Econometrica Supplementary Material

SUPPLEMENT TO "EFFICIENCY BOUNDS FOR MISSING DATA MODELS WITH SEMIPARAMETRIC RESTRICTIONS": PROOFS

(*Econometrica*, Vol. 79, No. 2, March 2011, 437–452)

BY BRYAN S. GRAHAM

THIS APPENDIX CONTAINS PROOFS of the results contained in the main paper. All notation is as defined in the main text unless explicitly noted otherwise. Equation numbering continues in sequence with that established in the main text. To simplify notation, let β denote the true parameter value β_0 unless explicitly stated otherwise; similarly, the 0 subscript is removed from other objects, such as the propensity score, when doing so does not cause confusion.

A. PROOF OF THEOREM 2.1

The proof closely follows that of Theorem 1 in Chamberlain (1992) and consists of three steps.

STEP 1—Demonstration of Equivalence With an Unconditional GMM Problem: The first step is to show that restrictions (4) and (5) are, in the multinomial case, equivalent to a finite set of unconditional moment restrictions. Under the multinomial assumption, we have $X \in \{x_1, \dots, x_L\}$ for some L. Let the $L \times 1$ vector B have a 1 in the lth row if $X = x_l$ and 0 elsewhere, and $\tau_l = \Pr(X = x_l)$ (observe that $\sum_{l=1}^{L} \tau_l = 1$). Denote the value of the selection probability at $X = x_l$ by ρ_l and define $\rho = \{\rho_1, \dots, \rho_L\}'$; this vector gives the values of $p(\cdot)$ at each of the mass points of X. Using this notation, we can write $p(X) = B'\rho$.

Under the multinomial assumption, restrictions (4) and (5) are equivalent to the $L + K \times 1$ vector of unconditional moment restrictions

$$\mathbb{E}[m(Z,\beta,\rho)] = \mathbb{E}\begin{bmatrix} m_1(Z,\rho) \\ m_2(Z,\beta,\rho) \end{bmatrix} = \mathbb{E}\begin{bmatrix} B\left(\dfrac{D}{B'\rho} - 1\right) \\ \dfrac{D}{B'\rho}\psi(Z,\beta) \end{bmatrix} = 0.$$

To verify that this is the case, note that by iterated expectations

$$\mathbb{E}[m_1(Z,\rho)] = \begin{pmatrix} \tau_1 \mathbb{E}\left[\left(\dfrac{D}{p(X)} - 1\right)\Big| X = x_1\right] \\ \vdots \\ \tau_L \mathbb{E}\left[\left(\dfrac{D}{p(X)} - 1\right)\Big| X = x_L\right] \end{pmatrix},$$

 DOI: 10.3982/ECTA7379

and hence $\mathbb{E}[m_1(Z, \rho)] = 0$ if and only if $\mathbb{E}[\frac{D}{p(X)} - 1|X] = 0$ for all $X \in \{x_1, \dots, x_L\}$. We also have

$$\mathbb{E}[m_2(Z, \beta, \rho)] = \mathbb{E}\left[\frac{D}{p(X)}\psi(Z, \beta)\right] = 0,$$

so $\mathbb{E}[m(Z, \beta, \rho)] = 0$ if and only if (4) and (5) are satisfied as claimed.

STEP 2—Application of Lemma 2 of Chamberlain (1987): Chamberlain (1987, Lemma 2) showed that for Z a multinomial random variable, the variance bound for β under the sole restriction that $\mathbb{E}[m(Z, \beta, \rho)] = 0$ is

$$\{(M'V^{-1}M)^{-1}\}_{22},$$

where $\{(M'V^{-1}M)^{-1}\}_{22}$ is the lower right $K \times K$ block of $(M'V^{-1}M)^{-1}$ with

$$V \stackrel{\text{def}}{\equiv} \mathbb{E}[m(Z, \beta, \rho)m(Z, \beta, \rho)'],$$
$$M \stackrel{\text{def}}{\equiv} \mathbb{E}\left[\frac{\partial m(Z, \beta, \rho)}{\partial \rho'}, \frac{\partial m(Z, \beta, \rho)}{\partial \beta'}\right].$$

The application of Chamberlain's result requires that M has full column rank and that V is nonsingular. The calculations made in Step 3 below demonstrate that these conditions are implied by the assumption that Γ has full column rank, $p(X)$ is bounded away from zero, and Λ is nonsingular.

