Two-step weak IV proof (MD approach)

February 5, 2018

Acknowledgement. I thank Isaiah Andrews for guidance on proofs contained in this section. I thank Kevin Li for helpful comments.

Contents

1 Test statistics based on a minimum-distance approach **1**
1.1 Non-robust confidence set 2
1.2 Robust confidence set 3

2 Validity of two-step confidence sets based on MD-LC statistics **4**

3 Two-step confidence sets for scalar parameters **5**

1 Test statistics based on a minimum-distance approach

We first define the minimum distance estimator for linear IV. We start by considering a linear IV model

$$\begin{aligned} Y &= X\theta_0 + \epsilon \\ X &= Z\pi_0 + V, \end{aligned}$$

which can equivalently be written as a reduced-form model

$$Y = Z\delta_0 + U \tag{1}$$
$$X = Z\pi_0 + V \tag{2}$$

for Z a $(N \times k)$ matrix of instruments, X a $(N \times m)$ vector of endogenous regressors, Y a $(N \times 1)$ vector of outcome variables. Assume that $E[Z_i U_i] = E[Z_i V_i^{'}] = 0$. We also assume that any exogenous regressors have already been partialled out.[1]

The reduced-form parameter vector is $\tau = (\delta, \pi)$ and its estimator is $\hat{\tau} = (\hat{\delta}, \hat{\pi})$. The structural parameter vector is θ. The structural function is then $r(\tau, \theta) = \delta - \pi\theta$. Note that $r(\hat{\tau}, \theta_0) = \hat{\delta} - \hat{\pi}\theta_0$ and $r(\tau_0, \theta_0) = \delta_0 - \pi_0\theta_0 = 0$. If $\frac{\partial r(\tau_0,\theta_0)}{\partial \theta}$ does not have full rank, then the model is under-identified. If $\frac{\partial r(\tau_0,\theta_0)}{\partial \theta}$ has full rank, then the model is identified. Weak identification arises when $\frac{\partial r(\tau_0,\theta_0)}{\partial \theta}$ is very close to a reduced rank matrix. Otherwise the model is (well) strongly identified.

[1] For exogenous regressors W and initial data $\left(\tilde{Y}, \tilde{X}, \tilde{Z}, W\right)$, we partial out W by letting $Y = M_W\tilde{Y}$,$X = M_W\tilde{X}$, $Z = M_W\tilde{Z}$ where $M_W = I - W\left(W^{'}W\right)^{-1}W^{'}$.

Suppose we observe one draw (X, Y, Z) from the distribution $F_N(Z, U, V; \pi_0, \theta_0)$, where $\theta_0 \in \Theta \subseteq \mathbb{R}^m$ and $\pi_0 \in \Pi \subseteq \mathbb{R}^k \times \mathbb{R}^m$. Assume that Θ is compact. We are interested in constructing a confidence set for the parameter θ, treating π as a nuisance parameter.

Assumption 1. *Let $\xi_0 = \{(\theta_{0,N}, \pi_{0,N})\}_{N=1}^{\infty} \in \Pi_{N=1}^{\infty} (\Theta \times \Pi)$ be a sequence of true parameter values. We define Ξ_S to be the set of potential parameter sequences $(\theta_{0,N}, \pi_{0,N}) \in (\Theta \times \Pi)$, where for all N, $(\theta_{0,N}, \pi_{0,N})$ is fixed at (θ_0, π_0) for some $\pi_0 \neq 0$. We refer to Ξ_S as strong identification and drop the index of true parameter value when we consider strong identification. Assume there exists a non-empty set Ξ_W of sequences $(\theta_{0,N}, \pi_{0,N})$ where $(\theta_{0,N}, \pi_{0,N}) \to (\theta_0, 0)$ for some θ_0 (for example, weak instrument asymptotics sets $\pi_{0,N} = \frac{1}{\sqrt{N}}\pi^*$,) which we refer to as potentially weak or simply weak identification. Assume that for all $\xi_0 \in \Xi_W \cup \Xi_S$, $\sqrt{N}\begin{pmatrix} \hat{\delta} - \delta_{0,N} \\ \hat{\pi} - \pi_{0,N} \end{pmatrix} \to_d N\left(0, \begin{pmatrix} \Sigma_\delta & \Sigma_{\delta\pi} \\ \Sigma_{\pi\delta} & \Sigma_\pi \end{pmatrix}\right)$ and we have consistent estimators $\hat{\Sigma}_\delta$, $\hat{\Sigma}_\pi$, and $\hat{\Sigma}_{\delta\pi}$ for the asymptotic variance-covariance matrix.*

Then by the delta method, we have $\sqrt{N}\left(r\left(\hat{\tau}, \theta_{0,N}\right) - r\left(\tau_{0,N}, \theta_{0,N}\right)\right) \to_d N\left(0, \Sigma_r\right)$ where $\Sigma_r = Avar(r(\hat{\tau}, \theta_{0,N}))$. Let $\hat{\Sigma}_r\left(\theta_{0,N}\right)$ be a consistent estimator for fixed Σ_r under $\xi_0 \in \Xi_W \cup \Xi_S$.

1.1 Non-robust confidence set

Suppose we are interested in inference on a p-dimensional parameter $(p \leq m)$ $\beta = f(\theta)$. Here f is a continuously differentiable function. Denote $\frac{\partial}{\partial\theta'} f(\theta) = F(\theta)$. Assume $F(\theta_{0,N})$ has full rank for all N. For example we may be interested in constructing a confidence set for the jth coordinate of the parameter vector e.g. $f(\theta) = \theta_j$. We focus on constructing a confidence set for a single coordinate of the parameter vector in Section 3. Below we focus on the general case where β is p-dimensional.

Assume we have some estimator $\tilde{\theta}$ for θ which under strong identification is first-order equivalent to $\hat{\theta}$, which solves $\min_{\theta\in\Theta} r(\hat{\tau}, \theta)' \hat{\Omega}(\theta) r(\hat{\tau}, \theta)$ where $\hat{\Omega}(\theta)$ is a symmetric p.d. weighting matrix which converges uniformly to some $\Omega(\theta)$ under strong identification.[2] Examples of such estimators $\tilde{\theta}$ are one-step MD, efficiently and inefficiently weighted two-step MD, and continuously updating MD.

Under strong identification and regularity conditions, $\tilde{\theta}$ is consistent and asymptotic normal with asymptotic variance V_θ. Given this, the delta method implies that

$$\sqrt{N}\left(f\left(\tilde{\theta}\right) - \beta_0\right) \to_d N\left(0, V_\beta\right)$$

where $\beta_0 = f(\theta_0)$ and $V_\beta = F(\theta_0) V_\theta F(\theta_0)'$.

Thus, if we consider the Wald statistic

$$W(\beta) = N \cdot \left(f\left(\tilde{\theta}\right) - \beta\right)' \hat{V}_\beta^{-1}\left(f\left(\tilde{\theta}\right) - \beta\right)$$

for $\hat{V}_\beta$ a consistent estimator of V_β, then the continuous mapping theorem implies that under $H_0 : \beta = \beta_0$ we have

$$W(\beta_0) \to_d \chi_p^2. \tag{3}$$

Collecting the set of values β such that the Wald test does not reject gives us a non-robust Wald confidence set for $\beta = f(\theta)$,

$$CS_N = \left\{\beta : W(\beta) \leq \chi_{p,1-\alpha}^2\right\}.$$

---

[2]By first-order asymptotic equivalence, we mean that $\sqrt{N}(\hat{\theta} - \tilde{\theta}) \to_p 0$ under $\xi_0 \in \Xi_S$.

