Econometrica Supplementary Material

SUPPLEMENT TO "DISTORTIONS OF ASYMPTOTIC CONFIDENCE SIZE IN LOCALLY MISSPECIFIED MOMENT INEQUALITY MODELS"

(*Econometrica*, Vol. 80, No. 4, July 2012, 1741–1768)

BY FEDERICO A. BUGNI, IVAN A. CANAY, AND PATRIK GUGGENBERGER

THIS SUPPLEMENT CONTAINS the lemmas (and their proofs) that are used in the proofs of Theorems 3.1 and 3.2 of the paper in Sections S1 and S2; the proof of Corollary 3.1 in Section S2; a missing data example and Monte Carlo simulations in Section S3; the verification of Assumptions A.6 and A.7 in two leading examples in Section S4; a discussion of the intuition behind Theorem 3.2 in Section S5; and details of the computations carried out in Table I in Section S6.

S1. AUXILIARY LEMMAS

LEMMA S1.1: *Assume that the parameter space is given by $\mathcal{F}_n$ in Eq.* (2.5) *and that S satisfies Assumption* A.1. *Under any sequence $\{\gamma_{\omega_n,h}\}_{n\geq 1} = \{\theta_{\omega_n,h}, F_{\omega_n,h}\}_{n\geq 1}$ defined in Definition* A.1 *for a subsequence $\{\omega_n\}_{n\geq 1}$ and $h = (h_1, h_2)$, it follows that*

$$T_{\omega_n}(\theta_{\omega_n,h}) \to_d J_h \sim S(h_2^{1/2} Z + h_1, h_2), \tag{S1.1}$$

where $T_n(\cdot)$ is the test statistic associated with S and $Z = (Z_1, \dots, Z_k) \sim N(0_k, I_k)$.

LEMMA S1.2: *For any $a \in (0,1)$ and $\rho \in [-1+a, 1-a]$, define*

$$\begin{aligned} f(z_1, z_2, \rho) &\equiv (1-\rho^2)^{-1} \\ &\quad \times \min_{(t_1,t_2)\in\mathbb{R}^2_{+,+\infty}} \{(z_1-t_1)^2 + (z_2-t_2)^2 - 2\rho(z_1-t_1)(z_2-t_2)\}. \end{aligned} \tag{S1.2}$$

Then $f(z_1, z_2, \rho)$ takes values according to the following four cases:

(i) *Let $z_1 \geq 0$ and $z_2 \geq 0$. Then, $f(z_1, z_2, \rho) = 0$.*

(ii) *Let $z_1 \geq 0$ and $z_2 < 0$. If $\rho \leq z_1/z_2$, then*

$$f(z_1, z_2, \rho) = (1-\rho^2)^{-1}[(z_1-z_2)^2 + 2(1-\rho)z_1 z_2]. \tag{S1.3}$$

If $\rho > z_1/z_2$, then $f(z_1, z_2, \rho) = z_2^2$.

(iii) *Let $z_1 < 0$ and $z_2 \geq 0$. If $\rho \leq z_2/z_1$, then Eq.* (S1.3) *holds. Otherwise, $f(z_1, z_2, \rho) = z_1^2$.*

 DOI: 10.3982/ECTA9604

(iv) *Let* $z_1 < 0$ *and* $z_2 < 0$. *If* $\rho \leq \min\{z_1/z_2, z_2/z_1\}$, *then Eq.* (S1.3) *holds. Otherwise*, $f(z_1, z_2, \rho) = \max\{z_1^2, z_2^2\}$.

LEMMA S1.3: *Suppose that* $k = p = 2$ *and fix* $\beta > 0$. *For* $\varepsilon > 0$, *let* $\rho_\varepsilon = -\sqrt{1-\varepsilon}$, $h_{2,\varepsilon} = \left[\begin{smallmatrix} 1 & \rho_\varepsilon \\ \rho_\varepsilon & 1 \end{smallmatrix}\right]$, *and*

$$H_\beta \equiv \{h_1 \in \mathbb{R}^2 : h_{1,1} \leq -\beta, h_{1,2} \leq 0\}. \tag{S1.4}$$

Define the set $A_{\varepsilon,\beta} \equiv A^a_{\varepsilon,\beta} \cup A^b_{\varepsilon,\beta} \cup A^c_{\varepsilon,\beta} \subseteq \mathbb{R}^2$, *where*

$$A^a_{\varepsilon,\beta} \equiv \{z \in \mathbb{R}^2 : z_1 \geq 0, z_2 < 0, 0 < z_1 - \rho_\varepsilon z_2 \leq \beta/2\}, \tag{S1.5}$$

$$A^b_{\varepsilon,\beta} \equiv \{z \in \mathbb{R}^2 : z_1 < 0, z_2 \geq 0, 0 < z_2 - \rho_\varepsilon z_1 \leq \beta/2\}, \tag{S1.6}$$

$$A^{c1}_{\varepsilon,\beta} \equiv \{z \in \mathbb{R}^2 : z_1 \geq 0, z_2 < 0, z_1 - \rho_\varepsilon z_2 \leq 0\}, \tag{S1.7}$$

$$A^{c2}_{\varepsilon,\beta} \equiv \{z \in \mathbb{R}^2 : z_1 < 0, z_2 \geq 0, z_2 - \rho_\varepsilon z_1 \leq 0\}, \tag{S1.8}$$

and $A^c_{\varepsilon,\beta} \equiv A^{c1}_{\varepsilon,\beta} \cup A^{c2}_{\varepsilon,\beta}$. *Let* $Z_{h_{2,\varepsilon}} \sim N(0, h_{2,\varepsilon})$. *Then there exists a real-valued function* $\tau_\varepsilon(z, h_1) : A_{\varepsilon,\beta} \times H_\beta \to \mathbb{R}_+$ *such that*

$$S_2(z + h_1, h_{2,\varepsilon}) = S_2(z, h_{2,\varepsilon}) + \frac{1}{1-\rho_\varepsilon^2}\tau_\varepsilon(z, h_1), \quad \forall z \in A_{\varepsilon,\beta}, \forall h_1 \in H_\beta, \tag{S1.9}$$

and $\forall \eta, C > 0, \exists \varepsilon > 0$ *such that*

$$\inf_{h_1 \in H_\beta} \Pr\left(\frac{1}{1-\rho_\varepsilon^2}\tau_\varepsilon\big(Z_{h_{2,\varepsilon}}, h_1\big) > C, Z_{h_{2,\varepsilon}} \in A_{\varepsilon,\beta}\right) \geq 1 - \eta. \tag{S1.10}$$

S2. PROOF OF LEMMAS AND COROLLARIES

PROOF OF LEMMA S1.1: The proof follows along the lines of the proof of Theorem 1 in Andrews and Guggenberger (2009; AG from now on). By Lemma 1 in AG, we have, for any $s \in \mathbb{N}$,

$$T_s(\theta_s) = S\big(\hat{D}_s^{-1/2}(\theta_s)s^{1/2}\bar{m}_s(\theta_s), \hat{D}_s^{-1/2}(\theta_s)\hat{\Sigma}_s(\theta_s)\hat{D}_s^{-1/2}(\theta_s)\big). \tag{S2.1}$$

For $j = 1, \ldots, k$, define $A_{s,j} = \sigma^{-1}_{F_s,j}(\theta_s)s^{1/2}(\bar{m}_{s,j}(\theta_s) - E_{F_s}\bar{m}_{s,j}(\theta_s))$. As in Lemma 2 in AG, we have

$$\text{(i)} \quad A_{\omega_n} = \big(A_{\omega_n,1}, \ldots, A_{\omega_n,k}\big)' \to_d Z_{h_2} = \big(Z_{h_2,1}, \ldots, Z_{h_2,k}\big)' \sim N(0_k, h_2) \quad \text{as} \quad n \to \infty, \tag{S2.2}$$

$$\text{(ii)}\quad \hat{\sigma}_{\omega_n,j}\big(\theta_{\omega_n,h}\big)/\sigma_{F_{\omega_n,h},j}\big(\theta_{\omega_n,h}\big) \to_p 1 \quad \text{as} \quad n\to\infty \quad \text{for} \quad j=1,\dots,k,$$

$$\text{(iii)}\quad \hat{D}_{\omega_n}^{-1/2}\big(\theta_{\omega_n,h}\big)\hat{\Sigma}_{\omega_n}\big(\theta_{\omega_n,h}\big)\hat{D}_{\omega_n}^{-1/2}\big(\theta_{\omega_n,h}\big) \to_p h_2 \quad \text{as} \quad n\to\infty,$$

under any sequence $\gamma_{\omega_n,h} = \{\theta_{\omega_n,h}, F_{\omega_n,h}\}_{n\geq 1}$. These results follow from completing the sequence $\gamma_{\omega_n,h}$, that is, from defining the sequence $\gamma_{s,h} = \{\theta_{s,h}, F_{s,h}\}_{s\geq 1}$ in the following fashion. For $s \in \mathbb{N}$, define the sequence $\{\theta_s, F_s\}_{s\geq 1}$ as follows. For any $s \leq \omega_1$, $(\theta_s, F_s) = (\theta_{\omega_1,h}, F_{\omega_1,h})$. For any $s > \omega_1$ and since $\{\omega_n\}_{n\geq 1}$ is a subsequence of $\mathbb{N}$, there exists a unique $m \in \mathbb{N}$ such that $\omega_{m-1} < s \leq \omega_m$. For every such s, set $(\theta_s, F_s) = (\theta_{\omega_m,h}, F_{\omega_m,h})$. Now let $\{W_i\}_{i\leq n}$ be i.i.d. under F_s. By construction, $\forall s \in \mathbb{N}$, $(\theta_s, F_s) \in \mathcal{F}_{\omega_m}$ for some $m \in \mathbb{N}$ and $\text{Corr}_{F_s}(m(W_i, \theta_s)) \to h_2$. Then, the results (i)–(iii) of Eq. (S2.2) hold by triangular array versions of central limit theorems and the law of large numbers with ω_n, $\theta_{\omega_n,h}$, and $F_{\omega_n,h}$ replaced by s, θ_s, and F_s, respectively. But the convergence results along $\{\theta_s, F_s\}_{s\geq 1}$ then imply convergence along the subsequence $\{\theta_{\omega_n,h}, F_{\omega_n,h}\}_{n\geq 1}$, as by construction the latter coincides with the former on indices $s = \omega_n$.