STEP 3—Calculation of the Bound: The final step is to solve for an explicit expression for $\{(M'V^{-1}M)^{-1}\}_{22}$. This requires some simple, albeit tedious, algebra. Partitioning V_0,

$$\underset{L+K\times L+K}{V} = \begin{pmatrix} V_{11} & V_{12} \\ V_{12}' & V_{22} \end{pmatrix},$$

we have the lower right-hand block, letting $\psi = \psi(Z, \beta)$ and $q(X) = \mathbb{E}[\psi|X]$, given by

$$\begin{aligned} \underset{K\times K}{V_{22}} &= \mathbb{E}[m_2(Z, \beta, \rho)m_2(Z, \beta, \rho)'] \\ &= \mathbb{E}\left[\frac{\mathbb{E}[\psi\psi'|X]}{p(X)}\right] \\ &= \mathbb{E}\left[\frac{\mathbb{V}(\psi|X)}{p(X)} + \frac{1-p(X)}{p(X)}q(X)q(X)' + q(X)q(X)'\right] \\ &= \sum_{l=1}^{L}\tau_l\left[\frac{\Sigma_l}{\rho_l} + \frac{1-\rho_l}{\rho_l}q_lq_l' + q_lq_l'\right], \end{aligned} \tag{9}$$

where $q_l = \mathbb{E}[\psi(Z,\beta)|x_l]$ and $\Sigma_l = \mathbb{V}(\psi|x_l)$.

The upper right-hand block is similarly derived as

$$\begin{aligned}
\underset{L\times K}{V_{12}} &= \mathbb{E}[m_1(Z,\beta)m_2(Z,\beta,\rho)'] \\
&= \mathbb{E}\left[B\left(\frac{D}{B'\rho}-1\right)\left\{\frac{D\psi(Z,\beta)}{B'\rho}\right\}'\right] \\
&= \mathbb{E}\left[B\left(\frac{1-p(X)}{p(X)}q(X)'\right)\right] \\
&= \left(\tau_1\frac{1-\rho_1}{\rho_1}q_1 \quad \cdots \quad \tau_L\frac{1-\rho_L}{\rho_L}q_L\right)'.
\end{aligned} \tag{10}$$

Finally the upper left-hand block is given by

$$\begin{aligned}
\underset{L\times L}{V_{11}} &= \mathbb{E}\left[B\left(\frac{D}{B'\rho}-1\right)\left(\frac{D}{B'\rho}-1\right)B'\right] \\
&= \mathbb{E}\left[BB'\left(\frac{1-p(X)}{p(X)}\right)\right] \\
&= \operatorname{diag}\left\{\tau_1\frac{1-\rho_1}{\rho_1} \quad \cdots \quad \tau_L\frac{1-\rho_L}{\rho_L}\right\}.
\end{aligned} \tag{11}$$

Now partition M

$$\underset{L+K\times L+K}{M} = \begin{pmatrix} M_{1\rho} & 0 \\ M_{2\rho} & M_{2\beta}\end{pmatrix},$$

where, from similar calculations to those made above, we have

$$\begin{aligned}
\underset{L\times L}{M_{1\rho}} &= -\operatorname{diag}\left\{\frac{\tau_1}{\rho_1} \quad \cdots \quad \frac{\tau_L}{\rho_L}\right\}, \\
\underset{K\times L}{M_{2\rho}} &= -\left(\tau_1\frac{q_1}{\rho_1} \quad \cdots \quad \tau_L\frac{q_L}{\rho_L}\right), \qquad \underset{K\times K}{M_{2\beta}} = \Gamma.
\end{aligned} \tag{12}$$

Applying standard results on partitioned inverses then yields

$$M^{-1} = \begin{pmatrix} M_{1\rho}^{-1} & 0 \\ -M_{2\beta}^{-1}M_{2\rho}M_{1\rho}^{-1} & M_{2\beta}^{-1}\end{pmatrix}.$$

Note that the existence of $M_{1\rho}^{-1}$ and $M_{2\beta}^{-1}$ follows from the assumptions that $p(X)$ is bounded away from zero and the assumption that Γ has full column rank.