1.2 Robust confidence set

In order to define a robust confidence set, we first define the K and S statistics, which are asymptotically pivotal even under weak identification. See Kleibergen (2005) for a detailed discussion on K statistic defined for GMM. Note that Magnusson (2010) has extended the K statistic for MD estimators. The K statistic defined below generalizes his result to allow inefficient weight matrices ($\hat{\Omega}(\theta) \neq \hat{\Sigma}_r(\theta)^{-1}$) to construct confidence sets for 2SLS estimators, which are inefficiently weighted MD estimators. First orthogonalize the Jacobian with respect to the structural function

$$\hat{D}(\theta) = \left[\frac{\partial r(\hat{\tau},\theta)}{\partial \theta_1} - \hat{\Sigma}_{\theta_1,r}(\theta)\hat{\Sigma}_r(\theta)^{-1} r(\hat{\tau},\theta), \ldots, \frac{\partial r(\hat{\tau},\theta)}{\partial \theta_m} - \hat{\Sigma}_{\theta_m,r}(\theta)\hat{\Sigma}_r(\theta)^{-1} r(\hat{\tau},\theta)\right]$$

where $\hat{\Sigma}_{\theta,r}(\theta)$ is a consistent estimator for $Acov\left(vec\left(\frac{\partial r(\hat{\tau},\theta)}{\partial \theta'}\right), r(\hat{\tau},\theta)\right)$ and $\hat{\Sigma}_{\theta_j,r}(\theta)$ is the $k \times k$ block of $\hat{\Sigma}_{\theta,r}(\theta)$ corresponding to θ_j. One can show that $\hat{D}(\theta)$ and $r(\hat{\tau},\theta_0)$ are asymptotically independent, regardless of the rank of $\frac{\partial r(\tau_0,\theta_0)}{\partial \theta}$, and $\hat{D}(\theta)$ is asymptotically equivalent to $\frac{\partial r(\hat{\tau},\theta)}{\partial \theta}$ when identification is strong. We then define the one-step estimator

$$\theta^*(\theta) = \theta - \left(\hat{D}(\theta)'\hat{\Omega}(\theta)\hat{D}(\theta)\right)^{-1}\hat{D}(\theta)'\hat{\Omega}(\theta) r(\hat{\tau},\theta)$$

and $\beta^*(\theta) = f(\theta) - F(\theta)\cdot(\theta^*(\theta) - \theta)$. For $M(\theta) = \hat{\Omega}(\theta)\hat{D}(\theta)\left(\hat{D}(\theta)'\hat{\Omega}(\theta)\hat{D}(\theta)\right)^{-1} F(\theta)'$, the K statistic is

$$K_{\Omega,f}(\theta) = N\cdot(\beta^*(\theta) - f(\theta))'\left(M(\theta)'\hat{\Sigma}_r(\theta)M(\theta)\right)^{-1}(\beta^*(\theta) - f(\theta)).$$

Also define the S statistic as $S(\theta) = N\cdot r(\hat{\tau},\theta)'\hat{\Sigma}_r(\theta)^{-1} r(\hat{\tau},\theta)$.

Theorem 1. *Under Assumption 1 and Assumption 4 of Andrews (Forthcoming), under all $\xi_0 \in \Xi_W \cup \Xi_S$, the K and S statistics derived above satisfy:*

$$\left(K_{\Omega,f}(\theta_{0,N}), S(\theta_{0,N}) - K_{\Omega,f}(\theta_{0,N})\right) \to_d \left(\chi^2_p, \chi^2_{k-p}\right)$$

and $K_{\Omega,f}(\theta_{0,N})$ and $S(\theta_{0,N}) - K_{\Omega,f}(\theta_{0,N})$ are asymptotically independent.

Given the results of Theorem 1, we can a construct robust confidence set for $\beta = f(\theta)$. In particular, define

$$\begin{aligned} CS_{R,\theta} &= \left\{\theta : K_{\Omega,f}(\theta) + a(\gamma)\cdot S(\theta) \leq H^{-1}(1-\alpha; a(\gamma), k, p)\right\} \\ CS_R &= \{f(\theta) : \theta \in CS_{R,\theta}\} \qquad (4) \\ &= \left\{\beta : \min_{\theta:\beta=f(\theta)} K_{\Omega,f}(\theta) + a(\gamma)\cdot S(\theta) \leq H^{-1}(1-\alpha; a(\gamma), k, p)\right\} \end{aligned}$$

where $H^{-1}(1-\alpha; a(\gamma), k, p)$ is the $1-\alpha$ quantile of a $(1+a(\gamma))\cdot\chi^2_p + a(\gamma)\cdot\chi^2_{k-p}$ distribution.[3] The initial confidence set $CS_{R,\theta}$ collects the set of values θ where the linear combination statistic

[3] By definition $a(\gamma)$ solves

$$Pr\left\{(1+a(\gamma))\cdot\chi^2_p + a(\gamma)\cdot\chi^2_{k-p} \leq \chi^2_{p,1-\alpha}\right\} = 1-\alpha-\gamma \qquad (5)$$

To find this value in practice, we take one million independent simulation draws from χ^2_p and χ^2_{k-p} distributions and solve numerically for the value a which sets the $1-\alpha-\gamma$ quantile of the corresponding linear combination of these draws to $\chi^2_{p,1-\alpha}$.

falls below the critical value, and so will cover $\theta_{0,N}$ with probability tending to α under both Ξ_W and Ξ_S by Theorem 1. CS_R then takes the image of the initial confidence set under $f(\cdot)$ to construct a confidence set for $\beta = f(\theta)$. This is known as the projection method, and ensures that CS_R has coverage at least $1-\alpha$ for $f(\theta_{0,N})$.[4]

2 Validity of two-step confidence sets based on MD-LC statistics

To summarize, we construct a two-step confidence set $CS_2(\gamma)$ for $\beta = f(\theta)$ as follows:

1. Construct the non-robust Wald confidence set
$$CS_N = \left\{\beta : W(\beta) \le \chi^2_{p,1-\alpha}\right\}.$$

2. Construct the robust confidence set
$$CS_R = \left\{\beta : \min_{\theta:\beta=f(\theta)} K_{\Omega,f}(\theta) + a(\gamma)\cdot S(\theta) \le H^{-1}(1-\alpha; a(\gamma), k, p)\right\} \tag{6}$$
where $H^{-1}(1-\alpha; a(\gamma), k, p)$ is the $1-\alpha$ quantile of a $(1+a(\gamma))\cdot\chi^2_p + a(\gamma)\cdot\chi^2_{k-p}$ distribution.

3. Define the preliminary robust confidence set through
$$CS_P(\gamma) = \left\{\beta : \min_{\theta:\beta=f(\theta)} K_{\Omega}(\theta) + a(\gamma)\cdot S(\theta) \le \chi^2_{p,1-\alpha}\right\}$$

Given such $(CS_R, CS_P(\gamma), CS_N)$, we construct a two-step confidence set $CS_2(\gamma)$, which is CS_N when $CS_P(\gamma)$ is contained in CS_N and CS_R otherwise. To show that $CS_2(\gamma)$ is valid i.e. has coverage at least $1-\alpha-\gamma$, we need to verify that $(CS_R, CS_P(\gamma), CS_N)$ satisfy the conditions for Assumption 1 and Equations (3) and (4) of Andrews (Forthcoming).

Equation (3) in Andrews (Forthcoming) says that the non-robust confidence set CS_N has sequential coverage at least $1-\alpha$ under strong identification. This holds as a consequence of asymptotic distribution of $W(\beta_0)$ under strong identification. Equation (4) in Andrews (Forthcoming) says that the robust confidence set CS_R has coverage at least $1-\alpha$ under both weak and strong identification. This holds as consequence of Theorem 1.