From Eq. (S2.2), the jth element of $\hat{D}_{\omega_n}^{-1/2}(\theta_{\omega_n,h})\omega_n^{1/2}\bar{m}_{\omega_n}(\theta_{\omega_n,h})$ equals $(A_{\omega_n,j} + \omega_n^{1/2}E_{F_{\omega_n,h}}\bar{m}_{\omega_n,j}(\theta_{\omega_n,h})/\sigma_{F_{\omega_n,h},j}(\theta_{\omega_n,h}))(1 + o_p(1))$. We next consider a k-vector-valued function of $\hat{D}_{\omega_n}^{-1/2}(\theta_{\omega_n,h})\omega_n^{1/2}\bar{m}_{\omega_n}(\theta_{\omega_n,h})$ that converges in distribution whether or not some elements of h_1 equal ∞. Write the right hand side (RHS) of Eq. (S2.1) as a continuous function of this k-vector and apply the continuous mapping theorem. Let $G(\cdot)$ be a strictly increasing continuous distribution function (d.f.) on $\mathbb{R}$, such as the standard normal d.f., and let $G(\infty) = 1$. For $j = 1, \dots, k$, we have

$$\text{(S2.3)}\qquad \begin{aligned} G_{\omega_n,j} &\equiv G\big(\hat{\sigma}_{\omega_n,j}^{-1}\big(\theta_{\omega_n,h}\big)\omega_n^{1/2}\bar{m}_{\omega_n,j}\big(\theta_{\omega_n,h}\big)\big) \\ &= G\big(\hat{\sigma}_{\omega_n,j}^{-1}\big(\theta_{\omega_n,h}\big)\sigma_{F_{\omega_n,h},j}\big(\theta_{\omega_n,h}\big) \\ &\qquad \times \big[A_{\omega_n,j} + \omega_n^{1/2}E_{F_{\omega_n,h}}\bar{m}_{\omega_n,j}\big(\theta_{\omega_n,h}\big)/\sigma_{F_{\omega_n,h},j}\big(\theta_{\omega_n,h}\big)\big]\big). \end{aligned}$$

If $h_{1,j} < \infty$, then

$$\text{(S2.4)}\qquad G_{\omega_n,j} \to_d G\big(Z_{h_2,j} + h_{1,j}\big)$$

by Eqs. (S2.3) and (S2.2), the definition of $\gamma_{\omega_n,h}$, and the continuous mapping theorem. If $h_{1,j} = \infty$ (which can only happen for $j = 1, \dots, p$), then

$$\text{(S2.5)}\qquad G_{\omega_n,j} = G\big(\hat{\sigma}_{\omega_n,j}^{-1}\big(\theta_{\omega_n,h}\big)\omega_n^{1/2}\bar{m}_{\omega_n,j}\big(\theta_{\omega_n,h}\big)\big) \to_p 1$$

by Eq. (S2.3), $A_{\omega_n,j} = O_p(1)$, and $G(x) \to 1$ as $x \to \infty$. The results in Eqs.

(S2.4) and (S2.5) hold jointly and combine to give

$$\text{(S2.6)}\qquad G_{\omega_n} \equiv (G_{\omega_n,1},\dots,G_{\omega_n,k})' \to_d \big(G(Z_{h_2,1}+h_{1,1}),\dots,G(Z_{h_2,k}+h_{1,k})\big)' \equiv G_\infty,$$

where $G(Z_{h_2,j}+h_{1,j})=1$ by definition when $h_{1,j}=\infty$. Let G^{-1} denote the inverse of G. For $x=(x_1,\dots,x_k)'\in\mathbb{R}^p_{+\infty}\times\mathbb{R}^v$, let $G_{(k)}(x)=(G(x_1),\dots,G(x_k))'\in(0,1]^p\times(0,1)^v$. For $y=(y_1,\dots,y_k)'\in(0,1]^p\times(0,1)^v$, let $G^{-1}_{(k)}(y)=(G^{-1}(y_1),\dots,G^{-1}(y_k))'\in\mathbb{R}^p_{+\infty}\times\mathbb{R}^v$. Define $S^*(y,\Omega)=S(G^{-1}_{(k)}(y),\Omega)$ for $y\in(0,1]^p\times(0,1)^v$ and $\Omega\in\Psi$. By Assumption A.1(d), $S^*(y,\Omega)$ is continuous at all (y,Ω) for $y\in(0,1]^p\times(0,1)^v$ and $\Omega\in\Psi$. We now have

$$\begin{aligned}\text{(S2.7)}\qquad T_{\omega_n}(\theta_{\omega_n,h}) &= S\big(G^{-1}_{(k)}(G_{\omega_n}),\hat{D}^{-1/2}_{\omega_n}(\theta_{\omega_n,h})\hat{\Sigma}_{\omega_n}(\theta_{\omega_n,h})\hat{D}^{-1/2}_{\omega_n}(\theta_{\omega_n,h})\big)\\ &= S^*\big(G_{\omega_n},\hat{D}^{-1/2}_{\omega_n}(\theta_{\omega_n,h})\hat{\Sigma}_{\omega_n}(\theta_{\omega_n,h})\hat{D}^{-1/2}_{\omega_n}(\theta_{\omega_n,h})\big)\\ &\to_d S^*(G_\infty,h_2)\\ &= S\big(G^{-1}_{(k)}(G_\infty),h_2\big)\\ &= S\big(Z_{h_2}+h_1,h_2\big)\sim J_h,\end{aligned}$$

where the convergence holds by Eqs. (S2.2) and (S2.6) and the continuous mapping theorem, the last equality holds by the definitions of $G^{-1}_{(k)}$ and G_∞, and the last line hold by definition of J_h. *Q.E.D.*

PROOF OF LEMMA S1.2: The FOC associated with the minimizers t_1 and t_2 in Eq. (S1.2) are

$$\text{(S2.8)}\qquad -(z_1-t_1)+\rho(z_2-t_2)\ge 0,\quad u_1[-(z_1-t_1)+\rho(z_2-t_2)]=0,\quad t_1\ge 0,$$

$$\text{(S2.9)}\qquad -(z_2-t_2)+\rho(z_1-t_1)\ge 0,\quad u_2[-(z_2-t_2)+\rho(z_1-t_1)]=0,\quad t_2\ge 0.$$

The SOC are immediately satisfied, as the function on the RHS of Eq. (S1.2) is strictly convex for $\rho\in[-1+a,1-a]$.

Consider Case (i). In this case, $t_1=z_1$ and $t_2=z_2$ satisfies Eqs. (S2.8) and (S2.9) and $f(z_1,z_2,\rho)=0$ regardless of the value of ρ.

Now consider Case (ii). First we note that $t_1\ge 0$, $t_2>0$ is not a feasible solution. This is because $t_2>0$ and Eq. (S2.9) imply $\rho^2(z_1-t_1)=\rho(z_2-t_2)$, and this, together with Eq. (S2.8), results in $\rho^2(z_1-t_1)\ge(z_1-t_1)$, which implies $t_1=z_1$. This results in $t_2=z_2<0$, which is a contradiction. The solution must then be of the form $t_1\ge 0$ and $t_2=0$. Then, it follows from the first conditions in Eq. (S2.8) that $t_1 \geq z_1 - \rho z_2$, so that $t_1 = \max\{z_1 - \rho z_2, 0\}$ and $t_2 = 0$ is the solution. This is a strictly convex optimization problem, and so the solution exists and is unique. Then, if $\rho \leq z_1/z_2$, the unique solution is $(t_1, t_2) = (0, 0)$ and the objective function is given by Eq. (S1.3). On the other hand, if $\rho < z_1/z_2$, $(t_1, t_2) = (z_1 - \rho z_2, 0)$ is the unique solution and

$$\text{(S2.10)} \quad f(z_1, z_2, \rho) = (1-\rho^2)^{-1}\{(z_1 - z_1 + \rho z_2)^2 + z_2^2 - 2\rho(z_1 - z_1 + \rho z_2)(z_2)\} = z_2^2.$$

Case (iii) is exactly analogous to Case (ii) by exchanging the subindices 1 and 2.

Consider Case (iv) then. First, we note again that $t_1 > 0$ and $t_2 > 0$ is not a feasible solution by the same arguments as before. Second, we note that $(t_1, t_2) = (0, 0)$ is a solution provided $\rho \leq \min\{z_1/z_2, z_2/z_1\}$, as this condition implies the correct sign of the derivatives in Eqs. (S2.8) and (S2.9). The remaining case is either $\rho > z_1/z_2$ or $\rho > z_2/z_1$. By steps similar to those used in Case (ii), it follows that the solutions for these cases are $(t_1, t_2) = (z_1 - \rho z_2, 0)$, $f(z_1, z_2, \rho) = z_2^2$ and $(t_1, t_2) = (0, z_2 - \rho z_1)$, $f(z_1, z_2, \rho) = z_1^2$, respectively. *Q.E.D.*

PROOF OF LEMMA S1.3: Note that, for $\bar{z}_j = z_j + h_{1,j}$, $j = 1, 2$,

$$\text{(S2.11)} \quad S_2(z + h_1, h_{2,\varepsilon}) = (1-\rho_\varepsilon^2)^{-1} \min_{t \in \mathbb{R}^2_{+,+\infty}} \{(\bar{z}_1 - t_1)^2 + (\bar{z}_2 - t_2)^2 - 2\rho_\varepsilon(\bar{z}_1 - t_1)(\bar{z}_2 - t_2)\}.$$

The RHS of Eq. (S2.11) is the same optimization problem as the one in Lemma S1.2. For $z \in A_{\varepsilon,\beta}$ and $h_1 \in H_\beta$, Case (i) of Lemma S1.2 cannot occur and Cases (ii)–(iv) always end up with the subcase that leads to Eq. (S1.3). This is because $\beta/2 \leq \rho_\varepsilon h_{1,2} - h_{1,1}$ and, for $\varepsilon < 1/2$, $\beta/2 \leq -h_{1,2} + \rho_\varepsilon h_{1,1}$. It follows from Lemma S1.2 that the solution of Eq. (S2.11) for $z \in A_{\varepsilon,\beta}$ and $h_1 \in H_\beta$ is

$$\text{(S2.12)} \quad S_2(z + h_1, h_{2,\varepsilon}) = (1-\rho_\varepsilon^2)^{-1} \times [(z_1 + h_{1,1} - z_2 - h_{1,2})^2 + 2(1-\rho_\varepsilon)(z_1 + h_{1,1})(z_2 + h_{1,2})].$$