Redundancy of knowledge of the propensity score suggests that $M^{-1}VM^{-1\prime}$ will be block diagonal. A sufficient condition for this is that (cf. Prokhorov and Schmidt (2009))

$$V'_{12} = M_{2\rho}M_{1\rho}^{-1}V_{11}. \tag{13}$$

To verify that this condition holds, use (11) and (12) to show that

$$M_{2\rho}M_{1\rho}^{-1}V_{11} = \begin{pmatrix} \tau_1 \frac{1-\rho_1}{\rho_1} q_1 & \cdots & \tau_L \frac{1-\rho_L}{\rho_L} q_L \end{pmatrix},$$

which equals V'_{12} as required. Exploiting the resulting simplifications yields

$$M^{-1}VM^{-1\prime} = \begin{pmatrix} M_{1\rho}^{-1}V_{11}M_{1\rho}^{-1} & 0 \\ 0 & M_{2\beta}^{-1}(V_{22} - V'_{12}V_{11}^{-1}V_{12})M_{2\beta}^{-1\prime} \end{pmatrix}$$

and hence

$$(M^{-1}VM^{-1\prime})_{22} = M_{2\beta}^{-1}(V_{22} - V'_{12}V_{11}^{-1}V_{12})M_{2\beta}^{-1\prime}.$$

By $M_{2\rho}M_{1\rho}^{-1} = (q_1, \dots, q_L)$ and (13), we have

$$\begin{aligned} V'_{12}V_{11}^{-1}V_{12} &= M_{2\rho}M_{1\rho}^{-1}V_{11}M_{1\rho}^{-1\prime}M'_{2\rho} \\ &= \sum_{l=1}^{L} \tau_l \frac{1-\rho_l}{\rho_l} q_l q'_l \\ &= \mathbb{E}\left[\frac{1-p(X)}{p(X)} q(X)q(X)'\right] \end{aligned}$$

and hence, using (9),

$$V_{22} - V'_{12}V_{11}^{-1}V_{12} = \mathbb{E}\left[\frac{\mathbb{V}(\psi|X)}{p(X)} + q(X)q(X)'\right] = \Lambda.$$

Using this result and taking the partitioned determinant gives

$$\det(V) = \det(V_{11})\det(V_{22} - V'_{12}V_{11}^{-1}V_{12}) = \mathbb{E}\left[\frac{1-p(X)}{p(X)}\right]\det\{\Lambda\},$$

and hence V is nonsingular under overlap (Assumption 1.4) and nonsingularity of Λ.

Since $M_{2\beta} = \Gamma$, we have $\mathcal{I}_{\mathrm{m}}(\beta_0) = \Gamma'\Lambda^{-1}\Gamma$ as claimed. For completeness, the upper left-hand portion of the full variance–covariance matrix is given by

$$\begin{aligned} M_{11}^{-1}V_{11}M_{11}^{-1\prime} &= \mathcal{I}_{\mathrm{m}}^{-1}(\rho_0) \\ &= \operatorname{diag}\left\{\frac{p(x_1)(1-p(x_1))}{f(x_1)}, \dots, \frac{p(x_L)(1-p(x_L))}{f(x_L)}\right\}, \end{aligned}$$

where $f(x) = \sum_{l=1}^{L} \tau_l \times \mathbf{1}(x = x_l)$.

B. PROOF OF THEOREM 3.1

The first two steps of the proof of Theorem 3.1 are analogous to those of Theorem 2.1 and are therefore omitted. The actual calculation of the bound, while conceptually straightforward, is considerably more tedious. Details of this step are provided here.

Assume that the marginal distributions of X_1 and X_2 have I and M points of support with probabilities $\pi_1, \dots, \pi_I$ and $\varsigma_1, \dots, \varsigma_M$. Let $L = I \times M$ and let τ_{im} denote the joint probability $\Pr(X_1 = x_{1,i}, X_2 = x_{2,m})$. Let $\lambda = (\lambda_1, \dots, \lambda_M)'$ be the values of $h(\cdot)$ at each of the mass points of X_2 (for simplicity I assume that $\dim(h(x_2)) = P = 1$ in the calculations below, but the results generalize). Let C be a $M \times 1$ vector with a 1 in the mth row if $X_2 = x_{2,m}$ and 0s elsewhere. Finally it is convenient to use the shorthand $\Psi = q(X)q(X)'$. In what follows I use both the single and double subscript notation to denote a point on the support of X as is convenient. We can map between the two notations by observing that $x_{im} = x_l$ for $l = (i-1)M + m$.