Assumption 1 part (1) requires that the preliminary confidence set $CS_P(\gamma)$ has sequential coverage at least $1-\alpha-\gamma$ when identification is weak. Assumption 1 part (2) requires that $CS_P(\gamma)$ is contained in CS_R with probability one. These hold by construction.

Assumption 1 part (3) requires that $CS_P(\gamma)$ is contained in CS_N with probability tending to one under all strongly identified sequences. We indirectly prove this by showing that under strong identification, the MD test statistics are asymptotically equivalent to their GMM counterparts in $\frac{1}{\sqrt{N}}$ neighborhoods of the true parameter value θ_0. Then we can apply proof of Corollary 2 and Theorem 3 in Andrews (Forthcoming).

To prove that the test statistics under MD and GMM are asymptotically equivalent, we first prove the following auxiliary claim. We include definitions for test statistics under GMM in the proofs included in the Appendix. Recall that the GMM moment function under linear IV is $g_N(\theta) = \frac{1}{N}\sum Z_i\left(Y_i - X_i'\theta\right) = \frac{1}{N}Z'(Y - X\theta)$.

[4] We can construct the robust confidence sets based on the K statistic $CS_{K,\theta} = \left\{\theta : K_{\Omega,f}(\theta) \le \chi^2_{p,1-\alpha}\right\}$ (and we do in `twostepweakiv`). However, the confidence set CS_K is not in general asymptotically equivalent to the confidence set CS_N. Thus, the corresponding preliminary robust confidence set does not satisfy Assumption 1 in Andrews (Forthcoming) and cannot be used to construct valid two-step confidence sets.

Claim 1. Under strong identification, let $\{A_{\theta,N}\}$ be a sequence of random closed subsets of Θ such that $\limsup_{N\to\infty} Pr_{\xi_0,N}\{A_{\theta,N}=\emptyset\}<1$ and $\sup_{\theta_N\in A_{\theta,N}}\|\theta_N-\theta_0\|=O_P\left(\frac{1}{\sqrt{N}}\right)$. Note that we leave out the index of true parameter value $\theta_{0,N}$ because we consider strong identification case here. We show that

1) $\sup_{\theta_N\in A_{\theta,N}}\left|\left(\frac{Z'Z}{N}\right)\hat{\Sigma}_r(\theta_N)\left(\frac{Z'Z}{N}\right)-\hat{\Sigma}_g(\theta_N)\right| = o_p(1)$, where $\hat{\Sigma}_g(\theta)$ is consistent estimator for $Avar(g_N(\theta))$ and $\hat{\Sigma}_r(\theta)$ is consistent estimator for $Avar(r(\hat{\tau},\theta))$;

2) $\sup_{\theta_N\in A_{\theta,N}}\left|\left(I_m\otimes\frac{Z'Z}{N}\right)\hat{\Sigma}_{\theta r}(\theta_N)\left(\frac{Z'Z}{N}\right)-\hat{\Sigma}_{\theta g}(\theta_N)\right| = o_p(1)$, where $\hat{\Sigma}_{\theta g}(\theta)$ is consistent estimator for $Acov\left(vec\left(\frac{\partial g_N(\theta)}{\partial\theta'}\right),g_N(\theta)\right)$ and $\hat{\Sigma}_{\theta r}(\theta)$ is consistent estimator for $Acov\left(vec\left(\frac{\partial r(\hat{\tau},\theta)}{\partial\theta'}\right),r(\hat{\tau},\theta)\right)$.

Using Claim 1, we prove that under strong identification, the MD test statistics are asymptotically equivalent to their GMM counterparts in $\frac{1}{\sqrt{N}}$ neighborhoods of the true parameter value θ_0.

Theorem 2. *Under strong identification, let* $\{A_{\theta,N}\}$ *be a sequence of random closed subsets of* Θ *such that* $\limsup_{N\to\infty} Pr_{\xi_0,N}\{A_{\theta,N}=\emptyset\}<1$ *and* $\sup_{\theta_N\in A_{\theta,N}}\|\theta_N-\theta_0\|=O_P\left(\frac{1}{\sqrt{N}}\right)$. *We have*

$$\sup_{\theta_N\in A_{\theta,N}}\left|W^{MD}(f(\theta_N))-W^{GMM}(f(\theta_N))\right|=o_p(1),$$

$$\sup_{\theta_N\in A_{\theta,N}}\left|K^{MD}_{\Omega,f}(\theta_N)-K^{GMM}_{\Omega,f}(\theta_N)\right|=o_p(1),$$

and $\sup_{\theta_N\in A_{\theta,N}}\left|S^{MD}(\theta_N)-S^{GMM}(\theta_N)\right|=o_p(1)$ *for both* Ω *the efficient and inefficient weight matrix.*

Theorem 3 of Andrews (Forthcoming) shows that $CS_P(\gamma)$ constructed using GMM test statistics is contained in $CS_{nonrobust}$ with probability tending to one under all strongly identified sequences. Theorem 2 allows us to apply the proof of Theorem 3 of Andrews (Forthcoming), and show that $CS_P(\gamma)$ constructed using MD test statistics is also contained in $CS_{nonrobust}$ with probability tending to one under all strongly identified sequences.

In short, we verify that $(CS_{robust},CS_P(\gamma),CS_{nonrobust})$ satisfy the conditions for Assumption 1 and Equations (3) and (4) of Andrews (Forthcoming) in this section. We conclude that the two-step confidence sets $CS_2(\gamma)$ constructed using an MD approach is valid i.e. has coverage at least $1-\alpha-\gamma$.

3 Two-step confidence sets for scalar parameters

Let $\theta=(\beta_1,\ldots,\beta_m)$ be the full parameter vector. Suppose we are interested in constructing a confidence set for the jth element β_j. This can be achieved by setting $f(\theta)=e_j'\cdot\theta$ where e_j is the j-th coordinate vector. Then we have $f(\theta)=e_j'\cdot\theta=\beta_j$ and $\frac{\partial}{\partial\theta}f(\theta)=e_j'$. Below we derive explicitly test statistics for a single element of the parameter vector.

The Wald statistics for β_j is

$$W(\beta_j)=N\cdot\frac{\left(\tilde{\beta}_j-\beta_j\right)'\left(\tilde{\beta}_j-\beta_j\right)}{\left(\hat{V}_\theta\right)_{jj}}$$

where $\left(\hat{V}_\theta\right)_{jj}$ is the jth diagonal entry of $\hat{V}_\theta$.

The K statistics for β_j is

$$K_{\Omega,e_j}\left(\theta\right) = N \cdot \frac{\left(\beta_j^*\left(\theta\right) - \beta_j\right)'\left(\beta_j^*\left(\theta\right) - \beta_j\right)}{M\left(\theta\right)'\hat{\Sigma}_r\left(\theta\right)M\left(\theta\right)}.$$

Note that here $M\left(\theta\right)'\hat{\Sigma}_r\left(\theta\right)M\left(\theta\right)$ is the jth diagonal entry of

$$\left(\hat{D}\left(\theta\right)'\hat{\Omega}\left(\theta\right)\hat{D}\left(\theta\right)\right)^{-1}\hat{D}\left(\theta\right)'\hat{\Omega}\left(\theta\right)\hat{\Sigma}_r\left(\theta\right)\hat{\Omega}\left(\theta\right)\hat{D}\left(\theta\right)\left(\hat{D}\left(\theta\right)'\hat{\Omega}\left(\theta\right)\hat{D}\left(\theta\right)\right)^{-1}.$$

The S statistics remains the same as in the full parameter case.

