Econometrica Supplementary Material

SUPPLEMENT TO "POOR (WO)MAN'S BOOTSTRAP"

(*Econometrica*, Vol. 85, No. 4, July 2017, 1277–1301)

BO E. HONORÉ
Department of Economics, Princeton University

LUOJIA HU
Economic Research Department, Federal Reserve Bank of Chicago

NOTE THAT THE NUMBERING OF SECTIONS, theorems, etc. is as in the paper.

S-1. IDENTIFICATION OF THE VARIANCE OF TWO-STEP ESTIMATORS

Consider the two-step estimation problem in equation (8) in Section 5. As mentioned, the asymptotic variance of $\widehat{\theta}_2$ is

$$R_2^{-1}R_1Q_1^{-1}V_{11}Q_1^{-1}R_1'R_2^{-1} - R_2^{-1}V_{21}Q_1^{-1}R_1'R_2^{-1} - R_2^{-1}R_1Q_1^{-1}V_{12}R_2^{-1} + R_2^{-1}V_{22}R_2^{-1},$$

where $\left(\begin{smallmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{smallmatrix}\right) = \operatorname{var}\left(\left(\begin{smallmatrix} q(z_i,\theta_1) \\ r(z_i,\theta_1,\theta_2) \end{smallmatrix}\right)\right)$, $Q_1 = E[\frac{\partial q(z_i,\theta_1)}{\partial \theta_1}]$, $R_1 = E[\frac{\partial r(z_i,\theta_1,\theta_2)}{\partial \theta_1}]$, and $R_2 = E[\frac{\partial r(z_i,\theta_1,\theta_2)}{\partial \theta_2}]$. It is often easy to estimate V_{11}, V_{22}, Q_1, and R_2 directly. When it is not, they can be estimated using the poor woman's bootstrap procedure above. We therefore focus on V_{12} and R_1.

Consider one-dimensional estimators of the form

$$\widehat{a}_1(\delta_1) = \arg\min_{a_1} \frac{1}{n}\sum Q(z_i, \theta_1 + a_1\delta_1),$$

$$\widehat{a}_2(\delta_1, \delta_2) = \arg\min_{a_2} \frac{1}{n}\sum R(z_i, \theta_1 + \widehat{a}_1\delta_1, \theta_2 + a_2\delta_2),$$

$$\widehat{a}_3(\delta_3) = \arg\min_{a_3} \frac{1}{n}\sum R(z_i, \theta_1, \theta_2 + a_3\delta_3).$$

The asymptotic variance of $(\widehat{a}_1(\delta_1), \widehat{a}_2(\delta_1, \delta_2), \widehat{a}_3(\delta_3))$ is

$$\begin{pmatrix} \delta_1'Q_1\delta_1 & 0 & 0 \\ \delta_1'R_1\delta_2 & \delta_2'R_2\delta_2 & 0 \\ 0 & 0 & \delta_3'R_2\delta_3 \end{pmatrix}^{-1} \begin{pmatrix} \delta_1'V_{11}\delta_1 & \delta_1'V_{12}\delta_2 & \delta_1'V_{12}\delta_3 \\ \delta_1'V_{12}'\delta_2 & \delta_2'V_{22}\delta_2 & \delta_2'V_{22}\delta_3 \\ \delta_1'V_{12}'\delta_3 & \delta_2'V_{22}\delta_3 & \delta_3'V_{22}\delta_3 \end{pmatrix}$$
$$\times \begin{pmatrix} \delta_1'Q_1\delta_1 & \delta_2'R_1'\delta_1 & 0 \\ 0 & \delta_2'R_2\delta_2 & 0 \\ 0 & 0 & \delta_3'R_2\delta_3 \end{pmatrix}^{-1}.$$

When $\delta_2 = \delta_3$, this has the form

$$\begin{pmatrix} q_1 & 0 & 0 \\ r_1 & r_2 & 0 \\ 0 & 0 & r_2 \end{pmatrix}^{-1} \begin{pmatrix} V_q & V_{qr} & V_{qr} \\ V_{qr} & V_r & V_r \\ V_{qr} & V_r & V_r \end{pmatrix} \begin{pmatrix} q_1 & r_1 & 0 \\ 0 & r_2 & 0 \\ 0 & 0 & r_2 \end{pmatrix}^{-1},$$

Bo E. Honoré: honore@Princeton.edu
Luojia Hu: lhu@frbchi.org

 DOI: 10.3982/ECTA13465

where $q_1 = \delta_1' Q_1 \delta_1$, $r_1 = \delta_1' R_1 \delta_2$, $r_2 = \delta_2' R_2 \delta_2$, $V_q = \delta_1' V_{11} \delta_1$, $V_{qr} = \delta V_{12} \delta_2$, and $V_r = \delta_2' V_{22} \delta_2$. This can be written as

$$\begin{pmatrix} \frac{V_q}{q_1^2} & \frac{1}{q_1 r_2} V_{qr} - \frac{V_q}{q_1^2}\frac{r_1}{r_2} & \frac{1}{q_1 r_2} V_{qr} \\ \frac{1}{q_1}\left(\frac{1}{r_2}V_{qr} - \frac{V_q}{q_1}\frac{r_1}{r_2}\right) & \frac{1}{r_2}\left(\frac{V_r}{r_2} - \frac{1}{q_1}\frac{r_1}{r_2}V_{qr}\right) - \frac{1}{q_1}\frac{r_1}{r_2}\left(\frac{1}{r_2}V_{qr} - \frac{V_q}{q_1}\frac{r_1}{r_2}\right) & \frac{1}{r_2}\left(\frac{V_r}{r_2} - \frac{1}{q_1}\frac{r_1}{r_2}V_{qr}\right) \\ \frac{1}{q_1 r_2} V_{qr} & \frac{V_r}{r_2^2} - \frac{1}{q_1}\frac{r_1}{r_2^2}V_{qr} & \frac{V_r}{r_2^2} \end{pmatrix}.$$

Normalize so $V_q = 1$, and parameterize $V_r = v^2$ and $V_{qr} = \rho\sqrt{V_q V_r} = \rho v$ gives the matrix

$$\begin{pmatrix} \frac{1}{q_1^2} & \frac{1}{q_1 r_2} \rho v - \frac{1}{q_1^2}\frac{r_1}{r_2} & \frac{1}{q_1 r_2} \rho v \\ \frac{1}{q_1}\left(\frac{1}{r_2}\rho v - \frac{1}{q_1}\frac{r_1}{r_2}\right) & \frac{1}{r_2}\left(\frac{v^2}{r_2} - \frac{1}{q_1}\frac{r_1}{r_2}\rho v\right) - \frac{1}{q_1}\frac{r_1}{r_2}\left(\frac{1}{r_2}\rho v - \frac{1}{q_1}\frac{r_1}{r_2}\right) & \frac{1}{r_2}\left(\frac{v^2}{r_2} - \frac{1}{q_1}\frac{r_1}{r_2}\rho v\right) \\ \frac{1}{q_1 r_2} \rho v & \frac{v^2}{r_2^2} - \frac{1}{q_1}\frac{r_1}{r_2^2}\rho v & \frac{v^2}{r_2^2} \end{pmatrix}.$$