In addition, it follows from Lemma S1.2 that the solution when $h_{1,1} = h_{1,2} = 0$ is given by $S_2(z, h_{2,\varepsilon}) = z_2^2$ for $z \in A^a_{\varepsilon,\beta}$, $S_2(z, h_{2,\varepsilon}) = z_1^2$ for $z \in A^b_{\varepsilon,\beta}$, and $S_2(z, h_{2,\varepsilon}) = z_1^2 + (z_2 - \rho_\varepsilon z_1)^2/(1-\rho_\varepsilon^2)$ for $z \in A^c_{\varepsilon,\beta}$. By some algebra, it follows that

$$\text{(S2.13)} \quad S_2(z + h_1, h_{2,\varepsilon}) = S_2(z, h_{2,\varepsilon}) + \frac{1}{1-\rho_\varepsilon^2}\tau_{\varepsilon,l}(z, h_1), \quad \forall z \in A^l_{\varepsilon,\beta}, \forall h_1 \in H_\beta, l \in \{a, b, c\},$$

where

$$\tau_{\varepsilon,a}(z,h_1)=(z_1+h_{1,1}-\rho_\varepsilon(z_2+h_{1,2}))^2+(1-\rho_\varepsilon^2)(h_{1,2}^2+2z_2h_{1,2}), \tag{S2.14}$$

$$\tau_{\varepsilon,b}(z,h_1)=(z_2+h_{1,2}-\rho_\varepsilon(z_1+h_{1,1}))^2+(1-\rho_\varepsilon^2)(h_{1,1}^2+2z_1h_{1,1}), \tag{S2.15}$$

$$\begin{aligned}\tau_{\varepsilon,c}(z,h_1)&=(h_{1,1}-h_{1,2})^2+2((z_2-\rho_\varepsilon z_1)(h_{1,2}-\rho_\varepsilon h_{1,1})\\&\quad+h_{1,1}z_1(1-\rho_\varepsilon^2)+(1-\rho_\varepsilon)h_{1,1}h_{1,2}).\end{aligned} \tag{S2.16}$$

Additional algebra shows that $\tau_{\varepsilon,a}(z,h_1)\geq\beta^2/4$ on $A^a_{\varepsilon,\beta}\times H_\beta$, $\tau_{\varepsilon,b}(z,h_1)\geq\beta^2/6$ on $A^b_{\varepsilon,\beta}\times H_\beta$ for $\varepsilon<1/2$, and $\tau_{\varepsilon,c}(z,h_1)\geq\beta^2$ on $A^c_{\varepsilon,\beta}\times H_\beta$. Thus, letting

$$\tau_\varepsilon(z,h_1)\equiv\sum_{l\in\{a,b,c\}}\tau_{\varepsilon,l}(z,h_1)I(z\in A^l_{\varepsilon,\beta}), \tag{S2.17}$$

it follows that $\tau_\epsilon(z,h_1)\geq\beta^2/6$ on $A_{\varepsilon,\beta}\times H_\beta$. We can conclude that, for any constant $C\in(0,\infty)$, there exists $\varepsilon\in(0,1/2)$ such that

$$\begin{aligned}\inf_{h_1\in H_\beta}\Pr&\left(\frac{1}{1-\rho_\varepsilon^2}\tau_\varepsilon\big(Z_{h_{2,\varepsilon}},h_1\big)>C,Z_{h_{2,\varepsilon}}\in A_{\varepsilon,\beta}\right)\\&\geq\Pr\left(\frac{\beta^2}{6(1-\rho_\varepsilon^2)}>C,Z_{h_{2,\varepsilon}}\in A_{\varepsilon,\beta}\right),\\&=\Pr\big(Z_{h_{2,\varepsilon}}\in A_{\varepsilon,\beta}\big).\end{aligned} \tag{S2.18}$$

Finally, define the set $A^{ac1}_{\varepsilon,\beta}\equiv A^a_{\varepsilon,\beta}\cup A^{c1}_{\varepsilon,\beta}$. Note that we can write $Z_{h_{2,\varepsilon},1}-\rho_\varepsilon Z_{h_{2,\varepsilon},2}=\sqrt{1-\rho_\varepsilon^2}W$ for $Z_{h_{2,\varepsilon},2}\perp W\sim N(0,1)$. Then,

$$\begin{aligned}&\Pr\big(Z_{h_{2,\varepsilon}}\in A^{ac1}_{\varepsilon,\beta}\big)\\&\quad=\Pr\left(Z_{h_{2,\varepsilon},2}\leq\min\left\{0,\frac{\sqrt{1-\rho_\varepsilon^2}W}{-\rho_\varepsilon}\right\},W\leq\frac{\beta}{2\sqrt{1-\rho_\varepsilon^2}}\right)\\&\quad\to1/2,\quad\text{as}\quad\rho_\varepsilon\to-1.\end{aligned} \tag{S2.19}$$

The same applies for the set $A^{bc2}_{\varepsilon,\beta}\equiv A^b_{\varepsilon,\beta}\cup A^{c2}_{\varepsilon,\beta}$. By continuity in ρ_ε, it follows that $\forall\eta>0$, $\exists\varepsilon>0$ such that

$$\Pr\big(Z_{h_{2,\varepsilon}}\in A_{\varepsilon,\beta}\big)\geq1-\eta. \tag{S2.20}$$

The result then follows from Eqs. (S2.18) and (S2.20). *Q.E.D.*

PROOF OF COROLLARY 3.1: Let $\varepsilon>0$. Define the test function $\tilde{S}_{2,\varepsilon}(m,\Sigma)$ as in Eq. (3.5) and the parameter space $\Psi_{2,\varepsilon}$. Let $f_\varepsilon:\Psi_1\to\mathcal{V}_{k\times k}$ be defined as

$$f_\varepsilon(\Omega)\equiv\Omega+\max\{\varepsilon-\det(\Omega),0\}I_k, \tag{S2.21}$$

and note that $f_\varepsilon(\cdot)$ is a continuous function. By construction, the matrix $f_\varepsilon(\Omega)$ has a determinant that is bounded away from zero, that is, $\exists C > 0$ such that $\inf_{\Omega\in\Psi_1} \det(f_\varepsilon(\Omega)) \geq C$. As

$$\text{(S2.22)} \quad \tilde{S}_{2,\varepsilon}(m, \Omega) = \inf_{t=(t_1,0_v):t_1\in\mathbb{R}^p_{+,+\infty}} (m-t)' f_\varepsilon(\Omega)^{-1}(m-t),$$

it follows that $\tilde{S}_{2,\varepsilon}(m, \Omega)$ is continuous at all $\Omega \in \Psi_1$, and so it satisfies Assumptions A.1–A.3. By Lemmas S1.1 and B.1, it follows that

$$\text{(S2.23)} \quad \text{AsySz}^{(\tilde{2},\varepsilon)}_{\text{PA}} = \inf_{h=(h_1,h_2)\in H} \Pr\bigl(\tilde{S}_{2,\varepsilon}\bigl(Z_{h_2} + h_1, h_2\bigr) \leq c_0(h_2, 1-\alpha)\bigr).$$

Next, note that, for all $h_2 \in \Psi_{2,\varepsilon}$, we have $f_\varepsilon(h_2) = h_2$, and so it follows that $\tilde{S}_{2,\varepsilon}(m, h_2) = S_2(m, h_2)$ for all $m \in \mathbb{R}^p_{+\infty} \times \mathbb{R}^v$ and $h_2 \in \Psi_{2,\varepsilon}$. Since $\Psi_{2,\varepsilon} \subset \Psi_1$, we have

$$\text{(S2.24)} \quad \text{AsySz}^{(\tilde{2},\varepsilon)}_{\text{PA}} \leq \text{AsySz}^{(2)}_{\text{PA}}.$$

It follows from Theorem 3.2 that, for every $\eta > 0$, $\exists \varepsilon > 0$ such that $\text{AsySz}^{(\tilde{2},\varepsilon)}_{\text{PA}} \leq \eta$. By the proof of Theorem 3.1, $\text{AsySz}^{(\tilde{2},\varepsilon)}_{\text{PA}} \geq \text{AsySz}^{(\tilde{2},\varepsilon)}_{SS} \geq \text{AsySz}^{(\tilde{2},\varepsilon)}_{\text{GMS}}$. *Q.E.D.*

S3. MISSING DATA EXAMPLE WITH SIMULATIONS

EXAMPLE S3.1—Missing Data: Suppose that the economic model indicates that

$$\text{(S3.1)} \quad E_{F_0}(Y|X = x) = G(x, \theta_0), \quad \forall x \in S_X,$$

where θ_0 is the true parameter value and $S_X = \{x_l\}_{l=1}^{d_x}$ is the (finite) support of X. The sample is affected by missing data on Y. Denote by Z the binary variable that takes a value of 1 if Y is observed and zero if Y is missing. Conditional on $X = x$, Y has logical lower and upper bounds given by $Y_L(x)$ and $Y_H(x)$, respectively. The observed data are $\{W_i\}_{i=1}^n$, where $\forall i = 1, \dots, n$, $W_i = (Y_i Z_i, Z_i, X_i)$. When the observed data come from the model in Eq. (S3.1), the true θ_0 satisfies the following inequalities for $l = 1, \dots, d_x$:

$$\text{(S3.2)} \quad \begin{aligned} &E_{F_0} m_{l,1}(W_i, \theta_0) \\ &\quad \equiv E_{F_0}\bigl[(YZ + Y_H(x_l)(1-Z) - G(x_l, \theta_0))I(X = x_l)\bigr] \geq 0, \\ &E_{F_0} m_{l,2}(W_i, \theta_0) \\ &\quad \equiv E_{F_0}\bigl[(G(x_l, \theta_0) - YZ - Y_L(x_l)(1-Z))I(X = x_l)\bigr] \geq 0. \end{aligned}$$

Notice that when $\{Y|X = x_l\}$ is fully observed (i.e., $\Pr(Z|X = x_l) = 1$), Eq. (S3.2) implies the equation

$$E_{F_0} m_l(W_i, \theta_0) \equiv E_{F_0}\big[(Y - G(x_l, \theta_0)) I(X = x_l)\big] = 0. \tag{S3.3}$$

Now suppose that, in fact, the data come from a local perturbation F_n of the hypothesized model F_0 such that

$$E_{F_n}(Y|X = x_l) = G_n(x_l, \theta_0), \quad \forall l = 1, \dots, d_x, \tag{S3.4}$$

and for a vector $r \in \mathbb{R}_+^{d_x}$,

$$|G_n(x_l, \theta_0) - G(x_l, \theta_0)| \leq r_l n^{-1/2}, \quad \forall l = 1, \dots, d_x. \tag{S3.5}$$

This last condition says that the true function G_n is not too distant from the model G used by the researcher. After a few manipulations, it follows that, for $l = 1, \dots, d_x$,

$$\begin{aligned} &E_{F_n} m_{l,1}(W_i, \theta_0) \\ &\quad = E_{F_n}\big[(YZ + Y_H(x_l)(1 - Z) - G(x_l, \theta_0)) I(X = x_l)\big] \geq -r_l n^{-1/2}, \\ &E_{F_n} m_{l,2}(W_i, \theta_0) \\ &\quad = E_{F_n}\big[(G(x_l, \theta_0) - YZ - Y_L(x_l)(1 - Z)) I(X = x_l)\big] \geq -r_l n^{-1/2}, \end{aligned} \tag{S3.6}$$

and when $\{Y|X = x_l\}$ is fully observed, Eq. (S3.5) implies the moment condition

$$E_{F_n} m_l(W_i, \theta_0) = E_{F_n}\big[(Y - G(x_l, \theta_0)) I(X = x_l)\big] \leq r_l n^{-1/2}. \tag{S3.7}$$

Therefore, under the perturbed distribution of the data, the original moment conditions, Eqs. (S3.2) and (S3.3), may be locally violated at θ_0.