Similarly, we can construct a valid two-step confidence set for β_j based on steps outlined in Section 2.

Appendix

This Appendix contains proofs for results stated in the paper.

Proof of Theorem 1

Proof. First note that for linear IV, Assumption 1 implies Assumption 2 and 3 in Andrews (Forthcoming). Denote the test statistics for MD by superscript MD. Denote the corresponding test statistics for GMM defined in Andrews (Forthcoming) by superscript GMM. Andrews (Forthcoming) proves asymptotic distribution for K and S statistics for GMM. To prove that K and S statistics have the same asymptotic distribution for MD, it suffices to show that $K_{\Omega}^{MD}\left(\theta_{0,N}\right) = K_{\Omega}^{GMM}\left(\theta_{0,N}\right) + o_p\left(1\right)$, and $S^{MD}\left(\theta_{0,N}\right) = S^{GMM}\left(\theta_{0,N}\right) + o_p\left(1\right)$ under all $\xi_0 \in \Xi_W \cup \Xi_S$. That is, we show that the test statistics derived using both approaches are asymptotically equivalent under $\xi_0 \in \Xi_W \cup \Xi_S$.

First note that $\left|\left(\frac{Z'Z}{N}\right)\hat{\Sigma}_r\left(\theta_{0,N}\right)\left(\frac{Z'Z}{N}\right) - \hat{\Sigma}_g\left(\theta_{0,N}\right)\right| = o_p\left(1\right)$. [5]

Consider the S statistic derived using a GMM approach

$$S^{GMM}\left(\theta_{0,N}\right) = N \cdot g_N\left(\theta_{0,N}\right)' \hat{\Sigma}_g\left(\theta_{0,N}\right)^{-1} g_N\left(\theta_{0,N}\right).$$

Substituting in $\left(\frac{Z'Z}{N}\right)\hat{\Sigma}_r\left(\theta_{0,N}\right)\left(\frac{Z'Z}{N}\right) = \hat{\Sigma}_g\left(\theta_{0,N}\right) + o_p\left(1\right)$, we obtain

$$S^{GMM}\left(\theta_{0,N}\right) = N \cdot g_N\left(\theta_{0,N}\right)' \left(\frac{Z'Z}{N}\right)^{-1} \hat{\Sigma}_r\left(\theta_{0,N}\right)^{-1} \left(\frac{Z'Z}{N}\right)^{-1} g_N\left(\theta_{0,N}\right) + o_p\left(1\right)$$

$$= N \cdot \left(Y - X\theta_{0,N}\right)' Z\left(Z'Z\right)^{-1} \hat{\Sigma}_r\left(\theta_{0,N}\right)^{-1} \left(Z'Z\right)^{-1} Z'\left(Y - X\theta_{0,N}\right) + o_p\left(1\right)$$

$$= N \cdot \left(\hat{\delta} - \hat{\pi}\theta_{0,N}\right)' \hat{\Sigma}_r\left(\theta_{0,N}\right)^{-1} \left(\hat{\delta} - \hat{\pi}\theta_{0,N}\right) + o_p\left(1\right) = S^{MD}\left(\theta_{0,N}\right) + o_p\left(1\right).$$

[5]Since both are consistent estimators, we have

$$\hat{\Sigma}_g\left(\theta_{0,N}\right) \to_p Avar\left(\frac{1}{N}\sum Z_i\left(Y_i - X_i'\theta_{0,N}\right)\right) = Avar\left(\frac{1}{N}\sum Z_i\left(U_i - V_i'\theta_{0,N}\right)\right)$$

and

$$\hat{\Sigma}_r\left(\theta_{0,N}\right) \to_p Avar\left(\hat{\delta} - \hat{\pi}\theta_{0,N}\right)$$

$$= \left(E\left[Z_iZ_i'\right]\right)^{-1} Avar\left(\frac{1}{N}\sum Z_i\left(U_i - V_i'\theta_{0,N}\right)\right)\left(E\left[Z_iZ_i'\right]\right)^{-1}.$$

By the continuous mapping theorem, we have $\left(\frac{Z'Z}{N}\right)\hat{\Sigma}_r\left(\theta_{0,N}\right)\left(\frac{Z'Z}{N}\right) \to_p Avar\left(\frac{1}{N}\sum Z_i\left(U_i - V_i'\theta_{0,N}\right)\right)$.

Now consider the K statistic. Note that K statistic depends on weight matrix $\hat{\Omega}(\theta_{0,N})$. Note that $\left|\left(I_m \otimes \frac{Z'Z}{N}\right)\hat{\Sigma}_{\theta r}(\theta_{0,N})\left(\frac{Z'Z}{N}\right) - \hat{\Sigma}_{\theta g}(\theta_{0,N})\right| = o_p(1)$.[6]

First consider the case of an inefficient weight matrix. That is, $\hat{\Omega}^{MD} = \frac{Z'Z}{N}$ for MD estimators and $\hat{\Omega}^{GMM} = \left(\frac{Z'Z}{N}\right)^{-1}$ for GMM estimators.

The Jacobian orthogonalized with respect to the GMM moment conditions is

$$D_N^{GMM}(\theta_{0,N}) = \left[\frac{\partial g_N(\theta_{0,N})}{\partial\theta_1} - \hat{\Sigma}_{\theta_1,g}(\theta_{0,N})\hat{\Sigma}_g(\theta_{0,N})^{-1}g_N(\theta_{0,N}), \ldots, \frac{\partial g_N(\theta_{0,N})}{\partial\theta_m} - \hat{\Sigma}_{\theta_m,g}(\theta_{0,N})\hat{\Sigma}_g(\theta_{0,N})^{-1}g_N(\theta_{0,N})\right]$$

where $\hat{\Sigma}_{\theta_j,g}(\theta_{0,N})$ is the $k \times k$ block of $\hat{\Sigma}_{\theta,g}(\theta_{0,N})$ corresponding to θ_j. Substituting in $\left(\frac{Z'Z}{N}\right)\hat{\Sigma}_r(\theta_{0,N})\left(\frac{Z'Z}{N}\right) = \hat{\Sigma}_g(\theta_{0,N}) + o_p(1)$ and $\left(I_m \otimes \frac{Z'Z}{N}\right)\hat{\Sigma}_{\theta r}(\theta_{0,N})\left(\frac{Z'Z}{N}\right) = \hat{\Sigma}_{\theta g}(\theta_{0,N}) + o_p(1)$ we have

$$\left(\frac{Z'Z}{N}\right)\hat{D}^{MD}(\theta_{0,N}) = D_N^{GMM}(\theta_{0,N}) + o_p(1).$$

We can show that the one-step estimators are asymptotically equivalent $\beta^{*MD}(\theta_{0,N}) = \beta^{*GMM}(\theta_{0,N}) + o_P(1)$ and thus $M^{MD}(\theta_{0,N})'\hat{\Sigma}_r(\theta_{0,N})M^{MD}(\theta_{0,N}) = M^{GMM}(\theta_{0,N})'\hat{\Sigma}_g(\theta_{0,N})M^{GMM}(\theta_{0,N}) + o_p(1)$. Recall that

$$K_{\Omega,f}^{MD}(\theta_{0,N}) = N \cdot \left(\beta^{*MD}(\theta_{0,N}) - f(\theta_{0,N})\right)'\left(M^{MD}(\theta_{0,N})'\hat{\Sigma}_r(\theta_{0,N})M^{MD}(\theta_{0,N})\right)^{-1}\left(\beta^{*MD}(\theta_{0,N}) - f(\theta_{0,N})\right)$$

$$K_{\Omega,f}^{GMM}(\theta_{0,N}) = N \cdot \left(\beta^{*GMM}(\theta_{0,N}) - f(\theta_{0,N})\right)'\left(M^{GMM}(\theta_{0,N})'\hat{\Sigma}_g(\theta_{0,N})M^{GMM}(\theta_{0,N})\right)^{-1}\left(\beta^{*GMM}(\theta_{0,N}) - f(\theta_{0,N})\right)$$

Thus we have $K_{\Omega,f}^{MD}(\theta_{0,N}) = K_{\Omega,f}^{GMM}(\theta_{0,N}) + o_p(1)$ for Ω the inefficient weight matrix.