Denoting the (ℓ, m)th element of this matrix by $\omega_{\ell m}$, we have

$$\omega_{33} - \omega_{32} = \frac{1}{q_1}\frac{r_1}{r_2^2}\rho v = \frac{r_1}{r_2}\omega_{31},$$

$$\frac{\omega_{33} - \omega_{32}}{\omega_{31}} = \frac{r_1}{r_2},$$

$$\rho = \frac{\omega_{31}}{\sqrt{\omega_{11}\omega_{33}}}.$$

Since r_2 is known, this gives r_1 and ρ. We also know v from ω_{33}.

This implies that the asymptotic variance of $(\widehat{a}_1(\delta_1), \widehat{a}_2(\delta_1, \delta_2), \widehat{a}_3(\delta_3))$ identifies $\delta_1' V_{12} \delta_2$ and $\delta_1' R_1 \delta_2$. Choosing $\delta_1 = e_j$ and $\delta_2 = e_m$ (for $j = 1, \ldots, k_1$ and $m = 1, \ldots, k_2$) recovers all the elements of V_{12} and R_1.

S-2. BOOTSTRAPPING WITH EASY SECOND-STEP ESTIMATOR

Consider the case of a two-step estimator like the one in Section 5, but where the first-step estimator is computationally challenging while it is feasible to recalculate the second-step estimator in each bootstrap sample. We again consider estimators of the form

$$\widehat{a}_1(\delta) = \arg\min_{a_1} \frac{1}{n}\sum Q(z_i, \theta_1 + a_1\delta),$$

$$\widehat{a}_2(\delta) = \arg\min_{a_2} \frac{1}{n}\sum R(z_i, \theta_1 + \widehat{a}_1\delta, \theta_2 + a_2),$$

$$\widehat{a}_3 = \arg\min_{a_3} \frac{1}{n}\sum R(z_i, \theta_1, \theta_2 + a_3),$$

but now $\widehat{a}_2$ is a vector of the same dimension as θ_2. The asymptotic variance of $(\widehat{a}_1(\delta), \widehat{a}_2(\delta), \widehat{a}_3)$ is

$$\begin{pmatrix} \delta' Q_1 \delta & 0 & 0 \\ R_1 \delta & R_2 & 0 \\ 0 & 0 & R_2 \end{pmatrix}^{-1} \begin{pmatrix} \delta' V_{11} \delta & \delta' V_{12} & \delta' V_{12} \\ V_{12}' \delta & V_{22} & V_{22} \\ V_{12}' \delta & V_{22} & V_{22} \end{pmatrix} \begin{pmatrix} \delta' Q_1 \delta & \delta' R_1' & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & R_2 \end{pmatrix}^{-1}. \tag{S1}$$

Multiplying (S1) yields a matrix with nine blocks. The upper-middle block is $-(\delta' Q_1 \delta)^{-1} \times (\delta' V_{11} \delta)(\delta' Q_1 \delta) \delta' R_1' R_2^{-1} + (\delta' Q_1 \delta)^{-1} \delta' V_{12} R_2^{-1}$, while the upper-right block is $(\delta' Q_1 \delta)^{-1} \times \delta' V_{12} R_2^{-1}$. The latter identifies $\delta' V_{12}$. When $\delta = e_j$, this is the jth row of V_{12}. The difference between the upper-middle block and the upper-right block gives $-(\delta' Q_1 \delta)^{-1} (\delta' V_{11} \delta)^{-1} \times (\delta' Q_1 \delta) \delta' R_1' R_2^{-1}$, which in turn gives $\delta' R_1'$ or $R_1' \delta$. When δ equals e_j, this is the jth column of R_1.

This approach requires calculation of only $2k_1$ one-dimensional estimators using the more difficult first-step objective function. Moreover, as above, the approach gives closed-form estimates of V_{12} and R_1.

S-3. MAXIMUM OF TWO LOGNORMALS

Let $(X_1, X_2)'$ have a bivariate normal distribution with mean $(\mu_1, \mu_2)'$ and variance $\left(\begin{smallmatrix} \sigma_1^2 & \tau\sigma_1\sigma_2 \\ \tau\sigma_1\sigma_2 & \sigma_2^2 \end{smallmatrix}\right)$ and let $(Y_1, Y_2)' = (\exp(X_1), \exp(X_2))'$. We are interested in $E[\max\{Y_1, Y_2\}]$.

Kotz, Balakrishnan, and Johnson (2000) presented the moment-generating function for $\min\{X_1, X_2\}$ as

$$\begin{aligned} M(t) &= E\big[\exp\big(\min\{X_1, X_2\}t\big)\big] \\ &= \exp\big(t\mu_1 + t^2\sigma_1^2/2\big)\Phi\left(\frac{\mu_2 - \mu_1 - t\big(\sigma_1^2 - \tau\sigma_1\sigma_2\big)}{\sqrt{\sigma_2^2 - 2\tau\sigma_1\sigma_2 + \sigma_1^2}}\right) \\ &\quad + \exp\big(t\mu_2 + t^2\sigma_2^2/2\big)\Phi\left(\frac{\mu_1 - \mu_2 - t\big(\sigma_2^2 - \tau\sigma_1\sigma_2\big)}{\sqrt{\sigma_2^2 - 2\tau\sigma_1\sigma_2 + \sigma_1^2}}\right). \end{aligned}$$

