorname{diag}(\nu) - \nu \nu'/n$, $(\tilde{Z}'\tilde{Z})^{-1} = \operatorname{diag}(\nu)^{-1} + \iota_{k_n}\iota'_{k_n}/n$, and

$$(ZZ')_{ii} = (\tilde{Z}(\tilde{Z}'\tilde{Z})^{-1}\tilde{Z}')_{ii} = 1/n_{i(i)} - 1/n,$$

where j(i) denotes the group index that individual *i* belongs to. Since $(W'(W'W)W)_{ii} = 1/n$, it follows that

$$H_{ii} = (ZZ')_{ii} - \frac{k_n}{n - 1 - k_n} \left(1 - (ZZ)_{ii} - 1/n\right) = \frac{n - 1}{n - 1 - k_n} \left(\frac{1}{n_{g(i)}} - \frac{1 + k_n}{n}\right)$$

Consequently,

$$\delta_n = \operatorname{diag}(H)' \operatorname{diag}(H) / k_n = \frac{(n-1)^2}{k_n (n-1-k_n)^2} \sum_{j=0}^{k_n} n_j \left(\frac{1}{n_j} - \frac{1+k_n}{n}\right)^2$$
$$= \frac{(n-1)^2}{(n-1-k_n)^2 k_n} \left(\sum_j \frac{1}{n_j} - \frac{(k_n+1)^2}{n}\right).$$

SA.1.3 Additional details for Section 5.2

I illustrate the minimization of the minimum distance objective function given in Equation (25) in the paper subject to the constraint $\Xi_{11,n} \ge \beta^2 \Xi_{22,n}$. For concreteness and simplicity, consider the random-effects weight matrix $\hat{W}_{RE} = D'_2(S^{-1} \otimes S^{-1})D_2$, and suppose that the errors are normal. The solution is given by

$$\begin{pmatrix} \hat{\Xi}_{11} & \hat{\Xi}_{22} & \hat{\beta} \end{pmatrix} = \begin{cases} \begin{pmatrix} \hat{\Xi}_{11,\text{UMD}} & \hat{\Xi}_{22,\text{UMD}} & \hat{\beta}_{\text{UMD}} \end{pmatrix} & \text{if } S - (k_n/n)T \text{ is positive semi-definite,} \\ \begin{pmatrix} \hat{\Xi}_{22,\text{RE}} \hat{\beta}_{\text{RE}}^2 & \hat{\Xi}_{22,\text{RE}} & \hat{\beta}_{\text{RE}} \end{pmatrix} & \text{otherwise.} \end{cases}$$

When Assumption PR does not hold, then $T - (k_n/n)S$ will be positive definite with probability approaching one so that the restriction will not bind asymptotically. Otherwise, under Assumptions N and MI, its distribution is given by

$$\sqrt{n}\left(\hat{\beta}-\beta\right) \Rightarrow \sqrt{V_{\text{LIML},N}}\mathcal{Z}_{2} + \frac{\sqrt{2\tau}(b'\Omega e_{2})}{\Xi_{22}}\max(\mathcal{Z}_{1},0), \qquad \begin{pmatrix}\mathcal{Z}_{1}\\\mathcal{Z}_{2}\end{pmatrix} \sim \mathcal{N}_{2}(0,I_{2}), \qquad (\text{SA-5})$$

where $V_{\text{LIML},N}$ is given in Equation (9) in the paper and $\tau = \frac{\alpha_k(1-\alpha_\ell)}{1-\alpha_k-\alpha_\ell}$. This result follows by verifying the conditions for Theorem 1 in Andrews (2002).

The asymptotic distribution is non-standard, and since $\mathbb{E} \max(\mathcal{Z}_1, 0) > 0$, $\hat{\beta}$ is asymptotically biased. Recall that for UMD,

$$\sqrt{n}\left(\hat{\beta}_{\text{UMD}} - \beta\right) \Rightarrow \sqrt{V_{\text{LIML}}}\mathcal{Z}_2 + \frac{\sqrt{2\tau}b'\Omega e_2}{\Xi_{22}}\mathcal{Z}_1.$$
 (SA-6)

The difference between this expression and the asymptotic distribution for the minimum distance estimator subject to the positive definiteness condition is that the term max(Z_1 , 0) in Equation SA–5 has been replaced by Z_1 . Lovell and Prescott (1970, Section 4) were the first

ones to point out that this increases the asymptotic mean squared error.