Next consider the case of efficient weight matrix. That is, $\hat{\Omega}^{MD}(\theta_{0,N}) = \hat{\Sigma}_r(\theta_{0,N})^{-1}$ for MD and $\hat{\Omega}^{GMM}(\theta_{0,N}) = \hat{\Sigma}_g(\theta_{0,N})^{-1}$ for GMM.

Note that with efficient weight matrix, the K statistics simplify to

$$K_{\Omega,f}^{MD}(\theta_{0,N}) = N \cdot \left(\beta^{*MD}(\theta_{0,N}) - f(\theta_{0,N})\right)'\left(\hat{D}^{MD}(\theta_{0,N})'\hat{\Sigma}_r(\theta_{0,N})\hat{D}^{MD}(\theta_{0,N})\right)^{-1}\left(\beta^{*MD}(\theta_{0,N}) - f(\theta_{0,N})\right)$$

$$K_{\Omega,f}^{GMM}(\theta_{0,N}) = N \cdot \left(\beta^{*GMM}(\theta_{0,N}) - f(\theta_{0,N})\right)'\left(D_N^{GMM}(\theta_{0,N})'\hat{\Sigma}_g(\theta_{0,N})D_N^{GMM}(\theta_{0,N})\right)^{-1}\left(\beta^{*GMM}(\theta_{0,N}) - f(\theta_{0,N})\right)$$

By the above result, we again have $K_{\Omega,f}^{MD}(\theta_{0,N}) = K_{\Omega,f}^{GMM}(\theta_{0,N}) + o_p(1)$ for Ω the efficient weight matrix.

[6]Since both are consistent estimators, we have

$$\hat{\Sigma}_{\theta g}(\theta_{0,N}) \to_p -Acov\left(vec\left(\frac{1}{N}\sum Z_i X_i'\right), \frac{1}{N}\sum Z_i\left(U_i - V_i'\theta_{0,N}\right)\right) - Acov\left(vec\left(\frac{1}{N}\sum Z_i V_i'\right), \frac{1}{N}\sum Z_i\left(U_i - V_i'\theta_{0,N}\right)\right)$$

and

$$\hat{\Sigma}_{\theta r}(\theta_{0,N}) \to_p -Acov\left(vec(\hat{\pi} - \pi_{0,N}), \left(\hat{\delta} - \delta_{0.N} - (\hat{\pi} - \pi_{0.N})\theta_{0,N}\right)\right)$$
$$= \left(I_m \otimes E\left[Z_i Z_i'\right]\right)^{-1} Acov\left(vec\left(\frac{Z'V}{N}\right), \frac{Z'U}{N} - \frac{Z'V}{N}\theta_{0,N}\right)\left(E\left[Z_i Z_i'\right]\right)^{-1}.$$

By the continuous mapping theorem, we have

$$\left(I_m \otimes \frac{Z'Z}{N}\right)\hat{\Sigma}_{\theta r}(\theta_{0,N})\left(\frac{Z'Z}{N}\right) \to_p Acov\left(vec\left(\frac{Z'V}{N}\right), \frac{Z'U}{N} - \frac{Z'V}{N}\theta_{0,N}\right).$$

$\square$

Proof of Claim 1

Proof. The assumption $\sup_{\theta \in A_{\theta,N}} \| \theta - \theta_0 \| = O_p\left(\frac{1}{\sqrt{N}}\right)$ and the assumption that $A_{\theta,N}$ is closed implies that $A_{\theta,N}$ is compact with probability tending to one. Stochastic equicontinuity of $\hat{\Sigma}_r(\theta)$ and $\hat{\Sigma}_g(\theta)$ on the set $A_{\theta,N}$ follows because $\hat{\Sigma}_r(\theta)$ and $\hat{\Sigma}_g(\theta)$ are quadratic in θ. Thus the uniform bound

$$\sup_{\theta_N \in A_{\theta,N}} \left| \left(\frac{Z'Z}{N}\right) \hat{\Sigma}_r(\theta_N) \left(\frac{Z'Z}{N}\right) - \hat{\Sigma}_g(\theta_N) \right| = o_p(1)$$

will follow if we can prove that the pointwise bounds

$$\left| \left(\frac{Z'Z}{N}\right) \hat{\Sigma}_r(\theta_N) \left(\frac{Z'Z}{N}\right) - \hat{\Sigma}_g(\theta_N) \right| = o_p(1)$$

hold whenever $\| \theta_N - \theta_0 \| = O_p(\frac{1}{\sqrt{N}})$ due to the stochastic Arzelà-Ascoli lemma.[7] Similarly, pointwise bounds on the difference in the covariance estimators imply a uniform bound. Thus, we will prove that $\left(\frac{Z'Z}{N}\right) \hat{\Sigma}_r(\theta_N) \left(\frac{Z'Z}{N}\right)$ and $\hat{\Sigma}_g(\theta_N)$ have the same probability limit, and that $\left(I_m \otimes \frac{Z'Z}{N}\right) \hat{\Sigma}_{\theta r}(\theta_N) \left(\frac{Z'Z}{N}\right)$ and $\hat{\Sigma}_{\theta g}(\theta_N)$ have the same probability limit.

We first derive the expressions for the probability limit of $\hat{\Sigma}_g(\theta_N)$. The GMM moment evaluated at $\theta_N \in A_{\theta,N}$ is

$$g_N(\theta_N) = \frac{1}{N} \sum Z_i \left(Y_i - X_i'\theta_N\right) = \frac{1}{N} \sum Z_i \left(U_i - V_i'\theta_N + Z_i'\pi_0 (\theta_0 - \theta_N)\right)$$

where we substitute in Equation 1 and 2.[8] Assume some form of CLT holds for the moment and its asymptotic variance exists and is finite. Then

$$\hat{\Sigma}_g(\theta_N) = A\hat{var}\left(\frac{1}{N} \sum Z_i \left(Y_i - X_i'\theta_N\right)\right) = A\hat{var}\left(\frac{1}{N} \sum Z_i \left(U_i - V_i'\theta_N + Z_i'\pi_0(\theta_0 - \theta_N)\right)\right)$$

$$= A\hat{var}\left(\frac{1}{N} \sum Z_i \left(U_i - V_i'\theta_N\right)\right) + A\hat{cov}\left(\frac{1}{N} \sum Z_i \left(U_i - V_i'\theta_N\right), \frac{1}{N} \sum Z_i Z_i'\pi_0(\theta_0 - \theta_N)\right) \quad (7)$$

$$+ A\hat{cov}\left(\frac{1}{N} \sum Z_i Z_i'\pi_0(\theta_0 - \theta_N), \frac{1}{N} \sum Z_i \left(U_i - V_i'\theta_N\right)\right) + A\hat{var}\left(\frac{1}{N} \sum Z_i Z_i'\pi_0(\theta_0 - \theta_N)\right)$$

$$= A\hat{var}\left(\frac{1}{N} \sum Z_i \left(U_i - V_i'\theta_N\right)\right)$$

$$+ A\hat{cov}\left(\frac{1}{N} \sum Z_i \left(U_i - V_i'\theta_N\right), \frac{1}{N} \sum vec\left(Z_i Z_i'\pi_0\right)\right)\left((\theta_0 - \theta_N) \otimes I_k\right)$$

$$+ \left((\theta_0 - \theta_N) \otimes I_k\right)' A\hat{cov}\left(\frac{1}{N} \sum vec\left(Z_i Z_i'\pi_0\right), \frac{1}{N} \sum Z_i \left(U_i - V_i'\theta_N\right)\right)$$

[7] For details, see proof of Theorem 2.7 in Newey and McFadden (1994).