Therefore,

$$\begin{aligned} E\big[\max\{Y_1, Y_2\}\big] &= E[Y_1] + E[Y_2] - E\big[\min\{Y_1, Y_2\}\big] \\ &= E\big[\exp(X_1)\big] + E\big[\exp(X_2)\big] - E\big[\min\big\{\exp(X_1), \exp(X_2)\big\}\big] \\ &= \exp\big(\mu_1 + \sigma_1^2/2\big) + \exp\big(\mu_2 + \sigma_2^2/2\big) - E\big[\exp\big(\min\{X_1, X_2\}\big)\big] \\ &= \exp\big(\mu_1 + \sigma_1^2/2\big) + \exp\big(\mu_2 + \sigma_2^2/2\big) \\ &\quad - \exp\big(\mu_1 + \sigma_1^2/2\big)\Phi\left(\frac{\mu_2 - \mu_1 - \big(\sigma_1^2 - \tau\sigma_1\sigma_2\big)}{\sqrt{\sigma_2^2 - 2\tau\sigma_1\sigma_2 + \sigma_1^2}}\right) \\ &\quad - \exp\big(\mu_2 + \sigma_2^2/2\big)\Phi\left(\frac{\mu_1 - \mu_2 - \big(\sigma_2^2 - \tau\sigma_1\sigma_2\big)}{\sqrt{\sigma_2^2 - 2\tau\sigma_1\sigma_2 + \sigma_1^2}}\right) \end{aligned}$$

$$= \exp\big(\mu_1 + \sigma_1^2/2\big)\left(1 - \Phi\left(\frac{\mu_2 - \mu_1 - \big(\sigma_1^2 - \tau\sigma_1\sigma_2\big)}{\sqrt{\sigma_2^2 - 2\tau\sigma_1\sigma_2 + \sigma_1^2}}\right)\right)$$
$$+ \exp\big(\mu_2 + \sigma_2^2/2\big)\left(1 - \Phi\left(\frac{\mu_1 - \mu_2 - \big(\sigma_2^2 - \tau\sigma_1\sigma_2\big)}{\sqrt{\sigma_2^2 - 2\tau\sigma_1\sigma_2 + \sigma_1^2}}\right)\right).$$

S-4. IMPLEMENTATION WITH TWO-STEP ESTIMATION

In the discussion in Section 5, we identified R_1 and V_{12} in closed forms using a subset of the information contained in the asymptotic variance of $(\widehat{a}_1, \widehat{a}_2, \widehat{a}_3)$. Here we present one way to use all the components of this variance to estimate R_1 and V_{12}. For simplicity, we consider the case where one recalculates the entire first-step estimator in each bootstrap sample.

Consider estimating the second-step parameter in J different directions in each bootstrap replication,

$$\widehat{a}_1 = \arg\min_{a_1} \frac{1}{n}\sum Q(z_i, \theta_1 + a_1),$$
$$\widehat{a}_2(\delta_j) = \arg\min_{a_2} \frac{1}{n}\sum R(z_i, \theta_1 + \widehat{a}_1, \theta_2 + a_2\delta_j),$$
$$\widehat{a}_3(\delta_j) = \arg\min_{a_3} \frac{1}{n}\sum R(z_i, \theta_1, \theta_2 + a_3\delta_j).$$

The asymptotic variance of $(\widehat{a}_1, \{\widehat{a}_2(\delta_j)\}_{j=1}^J, \{\widehat{a}_3(\delta_j)\}_{j=1}^J)$ is of the form $\Omega = A^{-1}B(A')^{-1}$, where

$$A = \begin{pmatrix} Q_1 & 0 & 0 \\ D'R_1 & C'(I\otimes R_2)C & 0 \\ 0 & 0 & C'(I\otimes R_2)C \end{pmatrix} \quad \text{and}$$
$$B = \begin{pmatrix} V_{11} & V_{12}D & V_{12}D \\ D'V_{12}' & D'V_{22}D & D'V_{22}D \\ D'V_{12}' & D'V_{22}D & D'V_{22}D \end{pmatrix}.$$

This gives

$$\begin{pmatrix} V_{11} & V_{12}D & V_{12}D \\ D'V_{12}' & D'V_{22}D & D'V_{22}D \\ D'V_{12}' & D'V_{22}D & D'V_{22}D \end{pmatrix}$$
$$= \begin{pmatrix} Q_1 & 0 & 0 \\ D'R_1 & C'(I\otimes R_2)C & 0 \\ 0 & 0 & C'(I\otimes R_2)C \end{pmatrix} \tag{S2}$$
$$\times\, \Omega \begin{pmatrix} Q_1 & R_1'D & 0 \\ 0 & C'(I\otimes R_2)C & 0 \\ 0 & 0 & C'(I\otimes R_2)C \end{pmatrix}.$$

This suggests estimating V_{12} and R_1 by minimizing

$$\sum_{i,j}\left(\left\{\begin{pmatrix}\widehat{V}_{11} & V_{12}D & V_{12}D\\ D'V'_{12} & D'\widehat{V}_{22}D & D'\widehat{V}_{22}D\\ D'V'_{12} & D'\widehat{V}_{22}D & D'\widehat{V}_{22}D\end{pmatrix}\right.\right.$$
$$-\begin{pmatrix}Q_1 & 0 & 0\\ D'R_1 & C'(I\otimes\widehat{R}_2)C & 0\\ 0 & 0 & C'(I\otimes\widehat{R}_2)C\end{pmatrix}$$
$$\left.\left.\times\,\widehat{\Omega}\begin{pmatrix}Q_1 & R'_1D & 0\\ 0 & C'(I\otimes\widehat{R}_2)C & 0\\ 0 & 0 & C'(I\otimes\widehat{R}_2)C\end{pmatrix}\right\}_{ij}\right)^2$$

over V_{12} and R_1.

When $\delta_j = e_j$, $D = I$ and $C'(I\otimes R_2)C = \operatorname{diag}(R_2) \overset{\text{def}}{=} M$. Using this and multiplying out the right-hand side of (S2) gives

$$\begin{pmatrix}V_{11} & V_{12} & V_{12}\\ V'_{12} & V_{22} & V_{22}\\ V'_{12} & V_{22} & V_{22}\end{pmatrix}$$
$$=\begin{pmatrix}Q_1 & 0 & 0\\ R_1 & M & 0\\ 0 & 0 & M\end{pmatrix}\begin{pmatrix}\Omega_{11} & \Omega_{12} & \Omega_{13}\\ \Omega_{21} & \Omega_{22} & \Omega_{23}\\ \Omega_{31} & \Omega_{32} & \Omega_{33}\end{pmatrix}\begin{pmatrix}Q_1 & R'_1 & 0\\ 0 & M & 0\\ 0 & 0 & M\end{pmatrix}$$
$$=\begin{pmatrix}Q_1\Omega_{11}Q_1 & Q_1\Omega_{11}R'_1 + Q_1\Omega_{12}M & Q_1\Omega_{13}M\\ R_1\Omega_{11}Q_1 + M\Omega_{21}Q_1 & R_1\Omega_{11}R'_1 + M\Omega_{21}R'_1 + R_1\Omega_{12}M + M\Omega_{22}M & R_1\Omega_{13}M + M\Omega_{23}M\\ M\Omega_{13}Q_1 & M\Omega_{31}R'_1 + M\Omega_{32}M & M\Omega_{33}M\end{pmatrix}.$$

The approach in Section 5 uses the last two parts of the first row to identify V_{12} and R_1. The upper left- and lower right-hand corners are not informative about V_{12} or R_1. Moreover, the matrix is symmetric. All the remaining information is therefore contained in the last two parts of the second row. R_1 enters the middle block nonlinearly, which leaves three blocks of equations that are linear in V_{12} and R_1:

$$V_{12} = Q_1\Omega_{11}R'_1 + Q_1\Omega_{12}M,$$
$$V_{12} = Q_1\Omega_{13}M,$$
$$V_{23} = R_1\Omega_{13}M + M\Omega_{23}M.$$

These overidentify V_{12} and R_1, but they could be combined through least squares.

