Supplementary Material

on

GMM Estimation and Uniform Subvector Inference with Possible Identification Failure

Donald W. K. Andrews
Cowles Foundation for Research in Economics
Yale University

Xu Cheng
Department of Economics
University of Pennsylvania

First Draft: August, 2007
Revised: January 30, 2013

8. Outline

This Supplement includes five Supplemental Appendices (denoted A-E) to the paper "GMM Estimation and Uniform Subvector Inference with Possible Identification Failure," denoted hereafter by AC3. Supplemental Appendix A verifies the assumptions of AC3 for the probit model with endogeneity. Supplemental Appendix B provides proofs of the GMM estimation results given in Section 4 of AC3. It also provides some results for minimum distance estimators. Supplemental Appendix C provides proofs of the Wald test and CS results given in Section 5 of AC3. Supplemental Appendix D gives some results that are used in the verification of the assumptions for the two examples of AC3. Supplemental Appendix E provides additional numerical results to those provided in AC3 for the nonlinear regression model with endogeneity.

9. Supplemental Appendix A: Probit Model with Endogeneity: Verification of Assumptions

In this Supplemental Appendix, we verify Assumptions GMM1-GMM5 and V1-V2 for the probit model with endogeneity and possibly weak instruments. Assumptions B1 and B2 hold immediately in this model given the definitions of Θ, Θ^*, and $\Phi^*(\theta)$ in Section 2.3 of AC3.

9.1. Verification of Assumption GMM1

Assumption GMM1(i) holds by (2.19) and (2.20) because $Z_i'\beta\pi$ does not depend on π when $\beta = 0$.

The quantity $g_0(\theta;\gamma)$ that appears in Assumptions GMM1(ii)-(v) is

$$\begin{aligned} g_0(\theta;\gamma_0) &= E_{\gamma_0} e_i(\theta) \otimes \overline{Z}_i = E_{\gamma_0} e_{0,i}(\theta) \otimes \overline{Z}_i, \text{ where} \\ e_{0,i}(\theta) &= \begin{pmatrix} w_{1,i}(\theta)(L_i(\theta_0) - L_i(\theta)) \\ Z_i'(\beta_0 - \beta) - X_i'(\zeta_{2,0} - \zeta_2) \end{pmatrix} \in R^2. \end{aligned} \tag{9.1}$$

The first uniform convergence condition in Assumption GMM1(ii) follows from the ULLN given in Lemma 12.1 in Supplemental Appendix D because $E_{\gamma_0}(y_i|X_i, Z_i) = L_i(\theta_0)$ when the true value is $\gamma_0 = (\theta_0, \phi_0)$.

When $\mathcal{W}_n(\theta)$ is the identity matrix, $\mathcal{W}(\theta;\gamma_0)$ in Assumption GMM1(ii) also is the identity matrix. When $\mathcal{W}_n(\theta)$ is the optimal weight matrix defined in (2.20), Assumption GMM1(ii) holds with

$$\begin{aligned}\mathcal{W}(\theta;\gamma_0) &= E_{\gamma_0}\left(e_i(\theta)e_i(\theta)'\right)\otimes(\overline{Z}_i\overline{Z}_i') = E_{\gamma_0}(\mathcal{W}_{e,i}(\theta;\gamma_0)\otimes(\overline{Z}_i\overline{Z}_i'), \text{ where}\\ \mathcal{W}_{e,i}(\theta;\gamma_0) &= E_{\gamma_0}\left(e_i(\theta)e_i(\theta)'|\overline{Z}_i\right) = \begin{pmatrix}\mathcal{W}_{11,i}(\theta) & \mathcal{W}_{12,i}(\theta)\\ \mathcal{W}_{12,i}(\theta) & \mathcal{W}_{22,i}(\theta)\end{pmatrix}\end{aligned} \tag{9.2}$$

and $\mathcal{W}_{11,i}(\theta), \mathcal{W}_{12,i}(\theta)$, and $\mathcal{W}_{22,i}(\theta)$ are defined in (9.4)-(9.5) below.[31] The convergence condition in Assumption GMM1(ii) holds for the optimal weight matrix $\mathcal{W}_n(\theta)$ by the ULLN given in Lemma 12.1 in Supplemental Appendix C.

Now we derive the elements of $\mathcal{W}_{e,i}(\theta;\gamma_0)$ in (9.2). Note that

$$P_{\gamma_0}(y_i = 1|\overline{Z}_i) = L_i(\theta_0) \text{ and } P_{\gamma_0}(y_i = 0|\overline{Z}_i) = 1 - L_i(\theta_0). \tag{9.3}$$

The upper left element of $\mathcal{W}_{e,i}(\theta;\gamma_0)$ is

$$\mathcal{W}_{11,i}(\theta) = E_{\gamma_0}(w_{1,i}(\theta)^2(y_i - L_i(\theta))^2|\overline{Z}_i) = w_{1,i}(\theta)^2(L_i(\theta_0) - 2L_i(\theta_0)L_i(\theta) + L_i(\theta)^2). \tag{9.4}$$

The lower-right element of $\mathcal{W}_{e,i}(\theta;\gamma_0)$ is

$$\mathcal{W}_{22,i}(\theta) = E_{\gamma_0}((Y_i - Z_i'\beta - X_i'\zeta_2)^2|\overline{Z}_i) = \sigma_v^2 + (Z_i'(\beta_0 - \beta) + X_i'(\zeta_{2,0} - \zeta_2))^2. \tag{9.5}$$

To calculate the off-diagonal term of $\mathcal{W}_{e,i}(\theta;\gamma_0)$, note that

$$\begin{aligned}E_{\gamma_0}(V_i|\overline{Z}_i, y_i = 1) &= E_{\gamma_0}(V_i|\overline{Z}_i, U_i > -(Z_i'\beta_0\pi_0 + X_i'\zeta_{1,0})) = \sigma_v\rho\frac{L_i'(\theta_0)}{L_i(\theta_0)} \text{ and}\\ E_{\gamma_0}(V_i|\overline{Z}_i, y_i = 0) &= E_{\gamma_0}(V_i|\overline{Z}_i, -U_i > Z_i'\beta_0\pi_0 + X_i'\zeta_{1,0}) = -\sigma_v\rho\frac{L_i'(\theta_0)}{1 - L_i(\theta_0)}.\end{aligned} \tag{9.6}$$

[31] Note that $\mathcal{W}_{11,i}(\theta), \mathcal{W}_{12,i}(\theta)$, and $\mathcal{W}_{22,i}(\theta)$ all depend on γ_0. We omit γ_0 from these terms for notational simplicity.

The off-diagonal term of $\mathcal{W}_{e,i}(\theta;\gamma_0)$ is

$$
\begin{aligned}
&\mathcal{W}_{12,i}(\theta)\\
&= E_{\gamma_0}(w_{1,i}(\theta)(y_i - L_i(\theta))(Y_i - Z_i'\beta - X_i'\zeta_2)|\overline{Z}_i)\\
&= w_{1,i}(\theta)\sum_{k=0,1}(k - L_i(\theta))\left[E_{\gamma_0}(V_i|\overline{Z}_i, y_i = k) + Z_i'(\beta_0 - \beta) + X_i'(\zeta_{2,0} - \zeta_2)\right]\\
&\quad \times P_{\gamma_0}(y_i = k|\overline{Z}_i)\\
&= w_{1,i}(\theta)\left[(1 - L_i(\theta))\sigma_v\rho\frac{L_i'(\theta_0)}{L_i(\theta_0)}L_i(\theta_0) + L_i(\theta)\sigma_v\rho\frac{L_i'(\theta_0)}{1 - L_i(\theta_0)}(1 - L_i(\theta_0)\right] +\\
&\quad w_{1,i}(\theta)\left[(1 - L_i(\theta))L_i(\theta_0) - L_i(\theta)(1 - L_i(\theta_0))\right]\left[Z_i'(\beta_0 - \beta) + X_i'(\zeta_{2,0} - \zeta_2)\right]\\
&= w_{1,i}(\theta)\left[\sigma_v\rho L_i'(\theta_0) + (L_i(\theta_0) - L_i(\theta))\left(Z_i'(\beta_0 - \beta) + X_i'(\zeta_{2,0} - \zeta_2)\right)\right]. \qquad (9.7)
\end{aligned}
$$

Now we verify Assumptions GMM1(iii) and GMM1(iv). We write $g_0(\theta;\gamma_0) = (g_{1,0}(\theta;\gamma_0)', g_{2,0}(\theta;\gamma_0)')'$ for $g_{j,0}(\theta;\gamma_0) \in R^{d_X + d_Z}$ for $j = 1, 2$. We have

$$
g_{2,0}(\theta;\gamma_0)\xi = \xi' E_{\gamma_0}\overline{Z}_i\overline{Z}_i'\xi > 0 \text{ for } \xi = ((\beta_0 - \beta)', (\zeta_{2,0} - \zeta_2))',
$$

where the inequality holds because $E_{\gamma_0}\overline{Z}_i\overline{Z}_i'$ is positive definite since $P_{\phi_0}(\overline{Z}_i'c = 0) < 1$ for any $c \neq 0$ by (2.21). Hence, $g_{2,0}(\theta;\gamma_0) = 0$ if and only if $\beta = \beta_0$ and $\zeta_2 = \zeta_{2,0}$. Now, for θ with $\beta = \beta_0$ and $\zeta_2 = \zeta_{2,0}$,

$$
g_{1,0}(\theta;\gamma_0) = E_{\gamma_0}w_{1,i}(\theta)(L_i(\theta_0) - L_i(\theta))\overline{Z}_i \text{ and } L_i(\theta) = L(Z_i'\beta_0\pi + X_i'\zeta_1). \qquad (9.8)
$$

If $\beta_0 \neq 0$, the conditions $g_{1,0}(\theta;\gamma_0) = 0$ are more restrictive than the populations first-order conditions for the standard probit ML estimator for a probit model with regression function $Z_i'\beta_0\pi + X_i'\zeta_1$ (because the latter has the multiplicative factor $(Z_i'\beta_0, X_i')'$, rather than $\overline{Z}_i$). The latter have a unique solution at the true parameter vector because, as is well known, the population log likelihood function of the probit model is strictly concave. Hence, $g_{1,0}(\theta;\gamma_0) = 0$ only if $\pi = \pi_0$ and $\zeta_1 = \zeta_{1,0}$ and Assumption GMM1(iv) holds. If $\beta_0 = 0$, then the same argument holds but with the regression function being $X_i'\zeta_1$, rather than $Z_i'\beta_0\pi + X_i'\zeta_1$. In this case, $g_{1,0}(\theta;\gamma_0) = 0$ only if $\zeta_1 = \zeta_{1,0}$ and Assumption GMM1(iii) holds.

The partial derivatives $g_\psi(\theta;\gamma_0)$ and $g_\theta(\theta;\gamma_0)$ in Assumptions GMM1(v) and GMM1

(viii) are

$$
\begin{aligned}
g_\psi(\theta;\gamma_0) &= E_{\phi_0}\begin{pmatrix}\overline{Z}_i a_i(\theta) d_{1\psi,i}(\pi)' \\ \overline{Z}_i d_{2\psi,i}'\end{pmatrix} \text{ and } g_\theta(\theta;\gamma_0) = E_{\phi_0}\begin{pmatrix}\overline{Z}_i a_i(\theta) d_{1,i}(\theta)' \\ \overline{Z}_i d_{2,i}'\end{pmatrix}, \text{ where} \\
d_{1\psi,i}(\pi) &= (\pi Z_i, X_i, 0_{d_X}) \in R^{d_Z+2d_X},\ d_{2\psi,i} = (Z_i, 0_{d_X}, X_i) \in R^{d_Z+2d_X}, \\
d_{1,i}(\theta) &= (d_{1\psi,i}(\pi), Z_i'\beta) \in R^{d_Z+2d_X+1},\ d_{2,i} = (d_{2\psi,i}, 0) \in R^{d_Z+2d_X+1}, \text{ and} \qquad (9.9) \\
a_i(\theta) &= \frac{L_i'(\theta)^2 + L_i''(\theta)(L_i(\theta) - L_i(\theta_0))}{L_i(\theta)(1 - L_i(\theta))} - \frac{L_i'(\theta)^2 (L_i(\theta) - L_i(\theta_0))(1 - 2L_i(\theta))}{L_i(\theta)^2 (1 - L_i(\theta))^2}.
\end{aligned}
$$

Assumptions GMM1(v) and GMM1(vi) hold by the continuity of $w_{1,i}(\theta)$ and $L_i(\theta)$ in θ and the moment conditions in (2.21).