[8] Unfortunately we don't have $Y_i - X_i'\theta_N = U_i - V_i'\theta_0 + V_i'(\theta_0 - \theta_N)$. We only have $Y_i - X_i'\theta_N = Y_i - X_i'\theta_0 + X_i'\theta_0 - X_i'\theta_N = U_i - V_i'\theta_0 + X_i'(\theta_0 - \theta_N)$. So I changed your derivation.

$$+\left(\left(\theta_0-\theta_N\right)\otimes I_k\right)'\hat{Avar}\left(\frac{1}{N}\sum vec\left(Z_iZ_i'\pi_0\right)\right)\left(\left(\theta_0-\theta_N\right)\otimes I_k\right).$$

Note that strong identification assumes π_0 is fixed. Since $\parallel \theta_N-\theta_0 \parallel\leq O_p\left(\frac{1}{\sqrt{N}}\right)$ by assumption, all the terms in the above expression but the first are negligible in the limit. Assume these asymptotic variance estimators are consistent, the probability limit of the GMM variance estimator is just

$$\hat{\Sigma}_g\left(\theta_N\right)\to_p Avar\left(\frac{1}{N}\sum Z_i\left(U_i-V_i'\theta_0\right)\right).$$

by the continuous mapping theorem and $\theta_N\to_p\theta_0$.

Similarly we derive the expression for the probability limit of $\hat{\Sigma}_{\theta g}\left(\theta_N\right)$. The Jacobian of the GMM moment evaluated at $\theta_N\in A_{\theta,N}$ is

$$\frac{\partial g_N\left(\theta_N\right)}{\partial\theta'}=-\frac{1}{N}\sum Z_iX_i'.$$

Assume some form of CLT holds such that the asymptotic covariance between the Jacobian and the moment function exists and is finite. Then

$$\hat{\Sigma}_{\theta g}\left(\theta_N\right)=-\hat{Acov}\left(\frac{1}{N}\sum vec\left(Z_iX_i'\right),\frac{1}{N}\sum Z_i\left(Y_i-X_i'\theta_N\right)\right)$$

$$=-\hat{Acov}\left(\frac{1}{N}\sum vec\left(Z_iX_i'\right),\frac{1}{N}\sum Z_i\left(U_i-V_i'\theta_N+Z_i'\pi_0\left(\theta_0-\theta_N\right)\right)\right)$$

$$=-\hat{Acov}\left(\frac{1}{N}\sum vec\left(Z_iX_i'\right),\frac{1}{N}\sum Z_i\left(U_i-V_i'\theta_N\right)\right)$$

$$-\hat{Acov}\left(\frac{1}{N}\sum vec\left(Z_iX_i'\right),\frac{1}{N}\sum Z_i\left(Z_i'\pi_0\left(\theta_0-\theta_N\right)\right)\right)$$

$$=-\hat{Acov}\left(\frac{1}{N}\sum vec\left(Z_iX_i'\right),\frac{1}{N}\sum Z_i\left(U_i-V_i'\theta_N\right)\right)$$

$$-\hat{Acov}\left(\frac{1}{N}\sum vec\left(Z_iX_i'\right),\frac{1}{N}\sum vec\left(Z_iZ_i'\pi_0\right)\right)\left(\left(\theta_0-\theta_N\right)\otimes I_k\right).$$

Since $\parallel \theta_N-\theta_0 \parallel\leq O_p\left(\frac{1}{\sqrt{N}}\right)$ by assumption, the second term in the above expression is negligible in the limit. Assuming these asymptotic covariance estimators are consistent, by the continuous mapping theorem and $\theta_N\to_p\theta_0$, the probability limit of $\hat{\Sigma}_{\theta g}\left(\theta_N\right)$ is

$$\hat{\Sigma}_{\theta g}\left(\theta_N\right)\to_p -Acov\left(vec\left(\frac{1}{N}\sum Z_iX_i'\right),\frac{1}{N}\sum Z_i\left(U_i-V_i'\theta_0\right)\right).$$

Substituting in Equation 2, the probability limit equals

$$-Acov\left(vec\left(\left(\frac{1}{N}\sum Z_iZ_i'\right)\pi_0+\left(\frac{1}{N}\sum Z_iV_i'\right)\right),\frac{1}{N}\sum Z_i\left(U_i-V_i'\theta_0\right)\right)$$

$$=-Acov\left(vec\left(\left(\frac{1}{N}\sum Z_iZ_i'\right)\pi_0\right),\frac{1}{N}\sum Z_i\left(U_i-V_i'\theta_0\right)\right)$$

$$-Acov\left(vec\left(\frac{1}{N}\sum Z_iV_i'\right),\frac{1}{N}\sum Z_i\left(U_i-V_i'\theta_0\right)\right)$$

$$= -\left(E\left[Z_i Z_i^{'}\right] \otimes I_m\right)^{'} Acov\left(vec\left(\pi_0\right), \frac{1}{N}\sum Z_i\left(U_i - V_i^{'}\theta_0\right)\right)$$
$$-Acov\left(vec\left(\frac{1}{N}\sum Z_i V_i^{'}\right), \frac{1}{N}\sum Z_i\left(U_i - V_i^{'}\theta_0\right)\right)$$
$$= -Acov\left(vec\left(\frac{1}{N}\sum Z_i V_i^{'}\right), \frac{1}{N}\sum Z_i\left(U_i - V_i^{'}\theta_0\right)\right).$$

Note that strong identification assumes π_0 is fixed. Thus $Acov\left(vec\left(\pi_0\right), \frac{1}{N}\sum Z_i\left(U_i - V_i^{'}\theta_N\right)\right) = 0$.

We next derive the expressions for the probability limit of $\hat{\Sigma}_r\left(\theta_N\right)$. Recall that the MD variance estimator is

$$\hat{\Sigma}_r\left(\theta_N\right) = A\hat{var}\left(\hat{\delta} - \hat{\pi}\theta_N\right)$$

for $\hat{\delta} = \left(\frac{1}{N}\sum Z_i Z_i^{'}\right)^{-1}\frac{1}{N}\sum Z_i Y_i$ and $\hat{\pi} = \left(\frac{1}{N}\sum Z_i Z_i^{'}\right)^{-1}\frac{1}{N}\sum Z_i X_i^{'}$, the OLS estimators. Assume we have

$$\sqrt{N}\begin{pmatrix} \hat{\delta} - \delta_0 \\ vec(\hat{\pi} - \pi_0) \end{pmatrix} \to_d N\left(0, \begin{pmatrix} \Sigma_\delta & \Sigma_{\delta\pi} \\ \Sigma_{\pi\delta} & \Sigma_\pi \end{pmatrix}\right)$$

where $\begin{pmatrix} \Sigma_\delta & \Sigma_{\delta\pi} \\ \Sigma_{\pi\delta} & \Sigma_\pi \end{pmatrix} =$

$$\begin{pmatrix} \left(E\left[Z_i Z_i^{'}\right]\right)^{-1} Avar(\frac{Z'U}{N})\left(E\left[Z_i Z_i^{'}\right]\right)^{-1} & \left(E\left[Z_i Z_i^{'}\right]\right)^{-1} Acov(\frac{Z'U}{N}, vec(\frac{Z'V}{N}))\left(E\left[Z_i Z_i^{'}\right] \otimes I_m\right)^{-1} \\ \left(E\left[Z_i Z_i^{'}\right] \otimes I_m\right)^{-1} Acov(vec(\frac{Z'V}{N}), \frac{Z'U}{N})\left(E\left[Z_i Z_i^{'}\right]\right)^{-1} & \left(E\left[Z_i Z_i^{'}\right] \otimes I_m\right)^{-1} Avar(vec(\frac{Z'V}{N}))\left(E\left[Z_i Z_i^{'}\right] \otimes I_m\right)^{-1} \end{pmatrix}.$$