S3.1. *Numerical Simulations*

In this section, we describe a small simulation study to assess the finite sample relevance of the asymptotic results in Theorem 3.1. We simulate data according to Example S3.1, using the following parameterization for the data generating process: $\theta_0 = (0.1, -0.5)$, $S_X = \{(1, 0), (0, 1)\}$ (i.e., $d_x = 2$), $\Pr(X = (1, 0)) = \Pr(X = (0, 1)) = 0.5$, $Y \in \{0, 1\}$ is a binary random variable, $G(x, \theta) = \Phi(\theta' x)$, where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function (i.e., F_0 implies that this is a Probit model), and $\Pr(Z = 1|X = x) = I(x = (1, 0)) + 0.5 I(x = (0, 1))$. According to the parameterization, Y is always observed when $X = (1, 0)$. The model then results in two moment inequalities and one moment equality (i.e., $p=2$ and $k=3$).[1] We compare PA, subsampling, and GMS critical values, together with the test function S_1 from Eq. (2.9).

The remaining parameters of the simulation are as follows. The sample size is $n=1{,}000$ and the number of Monte Carlo simulations is $\mathrm{MC}=500$. All of the critical values are approximated by simulation and, in every case, a total of $B=200$ simulations are used. The subsampling block size b is allowed to be $2n/25$, $(2n/25+n^{1/2})/2$, and $n^{1/2}$, resulting in three subsampling procedures: SS1, SS2, and SS3, respectively. The GMS tuning parameter κ_n is allowed to be $\ln\ln n$, $(\ln\ln n+\ln n)/2$, and $\ln n$, resulting in three GMS procedures: GMS1, GMS2, and GMS3, respectively. Finally, we take $r_l=r^*\geq 0$ for $l=1,2,3$, where $r^*n^{-1/2}$ is allowed to be any number in an equally spaced grid of numbers from 0 to 0.1.[2]

For simplicity, instead of focusing on the distortions in the confidence size of the CSs, we look at the distortions in the size of tests for the null hypothesis $H_0:\theta_0=(0.1,-0.5)$. This substitution simplifies the computations significantly and provides conclusions that can also be applied to the confidence size of CSs, as explained in Remark 2.3. Under the null hypothesis, it follows that

$$E_{F_0}(Y|X=(1,0))=0.54 \quad \text{and} \quad E_{F_0}(Y|X=(0,1))=0.31. \tag{S3.8}$$

A perturbation F' of F_0 results in different values of the expectations $E_{F'}(Y|X=x)$.[3] In our simulation exercise, we consider the set of all distributions F' that would result in values of $E_{F'}(Y|X=x)$ that are in a small neighborhood of the predictions of the model under the null hypothesis. In particular, for any $r^*\geq 0$, we consider the set of models $\mathcal{F}_{r^*}$ defined as

$$\mathcal{F}_{r^*}\equiv\left\{F'\in\mathcal{P}:\left\{\begin{array}{c}[E_{F'}m_1(W_i,\theta_0)/\sigma_{F',1}(\theta_0)]_-\leq r^*n^{-1/2}\\ [E_{F'}m_2(W_i,\theta_0)/\sigma_{F',2}(\theta_0)]_-\leq r^*n^{-1/2}\\ |E_{F'}m_3(W_i,\theta_0)/\sigma_{F',3}(\theta_0)|\leq r^*n^{-1/2}\end{array}\right\}\right\}, \tag{S3.9}$$

where $\mathcal{P}$ is the set of all distributions that satisfy the restrictions of the data generating process and

$$\begin{aligned} m_1(W_i,\theta_0)&=(YZ+(1-Z)-G(X_i,\theta_0))I(X_i=(0,1)),\\ m_2(W_i,\theta_0)&=(G(X_i,\theta_0)-YZ)I(X_i=(0,1)),\\ m_3(W_i,\theta_0)&=(Y-G(X_i,\theta_0))I(X_i=(1,0)). \end{aligned} \tag{S3.10}$$

[1]For simplicity, we also assume that the distributions of X and Z are known and fixed by design and, thus, the researcher is effectively conducting inference about the (conditional) distribution of Y.

[2]For $r^*n^{-1/2}\geq 0.1$, all of the inferential methods are 100% distorted.

[3]For example, if the model was a Logit model, then $E_{F'}(Y|X=(1,0))=\Lambda(0.1)=0.53$ and $E_{F'}(Y|X=(0,1))=\Lambda(-0.5)=0.37$, where Λ denotes the logistic cumulative distribution function.

TABLE S.I

MAXIMUM REJECTION PROBABILITY OVER CORRECTLY SPECIFIED MODELS (I.E., $r^* n^{-1/2} = 0$)[a]

PA	SS1	SS2	SS3	GMS1	GMS2	GMS3
0.07	0.065	0.035	0.005	0.105	0.105	0.105

[a]Simulation parameters: $n = 1{,}000$, $\mathrm{MC} = 500$, $B = 200$, $b = \{2n/25, (2n/25 + n^{1/2})/2, n^{1/2}\}$, $\kappa_n = \{\ln\ln n, (\ln\ln n + \ln n)/2, \ln n\}$, and $\alpha = 0.10$.

Given a value of $r^* \geq 0$, we explore *all* models that are in $\mathcal{F}_{r^*}$ and compare the maximum rejection probabilities across inferential methods. That is, we report

$$\text{(S3.11)} \quad \sup_{F \in \mathcal{F}_{r^*}} \Pr_F\left(T_n(\theta_0) > c_n(\theta_0, 1-\alpha)\right)$$

for each choice of critical value, which involves simulating data from all $F' \in \mathcal{F}_{r^*}$.[4]

Under correct specification (i.e., $r^* = 0$) and in the limit, the hypothesis tests based on PA are conservative, whereas the ones based on subsampling and GMS are size correct. Table S.I reports the empirical rejection rates under correct specification for each inferential method for a size of $\alpha = 0.10$. As expected, PA appears to be conservative and GMS appear to be size correct, whereas subsampling is unexpectedly very conservative. These differences between the rejection rates and size under correct specification will complicate the comparison across methods as we allow for misspecification. To make the results comparable, we size correct the maximum rejection probabilities so that they are all equal to the size. Finally, we also report the noncorrected results for PA, as it is actually conservative.

The results are reported in Table S.II. The results reveal that the noncorrected PA has a smaller maximum rejection probability than subsampling or GMS for all levels of misspecification. Furthermore, the corrected PA, subsampling, and GMS have very similar maximum rejection probabilities for all levels of misspecification. In particular, the finite sample rejection probabilities of GMS and subsampling are very similar, and the differences are not statistically significant given the $\mathrm{MC} = 500$ simulations. All these results are in line with Theorem 3.1. Finally, the table also illustrates that the robustness of PA

[4]The optimization problem in Eq. (S3.11) was solved numerically. The details of this computation are as follows. We selected a fine grid of 15,000 points in $\Theta = [0, 1]^2$; 10,201 of them were chosen equidistantly and the remaining ones were chosen at random using a distribution with a high concentration around F_0. For all of these points, we computed the rejection probability. Finally, to compute Eq. (S3.11) for a particular value of $r*$, we determined which subset of the 15,000 points belongs to $\mathcal{F}_{r^*}$ and reported the maximum value attained in this subset.

TABLE S.II

SIZE-CORRECTED MAXIMUM REJECTION PROBABILITY OVER MODELS WITH A MAXIMUM MISSPECIFICATION OF $r^* n^{-1/2}$[a]

$r^* n^{-1/2}$	PA noncorr.	PA	SS1	SS2	SS3	GMS1	GMS2	GMS3
0	0.07	0.10	0.10	0.10	0.10	0.10	0.10	0.10
0.005	0.08	0.12	0.11	0.15	0.35	0.12	0.12	0.12
0.010	0.11	0.14	0.14	0.16	0.37	0.14	0.14	0.14
0.015	0.16	0.18	0.21	0.21	0.37	0.19	0.19	0.19
0.020	0.17	0.23	0.23	0.25	0.37	0.25	0.25	0.25
0.025	0.24	0.30	0.31	0.30	0.38	0.30	0.30	0.30
0.030	0.28	0.35	0.35	0.34	0.38	0.35	0.35	0.35
0.035	0.39	0.43	0.44	0.42	0.38	0.43	0.43	0.43
0.040	0.40	0.49	0.48	0.48	0.38	0.52	0.52	0.52
0.045	0.52	0.61	0.62	0.61	0.40	0.62	0.62	0.62
0.050	0.63	0.68	0.69	0.69	0.49	0.69	0.69	0.69
0.055	0.71	0.76	0.77	0.77	0.60	0.78	0.78	0.78
0.060	0.75	0.81	0.81	0.81	0.64	0.82	0.82	0.82
0.065	0.88	0.91	0.91	0.91	0.75	0.91	0.91	0.91
0.070	0.91	0.93	0.95	0.93	0.81	0.93	0.93	0.93
0.075	0.93	0.96	0.96	0.94	0.86	0.96	0.96	0.96
0.080	0.97	0.98	0.98	0.98	0.93	0.99	0.99	0.99
0.085	0.98	0.99	1.00	0.99	0.96	1.00	1.00	1.00
0.090	0.99	1.00	1.00	1.00	0.97	1.00	1.00	1.00
0.095	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
0.100	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

[a]Simulation parameters: $n = 1{,}000$, MC $= 500$, $B = 200$, $b = \{2n/25, (2n/25 + n^{1/2})/2, n^{1/2}\}$, $\kappa_n = \{\ln\ln n, (\ln\ln n + \ln n)/2, \ln n\}$, and $\alpha = 0.10$.

is related to the fact that the method is conservative under correct specification.

S4. VERIFICATION OF ASSUMPTIONS IN THE EXAMPLES

S4.1. *Example S3.1*

We start by writing the example using the notation in Definition 2.1 and using the following primitive assumption. For the assumption, we use the following notation. $\Pr_n$ denotes the probability with respect to the distribution F_n, $I_l \equiv I(X = x_l)$, $p_{l,n} = \Pr_n(X = x_l)$, $\pi_{l,n} = \Pr_n(Z = 1|X = x_l)$, $E_{l,n} = E_{F_n}(Y|Z = 1, X = x_l)$, $H_{l,n} = E_{F_n}(Y^2|Z = 1, X = x_l)$, and $G_{l,n} = G(x_l, \theta_n)$.

ASSUMPTION S4.1: *Assume that, for* $c_1, c_2, c_3, c_4 \in \mathbb{R}$, (i) $p_{l,n} \geq c_1 > 0$, (ii) $H_{n,l} \leq c_2 < \infty$, $H_{n,l} - E^2_{n,l} \geq c_3 > 0$, *and* $\pi_{l,n} \geq c_4 > 0$ *for all* $n \geq 1$ *and* $l = 1, \dots, d_x$.