Next, we verify Assumption GMM1(vii). To show $\lambda_{\min}(\mathcal{W}(\psi_0,\pi;\gamma_0)) > 0$, $\forall \pi \in \Pi$, $\forall \gamma_0 \in \Gamma$, we show that for any $c = (c_1', c_2')'$ with $||c|| > 0$, $c'\mathcal{W}(\psi_0,\pi;\gamma_0)c > 0$, where $c_j \in R^{d_X+d_Z}$ for $j = 1, 2$. Let

$$
U_i^*(\theta) = w_{1,i}(\theta)(U_i + L_i(\theta_0) - L_i(\theta)). \qquad (9.10)
$$

For $\theta \in (\psi_0, \pi)$, we have

$$
\begin{aligned}
c'\mathcal{W}(\psi_0,\pi;\gamma_0)c &= c'\left[E_{\gamma_0}\begin{pmatrix}U_i^*(\theta) \\ V_i\end{pmatrix}\begin{pmatrix}U_i^*(\theta) \\ V_i\end{pmatrix}' \otimes \overline{Z}_i\overline{Z}_i'\right]c \\
&= E_{\gamma_0}E_{\gamma_0}((U_i^*(\theta)c_1'\overline{Z}_i + V_i c_2'\overline{Z}_i)^2|\overline{Z}_i) \\
&\geq E_{\gamma_0}E_{\gamma_0}((U_i w_{1,i}(\theta)c_1'\overline{Z}_i + V_i c_2'\overline{Z}_i)^2|\overline{Z}_i), \qquad (9.11)
\end{aligned}
$$

where the inequality holds because $E_{\gamma_0}(w_{1,i}(\theta)(L_i(\theta_0) - L_i(\theta))c_1'\overline{Z}_i V_i c_2'\overline{Z}_i|\overline{Z}_i) = 0$ a.s. since $E_{\gamma_0}(V_i|\overline{Z}_i) = 0$ a.s. and $E_{\gamma_0}((w_{1,i}(\theta)(L_i(\theta_0) - L_i(\theta))c_1'\overline{Z}_i)^2|\overline{Z}_i) \geq 0$ a.s. The rhs of (9.11) equals zero only if $E_{\gamma_0}((U_i w_{1,i}(\theta)c_1'\overline{Z}_i + V_i c_2'\overline{Z}_i)^2|\overline{Z}_i) = 0$ a.s. But,

$$
E_{\gamma_0}((U_i w_{1,i}(\theta)c_1'\overline{Z}_i + V_i c_2'\overline{Z}_i)^2|\overline{Z}_i) > 0 \qquad (9.12)
$$

for all $\overline{Z}_i$ for which $c_j'\overline{Z}_i \neq 0$ for $j = 1$ and $j = 2$ because $w_{1,i}(\theta) > 0$ a.s., (U_i, V_i) is independent of $\overline{Z}_i$, and $|Cov(U_i, V_i)| = |\rho| < 1$. By (2.21), $P_{\gamma_0}(c_j'\overline{Z}_i \neq 0$ for $j = 1$ and $j = 2) > 0$. Hence, we conclude that $c'\mathcal{W}(\psi_0,\pi;\gamma_0)c > 0$.

In addition, $\lambda_{\max}(\mathcal{W}(\psi_0,\pi;\gamma_0)) < \infty$ because $||\mathcal{W}(\psi_0,\pi;\gamma_0)|| = ||E_{\phi_0}[\mathcal{W}_{e,i}(\theta;\gamma_0) \otimes (\overline{Z}_i\overline{Z}_i')]|| < \infty$ using (9.4)-(9.5) and $E_{\phi_0}(||\overline{Z}_i||^{4+\varepsilon} + \overline{w}_{1,i}^{4+\varepsilon}) < \infty$ for some $\varepsilon > 0$ by (2.21), where $||\cdot||$ denotes the Frobenious norm. Thus, Assumption GMM1(vii) holds.

Assumption GMM1(viii) holds because $\mathcal{W}(\psi_0, \pi; \gamma_0)$ is non-singular $\forall \pi \in \Pi$ and $g_\psi(\psi_0, \pi; \gamma_0)$ has full column rank because $P_{\phi_0}(\overline{Z}_i' c = 0) < 1$ for all $c \neq 0$.

Assumption GMM1(ix) holds automatically by the Assumptions on the parameter space.

Assumption GMM1(x) holds because $\Psi(\pi)$ does not depend on π in this example.

9.2. Verification of Assumption GMM2

We verify Assumption GMM2 using the sufficient condition Assumption GMM2*. Assumption GMM2*(i) holds because $e_i(\theta)$ is continuously differentiable in θ. Assumption GMM2*(ii) holds by the ULLN given in Lemma 12.1 in Supplemental Appendix C. Assumption GMM2*(iii) holds by the uniform LLN given in Lemma 12.1 in Supplemental Appendix D using $||\beta||/||\beta_n|| = 1 + o(1)$ for $\theta \in \Theta_n(\delta_n)$ and $||\beta_n|| \neq 0$ for n large for $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$.

9.3. Verification of Assumption GMM3

Assumption GMM3(i) holds with

$$g(W_i, \theta) = e_i(\theta) \otimes \overline{Z}_i. \tag{9.13}$$

Assumption GMM3(ii) holds because $E_{\gamma^*} g(W_i, \psi^*, \pi) = E_{\gamma^*} e_{0,i}(\psi^*, \pi) \otimes \overline{Z}_i = 0$ when $\beta^* = 0$.

Assumption GMM3(iii) hold by the CLT for triangular arrays of row-wise i.i.d. random variables given in Lemma 12.3 of Supplemental Appendix D. The variance matrix is

$$\begin{aligned}\Omega_g(\gamma_0) &= E_{\gamma_0}\left(e_i(\theta_0) e_i(\theta_0)'\right) \otimes (\overline{Z}_i \overline{Z}_i') = \mathcal{W}(\theta_0; \gamma_0) \\ &= E_{\gamma_0}\begin{pmatrix} w_{1,i}(\theta_0) L_i'(\theta_0) & w_{1,i}(\theta_0) L_i'(\theta_0)\rho\sigma_v \\ w_{1,i}(\theta_0) L_i'(\theta_0)\rho\sigma_v & \sigma_v^2 \end{pmatrix} \otimes \left(\overline{Z}_i \overline{Z}_i'\right),\end{aligned} \tag{9.14}$$

where the second and third equalities follow from (9.2) and (9.4)-(9.5) with $\theta = \theta_0$ and $w_{1,i}(\theta_0)(L_i(\theta_0) - L_i(\theta_0)^2) = L_i'(\theta_0)$.

To verify Assumption GMM3(iv), first note that

$$E_{\gamma^*} g(W_i, \theta) = E_{\gamma^*} \begin{pmatrix} w_{1,i}(\theta)(L_i(\theta^*) - L_i(\theta)) \\ Z_i'(\beta^* - \beta) - X_i'(\zeta_2^* - \zeta_2) \end{pmatrix} \otimes \overline{Z}_i. \tag{9.15}$$

The derivative of $E_{\gamma^*} g(W_i, \theta)$ wrt β^* is

$$K_{n,g}(\theta; \gamma^*) = E_{\phi^*} \begin{pmatrix} w_{1,i}(\theta) L_i'(\theta^*) \pi^* \overline{Z}_i Z_i' \\ \overline{Z}_i Z_i' \end{pmatrix} \tag{9.16}$$

$\forall (\theta, \gamma^*) \in \Theta_\delta \times \Gamma_0$ and $\forall n \geq 1$. This verifies Assumption GMM3(iv)(a). Assumptions GMM3(iv)(b) and (c) hold with $K_g(\theta; \gamma_0) = K_{n,g}(\theta; \gamma_0)$.

To verify Assumption GMM3(v), note that $a_i(\psi_0, \pi) = w_{1,i}(\theta_0) L_i'(\theta_0)$ when $\beta_0 = 0$. Using (9.9) and (9.16), this yields

$$g_\psi(\psi_0, \pi; \gamma_0) = E_{\phi_0} M_i(\theta_0) \begin{pmatrix} d_{1\psi,i}(\pi)' \\ d_{2\psi,i}' \end{pmatrix}, \quad K_g(\psi_0, \pi; \gamma_0) = E_{\phi_0} M_i(\theta_0) \begin{pmatrix} \pi_0 Z_i' \\ Z_i' \end{pmatrix}, \text{ where}$$
$$M_i(\theta_0) = \begin{pmatrix} w_{1,i}(\theta_0) L_i'(\theta_0) \overline{Z}_i & 0_{d_Z} \\ 0_{d_Z} & \overline{Z}_i \end{pmatrix}. \tag{9.17}$$

Assumption GMM3(v) holds because (i) $M_i(\theta_0)$ has full rank a.s., (ii) $d_{2\psi,i} S = Z_i'$ for $S = (S_1, S_2, S_3) \in R^{d_Z \times d_X \times d_X}$ if and only if $S_1 = 1_{d_Z}$ and $S_3 = 0_{d_X}$, and (iii) $d_{1\psi,i}(\pi) S = \pi_0 Z_i$ for $S = (1_{d_Z}, S_2, 0_{d_X})$ if and only if $S_2 = 0_{d_X}$ and $\pi = \pi_0$.

Assumption GMM3(vi) holds by (9.15), (9.17), an exchange of "E" and "∂," the moment conditions in (2.21), and some calculations. The left-hand side does not depend on an average over n because the observations are identically distributed.

9.4. Verification of Assumption GMM4

When $d_Z > 1$, we do not have a proof that Assumption GMM4 holds. In this case, we just assume that it does. However, when $d_Z = 1$, Assumption GMM4 can be verified by verifying Assumption GMM4*. In this case, Assumption GMM4*(i) holds automatically. Using (9.17), we obtain

$$g_\psi^*(\psi_0, \pi_1, \pi_2; \gamma_0) = E_{\phi_0} M_i(\theta_0) \begin{pmatrix} \pi_1 Z_i', \pi_2 Z_i', X_i', 0_{d_X}' \\ Z_i', Z_i', 0_{d_X}', X_i' \end{pmatrix}, \tag{9.18}$$

where $M_i(\theta_0)$ is of full column rank a.s. Assumption GMM4*(ii) holds because $P_{\phi_0}(\overline{Z}_i'c = 0) < 1$ for $c \neq 0$ and $\pi_1 \neq \pi_2$. Assumption GMM4*(iii) holds with $\Omega_g(\gamma_0) = \mathcal{W}(\theta_0; \gamma_0)$ by (9.2) and (9.14) because $\mathcal{W}(\theta_0; \gamma_0)$ is positive definite by the verification of Assumption GMM1(vii) in (9.10)-(9.12).

9.5. Verification of Assumption GMM5

The verification of Assumption GMM5(i) is analogous to that of Assumption GMM3 (iii). The variance matrix $V_g(\gamma_0)$ is equal to $\Omega_g(\gamma_0)$ defined in (9.14).

Assumption GMM5(ii) holds with $g_\theta(\theta; \gamma_0)$ in (9.9) using $||\beta||/||\beta_n|| = 1 + o(1)$ for $\theta \in \Theta_n(\delta_n)$, $||\beta_n|| \neq 0$ for n large for $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, and the moment conditions in (2.21).

Assumption GMM5(iii) holds with

$$J_g(\gamma_0) = E_{\phi_0} M_i(\theta_0) \begin{pmatrix} \pi_0 Z_i', X_i', 0_{d_X}', Z_i'\omega_0 \\ Z_i', 0_{d_X}', X_i', 0 \end{pmatrix} \tag{9.19}$$

using (9.9) and (9.17) and $\beta_n/||\beta_n|| \to \omega_0$. The matrix $J_g(\gamma_0)$ has full column rank because $P_\phi(\overline{Z}_i'c = 0) < 1$ for $c \neq 0$.

9.6. Verification of Assumptions V1 and V2 (Vector β)

Here we verify Assumptions V1(i)-V1(iii) (vector β) and V2. We do not verify Assumption V1(iv) (vector β). However, it should hold because $\tau_\beta(\pi; \gamma_0, b)$ is a Gaussian process.