For the multinomial case, the conditional moment problem defined by (4), (5), and (6) is equivalent to the unconditional problem

$$\mathbb{E}[m(Z,\theta)] = \mathbb{E}\begin{bmatrix} m_1(Z,\rho) \\ m_2(Z,\rho,\lambda,\delta,\beta) \\ m_3(Z,\rho,\beta) \end{bmatrix} = 0,$$

with $\theta = (\rho', \lambda', \delta', \beta')'$ and

$$\underset{L\times 1}{m_1(Z,\rho)} = B\left(\frac{D}{B'\rho} - 1\right),$$

$$\underset{LK\times 1}{m_2(Z,\rho,\lambda,\delta,\beta)} = (B \otimes I_K)\left(\frac{D}{B'\rho}(\psi(Z,\beta) - q(X,\delta,C'\lambda;\beta))\right),$$

$$\underset{K\times 1}{m_3(Z,\rho,\beta)} = \frac{D}{B'\rho}\psi(Z,\beta).$$

Partition $V = \mathbb{E}[m(Z,\theta)m(Z,\theta)']$ as

$$\underset{L+KL+K\times L+KL+K}{V} = \begin{pmatrix} V_{11} & & \\ V_{21} & V_{22} & \\ V_{31} & V_{32} & V_{34} \end{pmatrix},$$

where, using calculations similar to those given in the proof of Theorem 2.1, we have

$$\underset{L\times L}{V_{11}} = \operatorname{diag}\left\{\tau_1 \frac{1-\rho_1}{\rho_1}, \dots, \tau_L \frac{1-\rho_L}{\rho_L}\right\},$$

$$\underset{L\times KL}{V_{12}} = (\underline{0}, \dots, \underline{0}),$$

$$\underset{KL\times KL}{V_{22}} = \operatorname{diag}\left\{\tau_1 \frac{\Sigma_1}{\rho_1}, \dots, \tau_L \frac{\Sigma_L}{\rho_L}\right\},$$

$$\underset{K\times L}{V_{31}} = \left(\tau_1 \frac{1-\rho_1}{\rho_1} q_1, \dots, \tau_L \frac{1-\rho_L}{\rho_L} q_L\right),$$

$$\underset{K\times KL}{V_{32}} = \left(\tau_1 \frac{\Sigma_1}{\rho_1}, \dots, \tau_L \frac{\Sigma_L}{\rho_L}\right),$$

$$\underset{K\times K}{V_{33}} = \sum_{l=1}^{L} \tau_l \left[\frac{\Sigma_l}{\rho_l} + \frac{1-\rho_l}{\rho_l} q_l q_l' + q_l q_l'\right].$$

We can partition the Jacobian matrix

$$\underset{L+KL+K\times L+M+J+K}{M} = \begin{pmatrix} M_{1\rho} & 0 & 0 & 0 \\ 0 & M_{2\lambda} & M_{2\delta} & 0 \\ M_{3\rho} & 0 & 0 & M_{3\beta} \end{pmatrix},$$

where

$$\underset{L\times L}{M_{1\rho}} = -\operatorname{diag}\left\{\frac{\tau_1}{\rho_1}, \dots, \frac{\tau_L}{\rho_L}\right\},$$

$$\underset{KL\times M}{M_{2\lambda}} = -(H_1', \dots, H_I')',$$

$$\underset{KL\times J}{M_{2\delta}} = -\begin{pmatrix} \tau_1 \nabla_\delta q_1 \\ \vdots \\ \tau_L \nabla_\delta q_L \end{pmatrix},$$

$$M_{3\rho} = -\begin{pmatrix} \tau_1 \frac{q_1}{\rho_1} & \cdots & \tau_L \frac{q_L}{\rho_L} \end{pmatrix},$$

$$M_{3\beta} = \Gamma,$$

where $H_i = \operatorname{diag}\{\tau_{i1}\nabla_h q_{i1}, \dots, \tau_{iM}\nabla_h q_{iM}\}$ for $i = 1, \dots, I$ with $q_{im} = q(x_{im}, \delta, h(x_{2,m}); \beta)$.