For simplicity, assume that $Y_L(x_l)$ and $Y_H(x_l)$ are both finite for all $l = 1, \dots, d_x$. Without loss of generality, assume that $Y_L(x_l) = 0$ and $Y_H(x_l) = 1$, so that

$$\begin{aligned}\gamma_{1,l,1,n} &\equiv \sigma_{F_n,l,1}^{-1} E_{F_n} m_{1,l}(W_i, \theta_n) \\ &= \sigma_{F_n,l,1}^{-1} E_{F_n}\big[(YZ - G(x_l, \theta_n) + 1 - Z)I_l\big] \\ &= \sigma_{F_n,l,1}^{-1}(\pi_{l,n}E_{l,n} - G_{l,n} + (1-\pi_{l,n}))p_{l,n} \geq -r_{l,1}n^{-1/2},\end{aligned} \tag{S4.1}$$

$$\begin{aligned}\gamma_{1,l,2,n} &\equiv \sigma_{F_n,l,2}^{-1} E_{F_n} m_{2,l}(W_i, \theta_n) = \sigma_{F_n,l,2}^{-1} E_{F_n}\big[(G(x_l, \theta_n) - YZ)I_l\big] \\ &= \sigma_{F_n,l,2}^{-1}(G_{l,n} - E_{l,n}\pi_{l,n})p_{l,n} \geq -r_{l,2}n^{-1/2},\end{aligned} \tag{S4.2}$$

where $\sigma_{F_n,l,j}^2 \equiv V_{F_n}(m_{j,l}(W_i, \theta_n))$, for $j = 1, 2$ and $l = 1, \dots, d_x$, is given by

$$\sigma_{F_n,l,1}^2 = p_{l,n}\pi_{l,n}[(H_{l,n} - E_{l,n}^2) + (1-\pi_{l,n})(1-E_{l,n})^2], \tag{S4.3}$$

$$\sigma_{F_n,l,2}^2 = p_{l,n}\pi_{l,n}[(H_{l,n} - E_{l,n}^2) + (1-\pi_{l,n})(E_{l,n} - 2G_{l,n})^2]. \tag{S4.4}$$

Also, for $l = 1, \dots, d_x$,

$$\begin{aligned}\rho_{12,l,n} &\equiv E_{F_n}(m_{1,l}(W_i, \theta_n)m_{2,l}(W_i, \theta_n)) \\ &= E_{F_n}[(YZ - G_{l,n} + 1 - Z)(G_{l,n} - YZ)I_l]\end{aligned} \tag{S4.5}$$

$$= (1-\pi_{l,n})p_{l,n}[G_{l,n}(1-p_{l,n}) + E_{l,n}\pi_{l,n}p_{l,n}] - \sigma_{F_n,l,2}^2. \tag{S4.6}$$

This model satisfies the relationship

$$m_{l,1}(W_i, \theta_n) + m_{l,2}(W_i, \theta_n) = (1-Z)I_l, \tag{S4.7}$$

for $l = 1, \dots, d_x$, so that

$$\gamma_{1,l,1,n} = \sigma_{F_n,l,1}^{-1}(1-\pi_{l,n})p_{l,n} - \sigma_{F_n,l,1}^{-1}\sigma_{F_n,l,2}\gamma_{1,l,2,n}. \tag{S4.8}$$

S4.1.1. *On Assumption A.5*

We begin with the case $d_x = 1$ and cover the case $d_x > 1$ afterward. By Definition A.1, $\gamma_{\omega_n,g_1,h}$ denotes a sequence of parameter vectors θ_{ω_n} and distributions F_{ω_n} for W_i such that $\omega_n^{1/2}\gamma_{1,j,\omega_n} \to h_{1,j}$ and $b_{\omega_n}^{1/2}\gamma_{1,j,\omega_n} \to g_{1,j}$ for $j \in \{1, 2\}$.

For a given $\gamma_{\omega_n,g_1,h}$, denote by J the set of $j \in \{1, 2\}$ that satisfy $h_{1,j} = \infty$ and $g_{1,j} < \infty$. By Assumption S4.1, there are constants $0 < B_1 < B_2 < \infty$ such that $\sigma_{F_n,j} \in [B_1, B_2]$ for all $j \in \{1, 2\}$ and $n \in \mathbb{N}$, which implies that $E_{F_n}m_j(W_i, \theta_n) = o(1)$ for all $j \in J$. When J is empty, there is nothing to show. We are therefore left with Cases (I) $J = \{1\}$, (II) $J = \{2\}$, and (III) $J = \{1, 2\}$. We start with the case $J = \{1\}$ and consider two subcases. In Case (a), $h_{1,2} < \infty$, while in Case (b), we have $h_{1,2} = \infty$ and $g_{1,2} = \infty$. To simplify notation, we write n rather than $\tilde{\omega}_n$ and b instead of $b_{\tilde{\omega}_n}$.

Case (I)(a). Since $h_{1,2} < \infty$, it follows by previous arguments that $E_{F_n}m_2(W_i, \theta_n) = o(1)$. By Eq. (S4.7), $(1-\pi_n) = o(1)$ and $E_n = G_n + o(1)$. It then follows that $\rho_{12,n} \to -1$. Consider an alternative sequence of parameters $\{\theta'_n, F'_n\}_{n\geq 1}$ such that $\theta'_n = \theta_n$ (so $G'_n = G_n$), $H'_n = H_n$, π'_n and E'_n given by

$$\pi'_n = \big(1 - (\ln b)b^{-1/2}\big) \to 1, \tag{S4.9}$$

$$\begin{aligned} E'_n &= \big(G_n + h_{1,2}\sigma_{F_n,2}n^{-1/2} - (\ln b)b^{-1/2}\big)\big(1 - (\ln b)b^{-1/2}\big)^{-1} \\ &= E_n + o(1). \end{aligned} \tag{S4.10}$$

This implies

$$(1 - \pi'_n) = (\ln b)b^{-1/2} = o(1), \tag{S4.11}$$

$$(G'_n - \pi'_n E'_n) = -h_{1,2}\sigma_{F_n,2}n^{-1/2} + (\ln b)b^{-1/2} = o(1), \tag{S4.12}$$

and $\sigma_{F'_n,j}\sigma^{-1}_{F_n,j} = 1 + o(1)$ for $j = 1, 2$. As a result,

$$b^{1/2}\sigma^{-1}_{F'_n,1}E_{F'_n}m_1(W_i, \theta'_n) = \sigma^{-1}_{F'_n,1}\big(-b^{1/2}\big(h_{1,2}\sigma_{F_n,1}\big)n^{-1/2} + \ln b\big) \to \infty, \tag{S4.13}$$

$$n^{1/2}\sigma^{-1}_{F'_n,2}E_{F'_n}m_2(W_i, \theta'_n) = \sigma^{-1}_{F'_n,2}\big(h_{1,2}\sigma_{F_n,1}\big) \to h_{1,2}. \tag{S4.14}$$

Finally, by $\pi'_n \to 1$ and Assumption S4.1, $\rho'_{12,n} \equiv \mathrm{Corr}_{F'_n}(m_1(W_i, \theta'_n), m_2(W_i, \theta'_n)) \to -1$.

Case (I)(b). Since $\sigma_{F_n,2} \in [B_1, B_2]$, it follows that $\lim E_{F_n}m_2(W_i, \theta_n) \in [0, \infty]$. In this case, $\lim(1 - \pi_n) \in [0, 1]$. Consider an alternative sequence of parameters $\{\theta'_n, F'_n\}_{n\geq 1}$ such that $\theta'_n = \theta_n$ (so $G'_n = G_n$), $H'_n = H_n$, π'_n and E'_n given by

$$\pi'_n = \pi_n - 2(\ln b)b^{-1/2} = \pi_n + o(1), \tag{S4.15}$$

$$E'_n = \big(\pi_n E_n - (\ln b)b^{-1/2}\big)\big(\pi_n - 2(\ln b)b^{-1/2}\big)^{-1} = E_n + o(1), \tag{S4.16}$$

where we used $\pi_n \geq c_4 > 0$. This implies $\sigma_{F'_n,j}\sigma^{-1}_{F_n,j} = 1 + o(1)$ for $j = 1, 2$. It then follows that

$$\begin{aligned} &b^{1/2}\sigma^{-1}_{F'_n,1}E_{F'_n}m_1(W_i, \theta'_n) \\ &\quad = b^{1/2}\sigma^{-1}_{F_n,1}E_{F_n}m_1(W_i, \theta_n) + \sigma^{-1}_{F'_n,1}\ln b + o(1) \to \infty, \end{aligned} \tag{S4.17}$$

$$\begin{aligned} &b^{1/2}\sigma^{-1}_{F'_n,2}E_{F'_n}m_2(W_i, \theta'_n) \\ &\quad = b^{1/2}\sigma^{-1}_{F_n,2}E_{F_n}m_2(W_i, \theta_n) + \sigma^{-1}_{F'_n,2}\ln b + o(1) \to \infty. \end{aligned} \tag{S4.18}$$

Finally, Assumption S4.1 and Eq. (S4.6) imply

$$\rho'_{12,n} \equiv \mathrm{Corr}_{F'_n}(m_1(W_i, \theta'_n), m_2(W_i, \theta'_n)) = \rho_{12,n} + o(1). \tag{S4.19}$$

Case (II). This case is analogous to Case (I) and is therefore omitted.

Case (III). By Eq. (S4.7), $(1-\pi_n)=o(1)$ and $E_n=G_n+o(1)$. As a consequence of this and Assumption S4.1, it follows that $\rho_{12,n}\to -1$. Consider an alternative sequence of parameters $\{\theta'_n,F'_n\}_{n\geq 1}$ such that $\theta'_n=\theta_n$ (so $G'_n=G_n$), $H'_n=H_n$, π'_n and E'_n given by Eqs. (S4.15) and (S4.16). Then, Eqs. (S4.17)–(S4.19) follow, and this concludes the proof for the case $d_x=1$.