We estimate $J(\gamma_0)$ and $V(\gamma_0)$ by $\widehat{J}_n = \widehat{J}_n(\widehat{\theta}_n^+)$ and $\widehat{V}_n = \widehat{V}_n(\widehat{\theta}_n^+)$, respectively, where

$$\begin{aligned}
\widehat{J}_n(\theta^+) &= \widehat{J}_{g,n}(\theta^+)'\mathcal{W}_n\widehat{J}_{g,n}(\theta^+),\ \widehat{V}_n(\theta^+) = \widehat{J}_{g,n}(\theta^+)'\mathcal{W}_n\widehat{V}_{g,n}(\theta^+)\mathcal{W}_n\widehat{J}_{g,n}(\theta^+), \\
\widehat{J}_{g,n}(\theta^+) &= n^{-1}\sum_{i=1}^n M_i(\theta) \begin{pmatrix} \pi Z_i', X_i', 0_{d_X}', Z_i'\omega \\ Z_i', 0_{d_X}', X_i', 0 \end{pmatrix} \text{ and} \\
\widehat{V}_{g,n}(\theta^+) &= n^{-1}\sum_{i=1}^n (e_i(\theta)e_i(\theta)') \otimes \left(\overline{Z}_i\overline{Z}_i'\right).
\end{aligned} \tag{9.20}$$

Assumption V1(i) (vector β) holds with

$$J(\theta^+;\gamma_0) = J_g(\theta^+;\gamma_0)'\mathcal{W}(\theta_0;\gamma_0)J_g(\theta^+;\gamma_0) \text{ and}$$
$$V(\theta^+;\gamma_0) = J_g(\theta^+;\gamma_0)'\mathcal{W}(\theta_0;\gamma_0)V_g(\theta^+;\gamma_0)\mathcal{W}(\theta_0;\gamma_0)J_g(\theta^+;\gamma_0), \tag{9.21}$$

where $J_g(\theta^+;\gamma_0)$ and $V_g(\theta^+;\gamma_0)$ are defined analogously to $\widehat{J}_g(\theta^+)$ and $\widehat{V}_g(\theta^+)$, respectively, but with $n^{-1}\sum_{i=1}^n$ replaced by E_{γ_0}. The uniform convergence conditions of Assumption V1(i) for $\widehat{J}_n(\theta^+)$ and $\widehat{V}_n(\theta^+)$ follow from the uniform convergence of $\widehat{J}_{g,n}(\theta^+)$ and $\widehat{V}_{g,n}(\theta^+)$ and $\mathcal{W}_n \to_p \mathcal{W}(\theta_0;\gamma_0)$. The former holds by the ULLN given in Lemma 12.1 in Supplemental Appendix C. When $\mathcal{W}_n$ is the identity matrix, the latter holds automatically. When $\mathcal{W}_n$ is the optimal weight matrix that involves a first step estimator $\overline{\theta}_n$ and $\overline{\theta}_n$ is based on the identity weight matrix, the convergence in probability of $\mathcal{W}_n$ holds by Lemma 3.1. The assumptions of Lemma 3.1 follow from Theorems 4.1(a) and 4.2(a).

Assumption V1(ii) (vector β) holds by the continuity of $M_i(\theta)$ and $e_i(\theta)$ in θ and the moment conditions in (2.21).

Assumption V1(iii) (vector β) holds provided that $J(\theta^+;\gamma_0)$ and $V(\theta^+;\gamma_0)$ are both finite and non-singular when $\beta_0 = 0$. To this end, we need that $J_g(\theta^+;\gamma_0)$, $V_g(\theta^+;\gamma_0)$, and $\mathcal{W}(\theta;\gamma_0)$ are all finite and non-singular. This holds using the forms of these matrices and $P_\phi(\overline{Z}_i'c = 0) < 1$ for $c \neq 0$ by the arguments used in the verifications of Assumptions GMM5(iii), GMM5(i), and GMM1(vii), respectively.

Assumption V2 follows from (i) the uniform convergence of $\widehat{J}_{g,n}(\theta^+)$ and $\widehat{V}_{g,n}(\theta^+)$, which holds by the ULLN given in Lemma 12.1 in Supplemental Appendix C, (ii) $\widehat{\theta}_n^+ \to_p \theta_0^+$ under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, which holds by Theorem 4.2(a) and $\widehat{\beta}_n/||\widehat{\beta}_n|| \to \omega_0$ (see Lemma 9.4(b) of Appendix B of AC1-SM), and (iii) $\mathcal{W}_n \to_p \mathcal{W}(\theta_0;\gamma_0)$, which holds by Lemma 3.1.

10. Supplemental Appendix B: Proofs of GMM Estimation Results

10.1. Lemmas

This Supplemental Appendix proves the results in Theorems 4.1 and 4.2 of AC3. The method of proof is to show that Assumptions B1, B2, and GMM1-GMM5 imply the high-level assumptions in AC1, viz., Assumptions A, B3, C1-C8, and D1-D3 of AC1. Given this, Theorems 3.1 and 3.2 of AC1 imply Theorems 4.1 and 4.2 because the results of these theorems are the same, just the assumptions differ.

Lemma 10.1. *Suppose Assumption* GMM1 *holds. Then,*
(a) *Assumption* A *of* AC1 *holds and*
(b) *Assumption* B3 *of* AC1 *holds with* $Q(\theta;\gamma_0) = g_0(\theta;\gamma_0)'\mathcal{W}(\theta;\gamma_0)g_0(\theta;\gamma_0)$.

Under Assumptions GMM1 and GMM2, Assumption GMM3 is used to show that the "C" assumptions of AC1 hold for the GMM estimator. As above, $\mathcal{W}(\psi_0;\gamma_0)$ abbreviates $\mathcal{W}(\psi_0,\pi;\gamma_0)$ when $\beta_0 = 0$.

Lemma 10.2. *Suppose Assumptions* GMM1-GMM3 *hold. Then, the following are true.*
(a) *Assumption* C1 *of* AC1 *holds with* $D_\psi Q_n(\theta) = g_\psi(\psi_0,\pi;\gamma_0)'\mathcal{W}(\psi_0;\gamma_0)\overline{g}_n(\theta)$ *and* $D_{\psi\psi}Q_n(\theta) = g_\psi(\psi_0,\pi;\gamma_0)'\mathcal{W}(\psi_0;\gamma_0)g_\psi(\psi_0,\pi;\gamma_0)$.
(b) *Assumption* C2 *of* AC1 *holds with* $m(W_i,\theta) = g_\psi(\psi_0,\pi;\gamma_0)'\mathcal{W}(\psi_0;\gamma_0)g(W_i,\theta)$.
(c) *Assumption* C3 *of* AC1 *holds with* $\Omega(\pi_1,\pi_2;\gamma_0) = g_\psi(\psi_0,\pi_1;\gamma_0)'\mathcal{W}(\psi_0;\gamma_0)\Omega_g(\gamma_0) \times \mathcal{W}(\psi_0;\gamma_0)g_\psi(\psi_0,\pi_2;\gamma_0)$.
(d) *Assumption* C4 *of* AC1 *holds with* $H(\pi;\gamma_0) = g_\psi(\psi_0,\pi;\gamma_0)'\mathcal{W}(\psi_0;\gamma_0)g_\psi(\psi_0,\pi;\gamma_0) = D_{\psi\psi}Q_n(\theta)$.
(e) *Assumption* C5 *of* AC1 *holds with* $K_n(\theta;\gamma^*) = g_\psi(\psi_0,\pi;\gamma_0)'\mathcal{W}(\psi_0;\gamma_0)K_{n,g}(\theta;\gamma^*) \in R^{d_\psi\times d_\beta}$, *and* $K(\psi_0,\pi;\gamma_0) = g_\psi(\psi_0,\pi;\gamma_0)'\mathcal{W}(\psi_0;\gamma_0)K_g(\psi_0,\pi;\gamma_0)$.
(f) *Assumption* C7 *of* AC1 *holds.*
(g) *Assumption* C8 *of* AC1 *holds.*

Comments. 1. To obtain Lemma 10.2(a), Assumption GMM3 is sufficient but not necessary. When $\overline{g}_n(\theta)$ is not a sample average, as occurs with the MD estimator, Assumption MD can be used in conjunction with Assumptions GMM1 and GMM2 to obtain Lemma 10.2(a). In this case, Assumptions C2-C5 of AC1 can be verified directly without using Assumption GMM3.

2. Lemma 10.2(c)-(e) provide the quantities that appear in Assumption C6 of AC1, which is the same as Assumption GMM4.

Lemma 10.3. *Suppose Assumptions* GMM1, GMM2, *and* GMM5 *hold.*
(a) *Assumption* D1 *of* AC1 *holds with* $DQ_n(\theta) = g_\theta(\theta_0;\gamma_0)'\mathcal{W}(\theta_0;\gamma_0)\overline{g}_n(\theta)$ *and* $D^2Q_n(\theta) = g_\theta(\theta_0;\gamma_0)'\mathcal{W}(\theta_0;\gamma_0)g_\theta(\theta_0;\gamma_0)$.
(b) *Assumption* D2 *of* AC1 *holds with* $J(\gamma_0) = J_g(\gamma_0)'\mathcal{W}(\theta_0;\gamma_0)J_g(\gamma_0)$.
(c) *Assumption* D3 *of* AC1 *holds with* $V(\gamma_0) = J_g(\gamma_0)'\mathcal{W}(\theta_0;\gamma_0)V_g(\gamma_0)\mathcal{W}(\theta_0;\gamma_0)J_g(\gamma_0)$.

10.2. Minimum Distance Estimators

For the MD estimator, Assumption MD can be used in place of Assumption GMM3 to obtain Assumption C1 of AC1.

Corollary 10.1. *Assumptions* GMM1, GMM2, *and* MD *imply that Assumption* C1 *of* AC1 *holds with* $D_\psi Q_n(\theta)$ *and* $D_{\psi\psi}Q_n(\theta)$ *defined as in Lemma* 10.2(a).

In addition to the result of Corollary 10.1, Lemmas 10.1 and 10.3 show that Assumptions A, B3, and D1-D3 of AC1 hold for the MD estimator under Assumptions GMM1, GMM2, and GMM5. Hence, in order to obtain the results of Theorems 3.1 and 3.2 of AC1 for MD estimators and other results concerning CS's, one just needs to verify Assumptions C2-C8 of AC1.

10.3. Proofs of Lemmas

Proof of Lemma 10.1. Assumption A of AC1 is implied by Assumption GMM1(i).

Assumption GMM1(ii) implies that Assumption B3(i) of AC1 holds with $Q(\theta;\gamma_0) = g_0(\theta;\gamma_0)'\mathcal{W}(\theta;\gamma_0)g_0(\theta;\gamma_0)$.

Now we verify Assumptions B3(ii) and B3(iii) of AC1 by using Lemma 8.1 in Appendix A of AC1-SM, which shows that Assumption B3* of AC1-SM is sufficient for Assumptions B3(ii) and B3(iii) of AC1. Assumption B3*(i) of AC1-SM holds by Assumptions GMM1(v) and GMM1(vi). Assumption B3*(ii) of AC1-SM holds by Assumptions GMM1(iii) and GMM1(vii). Assumption B3*(iii) of AC1-SM holds by Assumptions GMM1(iv) and GMM1(vii). Hence, Assumption B3 of AC1 holds. □

We prove Lemma 10.3 first and then prove Corollary 10.1 and Lemma 10.2.

Proof of Lemma 10.3. We start with the proof of part (a). For notational simplicity, in this proof $g_0(\theta;\gamma_0)$, $g_\theta(\theta;\gamma_0)$, $g_\psi(\theta;\gamma_0)$, and $\mathcal{W}(\theta;\gamma_0)$ are abbreviated to $g_0(\theta)$, $g_\theta(\theta)$, $g_\psi(\theta)$, and $\mathcal{W}(\theta)$, respectively.