The variance bound for β is given by the lower right-hand $K \times K$ block of $(M'V^{-1}M)^{-1}$. We begin by calculating V^{-1}. Partition V as

$$V = \begin{pmatrix} B_{11} & B_{12} \\ B_{12}' & B_{22} \end{pmatrix},$$

with

$$\underset{L+KL\times L+KL}{B_{11}} = \operatorname{diag}\{V_{11} \quad V_{22}\}, \qquad \underset{L+KL\times K}{B_{12}} = (V_{31} \quad V_{32})', \qquad B_{22} = V_{33}.$$

Now partition V^{-1} as

$$V_0^{-1} = \begin{pmatrix} C_{11} & C_{12} \\ C_{12}' & C_{22} \end{pmatrix}, \tag{14}$$

where the partitioned inverse formula gives

$$\begin{aligned} \underset{L+KL\times L+KL}{C_{11}} &= \operatorname{diag}\{V_{11}^{-1} \quad V_{22}^{-1}\} + D'\mathbb{E}[\Psi]^{-1}D, \\ \underset{K\times L+KL}{C_{12}'} &= -\mathbb{E}[\Psi]^{-1}D, \qquad \underset{K\times K}{C_{22}} = \mathbb{E}[\Psi]^{-1}, \end{aligned}$$

with $D = (A' \;\; (\iota_L \otimes I_K)') = B_{12}'B_{11}^{-1}$ and $A = (q_1 \;\; \cdots \;\; q_L)'$ a $L \times K$ matrix.

Expression (14) follows since

$$\begin{aligned} C_{22} &= (B_{22} - B_{12}'B_{11}^{-1}B_{12})^{-1} \\ &= \sum_{l=1}^{L} \tau_l \left[\frac{\Sigma_l}{\rho_l} + \frac{1-\rho_l}{\rho_l} q_l q_l' + q_l q_l' \right] - \sum_{l=1}^{L} \tau_l \left[\frac{1-\rho_l}{\rho_l} q_l q_l' + \frac{\Sigma_l}{\rho_l} \right] \\ &= \left\{ \sum_{l=1}^{L} \tau_l q_l q_l' \right\}^{-1} \\ &= \mathbb{E}[\Psi]^{-1}. \end{aligned}$$

We also have $C_{12}' = -C_{22}B_{12}'B_{11}^{-1} = -\mathbb{E}[\Psi]^{-1}D$ and

$$C_{11} = B_{11}^{-1} + B_{11}^{-1}B_{12}C_{22}B_{12}'B_{11}^{-1} = \operatorname{diag}\{V_{11}^{-1} \quad V_{22}^{-1}\} + D'\mathbb{E}[\Psi]^{-1}D.$$

We now evaluate $\mathcal{I}_{\mathrm{m}}^{\mathrm{f}}(\theta) = M'V^{-1}M$ to

$$\begin{pmatrix} M_{1\rho}'V_{11}^{-1}M_{1\rho} & 0 \\ 0 & M_{2\lambda}'[V_{22}^{-1} + (\iota_L \otimes I_K)\mathbb{E}[\Psi]^{-1}(\iota_L \otimes I_K)']M_{2\lambda} \\ 0 & M_{2\delta}'[V_{22}^{-1} + (\iota_L \otimes I_K)\mathbb{E}[\Psi]^{-1}(\iota_L \otimes I_K)']M_{2\lambda} \\ 0 & -M_{3\beta}'\mathbb{E}[\Psi]^{-1}(\iota_L \otimes I_K)'M_{2\lambda} \end{pmatrix.$$
$$\begin{matrix} 0 \\ M_{2\lambda}'[V_{22}^{-1} + (\iota_L \otimes I_K)\mathbb{E}[\Psi]^{-1}(\iota_L \otimes I_K)']M_{2\delta} \\ M_{2\delta}'[V_{22}^{-1} + (\iota_L \otimes I_K)\mathbb{E}[\Psi]^{-1}(\iota_L \otimes I_K)']M_{2\delta} \\ -M_{3\beta}'\mathbb{E}[\Psi]^{-1}(\iota_L \otimes I_K)'M_{2\delta} \end{matrix}$$
$$\left. \begin{matrix} 0 \\ -M_{2\lambda}'(\iota_L \otimes I_K)\mathbb{E}[\Psi]^{-1}M_{3\beta} \\ -M_{2\delta}'(\iota_L \otimes I_K)\mathbb{E}[\Psi]^{-1}M_{3\beta} \\ M_{3\beta}'\mathbb{E}[\Psi]^{-1}M_{3\beta} \end{matrix} \right),$$

where I have made use of the equality $M'_{1\rho}A = M'_{3\rho}$.