S-5. VALIDITY OF BOOTSTRAP

Hahn (1996) established that under random sampling, the bootstrap distribution of the standard GMM estimator converges weakly to the limiting distribution of the estimator in probability. In this section, we establish the same result under the same regularity conditions for estimators that treat part of the parameter vector as known. Whenever possible, we use the same notation and the same wording as Hahn (1996).

In particular, $o_p^\omega(\cdot)$, $O_p^\omega(\cdot)$, $o_B(\cdot)$, and $O_B(\cdot)$ are defined on page 190 of that paper. A number of papers have proved the validity of the bootstrap in different situations. We choose to tailor our derivation after Hahn (1996) because it so closely mimics the classic proof of asymptotic normality of GMM estimators presented in Pakes and Pollard (1989).

We first review Hahn's (1996) results. The parameter of interest θ_0 is the unique solution to $G(t)=0$, where $G(t)\equiv E[g(Z_i,t)]$, Z_i is the vector of data for observation i, and g is a known function. The parameter space is Θ.

Let $G_n(t)\equiv \frac{1}{n}\sum_{i=1}^n g(Z_i,t)$. The GMM estimator is defined by

$$\tau_n \equiv \arg\min_t \left|A_n G_n(t)\right|,$$

where A_n is a sequence of random matrices (constructed from $\{Z_i\}$) that converges to a nonrandom and nonsingular matrix A.

The bootstrap estimator is the GMM estimator defined in the same way as τ_n but from a bootstrap sample $\{\widehat{Z}_{n1},\dots,\widehat{Z}_{nn}\}$. Specifically,

$$\widehat{\tau}_n \equiv \arg\min_t \left|\widehat{A}_n \widehat{G}_n(t)\right|,$$

where $\widehat{G}_n(t)\equiv \frac{1}{n}\sum_{i=1}^n g(\widehat{Z}_{ni},t)$. $\widehat{A}_n$ is constructed from $\{\widehat{Z}_{ni}\}_{i=1}^n$ in the same way that A_n was constructed from $\{Z_i\}_{i=1}^n$.

Hahn (1996) proved the following results.

PROPOSITION 0—Hahn Proposition 1: *Assume that*

(i) θ_0 *is the unique solution to* $G(t)=0$;
(ii) $\{Z_i\}$ *is an i.i.d. sequence of random vectors*;
(iii) $\inf_{|t-\theta_0|\geq\delta}|G(t)|>0$ *for all* $\delta>0$;
(iv) $\sup_t|G_n(t)-G(t)|\longrightarrow 0$ *as* $n\longrightarrow\infty$ *a.s.*;
(v) $E[\sup_t|g(Z_i,t)|]<\infty$;
(vi) $A_n=A+o_p(1)$ *and* $\widehat{A}_n=A+o_B(1)$ *for some nonsingular and nonrandom matrix* A; *and*
(vii) $|A_nG_n(\tau_n)|\leq o_p(1)+\inf_t|A_nG_n(t)|$ *and* $|\widehat{A}_n\widehat{G}_n(\widehat{\tau}_n)|\leq o_B(1)+\inf_t|\widehat{A}_n\widehat{G}_n(t)|$.

Then $\tau_n=\theta_0+o_p(1)$ *and* $\widehat{\tau}_n=\theta_0+o_B(1)$.

THEOREM 0—Hahn Theorem 1: *Assume that*

(i) *Conditions* (i)–(vi) *in Proposition* 0 *are satisfied*;
(ii) τ_n *satisfies* $|A_nG_n(\tau_n)|\leq o_p(n^{-1/2})+\inf_t|A_nG_n(t)|$ *and* $\widehat{\tau}_n$ *satisfies* $|\widehat{A}_n\widehat{G}_n(\widehat{\tau}_n)|\leq o_B(n^{-1/2})+\inf_t|\widehat{A}_n\widehat{G}_n(t)|$;
(iii) $\lim_{t\to\theta_0}e(t,\theta_0)=0$ *where* $e(t,t')\equiv E[(g(Z_i,t)-g(Z_i,t'))^2]^{1/2}$;
(iv) *for all* $\varepsilon>0$,

$$\lim_{\delta\to 0}\limsup_{n\to\infty} P\Big(\sup_{e(t,t')\leq\delta}\left|G_n(t)-G(t)-G_n(t')+G(t')\right|\geq n^{-1/2}\varepsilon\Big)=0;$$

(v) $G(t)$ *is differentiable at* θ_0, *an interior point of the parameter space*, Θ, *with derivative* Γ *with full rank*; *and*
(vi) $\{g(\cdot,t):t\in\Theta\}\subset L_2(P)$ *and* Θ *is totally bounded under* $e(\cdot,\cdot)$.

Then

$$n^{1/2}(\tau_n - \theta_0) = -n^{-1/2}\big(\Gamma' A' A \Gamma\big)^{-1}\Gamma' A' A_n G_n(\theta_0) + o_p(1) \Longrightarrow N(0, \Omega)$$

and

$$n^{1/2}(\widehat{\tau}_n - \tau_n) \overset{p}{\Longrightarrow} N(0, \Omega),$$

where

$$\Omega = \big(\Gamma' A' A \Gamma\big)^{-1}\Gamma' A' A V A' A \Gamma \big(\Gamma' A' A \Gamma\big)^{-1}$$

and

$$V = E\big[g(Z_i, \theta_0) g(Z_i, \theta_0)'\big].$$

Our paper is based on the same GMM setting as in Hahn (1996). The difference is that we are primarily interested in an infeasible estimator that assumes that one part of the parameter vector is known. We will denote the true parameter vector by θ_0, which we partition as $\theta_0' = (\theta_0^1, \theta_0^2)$.