Assume $\hat{\Sigma}_r\left(\theta_N\right)$ is a consistent estimator. Then

$$\hat{\Sigma}_r\left(\theta_N\right) \to_p Avar\left(\hat{\delta} - \hat{\pi}\theta_0\right) = \left(E\left[Z_i Z_i^{'}\right]\right)^{-1} Avar\left(\frac{1}{N}\sum Z_i\left(U_i - V_i^{'}\theta_0\right)\right)\left(E\left[Z_i Z_i^{'}\right]\right)^{-1}.$$

where the convergence follows from the continuous mapping theorem and $\theta_N \to_p \theta_0$. Again, by the continuous mapping theorem, we have

$$\left(\frac{Z^{'}Z}{N}\right)\hat{\Sigma}_r\left(\theta_N\right)\left(\frac{Z^{'}Z}{N}\right) \to_p Avar\left(\frac{1}{N}\sum Z_i\left(U_i - V_i^{'}\theta_0\right)\right)$$

Similarly, we derive the expression for the probability limit of $\hat{\Sigma}_{\theta r}\left(\theta_N\right)$. The Jacobian of the structural function evaluated at $\theta_N \in A_{\theta,N}$ is

$$\frac{\partial r\left(\hat{\tau}, \theta_N\right)}{\partial\theta^{'}} = -\hat{\pi}.$$

Then

$$\hat{\Sigma}_{\theta r}\left(\theta_N\right) = A\hat{cov}\left(vec\left(\frac{\partial r\left(\hat{\tau}, \theta_N\right)}{\partial\theta^{'}}\right), \left(\hat{\delta} - \hat{\pi}\theta_N\right)\right)$$
$$= -A\hat{cov}\left(vec\left(\hat{\pi} - \pi_0\right), \left(\hat{\delta} - \delta_0 - \left(\hat{\pi} - \pi_0\right)\theta_N\right)\right).$$

Assume $\hat{\Sigma}_{\theta r}\left(\theta\right)$ is a consistent estimator for the asymptotic covariance, by the continuous mapping theorem and $\theta_N \to_p \theta_0$, we have

$$\hat{\Sigma}_{\theta r}\left(\theta_N\right) \to_p -Acov\left(vec\left(\hat{\pi} - \pi_0\right), \left(\hat{\delta} - \delta_0 - \left(\hat{\pi} - \pi_0\right)\theta_0\right)\right)$$

$$= \left(I_m \otimes E\left[Z_i Z_i^{'}\right]\right)^{-1} Acov\left(vec\left(\frac{Z'V}{N}\right), \frac{Z'U}{N} - \frac{Z'V}{N}\theta_0\right)\left(E\left[Z_i Z_i^{'}\right]\right)^{-1}.$$

Again, by the continuous mapping theorem, we have

$$\left(I_m \otimes \frac{Z^{'}Z}{N}\right)\hat{\Sigma}_{\theta r}\left(\theta_N\right)\left(\frac{Z^{'}Z}{N}\right) \to_p -Acov\left(vec\left(\frac{1}{N}\sum Z_i V_i^{'}\right), \frac{1}{N}\sum Z_i\left(U_i - V_i^{'}\theta_0\right)\right).$$

We have shown that $\hat{\Sigma}_g\left(\theta_N\right)$ and $\left(\frac{Z^{'}Z}{N}\right)\hat{\Sigma}_r\left(\theta_N\right)\left(\frac{Z^{'}Z}{N}\right)$ have the same probability limit. We have also shown that $\hat{\Sigma}_{\theta g}\left(\theta_N\right)$ and $\left(I_m \otimes \frac{Z^{'}Z}{N}\right)\hat{\Sigma}_{\theta r}\left(\theta_N\right)\left(\frac{Z^{'}Z}{N}\right)$ have the same probability limit. Thus we conclude that $\left|\left(\frac{Z^{'}Z}{N}\right)\hat{\Sigma}_r\left(\theta_N\right)\left(\frac{Z^{'}Z}{N}\right) - \hat{\Sigma}_g\left(\theta_N\right)\right| = o_p(1)$ and $\left|\left(I_m \otimes \frac{Z^{'}Z}{N}\right)\hat{\Sigma}_{\theta r}\left(\theta_N\right)\left(\frac{Z^{'}Z}{N}\right) - \hat{\Sigma}_{\theta g}\left(\theta_N\right)\right| = o_p(1)$. □

Proof of Theorem 2

Proof. Note that as a consequence of Claim 1, the proof for Theorem 1 can be adapted to show that $K_{\Omega,f}^{MD}\left(\theta_N\right)$, and $S^{MD}\left(\theta_N\right)$ are asymptotically equivalent to their GMM counterparts for $\theta_N \in A_{\theta,N}$. So here I only show that $W_{\Omega}^{MD}\left(\theta_N\right)$ is asymptotically equivalent to $W_{\Omega}^{GMM}\left(\theta_N\right)$ for $\theta_N \in A_{\theta,N}$. Note that $W_{\Omega}^{GMM}(\theta)$ and $W_{\Omega}^{MD}(\theta)$ are quadratic in θ. Then again, by stochastic Arzelà-Ascoli lemma, the pointwise convergence implies uniform convergence as desired.

The Wald statistic under GMM is defined as

$$W_{\Omega}^{GMM}\left(\beta_N\right) = N \cdot \left(f(\tilde{\theta}) - \beta_N\right)^{'}\hat{V}_{\beta}^{GMM-1}\left(f(\tilde{\theta}) - \beta_N\right)$$

where $\hat{V}_{\Omega,\beta}^{GMM} = F(\tilde{\theta}) \cdot \hat{V}_{\Omega,\theta}^{GMM} \cdot F(\tilde{\theta})^{'}$ and

$$\hat{V}_{\Omega,\theta}^{GMM} = \left(\frac{\partial g_N(\theta)}{\partial\theta}^{'}\hat{\Omega}^{GMM}\frac{\partial g_N(\theta)}{\partial\theta}\right)^{-1}\frac{\partial g_N(\theta)}{\partial\theta}^{'}\hat{\Omega}^{GMM}\hat{\Sigma}_g\left(\tilde{\theta}\right)\hat{\Omega}^{GMM}\frac{\partial g_N(\theta)}{\partial\theta}\left(\frac{\partial g_N(\theta)}{\partial\theta}^{'}\hat{\Omega}^{GMM}\frac{\partial g_N(\theta)}{\partial\theta}\right)^{-1}.$$