There are several possible approaches to inference on β using $\hat{\beta}$. I discuss two of them (see Andrews, 1999, for a discussion of the bootstrap and subsampling). The first approach is based on the observation that the conventional asymptotic standard errors based on the assumption that no parameters are on the boundary (i.e. standard errors for $\hat{\beta}_{\text{UMD}}$) yield conservative confidence intervals when, in fact Ξ is reduced rank (Andrews, 1999, p. 1369). The second approach suggested by Andrews (1999) is to do a pre-test of the hypothesis $H_0: \Xi_{11} = \Xi_{22}\beta^2$ against $H_1: \Xi_{11} > \Xi_{22}\beta^2$ to determine if the true parameter Ξ_{11} is at the boundary with critical values chosen such that the pre-test is consistent as $n \to \infty$. If the test rejects, then we conclude that we're not at the boundary, and we use UMD standard errors. Otherwise, we assume that we're at the boundary, and, we use the asymptotic distribution (SA–5) to obtain confidence intervals. Quantiles of the limiting distribution in Equation (SA–5) can be obtained by simulating draws of Z_1 and Z_2 . The pre-test used in this approach is, in fact, equivalent to a consistent test of overidentifying restrictions, so that the modified Cragg-Donald test can be used.

SA.1.4 Additional details for Section 6

I first derive the expression for \hat{J}_{MD} . First, observe that, since $Q_T(\hat{\beta}_{RE}, S) = m_{max}$,

$$\min_{\Xi_{11}=\Xi_{22}\beta^2} \mathcal{Q}_n(\beta, \Xi_{11}, \Xi_{22}, \hat{W}_{\text{RE}}) = \mathcal{Q}_n(\hat{\beta}_{\text{RE}}, \hat{\Xi}_{22,\text{RE}}, \hat{W}_{\text{RE}})
= \operatorname{tr}\left(\left((k_n/n)I_2 - S^{-1}T \right)^2 \right) - (m_{\max} - k_n/n)^2. \quad \text{(SA-7)}$$

Since tr $(((k_n/n)I_2 - S^{-1}T)^2) = (m_{\max} - k_n/n)^2 + (m_{\min} - k_n/n)^2$, it follows that (SA–7) can be written as

$$\mathcal{Q}_n(\hat{\beta}_{\text{RE}}, \hat{\Xi}_{22,\text{RE}}, \hat{W}_{\text{RE}}) = (m_{\min} - k_n / n)^2.$$

It follows from the results in Section SA.1.3 that if $S - (k_n/n)T$ is not positive semi-definite (which is equivalent to $m_{\min} < k_n/n$), then

$$\min_{\Xi_{11} \ge \Xi_{22}\beta^2} \mathcal{Q}_n(\beta, \Xi_{11}, \Xi_{22}; \hat{W}_{\text{RE}}) = \min_{\Xi_{11} = \Xi_{22}\beta^2} \mathcal{Q}_n(\beta, \Xi_{11}, \Xi_{22}, \hat{W}_{\text{RE}})$$

Otherwise, $\min_{\Xi_{11} \ge \Xi_{22}\beta^2} Q_n(\beta, \Xi_{11}, \Xi_{22}; \hat{W}_{RE}) = 0$, which yields the expression for \hat{J}_{MD} as stated in the main text.

Next, I derive the asymptotic properties of overidentification tests proposed by Sargan (1958), Cragg and Donald (1993), and Anderson and Rubin (1949). Let

$$\hat{J}_{\rm S} = \frac{\hat{b}_{\rm LIML}' T \hat{b}_{\rm LIML}}{\hat{b}_{\rm LIML}' (T - (k_n/n)S)\hat{b}_{\rm LIML}} = \frac{m_{\rm min}}{1 - k_n/n - \ell_n/n + m_{\rm min}}$$

The Sargan (1958) test rejects whenever $n\hat{J}_s > q_{1-ns}^{\chi^2_{k-1}}$, the 1 – ns quantile of a χ^2_{k-1} distribution where ns denotes the desired nominal size. The generalized likelihood ratio test based on the limited information likelihood of Anderson and Rubin (1949) replaces \hat{J}_s with $\hat{J}_{AR} = \log(nm_{\min}/(n-k_n-\ell_n)+1)$, and the Cragg and Donald (1993) test uses $\hat{J}_{CD} = m_{\min}$.