The infeasible estimator of θ_0, which assumes that θ_0^2 is known, is

$$\gamma_n = \arg\min_t \left| A_n G_n\left(\begin{pmatrix} t \\ \theta_0^2 \end{pmatrix}\right)\right| \tag{S3}$$

or

$$\gamma_n = \arg\min_t G_n\left(\begin{pmatrix} t \\ \theta_0^2 \end{pmatrix}\right)' A_n' A_n G_n\left(\begin{pmatrix} t \\ \theta_0^2 \end{pmatrix}\right).$$

Let the dimensions of θ_0^1 and θ_0^2 be k_1 and k_2, respectively. It is convenient to define $E_1 = (I_{k_1\times k_1} : 0_{k_1\times k_2})'$ and $E_2 = (0_{k_2\times k_1} : I_{k_2\times k_2})'$. Post-multiplying a matrix by E_1 or E_2 will extract the first k_1 or the last k_2 columns of the matrix, respectively.

Let

$$\big(\widehat{\theta}^1, \widehat{\theta}^2\big)' = \arg\min_{(t^1, t^2)} G_n\left(\begin{pmatrix} t^1 \\ t^2 \end{pmatrix}\right)' A_n' A_n G_n\left(\begin{pmatrix} t^1 \\ t^2 \end{pmatrix}\right)$$

be the usual GMM estimator of θ_0. We consider the bootstrap estimator

$$\widehat{\gamma}_n = \arg\min_t \left| \widehat{A}_n \widehat{G}_n\begin{pmatrix} t \\ \widehat{\theta}^2 \end{pmatrix}\right|, \tag{S4}$$

where $\widehat{G}_n(t) \equiv \frac{1}{n}\sum_{i=1}^n g(\widehat{Z}_{ni}, t)$. $\widehat{A}_n$ is constructed from $\{\widehat{Z}_{ni}\}_{i=1}^n$ in the same way that A_n was constructed from $\{Z_i\}_{i=1}^n$. Below, we adapt the derivations in Hahn (1996) to show that the distribution of $\widehat{\gamma}_n$ can be used to approximate the distribution of γ_n. We use exactly the same regularity conditions as Hahn (1996). The only exception is that we need an additional assumption to guarantee the consistency of $\widehat{\gamma}_n$. For this, it is sufficient that the moment function, G, is continuously differentiable and that the parameter space is compact. This additional stronger assumption would make it possible to state the conditions in Proposition 0 more elegantly. We do not restate those conditions because that would make it more difficult to make the connection to Hahn's (1996) result.

PROPOSITION 1—Adaption of Hahn's (1996) Proposition 1: *Suppose that the conditions in Proposition 0 are satisfied. In addition, suppose that G is continuously differentiable and that the parameter space is compact. Then $\gamma_n = \theta_0^1 + o_p(1)$ and $\widehat{\gamma}_n = \theta_0^1 + o_B(1)$.*

PROOF: As in Hahn (1996), the proof follows from standard arguments. The only difference is that we need

$$\sup_t \left| \widehat{G}_n\left(\begin{pmatrix} t \\ \widehat{\theta}^2 \end{pmatrix}\right) - G\left(\begin{pmatrix} t \\ \theta_0^2 \end{pmatrix}\right)\right| = o_p^{\omega}(1).$$

This follows from

$$\begin{aligned}
&\left| \widehat{G}_n\left(\begin{pmatrix} t \\ \widehat{\theta}^2 \end{pmatrix}\right) - G\left(\begin{pmatrix} t \\ \theta_0^2 \end{pmatrix}\right)\right| \\
&\quad = \left| \widehat{G}_n\left(\begin{pmatrix} t \\ \widehat{\theta}^2 \end{pmatrix}\right) - G\left(\begin{pmatrix} t \\ \widehat{\theta}^2 \end{pmatrix}\right) + G\left(\begin{pmatrix} t \\ \widehat{\theta}^2 \end{pmatrix}\right) - G\left(\begin{pmatrix} t \\ \theta_0^2 \end{pmatrix}\right)\right| \\
&\quad \leq \left| \widehat{G}_n\left(\begin{pmatrix} t \\ \widehat{\theta}^2 \end{pmatrix}\right) - G\left(\begin{pmatrix} t \\ \widehat{\theta}^2 \end{pmatrix}\right)\right| + \left| G\left(\begin{pmatrix} t \\ \widehat{\theta}^2 \end{pmatrix}\right) - G\left(\begin{pmatrix} t \\ \theta_0^2 \end{pmatrix}\right)\right|.
\end{aligned}$$

As in Hahn (1996), the first part is $o_p^{\omega}(1)$ by bootstrap uniform convergence. The second part is bounded by $\sup |\frac{\partial G(t_1,t_2)}{\partial t_2}||\widehat{\theta}^2 - \theta_0^2|$. This is $O_p(\widehat{\theta}^2 - \theta_0^2) = O_p(n^{-1/2})$ by the assumptions that G is continuously differentiable and that the parameter space is compact. *Q.E.D.*

THEOREM 3—Adaption of Hahn's (1996) Theorem 1: *Assume that the conditions in Proposition 1 and Theorem 0 are satisfied. Then*

$$n^{1/2}\big(\gamma_n - \theta_0^1\big) \Longrightarrow N(0, \Omega)$$

and

$$n^{1/2}(\widehat{\gamma}_n - \gamma_n) \overset{p}{\Longrightarrow} N(0, \Omega),$$

where

$$\Omega = \big(E_1'\Gamma' A' A \Gamma E_1\big)^{-1} E_1'\Gamma' A' A S A' A \Gamma E_1 \big(E_1'\Gamma' A' A \Gamma E_1\big)^{-1}$$

and

$$V = E\big[g(Z_i, \theta_0) g(Z_i, \theta_0)'\big].$$

PROOF: We start by showing that $\binom{\gamma_n}{\widehat{\theta}^2}$ is $\sqrt{n}$-consistent, and then move on to show asymptotic normality.