The Wald statistic under MD is defined as

$$W_{\Omega}^{MD}\left(\beta_N\right) = N \cdot \left(f(\tilde{\theta}) - \beta_N\right)^{'}\hat{V}_{\beta}^{MD-1}\left(f(\tilde{\theta}) - \beta_N\right)$$

where $\hat{V}_{\Omega,\beta}^{MD} = F(\tilde{\theta}) \cdot \hat{V}_{\Omega,\theta}^{MD} \cdot F(\tilde{\theta})^{'}$ and

$$\hat{V}_{\Omega,\theta}^{MD} = \left(\frac{\partial r(\hat{\tau},\theta)}{\partial\theta}^{'}\hat{\Omega}^{MD}\frac{\partial r(\hat{\tau},\theta)}{\partial\theta}\right)^{-1}\frac{\partial r(\hat{\tau},\theta)}{\partial\theta}^{'}\hat{\Omega}^{MD}\hat{\Sigma}_r\left(\tilde{\theta}\right)\hat{\Omega}^{MD}\frac{\partial r(\hat{\tau},\theta)}{\partial\theta}\left(\frac{\partial r(\hat{\tau},\theta)}{\partial\theta}^{'}\hat{\Omega}^{MD}\frac{\partial r(\hat{\tau},\theta)}{\partial\theta}\right)^{-1}.$$

Recall that under strong identification, we assume that we have an estimator $\tilde{\theta}$ satisfies $\sqrt{N}\left(\tilde{\theta} - \hat{\theta}\right) \to_p 0$ where $\hat{\theta}$ is $\sqrt{N}$-consistent. Thus $\tilde{\theta}$ is $\sqrt{N}$-consistent for θ_0 and is thus in $A_{\theta,N}$.

First consider the case of inefficient weight matrix. That is, $\hat{\Omega}^{MD} = \frac{Z^{'}Z}{N}$ for MD and $\hat{\Omega}^{GMM} = \left(\frac{Z^{'}Z}{N}\right)^{-1}$ for GMM. Note that

$$\frac{\partial r(\hat{\tau},\theta)}{\partial\theta}^{'}\hat{\Omega}^{MD}\frac{\partial r(\hat{\tau},\theta)}{\partial\theta} = -\hat{\pi}^{'}\hat{\Omega}^{MD}\left(-\hat{\pi}\right) = \left(\frac{Z^{'}X}{N}\right)\left(\frac{Z^{'}Z}{N}\right)^{-1}\left(\frac{Z^{'}Z}{N}\right)\left(\frac{Z^{'}Z}{N}\right)^{-1}\left(\frac{Z^{'}X}{N}\right)$$

$$= \left(\frac{Z^{'}X}{N}\right)\left(\frac{Z^{'}Z}{N}\right)^{-1}\left(\frac{Z^{'}X}{N}\right) = \frac{\partial g_N(\theta)}{\partial\theta}^{'}\hat{\Omega}^{GMM}\frac{\partial g_N(\theta)}{\partial\theta}$$

and

$$\frac{\partial r(\hat{\tau},\theta)}{\partial\theta}'\hat{\Omega}^{MD}\hat{\Sigma}_r\left(\tilde{\theta}\right)\hat{\Omega}^{MD}\frac{\partial r(\hat{\tau},\theta)}{\partial\theta} = -\hat{\pi}'\hat{\Omega}^{MD}\hat{\Sigma}_r\left(\tilde{\theta}\right)\hat{\Omega}^{MD}(-\hat{\pi})$$

$$= \left(\frac{Z'X}{N}\right)\left(\frac{Z'Z}{N}\right)^{-1}\left(\frac{Z'Z}{N}\right)\hat{\Sigma}_r\left(\tilde{\theta}\right)\left(\frac{Z'Z}{N}\right)\left(\frac{Z'Z}{N}\right)^{-1}\left(\frac{Z'X}{N}\right).$$

Given that $\left(\frac{Z'Z}{N}\right)\hat{\Sigma}_r\left(\tilde{\theta}\right)\left(\frac{Z'Z}{N}\right) = \hat{\Sigma}_g\left(\tilde{\theta}\right) + o_p(1)$, the above expression equals

$$= \left(\frac{Z'X}{N}\right)\left(\frac{Z'Z}{N}\right)^{-1}\left(\hat{\Sigma}_g\left(\tilde{\theta}\right) + o_p(1)\right)\left(\frac{Z'Z}{N}\right)^{-1}\left(\frac{Z'X}{N}\right)$$

$$= \frac{\partial g_N(\theta)}{\partial\theta}'\hat{\Omega}^{GMM}\hat{\Sigma}_g\left(\tilde{\theta}\right)\hat{\Omega}^{GMM}\frac{\partial g_N(\theta)}{\partial\theta} + o_p(1).$$

Thus we have $W_{\Omega}^{MD}(\beta_N) = W_{\Omega}^{GMM}(\beta_N) + o_p(1)$.

Next consider the case of efficient weight matrix. That is, $\hat{\Omega}^{MD}\left(\tilde{\theta}\right) = \hat{\Sigma}_r\left(\tilde{\theta}\right)^{-1}$ for MD and $\hat{\Omega}^{GMM}\left(\tilde{\theta}\right) = \hat{\Sigma}_g\left(\tilde{\theta}\right)^{-1}$ for GMM. Note that we can simplify

$$\hat{V}_{\Omega,\theta}^{GMM} = \left(\frac{\partial g_N(\theta)}{\partial\theta}'\hat{\Sigma}_g\left(\tilde{\theta}\right)^{-1}\frac{\partial g_N(\theta)}{\partial\theta}\right)^{-1}$$

Thus we have

$$\frac{\partial r(\hat{\tau},\theta)}{\partial\theta}'\hat{\Sigma}_r\left(\tilde{\theta}\right)^{-1}\frac{\partial r(\hat{\tau},\theta)}{\partial\theta} = -\hat{\pi}'\hat{\Sigma}_r\left(\tilde{\theta}\right)^{-1}(-\hat{\pi})$$

$$= \left(\frac{Z'X}{N}\right)\left(\frac{Z'Z}{N}\right)^{-1}\hat{\Sigma}_r\left(\tilde{\theta}\right)^{-1}\left(\frac{Z'Z}{N}\right)^{-1}\left(\frac{Z'X}{N}\right)$$

Given that $\left(\frac{Z'Z}{N}\right)\hat{\Sigma}_r\left(\tilde{\theta}\right)\left(\frac{Z'Z}{N}\right) = \hat{\Sigma}_g\left(\tilde{\theta}\right) + o_p(1)$, the above expression equals

$$= \left(\frac{Z'X}{N}\right)\left(\hat{\Sigma}_g\left(\tilde{\theta}\right) + o_p(1)\right)^{-1}\left(\frac{Z'X}{N}\right)$$

$$= \frac{\partial g_N(\theta)}{\partial\theta}'\hat{\Sigma}_g\left(\tilde{\theta}\right)^{-1}\frac{\partial g_N(\theta)}{\partial\theta} + o_p(1)$$

by continuous mapping theorem. Thus we have $W_{\Omega}^{MD}(\beta_N) = W_{\Omega}^{GMM}(\beta_N) + o_p(1)$. □

References

Andrews, Isaiah, "Valid Two-Step Identification-Robust Confidence Sets for GMM," *Review of Economics and Statistics*, Forthcoming.

Kleibergen, Frank, "Testing Parameters in GMM Without Assuming that They Are Identified," *Econometrica*, 2005, *73* (4), 1103–1123.

Magnusson, Leandro M., "Inference in limited dependent variable models robust to weak identification," *Econometrics Journal*, 2010, *13* (3), S56–S79.

Newey, Whitney K. and Daniel McFadden, "Chapter 36 Large sample estimation and hypothesis testing," *Handbook of Econometrics*, 1994, *4*, 2111 – 2245.