All three tests are equivalent in the sense that they all reject for large values of m_{\min} . Therefore, the only difference between them in finite samples is how well the chi-squared approximation controls size in each case. While under standard asymptotics their asymptotic distributions coincide and therefore do not provide any guidance as to which test has the best size control, allowing for α_k , $\alpha_\ell > 0$ reverses this conclusion.

Lemma SA.1. Under Assumptions PR, MI and RC, if $k_n \rightarrow \infty$ *,*

$$\mathbb{P}\left(n\hat{J}_{s} \geq q_{1-ns}^{\chi^{2}_{k_{n}-1}}\right) \rightarrow \begin{cases} \Phi\left(\frac{\Phi^{-1}(ns)}{\sqrt{(1-\alpha_{k})(1+(1-\alpha_{k})\kappa\delta/2)}}\right) & \text{if } \alpha_{\ell} = 0, \\ 1 & \text{otherwise.} \end{cases}$$
$$\mathbb{P}\left(n\hat{J}_{\text{CD}} \geq q_{1-ns}^{\chi^{2}_{k_{n}-1}}\right) \rightarrow \Phi\left(\frac{\Phi^{-1}(ns)}{\sqrt{(1-\alpha_{\ell})/(1-\alpha_{k}-\alpha_{\ell})+\kappa\delta/2}}\right), \end{cases}$$

and if $\alpha_k > 0$, then $\mathbb{P}\left(n\hat{J}_{AR} \ge q_{1-ns}^{\chi^2_{k_n-1}}\right) \to 1$, where $\Phi(\cdot)$ is the cdf of a standard normal distribution.

Proof of Lemma SA.1. Let $\tilde{\alpha} = (1 - \alpha_{\ell})/(1 - \alpha_{k} - \alpha_{\ell})$. By Proposition 4 and the delta method,

$$\begin{split} & \frac{n}{\sqrt{k_n}} \left(\hat{J}_{\mathsf{S}} - \frac{\alpha_k}{1 - \alpha_\ell} \right) \Rightarrow \mathcal{N} \left(0, \frac{2\tilde{\alpha} + \kappa \delta}{(1 - \alpha_\ell)^2 \tilde{\alpha}^2} \right) \\ & \frac{n}{\sqrt{k_n}} \left(\hat{J}_{\mathsf{AR}} - \log(\tilde{\alpha}) \right) \Rightarrow \mathcal{N} \left(0, \frac{2\tilde{\alpha} + \kappa \delta}{(1 - \alpha_\ell)^2} \right) \\ & \frac{n}{\sqrt{k_n}} \left(\hat{J}_{\mathsf{CD}} - \alpha_k \right) \Rightarrow \mathcal{N}(0, 2\tilde{\alpha} + \delta \kappa), \end{split}$$

I use the approximation from Peiser (1943) that as $k \to \infty$,

$$q_{1-\mathrm{ns}}^{\chi_k^2} = k + \Phi^{-1}(1-\mathrm{ns})\sqrt{2k} + O(1).$$

Therefore if $\tau > 0$,

$$\begin{split} \mathbb{P}\left(n\hat{J}_{\text{CD}} \geq q_{1-ns}^{\chi^2_{k_n-1}}\right) &= \mathbb{P}\left(\frac{n}{\sqrt{k_n}}(\hat{J}_{\text{CD}} - \alpha_k) \geq \Phi^{-1}(1-ns)\sqrt{2} + O(1/\sqrt{k_n})\right) \\ &= \mathbb{P}\left(\mathcal{N}(0,1) + o_p(1) \geq \frac{\Phi^{-1}(1-ns)}{\sqrt{\tilde{\alpha} + \kappa\delta/2}} + o(1)\right) \rightarrow \Phi\left(\frac{\Phi^{-1}(ns)}{\sqrt{\tilde{\alpha} + \kappa\delta/2}}\right). \end{split}$$