We start with the case in which $\mathcal{W}_n(\theta) = I_k$. When $DQ_n(\theta_n)$ and $D^2Q_n(\theta_n)$ take the form in Lemma 10.3(a), the remainder term in Assumption D1 becomes

$$R_n^*(\theta) = \|\overline{g}_n(\theta)\|^2/2 - \|\overline{g}_n(\theta_n)\|^2/2 - \overline{g}_n(\theta_n)'g_\theta(\theta_0)(\theta-\theta_n) - \|g_\theta(\theta_0)(\theta-\theta_n)\|^2/2. \quad (10.1)$$

We approximate $R_n^*(\theta)$ by replacing $g_\theta(\theta_0)(\theta-\theta_n)$ by $g_0(\theta) - g_0(\theta_n)$ and get

$$R_n^\dagger(\theta) = \|\overline{g}_n(\theta)\|^2/2 - \|\overline{g}_n(\theta_n)\|^2/2 - \overline{g}_n(\theta_n)'(g_0(\theta) - g_0(\theta_n)) - \|g_0(\theta) - g_0(\theta_n)\|^2/2. \quad (10.2)$$

Let a, c, and d be $k-$vectors for which $a = c + d$. By the Cauchy-Schwarz inequality,

$$\left|\|a\|^2 - \|c\|^2\right| = \left|\|d\|^2 + 2c'd\right| \leq \|d\|^2 + 2\|c\|\|d\|. \quad (10.3)$$

Let $a = g_0(\theta) - g_0(\theta_n)$ and $c = g_\theta(\theta_0)(\theta - \theta_n)$, then

$$\begin{aligned} d &= a - c = g_0(\theta) - g_0(\theta_n) - g_\theta(\theta_0)(\theta-\theta_n) \\ &= [(g_\theta(\theta_n^\dagger) - g_\theta(\theta_0))B^{-1}(\beta_n)]B(\beta_n)(\theta-\theta_n) = o(\|B(\beta_n)(\theta-\theta_n)\|), \end{aligned} \quad (10.4)$$

where the first two equalities hold by definition, the third equality follows from element-by-element mean-value expansions, where $\theta_n^\dagger$ is between θ and θ_n (and $\theta_n^\dagger$ may depend on the row), and the last equality follows from Assumption GMM5(ii). By Assumptions GMM5(ii) and GMM5(iii),

$$c = g_\theta(\theta_0)(\theta-\theta_n) = \left[g_\theta(\theta_0)B^{-1}(\beta_n)\right]B(\beta_n)(\theta-\theta_n) = O(\|B(\beta_n)(\theta-\theta_n)\|). \quad (10.5)$$

Hence,

$$
\begin{aligned}
&\sup_{\theta\in\Theta_n(\delta_n)} \frac{n|R_n^\dagger(\theta)-R_n^*(\theta)|}{(1+n^{1/2}||B(\beta_n)(\theta-\theta_n)||)^2} \\
&= \frac{1}{2}\sup_{\theta\in\Theta_n(\delta_n)} \frac{n\left|2\overline{g}_n(\theta_n)'d+\|g_0(\theta)-g_0(\theta_n)\|^2-\|g_\theta(\theta_0)(\theta-\theta_n)\|^2\right|}{(1+n^{1/2}||B(\beta_n)(\theta-\theta_n)||)^2} \qquad (10.6)\\
&\le \frac{1}{2}\sup_{\theta\in\Theta_n(\delta_n)} n\left(2\|\overline{g}_n(\theta_n)\|\,\|d\|+\|d\|^2+2\|c\|\,\|d\|\right)/(1+n^{1/2}||B(\beta_n)(\theta-\theta_n)||)^2=o_p(1),
\end{aligned}
$$

where the first equality follows from (10.1) and (10.2), the inequality holds by (10.3), and the second equality uses (10.4), (10.5), and $\overline{g}_n(\theta_n)=O_p(n^{-1/2})$, where the latter holds by Assumption GMM5(i). Thus, it suffices to show that Assumption D1(ii) holds with $R_n^*(\theta)$ replaced by $R_n^\dagger(\theta)$.

Note that

$$
\begin{aligned}
R_n^\dagger(\theta) &= \|\overline{g}_n(\theta)\|^2/2-\|\overline{g}_n(\theta_n)+g_0(\theta)-g_0(\theta_n)\|^2/2\\
&= \|\widetilde{g}_n(\theta)-\widetilde{g}_n(\theta_n)\|^2/2+(g_0(\theta)-g_0(\theta_n)+\overline{g}_n(\theta_n))'(\widetilde{g}_n(\theta)-\widetilde{g}_n(\theta_n)), \qquad (10.7)
\end{aligned}
$$

where the first equality follows from (10.2) and the second equality uses $\|a\|^2-\|c\|^2=\|a-c\|^2+2c'(a-c)$ with $a=\overline{g}_n(\theta)$, $c=\overline{g}_n(\theta_n)+g_0(\theta)-g_0(\theta_n)$, and $a-c=\widetilde{g}_n(\theta)-\widetilde{g}_n(\theta_n)$.

We have

$$
\eta_n=\sup_{\theta\in\Theta_n(\delta_n)}\frac{n^{1/2}\|\widetilde{g}_n(\theta)-\widetilde{g}_n(\theta_n)\|}{1+n^{1/2}\|B(\beta_n)(\theta-\theta_n)\|}=o_p(1), \qquad (10.8)
$$

where the $o_p(1)$ term holds by Assumption GMM2(ii). By (10.7), (10.8), and the triangle inequality,

$$
\begin{aligned}
&\sup_{\theta\in\Theta_n(\delta_n)}\frac{2n|R_n^\dagger(\theta)|}{(1+n^{1/2}\|B(\beta_n)(\theta-\theta_n)\|)^2}\\
&\le \eta_n^2+2\sup_{\theta\in\Theta_n(\delta_n)}\frac{n^{1/2}\|g_0(\theta)-g_0(\theta_n)\|+n^{1/2}||\overline{g}_n(\theta_n)||}{1+n^{1/2}\|B(\beta_n)(\theta-\theta_n)\|}\eta_n\\
&=\eta_n^2+O_p(1)\eta_n=o_p(1), \qquad (10.9)
\end{aligned}
$$

where the first equality holds because $\overline{g}_n(\theta_n)=O_p(n^{-1/2})$ and $\|g_0(\theta)-g_0(\theta_n)\|=O(||B(\beta_n)(\theta-\theta_n)||)$ uniformly on $\Theta_n(\delta_n)$. To see that the latter holds, element-by-element mean-value expansions give

$$g_0(\theta) - g_0(\theta_n) = (g_\theta(\theta_n^\dagger)B^{-1}(\beta_n))B(\beta_n)(\theta-\theta_n) = (J_g(\gamma_0)+o(1))B(\beta_n)(\theta-\theta_n), \tag{10.10}$$

where $\theta_n^\dagger$ lies between θ and θ_n and the last equality follows from Assumptions GMM5(ii) and GMM5(iii). This completes the proof of Lemma 10.3(a) for the case in which $\mathcal{W}_n(\theta) = I_k$.

Next, Lemma 10.3(a) is established for the case where $\mathcal{W}_n(\theta)$ is as in Assumption GMM1. By Assumptions GMM1(ii) and GMM1(vii), we know that $\mathcal{W}_n(\theta)$ is symmetric and positive definite in a neighborhood of θ_0. Hence, both $\mathcal{W}(\theta)$ and $\mathcal{W}_n(\theta)$ have square roots, denoted by $\mathcal{W}^{1/2}(\theta)$ and $\mathcal{W}_n^{1/2}(\theta)$, respectively. The idea is to use the same proof as above, but with $\overline{g}_n(\theta)$, $g_0(\theta)$, and $g_\theta(\theta_0)$ replaced by $\mathcal{W}_n^{1/2}(\theta)\overline{g}_n(\theta)$, $\mathcal{W}^{1/2}(\theta_0)g_0(\theta)$, and $\mathcal{W}^{1/2}(\theta_0)g_\theta(\theta_0)$. With these changes, $R_n^*(\theta)$ in (10.1) becomes

$$\begin{aligned} R_n^{**}(\theta) = {} & ||\mathcal{W}_n^{1/2}(\theta)\overline{g}_n(\theta)||^2/2 - ||\mathcal{W}_n^{1/2}(\theta_n)\overline{g}_n(\theta_n)||^2/2 - \\ & \overline{g}_n(\theta_n)'\mathcal{W}_n^{1/2}(\theta_n)'\mathcal{W}^{1/2}(\theta_0)g_\theta(\theta_0)(\theta-\theta_n) - ||\mathcal{W}^{1/2}(\theta_0)g_\theta(\theta_0)(\theta-\theta_n)||^2/2. \end{aligned} \tag{10.11}$$

To show the condition in Assumption D1(ii) holds for $R_n^{**}(\theta)$, the method used for the case $\mathcal{W}_n(\theta) = I_k$ works provided that Assumptions GMM2(ii) and GMM5, which are used in the foregoing proof, hold with the same changes. Assumption GMM5 obviously does with $V_g(\gamma_0)$ and $J_g(\gamma_0)$ adjusted to $\mathcal{W}^{1/2}(\theta_0)V_g(\gamma_0)\mathcal{W}^{1/2}(\theta_0)$ and $\mathcal{W}^{1/2}(\theta_0)J_g(\gamma_0)$, respectively.

We now show Assumption GMM2(ii) also holds with the changes above. For $\theta \in \Theta_n(\delta_n)$,

$$\begin{aligned} & ||\mathcal{W}_n^{1/2}(\theta)\overline{g}_n(\theta) - \mathcal{W}^{1/2}(\theta_0)g_0(\theta) - \mathcal{W}_n^{1/2}(\theta_n)\overline{g}_n(\theta_n) + \mathcal{W}^{1/2}(\theta_0)g_0(\theta_n)|| \\ \leq {} & ||\mathcal{W}^{1/2}(\theta_0)||\,\|\widetilde{g}_n(\theta)-\widetilde{g}_n(\theta_n)\| + ||\mathcal{W}_n^{1/2}(\theta)-\mathcal{W}^{1/2}(\theta_0)||\,\|\overline{g}_n(\theta)-\overline{g}_n(\theta_n)\| + \\ & ||\mathcal{W}_n^{1/2}(\theta)-\mathcal{W}_n^{1/2}(\theta_n)||\,\|\overline{g}_n(\theta_n)\| \\ \leq {} & O(1)\|\widetilde{g}_n(\theta)-\widetilde{g}_n(\theta_n)\| + o_p(1)(\|\widetilde{g}_n(\theta)-\widetilde{g}_n(\theta_n)\| + \|g_0(\theta)-g_0(\theta_n)\|) + \\ & o_p(1)\|\overline{g}_n(\theta_n)\| \\ = {} & o_p(n^{-1/2}\sup_{\theta\in\Theta_n(\delta_n)}(1+n^{1/2}||B(\beta_n)(\theta-\theta_n)||)) + O(\|(B(\beta_n)(\theta-\theta_n))\|) = o_p(1), \end{aligned} \tag{10.12}$$

where the first inequality follows from adding and subtracting $\mathcal{W}^{1/2}(\theta_0)\overline{g}_n(\theta)$,

$\mathcal{W}^{1/2}(\theta_0)\,\overline{g}_n(\theta_n)$, and $\mathcal{W}_n^{1/2}(\theta)\,\overline{g}_n(\theta_n)$ and invoking the triangle inequality, the second inequality holds by Assumptions GMM1(ii), GMM1(vi), and GMM1(vii), the first equality holds by Assumption GMM2(ii), (10.10), and $\overline{g}_n(\theta_n) = O_p(n^{-1/2})$, and the second equality holds by the definition of $\Theta_n(\delta_n)$ and $B(\beta_n)$. By (10.12), the condition in Assumption D1(ii) holds with $R_n^*(\theta)$ changed to $R_n^{**}(\theta)$.

When the random derivative matrices take the form in Lemma 10.3(a), the remainder term in Assumption D1(i) is

$$R_n^*(\theta) = ||\mathcal{W}_n^{1/2}(\theta)\,\overline{g}_n(\theta)||^2/2 - ||\mathcal{W}_n^{1/2}(\theta_n)\,\overline{g}_n(\theta_n)||^2/2 - \overline{g}_n(\theta_n)'\mathcal{W}(\theta_0)\,g_\theta(\theta_0)'(\theta-\theta_n) - ||\mathcal{W}^{1/2}(\theta_0)\,g_\theta(\theta_0)(\theta-\theta_n)||^2/2. \quad (10.13)$$

We now show the difference between $R_n^*(\theta)$ and $R_n^{**}(\theta)$ in (10.11) is small enough so that the condition in Assumption D1(ii) holds for $R_n^*(\theta)$ provided it holds for $R_n^{**}(\theta)$. For $\theta \in \Theta_n(\delta_n)$,

$$\begin{aligned} &|R_n^*(\theta) - R_n^{**}(\theta)| = |\overline{g}_n(\theta_n)'(\mathcal{W}_n^{1/2}(\theta_n) - \mathcal{W}^{1/2}(\theta_0))'\mathcal{W}^{1/2}(\theta_0)\,g_\theta(\theta_0)(\theta-\theta_n)| \\ &\leq \|\overline{g}_n(\theta_n)\| \cdot ||\mathcal{W}_n^{1/2}(\theta_n) - \mathcal{W}^{1/2}(\theta_0)|| \cdot ||\mathcal{W}^{1/2}(\theta_0)|| \cdot \left\|g_\theta(\theta_0)B^{-1}(\beta_n)\right\| \cdot \\ &\quad \|B(\beta_n)(\theta-\theta_n)\| \\ &= o_p(n^{-1/2}\,\|B(\beta_n)(\theta-\theta_n)\|) = o_p(1), \end{aligned} \quad (10.14)$$

where the second last equality holds by Assumptions GMM1 and GMM5. This completes the proof of part (a).