Now consider the case $d_x>1$. Notice that in the case with $d_x=1$, we considered a sequence of parameters $\{\theta'_n,F'_n\}_{n\geq 1}$ such that $\theta'_n=\theta_n$ and

$$\sigma^{-1}_{F'_n,1}E_{F'_n}m_1(W_i,\theta'_n)=\sigma^{-1}_{F_n,1}E_{F_n}m_1(W_i,\theta_n)+o(1), \tag{S4.20}$$

$$\sigma^{-1}_{F'_n,2}E_{F'_n}m_2(W_i,\theta'_n)=\sigma^{-1}_{F_n,2}E_{F_n}m_2(W_i,\theta_n)+o(1), \tag{S4.21}$$

$$\begin{aligned}&\lim \mathrm{Corr}_{F'_n}(m_1(W_i,\theta'_n),m_2(W_i,\theta'_n))\\&\quad=\lim \mathrm{Corr}_{F_n}(m_1(W_i,\theta_n),m_2(W_i,\theta_n)).\end{aligned} \tag{S4.22}$$

When $d_x>1$, we consider an alternative sequence of parameters $\{\theta'_n,F'_n\}_{n\geq 1}$ such that, for each $l=1,\dots,d_x$, we set $p'_{l,n}=p_{l,n}$, $\theta'_n=\theta_n$ (so $G'_{l,n}=G_{l,n}$), and the rest of the choices of the alternative distribution would be chosen according to the corresponding case in the previous part. According to this, it follows that, for every $l=1,\dots,d_x$,

$$\sigma^{-1}_{F'_n,1,l}E_{F'_n}m_{1,l}(W_i,\theta'_n)=\sigma^{-1}_{F_n,1,l}E_{F_n}m_{1,l}(W_i,\theta_n)+o(1), \tag{S4.23}$$

$$\sigma^{-1}_{F'_n,2,l}E_{F'_n}m_{2,l}(W_i,\theta'_n)=\sigma^{-1}_{F_n,2,l}E_{F_n}m_{2,l}(W_i,\theta_n)+o(1), \tag{S4.24}$$

$$\begin{aligned}&\lim \mathrm{Corr}_{F'_n}(m_{1,l}(W_i,\theta'_n),m_{2,l}(W_i,\theta'_n))\\&\quad=\lim \mathrm{Corr}_{F_n}(m_{1,l}(W_i,\theta_n),m_{2,l}(W_i,\theta_n)).\end{aligned} \tag{S4.25}$$

To conclude the proof, we notice that, for $l_1,l_2=1,\dots,d_x$ with $l_1\neq l_2$ and $a_1,a_2\in\{1,2\}$,

$$\begin{aligned}&\mathrm{Corr}_{F'_n}\big(m_{a_1,l_1}(W_i,\theta'_n),m_{a_2,l_2}(W_i,\theta'_n)\big)\\&\quad=-\sigma^{-1}_{F'_n,a_1,l_1}E_{F'_n}m_{a_1,l_1}(W_i,\theta'_n)\sigma^{-1}_{F'_n,a_2,l_2}E_{F'_n}m_{a_2,l_2}(W_i,\theta'_n)+o(1)\\&\quad=-\sigma^{-1}_{F_n,a_1,l_1}E_{F_n}m_{a_1,l_1}(W_i,\theta_n)\sigma^{-1}_{F_n,a_2,l_2}E_{F_n}m_{a_2,l_2}(W_i,\theta_n)+o(1)\\&\quad=\mathrm{Corr}_{F_n}\big(m_{a_1,l_1}(W_i,\theta_n),m_{a_2,l_2}(W_i,\theta_n)\big)+o(1).\end{aligned} \tag{S4.26}$$

S4.1.2. *On Assumption A.6*

We verify Assumption A.6 for $r^*>0$. For simplicity, consider the case $d_x=1$. Choose a sequence of parameters $\{\theta_n,F_n\}_{n\geq 1}$ with $1-\pi_n=o(1)$ and limiting parameter $h^*_{1,1}<0$. By Eq. (S4.8), $h^*_{1,2}=-h^*_{1,1}>0$ and h^*_2 is a 2×2 matrix equal to $[1,-1;-1,1]$.

First, consider the test function S_1. Let $c_0(h^*_2,1-\alpha)$ be the $1-\alpha$ quantile of

$$S_1\big(Z_{h^*_2},h^*_2\big)=[Z_1]^2_-+[-Z_1]^2_-=Z^2_1,\quad Z_{h^*_2}=(Z_1,Z_2)\sim N(0,h^*_2). \tag{S4.27}$$

Note that

$$S_1\big(Z_{h_2^*}+h_1^*,h_2^*\big)=[Z_1+h_{1,1}^*]_-^2+[-Z_1-h_{1,1}^*]_-^2=(Z_1+h_{1,1}^*)^2, \tag{S4.28}$$

and since $\Pr((Z_1+h_{1,1}^*)^2\le x)<\Pr(Z_1^2\le x)$ for $h_{1,1}^*<0$, $\Pr((Z_1+h_{1,1}^*)^2\le c_0(h_2^*,1-\alpha))<1-\alpha$. Assumption A.6 then holds.

Second, consider the test function S_2. We consider the version of S_2 in Eq. (3.5), as here the limit correlation matrix is singular (i.e., the test function should be defined on Ψ_1). Using the definition of $\tilde{\Sigma}_\varepsilon$ in Eq. (3.15), it follows that

$$\tilde{\Omega}_\varepsilon^*=\begin{bmatrix}1+\varepsilon & -1\\ -1 & 1+\varepsilon\end{bmatrix}\quad\text{and}\quad \tilde{\Omega}_\varepsilon^{*,-1}=a(\varepsilon)\begin{bmatrix}1+\varepsilon & 1\\ 1 & 1+\varepsilon\end{bmatrix}, \tag{S4.29}$$

where $a(\varepsilon)=[(1+\varepsilon)^2-1]^{-1}$. As a result,

$$\begin{aligned}&\tilde{S}_2\big(Z_{h_2^*},\tilde{\Omega}_\varepsilon^*\big)\\&\quad=(1+\varepsilon)a(\varepsilon)\\&\qquad\times\inf_{t_1\ge0,t_2\ge0}\{(Z_1-t_1)^2+(Z_2-t_2)^2+2(1+\varepsilon)^{-1}(Z_1-t_1)(Z_2-t_2)\}\\&\quad=(1+\varepsilon)a(\varepsilon)\\&\qquad\times\inf_{t_1\ge0,t_2\ge0}\{(Z_1-t_1)^2+(Z_1+t_2)^2-2(1+\varepsilon)^{-1}(Z_1-t_1)(Z_1+t_2)\}\\&\quad=\frac{1}{1+\varepsilon}Z_1^2,\end{aligned} \tag{S4.30}$$

where the first equality holds by definition, the second equality holds because $Z_{h_2^*}$ is such that $Z_2=-Z_1$, and the third equality holds by solving the optimization. Since $Z_2+h_{1,2}^*=-Z_1-h_{1,1}^*$ we also have

$$\tilde{S}_2\big(Z_{h_2^*}+h_1^*,\tilde{\Omega}_\varepsilon^*\big)=\frac{1}{1+\varepsilon}(Z_1+h_{1,1}^*)^2, \tag{S4.31}$$

where $h_{1,1}^*<0$. Let $c_0(h_2^*,1-\alpha)$ be the $1-\alpha$ quantile of $\tilde{S}_2(Z,\tilde{\Omega}_\varepsilon^*)$ in Eq. (S4.30). Since $\Pr((Z_1+h_{1,1}^*)^2\le x)<\Pr(Z_1^2\le x)$ for $h_{1,1}^*<0$, $\Pr((1+\varepsilon)^{-1}(Z_1+h_{1,1}^*)^2\le c_0(h_2^*,1-\alpha))<1-\alpha$. Assumption A.6 then holds. The general case where $d_x>1$ follows by applying the previous argument to each pair of moment inequalities.

S4.2. *Example 2.1*

We start again by writing the example using the notation in Definition 2.1. For simplicity of the argument, we assume that the distribution G is uniform, as stated below.

ASSUMPTION S4.2: *Under the distribution* G, $u_i = (u_{1,i}, u_{2,i})$ *is uniformly distributed on* $[0, 1]^2$.

Let $\theta_n = (\theta_{1,n}, \theta_{2,n})$ be the true parameter vector, $\Pr_n(\cdot)$ be the probability with respect to the distribution F_n of W_i, $p_{rs,n} \equiv \Pr_n(W_{1,i} = r, W_{2,i} = s)$ for $r, s \in \{0, 1\}$, and $\rho_{jj',n} \equiv \mathrm{Corr}_{F_n}[m_j(W_i, \theta_n), m_{j'}(W_i, \theta_n)]$ for $j, j' \in \{1, 2, 3\}$. As defined in the text, G_n denotes the true distribution of u_i for sample size n. Under Assumption S4.2, we have

$$
\begin{aligned}
\text{(S4.32)} \quad \gamma_{1,1,n} &\equiv \sigma_{F_n,1}^{-1} E_{F_n}[G_1(\theta_n) - W_{1,i}(1 - W_{2,i})] = \sigma_{F_n,1}^{-1}(\theta_{2,n} - p_{10,n}), \\
\gamma_{1,2,n} &\equiv \sigma_{F_n,2}^{-1} E_{F_n}[W_{1,i}(1 - W_{2,i}) - G_2(\theta_n)] \\
&= \sigma_{F_n,2}^{-1}(p_{10,n} - (1 - \theta_{1,n})\theta_{2,n}), \\
\gamma_{1,3,n} &\equiv \sigma_{F_n,3}^{-1} E_{F_n}[W_{1,i}W_{2,i} - G_3(\theta_n)] \\
&= \sigma_{F_n,3}^{-1}(p_{11,n} - (1 - \theta_{1,n})(1 - \theta_{2,n})).
\end{aligned}
$$

By simple calculations, we have

$$
\begin{aligned}
\text{(S4.33)} \quad & \sigma_{F_n,1}^2 = \sigma_{F_n,2}^2 = \mathrm{Var}_{F_n}[m_1(W_i, \theta_n)] = p_{10,n}(1 - p_{10,n}) \in (0, 1/4], \\
& \sigma_{F_n,3}^2 = \mathrm{Var}_{F_n}[m_3(W_i, \theta_n)] = p_{11,n}(1 - p_{11,n}) \in (0, 1/4], \\
& \rho_{12,n} = -1, \quad \rho_{13,n} = \frac{p_{10,n} p_{11,n}}{\sigma_{F_n,1}\sigma_{F_n,3}}, \quad \text{and} \quad \rho_{23,n} = -\rho_{13,n},
\end{aligned}
$$

where zero variances have been ruled out by Definition 2.1(iv). By Definition A.1, $\gamma_{\omega_n, g_1, h}$ denotes a sequence of parameter vectors θ_{ω_n} and distributions F_{ω_n} for W_i such that $\omega_n^{1/2}\gamma_{1,j,\omega_n} \to h_{1,j}$ and $b_{\omega_n}^{1/2}\gamma_{1,j,\omega_n} \to g_{1,j}$, for $j \in \{1, 2, 3\}$. Recall that $\gamma_{\omega_n, g_1, h}$ defines $\theta_{\omega_n} = (\theta_{1,\omega_n}, \theta_{2,\omega_n})$ and thus defines $G_1(\theta_{\omega_n})$, $G_2(\theta_{\omega_n})$, and $G_3(\theta_{\omega_n})$.

S4.2.1. *On Assumption A.5*

For a given $\gamma_{\omega_n, g_1, h}$, denote by J the set of $j \in \{1, 2\}$ that satisfy $h_{1,j} = \infty$ and $g_{1,j} < \infty$. When J is empty, there is nothing to show. We are therefore left with the cases $J = \{1\}$, $J = \{2\}$, and $J = \{1, 2\}$. We start with the case $J = \{1\}$ and consider two subcases. In Case (I), $h_{1,2} < \infty$, while in Case (II), we have $h_{1,2} = \infty$ and $g_{1,2} = \infty$. For each subcase, we consider two further subcases: in Case (a), $\rho_{13,n} \to 0$, while in Case (b), $\rho_{13,n} \to \rho_{13} \in (0, 1]$. To simplify notation, we write n rather than $\tilde{\omega}_n$ and b instead of $b_{\tilde{\omega}_n}$.