Similarly,

$$\begin{split} \mathbb{P}\left(n\hat{J}_{\mathsf{s}} \geq q_{\mathsf{n}\mathsf{s}}^{\chi^2_{k_n-1}}\right) &= \mathbb{P}\left(n\hat{J}_{\mathsf{s}} \geq k_n + \Phi^{-1}(1-\mathsf{n}\mathsf{s})\sqrt{2k_n} + O(1)\right) \\ &= \mathbb{P}\left(\sqrt{\frac{2\tilde{\alpha} + \kappa\delta}{(1-\alpha_\ell)^2\tilde{\alpha}^2}}\mathcal{N}(0,1) + o_p(1) \geq -\frac{\sqrt{k_n}\alpha_\ell}{(1-\alpha_\ell)} + \Phi^{-1}(1-\mathsf{n}\mathsf{s})\sqrt{2} + o(1)\right). \end{split}$$

Now, if $\alpha_{\ell} > 0$, then the right-hand side converges to $-\infty$, so that the rejection probability converges to one. If

 $\alpha_{\ell} = 0$, then

$$\mathbb{P}\left(n\hat{J}_{s} \geq q_{ns}^{\chi^{2}_{k_{n}-1}}\right) \to \Phi\left(\frac{\Phi^{-1}(ns)}{\sqrt{(1-\alpha_{k})(1+(1-\alpha_{k})\kappa\delta/2)}}\right)$$

Finally,

$$\mathbb{P}\left(n\hat{f}_{AR} \ge q_{ns}^{\chi^2_{k_n-1}}\right) = \mathbb{P}\left(\frac{n}{\sqrt{k_n}}(\hat{f}_{AR} - n\log(\tilde{\alpha})) \ge -\frac{n}{\sqrt{k_n}}\log(\tilde{\alpha}) + \sqrt{k_n} + \Phi^{-1}(1-ns)\sqrt{2} + o(1)\right)$$
$$= \mathbb{P}\left(\frac{\sqrt{2\tilde{\alpha} + \kappa\delta}}{1-\alpha_\ell}\mathcal{N}(0,1) + o_p(1) \ge \frac{n}{\sqrt{k_n}}\left(k_n/n - \log(\tilde{\alpha})\right) + \Phi^{-1}(1-ns)\sqrt{2} + o(1)\right)$$

Since $\alpha_k \leq -\log(1-\alpha_k)$,

$$\alpha_k - \log(\tilde{\alpha}) \le \log\left(\frac{1}{1-\alpha_k}\right) - \log\left(\frac{1-\alpha_\ell}{1-\alpha_k-\alpha_\ell}\right) = \log\left(\frac{1-\alpha_k-\alpha_\ell}{(1-\alpha_k)(1-\alpha_\ell)}\right) \le 0,$$

with equality only if $\alpha_k = 0$, so that the right-hand side of the previous display converges to $-\infty$ if $\alpha_k > 0$.

SA.2 Proof of Lemma A.2

Proof. Let $K_d = 2N_d - I_{d^2}$ denote the commutation matrix, which has the property that $K_d \operatorname{vec}(A) = \operatorname{vec}(A')$, where *A* is a $d \times d$ matrix. To show part (i), note that for any $v_1, v_2, v_3, v_4 \in \mathbb{R}^d$,

$$v_1 v_2' \otimes v_3 v_4' = K_d (v_3 v_2' \otimes v_1 v_4').$$
(SA-8)

This follows from $v_1v'_2 \otimes v_3v'_4 = v_1 \otimes (v_3v'_2 \otimes v'_4) = K_d(v_3v'_2 \otimes v'_4) \otimes v_1$, where the second equality uses the identity $K_d(A \otimes v) = v \otimes A$ for any $A \in \mathbb{R}^{d \times d'}$ and $v \in \mathbb{R}^d$ (Magnus and Neudecker, 1979, Theorem 3.1(ix)). Furthermore,

$$\operatorname{vec}(Q_n) = (I_{G^2} + K_G)(I_G \otimes M'_n P_n)\operatorname{vec}(U_n) + \operatorname{vec}(U'_n P_n U_n) + \operatorname{vec}(M'_n P_n M_n), \quad (SA-9a)$$