Part (b) follows from part (a) and Assumptions GMM5(ii) and GMM5(iii).

Part (c) follows from part (a) and Assumptions GMM5(i)-(iii). □

We now prove Corollary 10.1 and then use Corollary 10.1 to prove Lemma 10.2.

Proof of Corollary 10.1. The proof is analogous to the proof of Lemma 10.3(a) with (i) $DQ_n(\theta_n)$ and $D^2Q_n(\theta_n)$ in Lemma 10.3(a) changed to $D_\psi Q_n(\psi_{0,n},\pi)$ and $D_{\psi\psi}Q_n(\psi_{0,n},\pi)$ in Lemma 10.2(a), (ii) $R_n^*(\theta)$ changed to $R_n(\psi,\pi)$, (iii) θ_n and $\theta-\theta_n$ changed to $(\psi_{0,n},\pi)$ and $\psi-\psi_{0,n}$, (iv) $g_\theta(\cdot)$ changed to $g_\psi(\cdot)$, where as above $g_\theta(\cdot)$ and $g_\psi(\cdot)$ abbreviate $g_\theta(\cdot;\gamma_0)$ and $g_\psi(\cdot;\gamma_0)$, respectively, (v) $B(\beta_n)$ and $B^{-1}(\beta_n)$ deleted throughout, (vi) $\theta_n^\dagger$ changed to $(\psi_{0,n}^\dagger(\pi),\pi)$ with $\psi_{0,n}^\dagger(\pi)$ between ψ and $\psi_{0,n}$, (vii) $\theta \in \Theta_n(\delta_n)$ changed to $\psi \in \Psi(\pi)$ and $\left\|\psi-\psi_{0,n}\right\| \leq \delta_n$, and (viii) $O_p(1)$ and $o_p(1)$ changed to $O_{p\pi}(1)$ and $o_{p\pi}(1)$, where the uniformity over Π usually holds using the compactness of Π, and (ix) $\mathcal{W}(\theta_0)$ changed to $\mathcal{W}(\psi_0;\gamma_0)$. Note that Assumptions GMM3(iii) and MD hold with π_n replaced by π $\forall\pi\in\Pi$ under Assumption GMM1(i). The assumptions that are referenced in the proof also are changed accordingly. Specifically, the proof goes through with Assumption GMM2(ii) changed to Assumption GMM2(i), Assumption GMM5(i) changed to Assumption MD, Assumption GMM5(ii) changed to the continuity of $g_\psi(\theta,\pi)$ uniformly over Π, which is implied by Assumption GMM1(vii) and the compactness of Π, and Assumption GMM5(iii) changed to the continuity of $g_\psi(\theta)$. (The assumption that $J_g(\gamma_0)$ has full column rank is not used in the proof of Lemma 10.3(a).)

Assumption C1(iii) follows from the form of $D_\psi Q_n(\theta)$ and $D_{\psi\psi}Q_n(\theta)$ in Lemma 10.2 and Assumption GMM1(i). $\square$

Proof of Lemma 10.2. First we prove part (a). Under Assumption GMM3, we can show Assumption MD holds using a proof that is similar to the proof of Lemma 9.1 in Appendix B of AC1-SM with (i) $D_\psi Q_n\left(\psi_{0,n},\pi\right)$ changed to $\overline{g}_n(\psi_{0,n},\pi)$, (ii) $m\left(W_i,\theta\right)$ changed to $g\left(W_i,\theta\right)$, (iii) Assumptions C2, C3, and C5 of AC1 changed to the corresponding conditions in Assumptions GMM3. By Corollary 10.1, Lemma 10.2(a) holds under Assumptions GMM1-GMM3.

Part (b) follows from part (a) and Assumptions GMM3(i) and GMM3(ii).

Part (c) follows from part (b) and Assumptions GMM1(i) and GMM3(iii).

Part (d) follows from part (a), $H(\pi;\gamma_0)=D_{\psi\psi}Q_n(\psi_{0,n},\pi)$, and Assumption GMM1 (viii).

Part (e) follows from part (a) and Assumption GMM3(iv).

Now we verify part (f). Note that when $\beta_0=0$ as in Assumption C7, $K_g(\psi_0,\pi;\gamma_0)$ does not depend on π by Assumptions GMM1(i) and GMM3(i). Given the form of $H(\pi;\gamma_0)$ and $K(\pi;\gamma_0)$ in parts (d) and (e), for any $\pi\in\Pi$,

$$\begin{aligned}&\omega_0'K(\pi;\gamma_0)'H^{-1}(\pi;\gamma_0)K(\pi;\gamma_0)\omega_0=Y'X(\pi)(X(\pi)'X(\pi))^{-1}X(\pi)'Y\leq Y'Y,\text{ where}\\&X(\pi)=\mathcal{W}^{1/2}(\psi_0;\gamma_0)g_\psi(\psi_0,\pi;\gamma_0),\ Y=\mathcal{W}^{1/2}(\psi_0;\gamma_0)K_g(\psi_0,\pi;\gamma_0)\omega_0,\end{aligned}\tag{10.15}$$

and Y does not depend on π. The inequality in (10.15) holds because $X(\pi)(X(\pi)'X(\pi))^{-1}$ $X(\pi)'$ is a projection matrix. The inequality holds as an equality when $\mathcal{W}^{1/2}(\psi_0;\gamma_0)$ $\times K_g(\psi_0,\pi;\gamma_0)\omega_0=\mathcal{W}^{1/2}(\psi_0;\gamma_0)g_\psi(\psi_0,\pi;\gamma_0)S$ for some $S\in R^{d_\psi}$. By Assumptions GMM1(vii) and GMM3(v), the inequality in (10.15) holds as an equality iff $\pi=\pi_0$. This completes the verification of Assumption C7.

To verify Assumption C8 as in part (g), we have

$$\begin{aligned}\frac{\partial}{\partial\psi'}E_{\gamma_n}D_\psi Q_n(\psi_n,\pi_n) &= g_\psi(\psi_0,\pi_n;\gamma_0)'\mathcal{W}(\psi_0;\gamma_0)\frac{\partial}{\partial\psi'}E_{\gamma_n}\overline{g}_n(\theta_n)\\ &= g_\psi(\psi_0,\pi_n;\gamma_0)'\mathcal{W}(\psi_0;\gamma_0)\left(n^{-1}\sum_{i=1}^n\frac{\partial}{\partial\psi'}E_{\gamma_n}g(W_i,\theta_n)\right)\\ &\to g_\psi(\theta_0;\gamma_0)'\mathcal{W}(\psi_0;\gamma_0)g_\psi(\theta_0;\gamma_0)=H(\pi_0;\gamma_0), \qquad (10.16)\end{aligned}$$

where the first equality holds by Lemma 10.2(a), the second equality holds by Assumption GMM3(i), the convergence holds by Assumption GMM3(vi) and the continuity of $g_\psi(\theta;\gamma_0)$ in π in Assumption GMM1(v), and the third equality holds by Lemma 10.2(d). □

10.4. Proofs of Section 3 Lemmas

Proof of Lemma 3.1. By the triangle inequality,

$$\left\|\mathcal{W}_n(\overline{\theta}_n)-\mathcal{W}(\theta_0;\gamma_0)\right\|\le\left\|\mathcal{W}_n(\overline{\theta}_n)-\mathcal{W}(\overline{\theta}_n;\gamma_0)\right\|+\left\|\mathcal{W}(\overline{\theta}_n;\gamma_0)-\mathcal{W}(\theta_0;\gamma_0)\right\|, \quad (10.17)$$

where the first term on the rhs is $o_p(1)$ because $\mathcal{W}_n(\theta)$ converges to $\mathcal{W}(\theta;\gamma_0)$ uniformly over Θ. When $\beta_0\neq 0$, the second term on the rhs of (10.17) is $o_p(1)$ because $\mathcal{W}(\theta;\gamma_0)$ is continuous in θ and $\overline{\theta}_n\to_p\theta_0$. When $\beta_0=0$, to show the second term on the rhs of (10.17) is $o_p(1)$, we have

$$\begin{aligned}&\left\|\mathcal{W}(\overline{\theta}_n;\gamma_0)-\mathcal{W}(\theta_0;\gamma_0)\right\|\\ &\le\left\|\mathcal{W}(\overline{\psi}_n,\overline{\pi}_n,;\gamma_0)-\mathcal{W}(\psi_0,\overline{\pi}_n;\gamma_0)\right\|+\left\|\mathcal{W}(\psi_0,\overline{\pi}_n;\gamma_0)-\mathcal{W}(\psi_0,\pi_0;\gamma_0)\right\|\\ &\le\sup_{\pi\in\Pi}\left\|\mathcal{W}(\overline{\psi}_n,\pi,;\gamma_0)-\mathcal{W}(\psi_0,\pi;\gamma_0)\right\|, \qquad (10.18)\end{aligned}$$

where the first inequality holds by the triangle inequality, and the second inequality holds because $\mathcal{W}(\psi_0,\pi;\gamma_0)$ does not depend on π when $\beta_0=0$, which in turn holds by Assumptions GMM1(i) and GMM1(ii). The third line of (10.18) is $o_p(1)$ because $\overline{\psi}_n\to_p\psi_0$ and $\mathcal{W}(\psi,\pi,;\gamma_0)$ is continuous in ψ uniformly over $\pi\in\Pi$, where the latter holds because $\mathcal{W}(\theta;\gamma_0)$ is continuous in θ and Π is compact. This completes the proof. □

Proof of Lemma 3.2. First we show that Assumption GMM2(ii) holds under Assumption GMM2*. For $\theta \in \Theta_n(\delta_n)$,

$$\begin{aligned}
\widetilde{g}_n(\theta;\gamma_0) - \widetilde{g}_n(\theta_n;\gamma_0) &= \frac{\partial}{\partial\theta'}\widetilde{g}_n(\theta_n^{\dagger};\gamma_0)(\theta-\theta_n) \\
&= \left([\frac{\partial}{\partial\theta'}g_n(\theta_n^{\dagger};\gamma_0) - g_\theta(\theta_n^{\dagger};\gamma_0)]B^{-1}(\beta_n)\right)B(\beta_n)(\theta-\theta_n) \\
&= o_p(||B(\beta_n)(\theta-\theta_n)||), \qquad (10.19)
\end{aligned}$$

where the first equality holds by element-by-element mean-value expansions with $\theta_n^{\dagger}$ between θ and θ_n (and $\theta_n^{\dagger}$ may depend on the row), the second equality holds by the definition of $\widetilde{g}_n(\theta,\gamma_0)$, and the last equality holds uniformly over $\theta \in \Theta_n(\delta_n)$ by Assumption GMM2*(iii). Assumption GMM2(ii) follows from (10.19) using the "$||B(\beta_n)(\theta-\theta_n)||$" part of the denominator in Assumption GMM2(ii).