Part 1. $\sqrt{n}$-consistency. For $\widehat{\theta}^2$, root-n consistency follows from Pakes and Pollard (1989). Following Hahn (1996), we start with the observation that

$$\begin{aligned}
&\left| \widehat{A}_n \widehat{G}_n\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right) - A G\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right) - \widehat{A}_n \widehat{G}_n(\theta_0) + A G(\theta_0)\right| \\
&\quad \leq |\widehat{A}_n| \left| \widehat{G}_n\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right) - G\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right) - \widehat{G}_n(\theta_0) + G(\theta_0)\right| \qquad \text{(S5)}
\end{aligned}$$

$$+|\widehat{A}_n - A|\left|G\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - G(\theta_0)\right|$$
$$\leq o_B(n^{-1/2}) + o_B(1)\left|G\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - G(\theta_0)\right|.$$

Combining this with the triangular inequality, we have

$$\begin{aligned}\left|AG\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - AG(\theta_0)\right| &\leq \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - AG\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - \widehat{A}_n\widehat{G}_n(\theta_0) + AG(\theta_0)\right| \\ &\quad + \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - \widehat{A}_n\widehat{G}_n(\theta_0)\right| \\ &\leq o_B(n^{-1/2}) + o_B(1)\left|G\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - G(\theta_0)\right| \\ &\quad + \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - \widehat{A}_n\widehat{G}_n(\theta_0)\right|.\end{aligned} \tag{S6}$$

The nonsingularity of A implies the existence of a constant $C_1 > 0$ such that $|Ax| \geq C_1|x|$ for all x. Applying this fact to the left-hand side of (S6) and collecting the $G\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - G(\theta_0)$ terms yield

$$\left(C_1 - o_B(1)\right)\left|G\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - G(\theta_0)\right| \tag{S7}$$
$$\leq o_B(n^{-1/2}) + \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - \widehat{A}_n\widehat{G}_n(\theta_0)\right|$$
$$\leq o_B(n^{-1/2}) + \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right)\right| + \left|\widehat{A}_n\widehat{G}_n(\theta_0)\right|$$
$$\leq o_B(n^{-1/2}) + \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\theta_0^1\\ \widehat{\theta}^2\end{pmatrix}\right)\right| + \left|\widehat{A}_n\widehat{G}_n(\theta_0)\right|. \tag{S8}$$

Stochastic equicontinuity implies that

$$\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\theta_0^1\\ \widehat{\theta}^2\end{pmatrix}\right) = \widehat{A}_n\left(G\left(\begin{pmatrix}\theta_0^1\\ \widehat{\theta}^2\end{pmatrix}\right) - G(\theta_0)\right) + \widehat{A}_n\widehat{G}_n(\theta_0) + \widehat{A}_n o_B(n^{-1/2})$$

or

$$\left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\theta_0^1\\ \widehat{\theta}^2\end{pmatrix}\right)\right| \leq \left|\widehat{A}_n\left(G\left(\begin{pmatrix}\theta_0^1\\ \widehat{\theta}^2\end{pmatrix}\right) - G(\theta_0)\right)\right| + \left|\widehat{A}_n\widehat{G}_n(\theta_0)\right| + |\widehat{A}_n| o_B(n^{-1/2}),$$

so (S8) implies

$$\left(C_1 - o_B(1)\right)\left|G\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - G(\theta_0)\right|$$
$$\leq o_B(n^{-1/2}) + \left|\widehat{A}_n\left(G\left(\begin{pmatrix}\theta_0^1\\ \widehat{\theta}^2\end{pmatrix}\right) - G(\theta_0)\right)\right| + 2|\widehat{A}_n|\left|\widehat{G}_n(\theta_0)\right| + |\widehat{A}_n| o_B(n^{-1/2})$$

$$
\begin{aligned}
&\le o_B\big(n^{-1/2}\big) + |\widehat{A}_n| \left| \left(G\left(\begin{pmatrix} \theta_0^1 \\ \widehat{\theta}^2 \end{pmatrix} \right) - G(\theta_0) \right) \right| + 2|\widehat{A}_n| \big| \widehat{G}_n(\theta_0) - G_n(\theta_0) \big| \\
&\quad + 2|\widehat{A}_n| \big| G_n(\theta_0) \big| + |\widehat{A}_n| o_B\big(n^{-1/2}\big) \\
&= o_B\big(n^{-1/2}\big) + O_B(1) O_p\big(n^{-1/2}\big) + O_B(1) O_B\big(n^{-1/2}\big) \\
&\quad + O_B(1) O_p\big(n^{-1/2}\big) + O_B(1) o_B\big(n^{-1/2}\big).
\end{aligned}
\tag{S9}
$$

Note that

$$
G\left(\begin{pmatrix} \widehat{\gamma}_n \\ \theta_0^2 \end{pmatrix} \right) = \Gamma E_1\big(\widehat{\gamma}_n - \theta_0^1\big) + o_B(1) \big| \widehat{\gamma}_n - \theta_0^1 \big|.
$$

As above, the nonsingularity of Γ implies nonsingularity of ΓE_1, and hence, there exists a constant $C_2 > 0$ such that $|\Gamma E_1 x| \ge C_2 |x|$ for all x. Applying this to the equation above and collecting terms give

$$
\begin{aligned}
C_2 \big| \widehat{\gamma}_n - \theta_0^1 \big| &\le \big| \Gamma E_1\big(\widehat{\gamma}_n - \theta_0^1\big) \big| \\
&= \left| G\left(\begin{pmatrix} \widehat{\gamma}_n \\ \theta_0^2 \end{pmatrix} \right) - G(\theta_0) \right| + o_B(1) \big| \widehat{\gamma}_n - \theta_0^1 \big|.
\end{aligned}
\tag{S10}
$$

Combining (S10) with (S9) yields

$$
\begin{aligned}
&\big(C_1 - o_B(1)\big)\big(C_2 - o_B(1)\big) \big| \widehat{\gamma}_n - \theta_0^1 \big| \\
&\quad \le \big(C_1 - o_B(1)\big) \left| G\left(\begin{pmatrix} \widehat{\gamma}_n \\ \theta_0^2 \end{pmatrix} \right) - G(\theta_0) \right| \\
&\quad \le o_B\big(n^{-1/2}\big) + O_B(1) O_p\big(n^{-1/2}\big) + O_B(1) O_B\big(n^{-1/2}\big) \\
&\qquad + O_B(1) O_p\big(n^{-1/2}\big) + O_B(1) o_B\big(n^{-1/2}\big)
\end{aligned}
$$

or

$$
\big| \widehat{\gamma}_n - \theta_0^1 \big| \le O_B(1) \big(O_p\big(n^{-1/2}\big) + O_B\big(n^{-1/2}\big) \big).
$$

Part 2: *Asymptotic Normality*. Let

$$
\widetilde{L}_n(t) = A\Gamma \left(\begin{pmatrix} t \\ \widehat{\theta}^2 \end{pmatrix} - \begin{pmatrix} \theta_0^1 \\ \theta_0^2 \end{pmatrix} \right) + \widehat{A}_n \widehat{G}_n(\theta_0).
$$