$$\mathbb{E}[U'_n P_n U_n] = \operatorname{tr}(P_n)\Omega_n, \tag{SA-9b}$$

$$\mathbb{E}[\operatorname{vec}(Q_n)] = \operatorname{vec}(M'_n P_n M_n + \operatorname{tr}(P_n)\Omega_n), \qquad (SA-9c)$$

$$\mathbb{E}[\operatorname{vec}(U_n)\operatorname{vec}(U_n)'] = \Omega_n \otimes I_n, \tag{SA-9d}$$

$$\mathbb{E}[\operatorname{vec}(U_n)\operatorname{vec}(U'_nP_nU_n)'] = \mathbb{E}[u_{in}\otimes\operatorname{diag}(P_n)\otimes u'_{in}\otimes u'_{in}],$$
(SA-9e)

$$\mathbb{E}\operatorname{vec}(U'_n P_n U_n)\operatorname{vec}(U'_n P_n U_n)' = \delta_n \mathbb{E}[u_{in}u'_{in} \otimes u_{in}u'_{in}] + (\operatorname{tr}(P_n)^2 - \delta_n)\operatorname{vec}(\Omega_n)\operatorname{vec}(\Omega_n)'$$
(SA-9f)

$$+(\mathrm{tr}(P_n^2)-\delta_n)(I_{G^2}+K_G)\Omega_n\otimes\Omega_n$$

where (SA–9a) follows by the definition of the commutation matrix, (SA–9b) follows from the expansion $U'_n P_n U_n = \sum_{i,j} p_{ij} u_{in} u'_{jn}$, (SA–9f) also follows from this expansion and from (SA–8), (SA–9c) follows from (SA–9b). Equations (SA–9d) and (SA–9e) follow by direct calculation. Therefore,

$$\begin{aligned} \operatorname{var}[\operatorname{vec}(Q_n)] &= (I_{G^2} + K_G)(I_G \otimes M'_n P_n) \mathbb{E}[\operatorname{vec}(U_n) \operatorname{vec}(U_n)'](I_G \otimes P_n M_n)(I_{G^2} + K_G) \\ &+ (I_{G^2} + K_G)(I_G \otimes M'_n P_n) \mathbb{E}[\operatorname{vec}(U_n) \operatorname{vec}(U'_n P_n U_n)'] \\ &+ \mathbb{E}[\operatorname{vec}(U'_n P_n U_n) \operatorname{vec}(U_n)'](I_G \otimes P_n M_n)(I_{G^2} + K_G) + \mathbb{E}\operatorname{vec}(U'_n P_n U_n) \operatorname{vec}(U'_n P_n U_n)' \\ &- \operatorname{tr}(P_n)^2 \operatorname{vec}(\Omega_n) \operatorname{vec}(\Omega_n)' \end{aligned}$$
$$= (I_{G^2} + K_G)(\Omega_n \otimes \Psi_n)(I_{G^2} + K_G) + \operatorname{tr}(P_n^2)(I_{G^2} + K_{GG})\Omega_n \otimes \Omega_n \\ &+ (I_{G^2} + K_G)\mathbb{E}[u_{in}u'_{in} \otimes \overline{m}_n u'_{in}] + \mathbb{E}[u_{in}u'_{in} \otimes u_{in}\overline{m}'_n](I_{G^2} + K_G) \\ &+ \delta_n \left(\mathbb{E}[u_{in}u'_{in} \otimes u_{in}u'_{in}] - \operatorname{vec}(\Omega_n)\operatorname{vec}(\Omega_n)' - (I_{G^2} + K_{GG})\Omega_n \otimes \Omega_n\right),\end{aligned}$$

where the first equality uses (SA–9a)–(SA–9c), and the second equality uses (SA–9d)–(SA–9f). The result then follows by applying the identities (SA–8) and $A_1 \otimes A_2 = K_d(A_2 \otimes A_1)$ for any $A_1, A_2 \in \mathbb{R}^{d \times d}$ (Magnus and Neudecker, 1979, Theorem 3.1(ix)).