The proof for Assumption GMM2(i) is analogous to the proof of Assumption GMM2(ii). For $\psi \in \Psi(\pi) : ||\psi - \psi_{0,n}|| \leq \delta_n$,

$$\begin{aligned}
\widetilde{g}_n(\psi,\pi;\gamma_0) - \widetilde{g}_n(\psi_{0,n},\pi;\gamma_0) &= \left(\frac{\partial}{\partial\psi'}g_n(\psi_{0,n}^{\dagger}(\pi),\pi;\gamma_0) - g_\psi(\psi_{0,n}^{\dagger}(\pi),\pi;\gamma_0)\right)(\psi-\psi_{0,n}) \\
&= o_{p\pi}(||\psi-\psi_{0,n}||), \qquad (10.20)
\end{aligned}$$

where the first equality holds by element-by-element mean-value expansions with $\psi_{0,n}^{\dagger}(\pi)$ between ψ and $\psi_{0,n}$ (and $\psi_{0,n}^{\dagger}(\pi)$ may depend on the row), and the second equality holds uniformly over $\psi \in \Psi(\pi) : ||\psi - \psi_{0,n}|| \leq \delta_n$ by Assumption GMM2*(ii). Assumption GMM2(i) follows from (10.20) using the "$||\psi - \psi_{0,n}||$" part of the denominator in Assumption GMM2(i). □

Proof of Lemma 3.3. Assumption GMM4 is the same as Assumption C6 of AC1. Hence, it suffices to verify the latter. We verify Assumption C6 of AC1 by verifying the sufficient condition Assumption C6** given in Lemma 8.5 in Appendix A of AC1-SM. Because β is a scalar, it remains to show Assumption C6**(ii) of AC1 holds. By Lemma 10.2(c), the covariance matrix $\Omega_G(\pi_1,\pi_2;\gamma_0)$ in Assumption C6**(ii) is

$$\begin{aligned}
\Omega_G(\pi_1,\pi_2;\gamma_0) &= g_\psi^*(\psi_0,\pi_1,\pi_2;\gamma_0)'\widetilde{\Omega}_g(\gamma_0)g_\psi^*(\psi_0,\pi_1,\pi_2;\gamma_0)', \text{ where} \\
\widetilde{\Omega}_g(\gamma_0) &= \mathcal{W}(\psi_0;\gamma_0)\Omega_g(\gamma_0)\mathcal{W}(\psi_0;\gamma_0) \qquad (10.21)
\end{aligned}$$

and $\widetilde{\Omega}_g(\gamma_0)$ does not depend on π_1 and π_2 by Assumptions GMM1(i) and GMM3(i). Because $g_\psi^*(\psi_0,\pi_1,\pi_2;\gamma_0) \in R^{k\times(d_\zeta+2)}$ and $k \geq d_\theta \geq d_\zeta + 2$, Assumption C6**(ii) is implied by Assumptions GMM1*(vii), GMM4*(ii), and GMM4*(iii). □

11. Supplemental Appendix C: Proofs for Wald Tests

11.1. Proofs of Asymptotic Distributions

Most of the results in Section 5 of AC3 are stated to hold under some combination of Assumptions GMM1-GMM5 or under certain assumptions from AC1 (plus some other assumptions). We prove the results of this section using the stated assumptions from AC1. Lemmas 10.1-10.3 in Supplemental Appendix B show that the appropriate combination of Assumptions GMM1-GMM5 imply the corresponding assumptions from AC1.

Proof of Lemma 5.1. (i) When $d_\pi^* = d_r$, $\eta_n(\widehat{\theta}_n) = 0$ by definition in (5.10).

(ii) When $d_r = 1$, $d_\pi^* = 0$ or $d_\pi^* = 1$ by Assumption R1(iii). If $d_\pi^* = 1$, $\eta_n(\widehat{\theta}_n) = 0$ by definition in (5.10). If $d_\pi^* = 0$, $r_\pi(\theta) = 0$ for $\theta \in \Theta_\delta$ by Assumption R1(iii). By the mean-value expansion, we have

$$r(\psi_n, \widehat{\pi}_n) - r(\psi_n, \pi_n) = r_\pi(\psi_n, \widetilde{\pi}_n)(\widehat{\pi}_n - \pi_n), \tag{11.1}$$

where $\widetilde{\pi}_n$ is between $\widehat{\pi}_n$ and π_n. For n large enough that $||\beta_n|| < \delta$, $(\psi_n, \widetilde{\pi}_n) \in \Theta_\delta$ and $r_\pi(\psi_n, \widetilde{\pi}_n) = 0$, which implies $\eta_n(\widehat{\theta}_n) = o_p(1)$.

(iii) From (11.1), we have

$$\eta(\widehat{\theta}_n) = n^{1/2} A_1(\widehat{\theta}_n) r_\pi(\psi_n, \widetilde{\pi}_n)(\widehat{\pi}_n - \pi_n). \tag{11.2}$$

Under Assumption R2*(iii), $A_1(\widehat{\theta}_n) r_\pi(\psi_n, \widetilde{\pi}_n) \rightarrow_p 0$ because the column space of $r_\pi(\theta)$ is the same for all $\theta \in \Theta_\delta$, by definition the rows of $A_1(\theta)$ are in the null space of $r_\pi(\theta)'$ $\forall \theta \in \Theta_\delta$, and $\widehat{\theta}_n \in \Theta_\delta$ holds with probability that goes to one by Lemma 3.1(a) of AC1 using Assumptions A and B3(i)-(ii) of AC1. This gives the desired result. □

Proof of Lemma 5.2. Under Assumption $\mathrm{R_L}$, $r_\theta(\theta) = R$ $\forall \theta \in \Theta$ and R has full row rank. Assumption R1 is satisfied directly. Moreover, under Assumption $\mathrm{R_L}$, $r_\pi(\theta)$ does not depend on θ. This implies Assumption R2*(iii), which is a sufficient condition of Assumption R2 by Lemma 5.1. □

The proof of Theorem 5.1 below uses the following Lemma. Define $\widehat{\omega}_n = \widehat{\beta}_n / ||\widehat{\beta}_n||$.

Lemma 11.1. *Suppose Assumption* V1 (*vector* β) *holds. In addition, suppose Assumptions* GMM1-GMM4 *hold* (*or Assumptions* A, B1-B3, *and* C1-C8 *of* AC1 *hold*).
(a) *Under* $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ *with* $||b|| < \infty$, $\widehat{\omega}_n \to_d \omega^*(\pi^*(\gamma_0, b); \gamma_0, b)$.
(b) *Under* $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, $\widehat{\omega}_n \to_p \omega_0$.

Proof of Lemma 11.1. To prove Lemma 11.1(a), we have

$$\widehat{\omega}_n = n^{1/2}\widehat{\beta}_n/||n^{1/2}\widehat{\beta}_n|| \to_d \frac{\tau_\beta(\pi^*(\gamma_0, b); \gamma_0, b)}{||\tau_\beta(\pi^*(\gamma_0, b); \gamma_0, b)||} = \omega^*(\pi^*(\gamma_0, b); \gamma_0, b) \quad (11.3)$$

by the continuous mapping theorem, because $n^{1/2}\widehat{\beta}_n \to_d \tau_\beta(\pi^*(\gamma_0, b); \gamma_0, b)$ by Theorem 4.1(a) and Comment 2 to Theorem 4.1(a) and $P(\tau_\beta(\pi^*; \gamma_0, b) = 0) = 0$ by Assumption V1(iv) (vector β).

Next, we prove that Lemma 11.1(b) holds when $\beta_0 = 0$. By Lemma 3.4 in AC1, $||\beta_n||^{-1}(\widehat{\beta}_n - \beta_n) = o_p(1)$. This implies that $\widehat{\beta}_n = \beta_n + ||\beta_n||o_p(1)$ and $||\widehat{\beta}_n||/||\beta_n|| = 1 + o_p(1)$. Hence,

$$\widehat{\omega}_n = \frac{\widehat{\beta}_n}{||\widehat{\beta}_n||} = \frac{\widehat{\beta}_n - \beta_n}{||\beta_n||}\frac{||\beta_n||}{||\widehat{\beta}_n||} + \frac{\beta_n}{||\beta_n||}\frac{||\beta_n||}{||\widehat{\beta}_n||} \to_p \omega_0. \quad (11.4)$$

Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ with $\beta_0 \neq 0$, $\widehat{\omega}_n \to \omega_0$ by the continuous mapping theorem given that $\widehat{\beta}_n \to_p \beta_0$ by Lemma 3.3(b) in AC1. □

Proof of Theorem 5.1. Under the null hypothesis $H_0 : r(\theta_n) = v_n$, the Wald statistic defined in (5.2) with $v = v_n$ becomes

$$W_n = n(r(\widehat{\theta}_n) - r(\theta_n))'(r_\theta(\widehat{\theta}_n)B^{-1}(\widehat{\beta}_n)\widehat{\Sigma}_n B^{-1}(\widehat{\beta}_n)r_\theta(\widehat{\theta}_n)')^{-1}(r(\widehat{\theta}_n) - r(\theta_n)). \quad (11.5)$$

Before proving the specific results in parts (a) and (b), we analyze the Wald statistic under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$. With the rotation represented by $A(\widehat{\theta}_n)$, the Wald statistic in (11.5) can be written as

$$W_n = n(r(\widehat{\theta}_n) - r(\theta_n))'A(\widehat{\theta}_n)'(r_\theta^A(\widehat{\theta}_n)B^{-1}(\widehat{\beta}_n)\widehat{\Sigma}_n B^{-1}(\widehat{\beta}_n)r_\theta^A(\widehat{\theta}_n)')^{-1}A(\widehat{\theta}_n)(r(\widehat{\theta}_n) - r(\theta_n)). \quad (11.6)$$

To deal with the normalizing matrix $B^{-1}(\widehat{\beta}_n)$, part of which diverges as $n \to \infty$ and

$\beta_n \to 0$, we define a $d_r \times d_r$ matrix

$$B^*(\widehat{\beta}_n) = \begin{bmatrix} I_{(d_r - d_\pi^*)} & 0 \\ 0 & \iota(\widehat{\beta}_n) I_{d_\pi^*} \end{bmatrix} \tag{11.7}$$

where $\iota(\beta) = \beta$ when β is a scalar and $\iota(\beta) = ||\beta||$ when β is a scalar . We write the Wald statistic in (11.6) as

$$\begin{aligned} W_n &= \varrho(\widehat{\theta}_n)'(\overline{r}_\theta(\widehat{\theta}_n)\widehat{\Sigma}_n \overline{r}_\theta(\widehat{\theta}_n)')^{-1}\varrho(\widehat{\theta}_n), \text{ where} \\ \varrho(\widehat{\theta}_n) &= n^{1/2} B^*(\widehat{\beta}_n) A(\widehat{\theta}_n)(r(\widehat{\theta}_n) - r(\theta_n)) \text{ and } \overline{r}_\theta(\widehat{\theta}_n) = B^*(\widehat{\beta}_n) r_\theta^A(\widehat{\theta}_n) B^{-1}(\widehat{\beta}_n). \end{aligned} \tag{11.8}$$

Note that

$$\overline{r}_\theta(\widehat{\theta}_n) = \begin{bmatrix} r_\psi^*(\widehat{\theta}_n) & 0 \\ \iota(\widehat{\beta}_n) r_\psi^0(\widehat{\theta}_n) & r_\pi^*(\widehat{\theta}_n) \end{bmatrix} = r_\theta^*(\widehat{\theta}_n) + \begin{bmatrix} 0 & 0 \\ \iota(\widehat{\beta}_n) r_\psi^0(\widehat{\theta}_n) & 0 \end{bmatrix} = r_\theta^*(\widehat{\theta}_n) + o_p(1), \tag{11.9}$$

where the $o_p(1)$ term holds because $\iota(\widehat{\beta}_n) = o_p(1)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ and $r_\psi^0(\widehat{\theta}_n) = O_p(1)$ under Assumption R1(i).

The next step is to derive the asymptotic distribution of $\varrho(\widehat{\theta}_n)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$. Note that

$$\begin{aligned} r(\widehat{\theta}_n) - r(\theta_n) &= r(\widehat{\psi}_n, \widehat{\pi}_n) - r(\psi_n, \widehat{\pi}_n) + r(\psi_n, \widehat{\pi}_n) - r(\psi_n, \pi_n) \\ &= r_\psi(\widehat{\theta}_n)(\widehat{\psi}_n - \psi_n) + (r(\psi_n, \widehat{\pi}_n) - r(\psi_n, \pi_n)) + o_p(n^{-1/2}), \end{aligned} \tag{11.10}$$

where the first equality is trivial and the second equality holds by a mean-value expansion, $\widehat{\psi}_n - \psi_n = O_p(n^{-1/2})$, and Assumption R1(i). From (11.7) and $A(\theta) = [A_1'(\theta) : A_2'(\theta)]'$, we have

$$\begin{aligned} \varrho(\widehat{\theta}_n) &= \begin{pmatrix} n^{1/2} A_1(\widehat{\theta}_n)(r(\widehat{\theta}_n) - r(\theta_n)) \\ n^{1/2}\iota(\widehat{\beta}_n) A_2(\widehat{\theta}_n)(r(\widehat{\theta}_n) - r(\theta_n)) \end{pmatrix} = \varrho_1(\widehat{\theta}_n) + \varrho_2(\widehat{\theta}_n) + o_p(1), \text{ where} \\ \varrho_1(\widehat{\theta}_n) &= \begin{pmatrix} n^{1/2} A_1(\widehat{\theta}_n) r_\psi(\widehat{\theta}_n)(\widehat{\psi}_n - \psi_n) \\ n^{1/2}\iota(\widehat{\beta}_n) A_2(\widehat{\theta}_n)(r(\psi_n, \widehat{\pi}_n) - r(\psi_n, \pi_n)) \end{pmatrix}, \\ \varrho_2(\widehat{\theta}_n) &= \begin{pmatrix} \eta_n(\widehat{\theta}_n) \\ n^{1/2}\iota(\widehat{\beta}_n) A_2(\widehat{\theta}_n) r_\psi(\widehat{\theta}_n)(\widehat{\psi}_n - \psi_n) \end{pmatrix} = \begin{pmatrix} \eta_n(\widehat{\theta}_n) \\ o_p(1) \end{pmatrix}, \end{aligned} \tag{11.11}$$

the second equality in $\varrho(\widehat{\theta}_n)$ uses (11.10), and the $o_p(1)$ term associated with $\varrho_2(\widehat{\theta}_n)$ holds by $n^{1/2}(\widehat{\psi}_n - \psi_n) = O_p(1)$ and $\iota(\widehat{\beta}_n) = o_p(1)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$. Under Assumption R2, $\eta_n(\widehat{\theta}_n) = o_p(1)$, and, hence, $\varrho_2(\widehat{\theta}_n) = o_p(1)$.