Define

$$
\begin{aligned}
\widehat{\sigma}_n &= \arg\min_t \big| \widetilde{L}_n(t) \big| \\
&= \arg\min_t \left(A\Gamma \left(\begin{pmatrix} t \\ \widehat{\theta}^2 \end{pmatrix} - \begin{pmatrix} \theta_0^1 \\ \theta_0^2 \end{pmatrix} \right) + \widehat{A}_n \widehat{G}_n(\theta_0) \right)' \\
&\qquad \times \left(A\Gamma \left(\begin{pmatrix} t \\ \widehat{\theta}^2 \end{pmatrix} - \begin{pmatrix} \theta_0^1 \\ \theta_0^2 \end{pmatrix} \right) + \widehat{A}_n \widehat{G}_n(\theta_0) \right).
\end{aligned}
$$

Solving for $\widehat{\sigma}_n$ gives

$$\begin{aligned}\widehat{\sigma}_n &= \theta_0^1 - B_{11}^{-1}\big(B_{21}'x + C_1'\big)\\ &= \theta_0^1 - \big((\Gamma E_1)'A'A\Gamma E_1\big)^{-1}\\ &\quad\times\big((\Gamma E_1)'A'A\Gamma E_2\big(\widehat{\theta}^2 - \theta_0^2\big) + (\Gamma E_1)'A'\widehat{A}_n\widehat{G}_n(\theta_0)\big)\\ &= \theta_0^1 - \big((\Gamma E_1)'A'A\Gamma E_1\big)^{-1}(\Gamma E_1)'A'\\ &\quad\times\big(A\Gamma E_2\big(\widehat{\theta}^2 - \theta_0^2\big) + \widehat{A}_n\widehat{G}_n(\theta_0)\big).\end{aligned}$$

Mimicking the calculation on the top of page 195 of Hahn (1996),

$$\begin{aligned}(\widehat{\sigma}_n - \gamma_n) &= -\big((\Gamma E_1)'A'A\Gamma E_1\big)^{-1}(\Gamma E_1)'A'\big(A\Gamma E_2\big(\widehat{\theta}^2 - \theta_0^2\big) + \widehat{A}_n\widehat{G}_n(\theta_0)\big)\\ &\quad + \big(E_1'\Gamma'A'A\Gamma E_1\big)^{-1}E_1'\Gamma'A'AG_n(\theta_0)\\ &= -\big((\Gamma E_1)'A'A\Gamma E_1\big)^{-1}(\Gamma E_1)'A'\\ &\quad\times\big(A\Gamma E_2\big(\widehat{\theta}^2 - \theta_0^2\big) + \widehat{A}_n\widehat{G}_n(\theta_0) - AG_n(\theta_0)\big)\\ &= -\Delta\big(\rho_n + \widehat{A}_n\widehat{G}_n(\theta_0) - AG_n(\theta_0)\big),\end{aligned}$$

where $\Delta = ((\Gamma E_1)'A'A\Gamma E_1)^{-1}(\Gamma E_1)'A'$ and $\rho_n = A\Gamma E_2(\widehat{\theta}^2 - \theta_0^2)$. Or

$$(\widehat{\sigma}_n - \gamma_n + \Delta\rho_n) = -\Delta\big(\widehat{A}_n\widehat{G}_n(\theta_0) - AG_n(\theta_0)\big).$$

From this, it follows that $\widehat{\sigma}_n - \gamma_n = O_B(n^{-1/2})$.

Next we want to argue that $\sqrt{n}(\widehat{\sigma}_n - \widehat{\gamma}_n) = o_B(1)$.

We next proceed as in Hahn (1996) (page 194). First we show that

$$\left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - \widetilde{L}_n(\widehat{\gamma}_n)\right| = o_B\big(n^{-1/2}\big). \tag{S11}$$

It follows from Hahn that

$$\left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - AG\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - \widehat{A}_n\widehat{G}_n(\theta_0) + AG(\theta_0)\right| = o_B\big(n^{-1/2}\big).$$

We thus have

$$\begin{aligned}\left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - \widetilde{L}_n(\widehat{\gamma}_n)\right| &= \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - A\Gamma\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix} - \begin{pmatrix}\theta_0^1\\ \theta_0^2\end{pmatrix}\right) - \widehat{A}_n\widehat{G}_n(\theta_0)\right|\\ &\le \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - AG\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - \widehat{A}_n\widehat{G}_n(\theta_0) + AG(\theta_0)\right|\\ &\quad + \left|AG\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right) - AG(\theta_0) - A\Gamma\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix} - \theta_0\right)\right|\\ &= o_B\big(n^{-1/2}\big) + o\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix} - \theta_0\right)\\ &= o_B\big(n^{-1/2}\big).\end{aligned}$$

This uses the fact that $\binom{\widehat{\gamma}_n}{\widehat{\theta}^2}$ is $\sqrt{n}$-consistent.
Next, we will show that

$$\left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}\right)-\widetilde{L}_n(\widehat{\sigma}_n)\right|=o_B\big(n^{-1/2}\big). \tag{S12}$$

We have

$$\begin{aligned}\left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}\right)-\widetilde{L}_n(\widehat{\sigma}_n)\right| &= \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}\right)-A\Gamma\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}-\theta_0\right)-\widehat{A}_n\widehat{G}_n(\theta_0)\right| \\ &\le \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}\right)-AG\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}\right)-\widehat{A}_n\widehat{G}_n(\theta_0)+AG(\theta_0)\right| \\ &\quad+\left|AG\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}\right)-AG(\theta_0)-A\Gamma\left(\begin{pmatrix}\widehat{\sigma}_1\\ \widehat{\theta}^2\end{pmatrix}-\theta_0\right)\right| \\ &= o_B\big(n^{-1/2}\big)+o\left(\begin{pmatrix}\widehat{\sigma}_1\\ \widehat{\theta}^2\end{pmatrix}-\theta_0\right) \\ &= o_B\big(n^{-1/2}\big).\end{aligned}$$

For the last step, we use $\widehat{\sigma}_n-\theta_0^1=(\widehat{\sigma}_n-\gamma_n)+(\gamma_n-\theta_0^1)=O_B(n^{-1/2})+O_p(n^{-1/2})$.
Combining (S11) and (S12) with the definitions of $\widehat{\gamma}_n$ and $\widehat{\sigma}_n$, we get

$$\big|\widetilde{L}_n(\widehat{\gamma}_n)\big|=\big|\widetilde{L}_n(\widehat{\sigma}_n)\big|+o_B\big(n^{-1/2}\big). \tag{S13}$$