REMARK S4.1: Note that, for any positive numbers a_{10}, a_{01}, a_{11} whose sum equals 1 and $\theta = (\theta_1, \theta_2) \in (0, 1)^2$, there exists a random variable $u_i = (u_{1,i}, u_{2,i})$ on $[0, 1]^2$ with continuous distribution such that $\Pr(u_{1,i} > \theta_1$ and $u_{2,i} < \theta_2) = a_{10}$, $\Pr(u_{1,i} < \theta_1$ and $u_{2,i} > \theta_2) = a_{01}$, and $\Pr(u_{1,i} > \theta_1$ and $u_{2,i} > \theta_2) = a_{11}$ (and consequently $\Pr(u_{1,i} < \theta_1$ and $u_{2,i} < \theta_2) = 0$).

Letting a_{10}, a_{01}, a_{11} play the role of $p_{10,n}$, $p_{01,n}$, $p_{11,n}$, Remark S4.1 implies that, for a given vector $\theta_n = (\theta_{1,n}, \theta_{2,n})$, any desired outcome probabilities $p_{10,n}$, $p_{01,n}$, $p_{11,n}$ can be generated by a random variable $u_i = (u_{1,i}, u_{2,i})$ that has a continuous distribution G_n.

Case (I)(a). We have to produce a sequence $\gamma_{n,\tilde{g}_1,h}$ for which $\tilde{g}_{1,1} = \infty$, $h_{1,2}$ is a specific finite number, and the upper right element of h_2 equals 0. Define

$$p'_{10,n} = b^{-3/7}. \tag{S4.34}$$

Let $\theta'_n = (\theta'_{1,n}, \theta'_{2,n})$ for $\theta'_{1,n}$ and $\theta'_{2,n}$ defined next. Pick $\theta'_{2,n} \in (0,1)$ such that

$$G_1(\theta'_n) = \theta'_{2,n} = p'_{10,n} + b^{-1/2}(p'_{10,n}(1 - p'_{10,n}))^{1/2} b^{2/7}, \tag{S4.35}$$

and pick $\theta'_{1,n} \in (0,1)$ such that

$$G_2(\theta'_n) = (1 - \theta'_{1,n})\theta'_{2,n} = p'_{10,n} - n^{-1/2}(p'_{10,n}(1 - p'_{10,n}))^{1/2} h_{1,2}. \tag{S4.36}$$

This is clearly possible because $p'_{10,n} \to 0$, $|n^{-1/2}(p'_{10,n}(1 - p'_{10,n}))^{1/2} h_{1,2}| < p'_{10,n}$, and $b^{-1/2}(p'_{10,n}(1 - p'_{10,n}))^{1/2} b^{2/7} \to 0$. We have $G_1(\theta'_n) = \theta'_{2,n} = 2b^{-3/7}(1 + o(1))$, $G_2(\theta'_n) = b^{-3/7}(1 + o(1))$. Now

$$b^{-3/7}(1 + o(1)) = G_2(\theta'_n) = (1 - \theta'_{1,n})\theta'_{2,n} = 2(1 - \theta'_{1,n}) b^{-3/7}(1 + o(1)), \tag{S4.37}$$

which implies that $\theta'_{1,n}$ cannot converge to 1. Without loss of generality, we can therefore assume that $\theta'_{1,n} \to \theta'_1$ for some $\theta'_1 \in [0,1)$. We then have

$$G_3(\theta'_n) = (1 - \theta'_{1,n})(1 - \theta'_{2,n}) \to (1 - \theta'_1). \tag{S4.38}$$

Consider the function

$$f(x) \equiv x - h_{1,3} n^{-1/2}(x(1 - x))^{1/2} \tag{S4.39}$$

for $x \in [0,1]$. The function f is continuous and satisfies $f(0) = 0$ and $f(1) = 1$. Therefore, for given $G_3(\theta'_n)$, the intermediate value theorem implies that there exists a value $p'_{11,n}$ such that

$$G_3(\theta'_n) = p'_{11,n} - h_{1,3} n^{-1/2}(p'_{11,n}(1 - p'_{11,n}))^{1/2}. \tag{S4.40}$$

Define $p'_{11,n}$ to be any value in $(0,1)$ that satisfies Eq. (S4.40). It cannot be the case that $p'_{11,n} \to 1$, as otherwise we would have $G_3(\theta'_n) \to 1$, contradicting Eq. (S4.38). Therefore, without loss of generality, $p'_{11,n} \to p'_{11}$ for some $p'_{11} \in [0,1)$. Note that $p'_{10,n} \to 0$ and $p'_{11} \in [0,1)$ imply that

$$\rho'_{13,n} = \frac{p'_{10,n} p'_{11,n}}{\sigma'_{F_n,1} \sigma'_{F_n,3}} \to 0. \tag{S4.41}$$

For these given choices of $p'_{10,n}$, $p'_{11,n}$, and $\theta'_n = (\theta'_{1,n}, \theta'_{2,n})$, Remark S4.1 implies that there exists a continuous distribution G'_n for the random variable $u_i = (u_{1,i}, u_{2,i})$ such that $\Pr(u_{1,i} > \theta'_{1,n}$ and $u_{2,i} < \theta'_{2,n}) = p'_{10,n}$, $\Pr(u_{1,i} < \theta'_{1,n}$ and $u_{2,i} > \theta'_{2,n}) = 1 - p'_{10,n} - p'_{11,n}$, and $\Pr(u_{1,i} > \theta'_{1,n}$ and $u_{2,i} > \theta'_{2,n}) = p'_{11,n}$. By construction, all requirements are fulfilled under the sequence θ'_n and G'_n.

Case (I)(b). We have to produce a sequence $\gamma_{n,\tilde{g}_1,h}$ for which $\tilde{g}_{1,1} = \infty$, $h_{1,2} < \infty$, and the upper right element of h_2 equals $\rho_{13} \in (0, 1]$. Assume first that $\rho_{13} \in (0, 1)$. Define

$$p'_{10,n} = cb^{-1/7} \tag{S4.42}$$

for some constant $c > 0$, and define $(\theta'_{1,n}, \theta'_{2,n}) \in (0, 1)^2$ as in Eqs. (S4.35) and (S4.36). We then have $\theta'_{2,n} = cb^{-1/7} + c^{1/2}b^{-2/7}(1 + o(1))$ and $(1 - \theta'_{1,n})\theta'_{2,n} = cb^{-1/7} + o(b^{-2/7})$ and thus

$$\theta'_{1,n} = 1 - \frac{cb^{-1/7} + o(b^{-2/7})}{cb^{-1/7} + c^{1/2}b^{-2/7}(1 + o(1))} = c^{-1/2}b^{-1/7}(1 + o(1)). \tag{S4.43}$$

Next,

$$\begin{aligned} G_3(\theta'_n) &= (1 - \theta'_{1,n})(1 - \theta'_{2,n}) \\ &= \bigl(1 - c^{-1/2}b^{-1/7}(1 + o(1))\bigr)\bigl(1 - cb^{-1/7}(1 + o(1))\bigr) \\ &= 1 - \bigl(c^{-1/2} + c\bigr)b^{-1/7}(1 + o(1)). \end{aligned} \tag{S4.44}$$

Arguing as in Case (I)(a), there is a value $p'_{11,n} \in (0, 1)$ such that

$$G_3(\theta'_n) = p'_{11,n} - h_{1,3}n^{-1/2}(p'_{11,n}(1 - p'_{11,n}))^{1/2}. \tag{S4.45}$$

As $G_3(\theta'_n) \to 1$, we have $p'_{11,n} \to 1$. More precisely,

$$\begin{aligned} p'_{11,n} &= G_3(\theta'_n) + h_{1,3}n^{-1/2}(p'_{11,n}(1 - p'_{11,n}))^{1/2} \\ &= 1 - \bigl(c^{-1/2} + c\bigr)b^{-1/7}(1 + o(1)). \end{aligned} \tag{S4.46}$$

Therefore,

$$\begin{aligned} \rho'_{13,n} &\equiv \left(\frac{p'_{10,n}p'_{11,n}}{(1 - p'_{10,n})(1 - p'_{11,n})}\right)^{1/2} \\ &= \left(\frac{cb^{-1/7}(1 - (c^{-1/2} + c)b^{-1/7})}{(1 - cb^{-1/7})((c^{-1/2} + c)b^{-1/7})}\right)^{1/2}(1 + o(1)) \\ &\to \bigl(c/\bigl(c^{-1/2} + c\bigr)\bigr)^{1/2}. \end{aligned} \tag{S4.47}$$

The function $(c/(c^{-1/2}+c))^{1/2}$ is continuous for $c>0$ and converges to 1 as $c\to\infty$ and to 0 as $c\to 0$. There is therefore $c>0$ such that $(c/(c^{-1/2}+c))^{1/2}=\rho_{13}$. The proof is then concluded as in Case (I)(a). If $\rho_{13}=1$, the same proof applies once the constant c in Eq. (S4.42) is replaced by the sequence $c_n=\ln b$ that slowly converges to infinity.

Case (II)(a). We have to produce a sequence $\gamma_{n,\tilde{g}_1,h}$ for which $\tilde{g}_{1,1}=\tilde{g}_{1,2}=\infty$ and the upper right element of h_2 equals zero. Define

$$p'_{10,n}=b^{-3/7}. \tag{S4.48}$$

Let $\theta'_n=(\theta'_{1,n},\theta'_{2,n})\in(0,1)^2$ be defined as follows. Let $\theta'_{2,n}\in(0,1)$ be such that

$$G_1(\theta'_n)=\theta'_{2,n}=p'_{10,n}+b^{-1/2}(p'_{10,n}(1-p'_{10,n}))^{1/2}b^{2/7}, \tag{S4.49}$$

and pick $\theta'_{1,n}\in(0,1)$ such that

$$G_2(\theta'_n)=(1-\theta'_{1,n})\theta'_{2,n}=p'_{10,n}-b^{-1/2}(p'_{10,n}(1-p'_{10,n}))^{1/2}b^{1/7}. \tag{S4.50}$$

As in Case (I)(a), we have $G_1(\theta'_n)=2b^{-3/7}(1+o(1))$ and $G_2(\theta'_n)=b^{-3/7}(1+o(1))$. Using the same steps as in Case (I)(a), we have $\theta'_{1,n}\to\theta'_1$ for some $\theta'_1\in[0,1)$ and thus that $G_3(\theta'_n)$ converges to a number smaller than 1. Then again, there exists $p'_{11,n}$ such that $G_3(\theta'_n)=p'_{11,n}-h_{1,3}n^{-1/2}(p'_{11,n}(1-p'_{11,n}))^{1/2}$ and $p'_{11,n}\to p'_{11}$ for some $p'_{11}\in[0,1)$. Therefore, we have again that $\rho'_{13,n}\to 0$ and the proof concludes as in Case (I)(a).