The proof of part (ii) adapts the arguments in Chao, Swanson, Hausman, Newey and Woutersen (2012) and Hansen, Hausman and Newey (2008). By the Cramér–Wold device, it suffices to prove the result for

$$\operatorname{vec}(A)'\operatorname{vec}(Q_n) = \operatorname{tr}(A'Q_n)$$

where $A \in \mathbb{R}^{G \times G}$ is an arbitrary matrix of constants. Since Q_n is symmetric, we can without loss of generality assume that A is also symmetric. Expanding the expression, and using symmetry of P_n yields

$$\operatorname{tr}(AQ_n - \mathbb{E}[AQ_n]) = \sum_{i=1}^n \sum_{j=1}^n (m_{jn} + u_{jn})' A(m_{in} + u_{in}) p_{ij} - \sum_{i=1}^n \sum_{j=1}^n m'_{jn} Am_{in} p_{ij} - \sum_{i=1}^n p_{ii} \operatorname{tr}(A\Omega_n)$$
$$= \sum_{i=1}^n W_{in} + \sum_{i=2}^n \sum_{j=1}^{i-1} 2p_{ij} u'_{in} Au_{jn} = \sum_{i=1}^n y_{in},$$

where $y_{in} = W_{in} + 2S_{in}$ for $i \ge 2$, $y_{1n} = W_{1n}$ and

$$W_{in} = 2e'_{in}P_nM_nAu_{in} + p_{ii}(u'_{in}Au_{in} - \operatorname{tr}(A\Omega_n)),$$

$$S_{in} = \sum_{j=1}^{i-1} p_{ij}u'_{in}Au_{jn},$$

Note that y_{in} is a martingale difference array with respect to the filtration $\mathcal{F}_{in} = \sigma(u_{1n}, \dots, u_{i-1,n})$. By the martingale central limit theorem, it therefore suffices to show that for some $\varepsilon > 0$,

$$\sum_{i=1}^{n} \mathbb{E}[|y_{in}|^{2+\varepsilon}] = o(1), \tag{SA-10}$$

and that the conditional variance $\sum_{i=1}^{n} \mathbb{E}[y_{in}^2 | \mathcal{F}_{i-1,n}]$ converges. By the Loève c_r -inequality if

$$\mathbb{E}[|u_{in}'Au_{in} - \operatorname{tr}(A\Omega_n)|^4] \sum_{i=1}^n p_{ii}^4 = o(1),$$
(SA-11)

$$\sum_{i=2}^{n} \mathbb{E}[S_{in}^{4}] = o(1), \quad \text{and} \quad (SA-12)$$

$$\sum_{i=1}^{n} \mathbb{E}[(e'_{in}P_n M_n A u_{in})^4] = o(1),$$
(SA-13)

then (SA–10) holds with ε = 2. Now, (SA–11) follows from Assumptions (ii)b and (ii)c. To show (SA–12), note that expanding the expression yields

$$\sum_{i=2}^{n} \mathbb{E}[S_{in}^{4}] = 2 \sum_{i=2}^{n} \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} p_{ij}^{2} p_{ik}^{2} \mathbb{E}[(u_{jn}' A u_{in})^{2} (u_{in}' A u_{kn})^{2}] \le C \sum_{i=2}^{n} \sum_{j=1}^{i-1} \sum_{k=1}^{i-1} p_{ij}^{2} p_{ik}^{2} \le C \sum_{i=1}^{n} \left(\sum_{j=1}^{n} p_{ij}^{2}\right)^{2},$$

for some constant C, which is o(1) by Assumption (ii)c. Next, to show (SA-13), note that

$$\sum_{i=1}^{n} \mathbb{E}[(e_{in}'P_n M_n A u_{in})^4] \le \mathbb{E}[\|A u_{in}\|^4] \sum_{i=1}^{n} \|e_{in}'P_n M_n\|^4,$$

which is also o(1) by Assumptions (ii)b and (ii)d.