In part (a), in which case $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ and $||b|| < \infty$, we have

$$\begin{aligned}
\varrho_1(\widehat{\theta}_n) &= \overline{B}_n(\widehat{\pi}_n)\tau_n^A(\widehat{\pi}_n), \text{ where} \\
\tau_n^A(\pi) &= \begin{pmatrix} r_\psi^*(\widehat{\psi}_n(\pi), \pi)n^{1/2}(\widehat{\psi}_n(\pi) - \psi_n) \\ A_2(\widehat{\psi}_n(\pi), \pi)(r(\psi_n, \pi) - r(\psi_n, \pi_n)) \end{pmatrix} \text{ and} \\
\overline{B}_n(\pi) &= \begin{bmatrix} I_{(d_r - d_\pi^*)} & 0 \\ 0 & \iota(n^{1/2}\widehat{\beta}_n(\pi))I_{d_\pi^*} \end{bmatrix}. \qquad (11.12)
\end{aligned}$$

Using Assumption R1(i), Lemma 3.1(a) of AC1, Lemma 9.2(b) in Appendix B of AC1-SM, and $\tau_n(\pi) = n^{1/2}(\widehat{\psi}_n(\pi) - \psi_n) \Rightarrow \tau(\pi; \gamma_0, b)$ in (9.21) of AC1-SM, we have

$$\begin{pmatrix} \tau_n^A(\cdot) \\ \overline{B}_n(\cdot) \end{pmatrix} \Rightarrow \begin{pmatrix} \tau^A(\cdot; \gamma_0, b) \\ \overline{B}(\cdot; \gamma_0, b) \end{pmatrix} \qquad (11.13)$$

under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $||b|| < \infty$. From (11.8), (11.9), (11.11), and (11.12), in the case of a scalar β, we have

$$\begin{aligned}
W_n &= \tau_n^A(\widehat{\pi}_n)'\overline{B}_n(\widehat{\pi}_n)(r_\theta^*(\widehat{\theta}_n)\widehat{\Sigma}_n r_\theta^*(\widehat{\theta}_n)')^{-1}\overline{B}_n(\widehat{\pi}_n)\tau_n^A(\widehat{\pi}_n) + o_p(1) \\
&= \lambda_n(\widehat{\pi}_n) + o_p(1) \to_d \lambda(\pi^*(\gamma_0, b); \gamma_0, b), \qquad (11.14)
\end{aligned}$$

where $\lambda_n(\pi)$ is defined implicitly, $\widehat{\Sigma}_n = \widehat{\Sigma}_n(\widehat{\theta}_n) = \widehat{J}_n(\widehat{\theta}_n)^{-1}\widehat{V}_n(\widehat{\theta}_n)\widehat{J}_n(\widehat{\theta}_n)^{-1}$ by Assumption V1 (scalar β), and the convergence follows from the joint convergence $(\lambda_n(\cdot), \widehat{\pi}_n) \Rightarrow (\lambda(\cdot; \gamma_0, b), \pi^*(\gamma_0, b))$ and the continuous mapping theorem. The latter joint convergence holds by (11.13), Assumptions V1 (scalar β) and R1, Theorem 4.1(a), the uniform consistency of $\widehat{\psi}_n(\pi)$ over $\pi \in \Pi$, and the fact that $\tau_n^A(\cdot)$, $\overline{B}_n(\cdot)$, and $\widehat{\pi}_n$ are continuous functions of the empirical process $G_n(\cdot)$ with probability one.

In the case of a vector β, (11.14) holds with $\widehat{\Sigma}_n(\widehat{\theta}_n)$ replaced by $\widehat{\Sigma}_n(\widehat{\theta}_n^+) = \widehat{J}_n^{-1}(\widehat{\theta}_n^+)$ $\widehat{V}_n(\widehat{\theta}_n^+)\widehat{J}_n^{-1}(\widehat{\theta}_n^+)$ using Assumption V1 (vector β) and with $\lambda_n(\widehat{\pi}_n)$ replaced by $\lambda_n(\widehat{\pi}_n, \widehat{\omega}_n)$, which is defined implicitly. In this case, the convergence in (11.14) follows from the joint convergence $(\lambda_n(\cdot), \widehat{\pi}_n, \widehat{\omega}_n) \Rightarrow (\lambda(\cdot; \gamma_0, b),\ \pi^*(\gamma_0, b),\ \omega^*(\pi^*(\gamma_0, b); \gamma_0, b)$, which holds by the same argument as above plus Lemma 11.1(a) and Assumption V1 (vector β). This completes the proof of part (a).

The proof of part (b) is the same for the scalar and vector β cases because it relies on Assumption V2 which applies in both cases. To prove part (b), we first analyze the case where $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ and $\beta_0 = 0$. In this case, $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ with $b \notin R^{d_\beta}$, so (11.6)-(11.11) apply. As in (5.1), $\Sigma(\gamma_0) = J^{-1}(\gamma_0)V(\gamma_0)J^{-1}(\gamma_0)$. We have

$$\begin{aligned}
\varrho_1(\widehat{\theta}_n) &= \begin{pmatrix} n^{1/2} r_\psi^*(\widehat{\theta}_n)(\widehat{\psi}_n - \psi_n) \\ n^{1/2}\iota(\widehat{\beta}_n)A_2(\widehat{\theta}_n)(r_\pi(\widehat{\theta}_n) + o_p(1))(\widehat{\pi}_n - \pi_n) \end{pmatrix} \\
&= \begin{pmatrix} n^{1/2} r_\psi^*(\widehat{\theta}_n)(\widehat{\psi}_n - \psi_n) \\ n^{1/2}\iota(\beta_n)A_2(\widehat{\theta}_n)r_\pi(\widehat{\theta}_n)(\widehat{\pi}_n - \pi_n) + o_p(1) \end{pmatrix} \\
&= r_\theta^*(\widehat{\theta}_n)n^{1/2}B(\beta_n)(\widehat{\theta}_n - \theta_n) + o_p(1) \\
&\rightarrow_d N(0, r_\theta^*(\theta_0)\Sigma(\gamma_0)r_\theta^*(\theta_0)), \qquad (11.15)
\end{aligned}$$

where the first equality holds by a mean-value expansion, the fact that $\widehat{\pi}_n$ is consistent under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, and the continuity of $r_\pi(\theta)$ which holds by Assumption R1, the second equality holds by $n^{1/2}(\widehat{\beta}_n - \beta_n) = O_p(1)$ and $||\beta_n||n^{1/2}(\widehat{\pi}_n - \pi_n) = O_p(1)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$, the third equality holds by the definitions of $B(\beta)$ and $r_\theta^*(\theta)$, and the convergence in distribution holds by Theorem 4.2(a). The result of part (b) follows from (11.8), (11.9), (11.11), (11.15), and Assumptions D2 and D3(ii) of AC1 and Assumption V2.

Under $\{\gamma_n\} \in \Gamma(\gamma_0, \infty, \omega_0)$ and $\beta_0 \neq 0$,

$$n^{1/2}(r(\widehat{\theta}_n) - r(\theta_n)) \rightarrow_d N(0, r_\theta(\theta_0)B^{-1}(\beta_0)\Sigma(\gamma_0)B^{-1}(\beta_0)r_\theta(\theta_0)') \qquad (11.16)$$

by Theorem 4.2(a) and the delta method. By Assumptions R1(i) and V2,

$$r_\theta(\widehat{\theta}_n)B^{-1}(\widehat{\beta}_n)\widehat{\Sigma}_n B^{-1}(\widehat{\beta}_n)r_\theta(\widehat{\theta}_n)' \rightarrow_p r_\theta(\theta_0)B^{-1}(\beta_0)\Sigma(\gamma_0)B^{-1}(\beta_0)r_\theta(\theta_0)'. \qquad (11.17)$$

The desired result follows from (11.5), (11.16), and (11.17). $\square$

Proof of Corollary 5.1. By Lemma 5.2, Assumption R2 is satisfied. Based on Theorem 5.1, it suffices to show that the stochastic process $\{\lambda(\pi; \gamma_0, b) : \pi \in \Pi\}$ can be written as $\{\lambda_L(\pi; \gamma_0, b) : \pi \in \Pi\}$ under Assumption R_L. Under Assumption R_L, $r_\theta(\theta)$,

$A(\theta)$, and $r_\theta^*(\theta)$ do not depend on θ, and, hence,

$$\tau^A(\pi;\gamma_0,b) = \begin{pmatrix} r_\psi^* \tau(\pi;\gamma_0,b) \\ A_2 r_\pi \cdot (\pi - \pi_0) \end{pmatrix} = \begin{pmatrix} r_\psi^* \tau(\pi;\gamma_0,b) \\ r_\pi^* \cdot (\pi - \pi_0) \end{pmatrix} = R^* \overline{\tau}(\pi;\gamma_0,b). \quad (11.18)$$

The desired result follows from (11.18) and $r_\theta^*(\pi) = R^*\ \forall \pi \in \Pi$. □

Proof of Theorem 5.2. From the proof of Theorem 5.1, we know that $\varrho_1(\widehat{\theta}_n) = O_p(1)$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$. Therefore, when $||\eta_n(\widehat{\theta}_n)|| \to_p \infty$, it follows from (11.11) that $||\varrho(\widehat{\theta}_n)|| \to_p \infty$. This result, together with (11.8), (11.9), and Assumptions R1 and V1, completes the proof. □

11.2. Proofs of Asymptotic Size Results

Proof of Theorem 5.3. The proof is the same as the proof of Theorem 4.4 of AC1, which is given in Appendix B of AC1-SM, but with $|T_n|$, $|T(h)|$, and $z_{1-\alpha/2}$ replaced by W_n, $W(h)$, and $\chi^2_{d_r,1-\alpha}$, respectively; with Theorem 4.1 of AC1 replaced by Theorem 5.1; and with Assumption V3 of AC1 replaced by Assumption V4. □

Proof of Corollary 5.2. By Theorem 5.2, $P_{\gamma_n}(W_n \leq \chi^2_{d_r,1-\alpha}) \to_p 0$ under $\{\gamma_n\} \in \Gamma(\gamma_0, 0, b)$ for which $||\eta_n(\widehat{\theta}_n)|| \to_p \infty$. As a result, the nominal $1-\alpha$ Wald CS has $AsySz = 0$ by the definition of asymptotic size. □

Proof of Theorem 5.4. The proof of Theorem 5.4 is the same as the proof of Theorem 5.1 of AC1, which is given in Appendix B of AC1-SM, but with $|T_n|$, $|T(h)|$, and $z_{1-\alpha/2}$ replaced by W_n, $W(h)$, and $\chi^2_{d_r,1-\alpha}$, respectively; with $c^{LF}_{|t|,1-\alpha}$, $c_{|t|,1-\alpha}(h), \ldots$ replaced by $c^{LF}_{W,1-\alpha}$, $c_{W,1-\alpha}(h), \ldots$ throughout; with Theorem 4.1 of AC1 replaced by Theorem 5.1; and with Assumption V3 of AC1 replaced by Assumption V4. □

12. Supplemental Appendix D: Uniform LLN and CLT

In this Supplemental Appendix, we state a uniform convergence result, a uniform LLN, and a CLT that are used in the verification of Assumptions GMM1-GMM5 in the two examples considered in the paper. Specifically, Lemma 12.1 is a uniform convergence result for non-stochastic functions, Lemma 12.2 is a uniform LLN, and Lemma 12.3 is a

CLT. The latter two results are for strong mixing triangular arrays. These are standard sorts of results. The proofs of these Lemmas are given in Appendix A of Andrews and Cheng (2011).