Observe that, as in the standard semiparametric missing data model, $\mathcal{I}_{\mathrm{m}}^{\mathrm{f}}(\theta)$ satisfies Stein's condition for redundancy of knowledge of the propensity score for β. However, the structure of the bound does indicate that knowledge of the finite-dimensional parameters and nonparametric portions of the CEF of $\psi(Z,\beta)$ given X does increase the precision with which β can be estimated.

The variance bound for β_0 is given by the lower right-hand $K \times K$ block of the inverse of this matrix. Because of block diagonality, we only need to consider the lower right-hand block. Partition this block as

$$\begin{pmatrix} B_{11} & B_{12} \\ B'_{12} & B_{22} \end{pmatrix},$$

where B_{11}, B_{12}, and B_{22} are redefined to equal

$$\begin{aligned}
B_{11} &= \begin{pmatrix} M'_{2\lambda}V_{22}^{-1}M_{2\lambda} & M'_{2\lambda}V_{22}^{-1}M_{2\delta} \\ M'_{2\delta}V_{22}^{-1}M_{2\lambda} & M'_{2\delta}V_{22}^{-1}M_{2\delta} \end{pmatrix} \\
&\quad + \begin{pmatrix} M'_{2\lambda}(\iota_L \otimes I_K) \\ M'_{2\delta}(\iota_L \otimes I_K) \end{pmatrix} \mathbb{E}[\Psi]^{-1} \left((\iota_L \otimes I_K)'M_{2\lambda} \quad (\iota_L \otimes I_K)'M_{2\delta} \right) \\
B_{12} &= \begin{pmatrix} M'_{2\lambda}(\iota_L \otimes I_K) \\ M'_{2\delta}(\iota_L \otimes I_K) \end{pmatrix} \mathbb{E}[\Psi]^{-1}M_{3\beta}, \\
B_{33} &= M'_{3\beta}\mathbb{E}[\Psi]^{-1}M_{3\beta}.
\end{aligned}$$

The information bound is, therefore, given by

$$\begin{aligned}
\mathcal{I}_{\mathrm{m}}^{\mathrm{f}}(\theta) &= B_{22} - B'_{12}B_{11}^{-1}B_{12} \\
&= M'_{3\beta}\mathbb{E}[\Psi]^{-1}M_{3\beta} \\
&\qquad - M'_{3\beta}\mathbb{E}[\Psi_0]^{-1} \left((\iota_L \otimes I_K)'M_{2\lambda} \quad (\iota_L \otimes I_K)'M_{2\delta} \right) \\
&\qquad \times \Bigg\{ \begin{pmatrix} M'_{2\lambda}V_{22}^{-1}M_{2\lambda} & M'_{2\lambda}V_{22}^{-1}M_{2\delta} \\ M'_{2\delta}V_{22}^{-1}M_{2\lambda} & M'_{2\delta}V_{22}^{-1}M_{2\delta} \end{pmatrix} \\
&\qquad + \begin{pmatrix} M'_{2\lambda}(\iota_L \otimes I_K) \\ M'_{2\delta}(\iota_L \otimes I_K) \end{pmatrix} \mathbb{E}[\Psi]^{-1} \\
&\qquad \times \left((\iota_L \otimes I_K)'M_{2\lambda} \quad (\iota_L \otimes I_K)'M_{2\delta} \right) \Bigg\}^{-1} \\
&\qquad \times \begin{pmatrix} M'_{2\lambda}(\iota_L \otimes I_K) \\ M'_{2\delta}(\iota_L \otimes I_K) \end{pmatrix} \mathbb{E}[\Psi]^{-1}M_{3\beta}
\end{aligned}$$

$$= M'_{3\beta}\left[\mathbb{E}[\Psi] + \left(\,(\iota_L \otimes I_K)'M_{2\lambda} \quad (\iota_L \otimes I_K)'M_{2\delta}\,\right)\right.$$
$$\left.\times \begin{pmatrix} M'_{2\lambda}V_{22}^{-1}M_{2\lambda} & M'_{2\lambda}V_{22}^{-1}M_{2\delta} \\ M'_{2\delta}V_{22}^{-1}M_{2\lambda} & M'_{2\delta}V_{22}^{-1}M_{2\delta} \end{pmatrix}^{-1} \begin{pmatrix} M'_{2\lambda}(\iota_L \otimes I_K) \\ M'_{2\delta}(\iota_L \otimes I_K) \end{pmatrix}\right]^{-1}$$
$$\times M_{3\beta},$$

where I have used the identity $A^{-1} - A^{-1}U(B^{-1} + U'A^{-1}U)^{-1}U'A^{-1} = (A + UBU')^{-1}$.