It remains to show convergence of the conditional variance. By Assumption (ii)a, it suffices to show that

$$\sum_{i=1}^{n} \mathbb{E}[y_{in}^2 \mid \mathcal{F}_{i-1,n}] - \operatorname{var}(\operatorname{tr}(AQ_n)) \xrightarrow{p} 0.$$
(SA-14)

Since $\mathbb{E}[W_{in}^2 | \mathcal{F}_{i-1,n}] = \mathbb{E}[W_{in}^2]$, and since $\operatorname{var}(\operatorname{tr}(AQ_n)) = \sum_{i=1}^n \mathbb{E}[W_{in}^2] + 4\sum_{i=2}^n \mathbb{E}[S_{in}^2]$, the left-hand side of (SA-14) can be written as

$$\sum_{i=1}^{n} \mathbb{E}[y_{in}^2 \mid \mathcal{F}_{i-1,n}] - \operatorname{var}(\operatorname{tr}(AQ_n)) = 4 \sum_{i=2}^{n} \left(\mathbb{E}[S_{in}^2 \mid \mathcal{F}_{i-1,n}] - \mathbb{E}[S_{in}^2] \right) + 4 \sum_{i=2}^{n} \mathbb{E}[W_{in}S_{in} \mid \mathcal{F}_{i-1,n}].$$
(SA-15)

Letting P_n^L denote the lower triangular matrix with elements $p_{ij}1\{i > j\}$, we can write the second sum in (SA–15) as

$$\sum_{i=2}^{n} \mathbb{E}[W_{in}S_{in} \mid \mathcal{F}_{i-1}] = \sum_{i=1}^{n} \sum_{j=1}^{n} (P_{n}^{L})_{ij} \mathbb{E}[W_{in}u_{in}'] A u_{jn} = \operatorname{tr}\left(U'P_{n}^{L'}\mathbb{E}[D_{W}U]A\right) = \sum_{g=1}^{G} e'_{gG}U'P_{n}^{L'}\overline{w}_{g}.$$

where D_W denotes a diagonal matrix with elements $(D_W)_{ii} = W_{in}$, and $\overline{w}_g = \mathbb{E}[D_W U]Ae_{gG}$. The variance of the summand on the right-hand side is given by

$$\mathbb{E}[(e'_{g}U'P_{n}^{L'}\overline{w}_{g})^{2}] = \Omega_{gg}\overline{w}'_{g}P_{n}^{L}P_{n}^{L'}\overline{w}_{g} \leq \Omega_{gg}\|P_{n}^{L}P_{n}^{L'}\|_{F}\|\overline{w}_{g}\|^{2} \\ = \Omega_{gg}\|P_{n}^{L}P_{n}^{L'}\|_{F}\sum_{i=1}^{n}\mathbb{E}[W_{in}u'_{in}Ae_{g}]^{2} \leq \Omega_{gg}e'_{g}A\Omega_{n}Ae_{g}\|P_{n}^{L}P_{n}^{L'}\|_{F}\sum_{i=1}^{n}\mathbb{E}[W_{in}^{2}],$$

where $\|\cdot\|_F$ denotes the Frobenius norm. It follows by Loève C_r inequality and Assumption (ii)a that $\sum_{i=1}^{n} \mathbb{E}[W_{in}^2]$ is bounded, and

$$\begin{split} \|P_n^L P_n^{L'}\|_F^2 &= \sum_{k < i} p_{ik}^4 + 2\sum_{k < \ell < i} p_{ik}^2 p_{i\ell}^2 + 2\sum_{k < i < j} p_{ik}^2 p_{jk}^2 + 4\sum_{k < \ell < i < j} p_{ik} p_{jk} p_{i\ell} p_{j\ell} \\ &\leq 5\sum_i (\sum_j p_{ij}^2)^2 + 4\sum_{k < \ell < i < j} p_{ik} p_{jk} p_{i\ell} p_{j\ell} = o(1), \end{split}$$

where the last equality follows by Assumption (ii)c. Hence, by Markov inequality, the second term in (SA–15) is $o_p(1)$. Next, consider the first term in (SA–15), which can be written as

$$\sum_{i=2}^{n} \left(\mathbb{E}[S_{in}^{2} \mid \mathcal{F}_{i-1,n}] - \mathbb{E}[S_{in}^{2}] \right) = \sum_{i=2}^{n} \left(\sum_{k=1}^{i-1} \sum_{j=1}^{i-1} p_{ij} p_{ik} u_{kn}' A \Omega A u_{jn} - \sum_{j=1}^{i-1} p_{ij}^{2} \operatorname{tr}(\Omega A \Omega A) \right)$$

$$= \sum_{i=2}^{n} \sum_{j=1}^{i-1} p_{ij}^{2} (u_{jn}' A \Omega A u_{jn} - \operatorname{tr}(\Omega A \Omega A)) + 2 \sum_{i=2}^{n} \sum_{k=1}^{i-1} \sum_{j=1}^{k-1} p_{ij} p_{ik} u_{kn}' A \Omega A u_{jn}.$$
(SA-16)