Exactly as in Hahn (1996) and Pakes and Pollard (1989), we start with

$$\begin{aligned}\big|\widetilde{L}_n(\widehat{\sigma}_n)\big| &\le \left|A\Gamma\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}-\theta_0\right)\right|+\big|\widehat{A}_n\widehat{G}_n(\theta_0)\big| \\ &\le \left|A\Gamma\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}-\begin{pmatrix}\gamma_n\\ \widehat{\theta}^2\end{pmatrix}\right)\right|+\big|\widehat{A}_n\widehat{G}_n(\theta_0)-\widehat{A}_nG_n(\theta_0)\big| \\ &\quad+\left|A\Gamma\left(\begin{pmatrix}\gamma_n\\ \widehat{\theta}^2\end{pmatrix}-\theta_0\right)\right|+\big|\widehat{A}_nG_n(\theta_0)\big| \\ &= O_B\big(n^{-1/2}\big)+O_B(1)O_B\big(n^{-1/2}\big)+O_p\big(n^{-1/2}\big)+O_B(1)O_p\big(n^{-1/2}\big).\end{aligned} \tag{S14}$$

Squaring both sides of (S13), we have

$$\big|\widetilde{L}_n(\widehat{\gamma}_n)\big|^2=\big|\widetilde{L}_n(\widehat{\sigma}_n)\big|^2+o_B\big(n^{-1}\big) \tag{S15}$$

because (S14) implies that the cross-product term can be absorbed in the $o_B(n^{-1})$. On the other hand, for any t,

$$\widetilde{L}_n(t)=A\Gamma\left(\begin{pmatrix}t\\ \widehat{\theta}^2\end{pmatrix}-\begin{pmatrix}\theta_0^1\\ \theta_0^2\end{pmatrix}\right)+\widehat{A}_n\widehat{G}_n(\theta_0)$$

has the form $\widetilde{L}_n(t)=y-Xt$, where $X=-A\Gamma E_1$ and $y=-A\Gamma E_1\theta_0^1+A\Gamma E_2(\widehat{\theta}^2-\theta_0^2)+\widehat{A}_n\widehat{G}_n(\theta_0)$.

$\widehat{\sigma}_n$ solves a least squares problem with first-order condition $X'\widetilde{L}_n(\widehat{\sigma}_n) = 0$. Also

$$\begin{aligned}
\left|\widetilde{L}_n(t)\right|^2 &= (y - Xt)'(y - Xt) \\
&= \big((y - X\widehat{\sigma}_n) - X(t - \widehat{\sigma}_n)\big)'\big((y - X\widehat{\sigma}_n) - X(t - \widehat{\sigma}_n)\big) \\
&= (y - X\widehat{\sigma}_n)'(y - X\widehat{\sigma}_n) + (t - \widehat{\sigma}_n)'X'X(t - \widehat{\sigma}_n) \\
&\quad - 2(t - \widehat{\sigma}_n)'X'(y - X\widehat{\sigma}_n) \\
&= \left|\widetilde{L}_n(\widehat{\sigma}_n)\right|^2 + \left|X(t - \widehat{\sigma}_n)\right|^2 - 2(t - \widehat{\sigma}_n)'X'\widetilde{L}_n(\widehat{\sigma}_n) \\
&= \left|\widetilde{L}_n(\widehat{\sigma}_n)\right|^2 + \left|(A\Gamma E_1)(t - \widehat{\sigma}_n)\right|^2.
\end{aligned}$$

Plugging in $t = \widehat{\gamma}_n$, we have

$$\left|\widetilde{L}_n(\widehat{\gamma}_n)\right|^2 = \left|\widetilde{L}_n(\widehat{\sigma}_n)\right|^2 + \left|(A\Gamma E_1)(\widehat{\gamma}_n - \widehat{\sigma}_n)\right|^2.$$

Compare this to (S15) to conclude that

$$(A\Gamma E_1)(\widehat{\gamma}_n - \widehat{\sigma}_n) = o_B\big(n^{-1/2}\big).$$

$A\Gamma E_1$ has full rank by assumption, so $(\widehat{\gamma}_n - \widehat{\sigma}_n) = o_B(n^{-1/2})$ and $n^{1/2}(\widehat{\gamma}_n - \gamma_n) = n^{1/2}(\widehat{\sigma}_n - \gamma_n) + o_B(n^{-1/2})$, and since $n^{1/2}(\widehat{\sigma}_n - \gamma_n) \overset{p}{\Longrightarrow} N(0, \Omega)$, we obtain $n^{1/2}(\widehat{\gamma}_n - \gamma_n) \overset{p}{\Longrightarrow} N(0, \Omega)$. *Q.E.D.*

Theorem 3 is stated for GMM estimators. This covers extremum estimators and the two-step estimators as special cases. Theorem 3 also covers the case where one is interested in different infeasible lower-dimensional estimators as in Section 4.2. To see this, consider two estimators of the form

$$\widehat{a}(\delta_1) = \arg\min_a \left(\frac{1}{n}\sum_{i=1}^n f(x_i, \theta_0 + a\delta_1)\right)' W_n \left(\frac{1}{n}\sum_{i=1}^n f(x_i, \theta_0 + a\delta_1)\right)$$

and

$$\widehat{a}(\delta_2) = \arg\min_a \left(\frac{1}{n}\sum_{i=1}^n f(x_i, \theta_0 + a\delta_2)\right)' W_n \left(\frac{1}{n}\sum_{i=1}^n f(x_i, \theta_0 + a\delta_2)\right),$$

and let A_n denote the matrix-square root of W_n. We can then write

$$\big(\widehat{a}(\delta_1), \widehat{a}(\delta_2)\big) = \arg\min \left| \begin{pmatrix} A_n & 0 \\ 0 & A_n \end{pmatrix} \frac{1}{n}\sum_{i=1}^n \begin{pmatrix} f(x_i, \theta_0 + a\delta_1) \\ f(x_i, \theta_0 + a\delta_2) \end{pmatrix} \right|,$$

which has the form of (S3).

REFERENCES

HAHN, J. (1996): "A Note on Bootstrapping Generalized Method of Moments Estimators," *Econometric Theory*, 12 (1), 187–197. [5-8,11,12]

KOTZ, S., N. BALAKRISHNAN, AND N. L. JOHNSON (2000): *Continuous Multivariate Distributions, Models and Applications* (Second Ed.). Continuous Multivariate Distributions. Wiley. [3]

PAKES, A., AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057. [6,8,12]

Co-editor Elie Tamer handled this manuscript.

Manuscript received 15 May, 2015; final version accepted 17 January, 2017; available online 13 April, 2017.