Case (II)(b). We have to produce a sequence $\gamma_{n,\tilde{g}_1,h}$ for which $\tilde{g}_{1,1}=\tilde{g}_{1,2}=\infty$ and the upper right element of h_2 equals $\rho_{13}\in(0,1]$. The proof follows along the same lines as Case (I)(b) with the one difference that $G_2(\theta'_n)$ is defined as in Eq. (S4.50).

That concludes the verification of the assumption for the case $J=\{1\}$. Regarding the other cases, note that the case $J=\{1,2\}$ is covered by Cases (II)(a) and (II)(b) above. The verification of the assumption in case $J=\{2\}$ is also partially covered by Cases (II)(a) and (II)(b), and the remaining cases for $J=\{2\}$ are similar to Cases (I)(a) and (I)(b) above for $J=\{1\}$, and therefore omitted.

S4.2.2. *On Assumption A.6*

The verification of Assumption A.6 follows almost identical steps to those used in Section S4.1.2 and is therefore omitted.

S5. ADDITIONAL DISCUSSION ON THEOREM 3.2

To understand the intuition behind Theorem 3.2, it is enough to consider the case with two moment inequalities, $p=k=2$, together with the limit of the PA critical value. In this case it follows from Lemma S1.1 that

$$\mathrm{AsySz}^{(1)}_{\mathrm{PA}}\le\Pr([Z^\star_1-r_1]^2_-+[-Z^\star_1]^2_-\le c_0(\Omega,1-\alpha)), \tag{S5.1}$$

where $Z^\star \sim N(0, \Omega)$ and $\Omega \in \Psi_1$ is a correlation matrix with off-diagonal elements $\rho = -1$. Theorem 3.2 shows that $\text{AsySz}^{(1)}_{\text{PA}}$ is strictly positive provided the amount of misspecification is not too big, that is, $r^* \leq B$. The reason why some condition on r^* must be placed is evident: if the amount of misspecification is really big, there is no way to bound the asymptotic distortion. To illustrate this, suppose $r_1 > (2c_0(\Omega, 1-\alpha))^{1/2}$ and let $A \equiv [Z^\star_1 - r_1]^2_-$ and $B \equiv [-Z^\star_1]^2_-$ so that the RHS of Eq. (S5.1) is $\Pr(A + B \leq c_0(\Omega, 1-\alpha))$. On the one hand, if $Z^\star_1 \notin [0, r_1]$, it follows that either $B = 0$ and $A > c_0(\Omega, 1-\alpha)$ or $A = 0$ and $B > c_0(\Omega, 1-\alpha)$. On the other hand, if $Z^\star_1 \in [0, r_1]$, $A + B = (Z^\star_1 - r_1)^2 + Z^{\star 2}_1 \geq r_1^2/2 > c_0(\Omega, 1-\alpha)$. We can then conclude that

$$\Pr([Z^\star_1 - r_1]^2_- + [-Z^\star_1]^2_- \leq c_0(\Omega, 1-\alpha)) = 0, \tag{S5.2}$$

meaning that $\text{AsySz}^{(1)}_{\text{PA}} = 0$ when $r^* > (2c_0(\Omega, 1-\alpha))^{1/2}$. For this level of r^*, $\text{AsySz}^{(2)}_{\text{PA}} = 0$ as well, so both test statistics suffer from the maximum amount of distortion. Therefore, to get nontrivial results, we must restrict the magnitude of r^* as in Theorem 3.2.

In addition, Theorem 3.2 shows that $\text{AsySz}^{(2)}_{\text{PA}}$ can be arbitrarily close to zero when ε in the space $\Psi_{2,\varepsilon}$ is small. What drives this result is the possibility that at least two inequalities are violated (or one is violated and the other one is binding) and strongly negatively correlated. To illustrate this, consider again the case where $p = k = 2$ together with the limit of the PA critical value. By $\Omega \in \Psi_{2,\varepsilon}$, the off-diagonal element ρ of the correlation matrix Ω has to lie in $[-(1-\varepsilon)^{1/2}, (1-\varepsilon)^{1/2}]$. It follows from Lemma S1.1 that

$$\text{AsySz}^{(2)}_{\text{PA}} \leq \Pr(S_2(Z^\star, r_1, \Omega_\varepsilon) \leq c_0(\Omega_\varepsilon, 1-\alpha)), \tag{S5.3}$$

where $Z^\star \sim N(0, \Omega_\varepsilon)$, Ω_ε is a matrix with $\rho = -(1-\varepsilon)^{1/2}$, and

$$\begin{aligned} S_2(Z^\star, r_1, \Omega_\varepsilon) = \frac{1}{\varepsilon} \inf_{t \in \mathbb{R}^2_{+,+\infty}} \Bigg\{ &\sum_{j=1}^{2} (Z^\star_j - r_1 - t_j)^2 \\ &+ 2(1-\varepsilon)^{1/2}(Z^\star_1 - r_1 - t_1)(Z^\star_2 - r_1 - t_2) \Bigg\}. \end{aligned} \tag{S5.4}$$

The solution to the above optimization problem can be divided in four cases (see Lemma S1.2 for details), depending on the value of the realizations $(Z^\star_1, Z^\star_2)$. However, there exists a set $A \subset \mathbb{R}^2$ such that, for all $(z_1, z_2) \in A$,

$$S_2(z, r_1, \Omega_\varepsilon) \geq S_2(z, 0, \Omega_\varepsilon) + \frac{2}{\varepsilon}[r_1^2 - z_1 - z_2], \tag{S5.5}$$

with $[r_1^2 - z_1 - z_2] > 0$, and $\Pr(Z^\star \in A) \to 1$ as $\varepsilon \to 0$. It is immediate from Eq. (S5.5) that small distortions $r_1 > 0$ can produce a value of $S_2(Z^\star, r_1, \Omega_\varepsilon)$

that is arbitrarily high on the set A by allowing ε to be arbitrarily close to zero, that is, correlation close to -1. Since $c_0(\Omega_\varepsilon, 1-\alpha)$ can be shown to be bounded in $\Psi_{2,\varepsilon}$, it follows that

$$\Pr(S_2(Z^\star, r_1, \Omega_\varepsilon) \le c_0(\Omega_\varepsilon, 1-\alpha)|A) \to 0, \tag{S5.6}$$

as $\varepsilon \to 0$. Therefore, Eq. (S5.3) implies that CSs based on S_2 have asymptotic confidence size arbitrarily close to zero when ε is small.

S6. DETAILS OF THE NUMERICAL COMPUTATIONS IN TABLE I

Table I reports a numerical approximation to the AsySz of the CSs based on S_1 and S_2 (with a PA critical value) using the formula in Lemma B.1, that is,

$$\text{AsySz}_{\text{PA}} = \inf_{h=(h_1,h_2)\in H} \Pr(S(Z_h + h_1, h_2) \le c_0(h_2, 1-\alpha)), \tag{S6.1}$$

where $Z_h \sim N(0_p, h_2)$. Table I reports the cases where $p \in \{2, 4, 8, 10\}$, $k = p$, $\varepsilon \in \{0.10, 0.05\}$, and $r^* \in \{0.25, 0.50, 1.00\}$. Therefore, the parameter space H is given by all the p vectors h_1 such that $h_{1,j} \ge -r^*$ (i.e., $r_j = r^*$) for all $j = 1, \dots, p$, and $h_2 \in \Psi_1$ (or $\Psi_{2,\epsilon}$) for S_1 (or S_2).

Having defined the parameter space H, we now describe how we compute the minimizer in Eq. (S6.1). By the nature of the parameter space H, the problem can be broken down into finding the worst subvectors h_1 and h_2. In the case of h_1, it is not hard to show that the worst subvector is given by $h_{1,j} = -r^*$ (i.e., all the inequalities are as violated as possible). The case of h_2 is not as simple, as it is not clear which correlation matrix is the worst case correlation when $p > 2$. In the results presented in Table I we use the following three types of correlation matrix structures:

(a) Matrices h_2 of the form

$$h_2 = \begin{bmatrix} A & 0_{2\times(p-2)} \\ 0_{(p-2)\times 2} & I_{p-2} \end{bmatrix}, \quad \text{where} \quad A = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \tag{S6.2}$$

and $0_{l,s}$ denotes an $l \times s$ matrix of zeros. This is the structure used in the proof of Theorem 3.2. For the test function S_2, we set $\rho = -\sqrt{1-\epsilon}$, as we know this is the worse correlation structure for this test function. For the test function S_1, we consider a grid of 20 points in $[-1, 1]$ for ρ and compute Eq. (S6.1) taking the minimum on this grid. This method generates one probability for S_1 and another one for S_2.

(b) Matrices h_2 of the form

$$h_2 = \begin{bmatrix} A & 0 & \cdots & 0 \\ 0 & A & \cdots & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & A \end{bmatrix}. \tag{S6.3}$$

For the test function S_2, we set $\rho = -\sqrt{1-\epsilon^{2/p}}$, as we know this is the worse correlation structure for this test function. For the test function S_1, we consider a grid of 20 points in $[-1, 1]$ for ρ and compute Eq. (S6.1) taking the minimum on this grid. This method again generates one probability for S_1 and another one for S_2.

(c) Matrices h_2 equal to a random matrix generated by the method of Marsaglia and Olkin (1984). We generate 15,000 random correlation matrices from Ψ_1 for the test function S_1, and for the test function S_2 we adjust the correlation matrix so as to satisfy the restriction $\det(h_2) \geq \epsilon$ if the randomly generated matrix has a determinant below ϵ. We then compute Eq. (S6.1) taking the minimum of the 15,000 randomly generated matrices. Once more, the end result is one probability for S_1 and another one for S_2.

What Table I presents is the minimum value, for each test function S_1 and S_2, of the three probabilities derived in each of the above cases. Note that for each matrix h_2 in cases (a), (b), and (c) above (e.g., for each of the 15,000 in case (c)), we approximate the probability in Eq. (S6.1) by generating 2,000 random variables with distribution $N(0_p, h_2)$.

REFERENCES

ANDREWS, D. W. K., AND P. GUGGENBERGER (2009): "Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669–709. [2]

MARSAGLIA, G., AND I. OLKIN (1984): "Generating Correlation Matrices," *SIAM Journal on Scientific Computing*, 5, 470–475. [22]

Dept. of Economics, Duke University, Durham, NC 27708, U.S.A.; federico.bugni@duke.edu,

Dept. of Economics, Northwestern University, Evanston, IL 60208, U.S.A.; iacanay@northwestern.edu,

and

Dept. of Economics, University of California, San Diego, La Jolla, CA 92093, U.S.A.; pguggenberger@ucsd.edu.

Manuscript received October, 2010; final revision received June, 2011.