Variance of the first term in (SA-16) is given by

$$\operatorname{var}\left(\sum_{i=2}^{n}\sum_{j=1}^{i-1}p_{ij}^{2}(u_{jn}^{\prime}A\Omega A u_{jn} - \operatorname{tr}(\Omega A \Omega A))\right) = \mathbb{E}[(u_{in}^{\prime}A\Omega A u_{in} - \operatorname{tr}(\Omega A \Omega A))^{2}]\sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{\ell=j+1}^{n}p_{\ell j}^{2}p_{ij}^{2}$$

which converges to zero since the triple sum $\sum_{i=2}^{n} \sum_{j=1}^{i-1} \sum_{\ell=j+1}^{n} p_{\ell j}^2 p_{i j}^2$ is bounded by

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{\ell=1}^{n} p_{\ell j}^{2} p_{i j}^{2} = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} p_{i j}^{2} \right)^{2} = o(1),$$
(SA-17)

where the last equality follows by Assumption (ii)c. Variance of the second term in (SA-16) given by

$$\operatorname{var}\left(\sum_{i=2}^{n}\sum_{k=1}^{i-1}\sum_{j=1}^{k-1}p_{ij}p_{ik}u_{kn}'A\Omega Au_{jn}\right) = \operatorname{tr}((A\Omega)^{4})\sum_{i=2}^{n}\sum_{k=1}^{i-1}\sum_{j=1}^{k-1}\sum_{\ell=k+1}^{n}p_{ij}p_{ik}p_{\ell j}p_{\ell k}$$
$$= \operatorname{tr}((A\Omega)^{4})\left[\sum_{j< k< i}p_{ij}^{2}p_{ik}^{2} + 2\sum_{j< k< i<\ell}p_{ji}p_{ik}p_{\ell j}p_{\ell k}\right]$$

where the first sum is again bounded by (SA–17), and the second term equals $\sum_{i < j < k < \ell} p_{ik} p_{i\ell} p_{jk} p_{j\ell}$, which is o(1) by Assumption (ii)c. Therefore, by Markov inequality, the first term in (SA–15) is $o_p(1)$, so that (SA–14) holds, which proves the theorem.

References

- ANDERSON, T. W. and RUBIN, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, **20** (1), 46–63. 4, 5
- ANDREWS, D. W. K. (1999). Estimation when a parameter is on a boundary. Econometrica, 67 (6), 1341–1383. 4
- (2002). Generalized method of moments estimation when a parameter is on a boundary. Journal of Business & Economic Statistics, 20 (4), 530–544. 3
- CHAO, J. C., SWANSON, N. R., HAUSMAN, J. A., NEWEY, W. K. and WOUTERSEN, T. (2012). Asymptotic distribution of jive in a heteroskedastic iv regression with many instruments. *Econometric Theory*, **12** (1), 42–86. 7
- CRAGG, J. G. and DONALD, S. G. (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory*, **9** (2), 222–240. 4, 5
- HANSEN, C. B., HAUSMAN, J. A. and NEWEY, W. K. (2008). Estimation with many instrumental variables. *Journal of Business & Economic Statistics*, **26** (4), 398–422. 7
- LOVELL, M. C. and PRESCOTT, E. (1970). Multiple regression with inequality constraints: Pretesting bias, hypothesis testing and efficiency. *Journal of the American Statistical Association*, **65** (330), 913–925. 3
- MAGNUS, J. R. and NEUDECKER, H. (1979). The commutation matrix: Some properties and applications. *The Annals of Statistics*, 7 (2), 381–394. 6, 7
- PEISER, A. M. (1943). Asymptotic formulas for significance levels of certain distributions. *The Annals of Mathematical Statistics*, **14** (1), 56–62. 5
- SARGAN, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica*, **26** (3), 393–415. 4, 5