Lemma 12.1. *Let* $\{q_n(\theta) : n \geq 1\}$ *be non-stochastic functions on* Θ. *Suppose* (i) $q_n(\theta) \to 0$ $\forall \theta \in \Theta$, (ii) $||q_n(\theta_1) - q_n(\theta_2)|| \leq C\delta$ $\forall \theta_1, \theta_2 \in \Theta$ *with* $||\theta_1 - \theta_2|| \leq \delta$, $\forall n \geq 1$, *for some* $C < \infty$ *and all* $\delta > 0$, *and* (iii) Θ *is compact. Then,* $\sup_{\theta \in \Theta} ||q_n(\theta)|| \to 0$.

Assumption S1. Under any $\gamma_0 \in \Gamma$, $\{W_i : i \geq 1\}$ is a strictly stationary and strong mixing sequence with mixing coefficients $\alpha_m \leq Cm^{-A}$ for some $A > d_\theta q/(q - d_\theta)$ and some $q > d_\theta \geq 2$, or $\{W_i : i \geq 1\}$ is an i.i.d. sequence and the constant q equals $2 + \delta$ for some $\delta > 0$.

Lemma 12.2. *Suppose* (i) *Assumption* S1 *holds,* (ii) *for some function* $M_1(w) : \mathcal{W} \to R^+$ *and all* $\delta > 0$, $||s(w, \theta_1) - s(w, \theta_2)|| \leq M_1(w)\delta$, $\forall \theta_1, \theta_2 \in \Theta$ *with* $||\theta_1 - \theta_2|| \leq \delta$, $\forall w \in \mathcal{W}$, (iii) $E_\gamma \sup_{\theta \in \Theta} ||s(W_i, \theta)||^{1+\varepsilon} + E_\gamma M_1(W_i) \leq C$ $\forall \gamma \in \Gamma$ *for some* $C < \infty$ *and* $\varepsilon > 0$, *and* (iv) Θ *is compact. Then,* $\sup_{\theta \in \Theta} ||n^{-1} \sum_{i=1}^n s(W_i, \theta) - E_{\gamma_0} s(W_i, \theta)|| \to_p 0$ *under* $\{\gamma_n\} \in \Gamma(\gamma_0)$ *and* $E_{\gamma_0} s(W_i, \theta)$ *is uniformly continuous on* Θ $\forall \gamma_0 \in \Gamma$.

Comment. Note that the centering term in Lemma 12.2 is $E_{\gamma_0} s(W_i, \theta)$, rather than $E_{\gamma_n} s(W_i, \theta)$.

Lemma 12.3. *Suppose* (i) *Assumption* S1 *holds,* (ii) $s(w) \in R$ *and* $E_\gamma |s(W_i)|^q \leq C$ $\forall \gamma \in \Gamma$ *for some* $C < \infty$ *and* q *as in Assumption* S1. *Then,* $n^{-1/2} \sum_{i=1}^n (s(W_i) - E_{\gamma_n} s(W_i)) \to_d N(0, V_s(\gamma_0))$ *under* $\{\gamma_n\} \in \Gamma(\gamma_0)$ $\forall \gamma_0 \in \Gamma$, *where* $V_s(\gamma_0) = \sum_{m=-\infty}^{\infty} Cov_{\gamma_0}(s(W_i), s(W_{i+m}))$.

13. Supplemental Appendix E: Numerical Results

Here we report some additional numerical results for the nonlinear regression model with endogeneity.

Figures S-1 and S-2 report asymptotic and finite-sample ($n = 500$) densities of the estimators for β and π when $\pi_0 = 3.0$. Figures S-3 to S-6 report asymptotic and finite-sample ($n = 500$) densities of the t and QLR statistics for β and π when $\pi_0 = 1.5$. Figures S-7 and S-8 report CP's of nominal 0.95 standard and robust $|t|$ and QLR CI's for β and π when $\pi_0 = 3.0$.

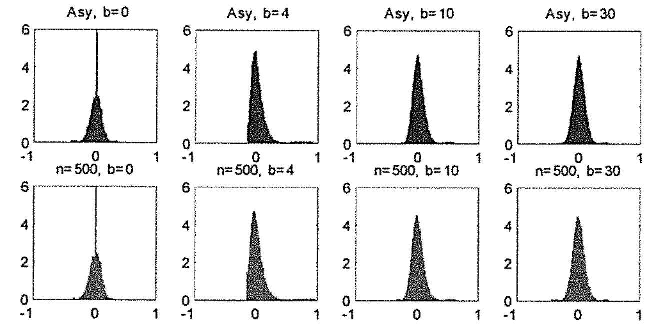

Figure S-1. Asymptotic and Finite-Sample ($n = 500$) Densities of the Estimator of β in the Nonlinear Regression Model with Endogeneity when $\pi_0 = 3.0$.

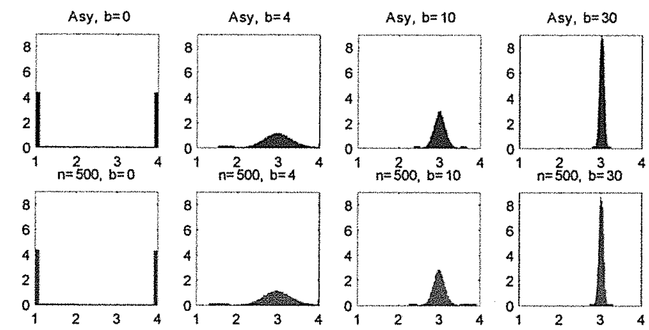

Figure S-2. Asymptotic and Finite-Sample ($n = 500$) Densities of the Estimator of π in the Nonlinear Regression Model with Endogeneity when $\pi_0 = 3.0$.

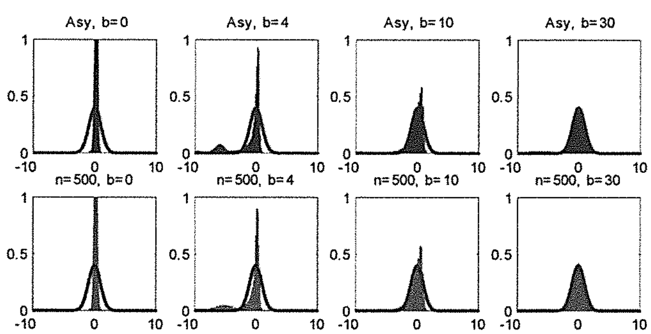

Figure S-3. Asymptotic and Finite-Sample ($n = 500$) Densities of the t Statistic for β in the Nonlinear Regression Model with Endogeneity when $\pi_0 = 1.5$ and the Standard Normal Density (Black Line).

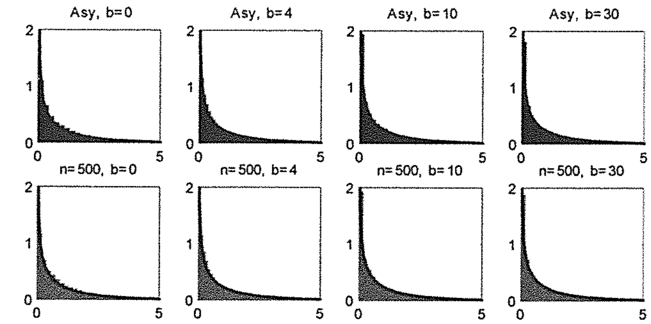

Figure S-4. Asymptotic and Finite-Sample (n=500) Densities of the QLR Statistic for β in the Nonlinear Regression Model with Endogeneity when $\pi_0 = 1.5$ and the χ_1^2 Density (Black Line).

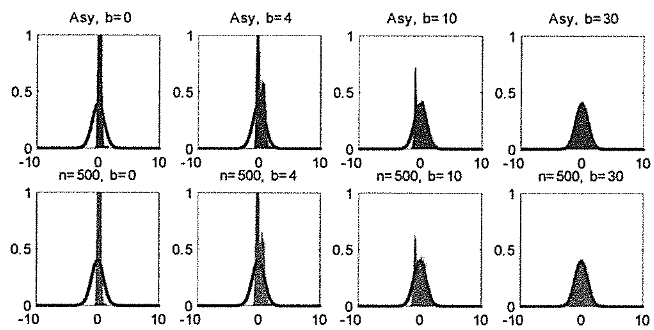

Figure S-5. Asymptotic and Finite-Sample ($n = 500$) Densities of the t Statistic for π in the Nonlinear Regression Model with Endogeneity when $\pi_0 = 1.5$ and the Standard Normal Density (Black Line).

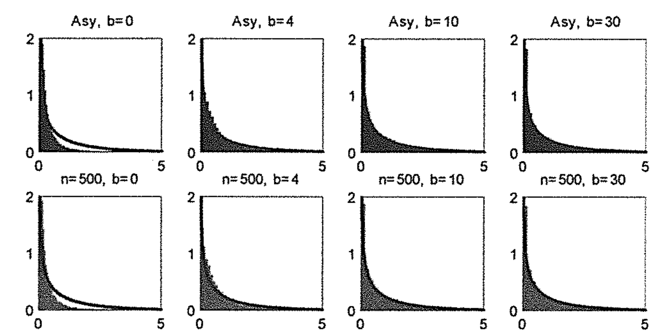

Figure S-6. Asymptotic and Finite-Sample (n=500) Densities of the QLR Statistic for π in the Nonlinear Regression Model with Endogeneity when $\pi_0 = 1.5$ and the χ_1^2 Density (Black Line).

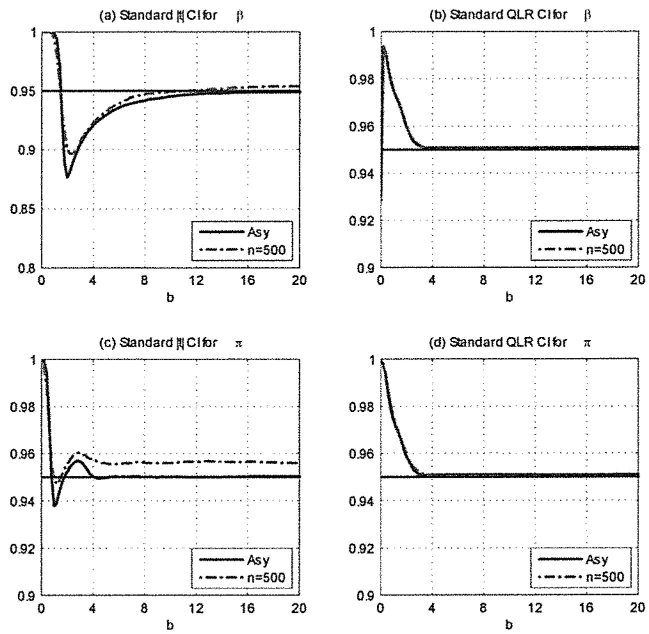

Figure S-7. Coverage Probabilities of Standard $|t|$ and QLR CI's for β and π in the Nonlinear Regression Model with Endogeneity when $\pi_0 = 3.0$.

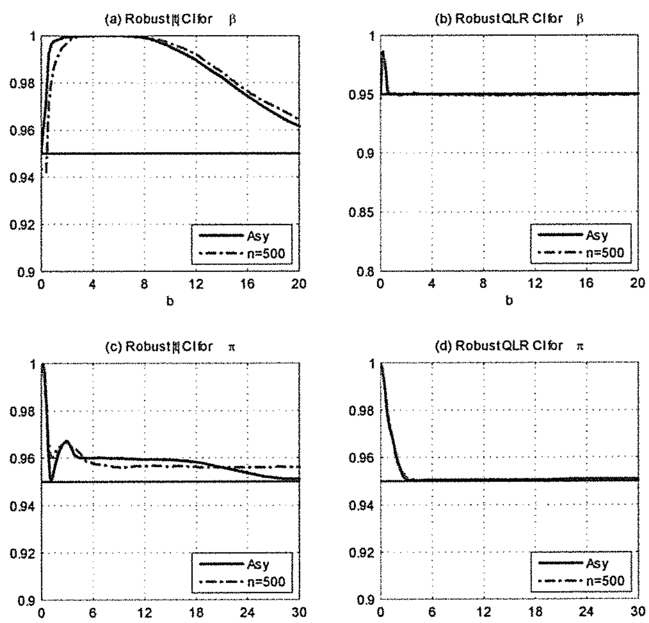

Figure S-8. Coverage Probabilities of Robust $|t|$ and QLR CI's for β and π in the Nonlinear Regression Model with Endogeneity when $\pi_0 = 3.0$, $\kappa = 1.5, D = 1$, and $s(x) = \exp(-2x)$.

Reference

Andrews, D.W.K. & X. Cheng (2011) Supplemental Appendix to Maximum likelihood estimation and uniform inference with sporadic identification failure. Cowles Foundation Discussion Paper No. 1824R.