Using the partitioned inverse formula and multiplying out the expression in $[\cdot]$ above then gives

$$\mathcal{I}_{\mathrm{m}}^{\mathrm{f}}(\theta) = M'_{3\beta} \times \Big[\mathbb{E}[\Psi] + (\iota_L \otimes I_K)'\big[M_{2\lambda}(M'_{2\lambda}V_{22}^{-1}M_{2\lambda})M'_{2\lambda}$$
$$+ (M_{2\delta} - M_{2\lambda}(M'_{2\lambda}V_{22}^{-1}M_{2\lambda})^{-1}M'_{2\lambda}V_{22}^{-1}M_{2\delta})$$
$$\times (M'_{2\delta}V_{22}^{-1}M_{2\delta} - M'_{2\delta}V_{22}^{-1}M_{2\lambda}(M'_{2\lambda}V_{22}^{-1}M_{2\lambda})^{-1}M'_{2\lambda}V_{22}^{-1}M_{2\delta})^{-1}$$
$$\times (M_{2\delta} - M_{2\lambda}(M'_{2\lambda}V_{22}^{-1}M_{2\lambda})^{-1}M'_{2\lambda}V_{22}^{-1}M_{2\delta})'\big](\iota_L \otimes I_K)\Big]^{-1}$$
$$\times M_{3\beta}.$$

We can now use the explicit expressions for V_0 and M_0 given above to generate an interpretable bound. The required calculations are tedious but straightforward (details are available from the author on request); they give an information bound of $\mathcal{I}_{\mathrm{m}}^{\mathrm{f}}(\theta)$ as defined in the main text of the paper.

C. PROOF OF THEOREM 3.2

In calculating the efficiency bound for the semiparametric missing data model defined by restriction (1) and Assumptions 1.1–1.5 above, I follow the general approach outlined by Bickel, Klaassen, Ritov, and Wellner (1993) and, especially, Newey (1990, Section 3). First, I characterize the nuisance tangent space. Second, I demonstrate pathwise differentiability of the parameter of interest, β. The efficient influence function for this model equals the projection of the pathwise derivative onto the tangent space. In the present example, the direct calculation of this projection appears to be particularly difficult. However, inspection of the variance bound associated with the conditional moment problem defined by restrictions (4), (5), and (6) provides a conjecture for the form of the efficient influence function. The third and final step of the proof therefore involves demonstrating that (i) this conjectured influence function lies in the model tangent space and (ii) that it is indeed the required projection (i.e., that it satisfies equation (9) in Newey (1990, p. 106)). The result then follows from an application of Theorem 3.1 in Newey (1990).

STEP 1—Characterization of the Nuisance Tangent Space: Recalling that $Y = DY_1$, the joint density function for (Y, X, D), making use of Assumption 1.3, is given by

$$f(y, x, d) = f(y_1|x)^d p(x)^d [1 - p(x)]^{1-d} f(x).$$

Assumption 1.5 also requires that $f(y_1|x)$ satisfy the restriction

$$\int \rho(z, \delta_0, h_0(x_2); \beta_0) f(y_1|x)\, dy_1 = 0,$$

where $\psi(z, \beta) = \psi(x, y_1, \beta)$ and

$$\rho(z, \delta, h(x_2); \beta) = \psi(x, y_1, \beta) - q(x, \delta, h(x_2); \beta).$$

Consider a regular parametric submodel with $f(y, x, d; \eta) = f(y, x, d)$ at $\eta = \eta_0$. The submodel joint density is given by

$$f(y, x, d; \eta) = f(y_1|x; \eta)^d p(x; \eta)^d [1 - p(x; \eta)]^{1-d} f(x; \eta) \tag{15}$$

and satisfies the restriction

$$\int \rho(z, \delta(\eta), h(x_2; \eta); \beta_0) f(y_1|x; \eta)\, dy_1 = 0. \tag{16}$$

The submodel score vector equals

$$\begin{aligned} s_\eta(y, x, d; \eta) = {} & d s_\eta(y_1|x; \eta) + \frac{d - p(x; \eta)}{p(x; \eta)[1 - p(x; \eta)]} \nabla_\eta p(x; \eta) \\ & + t_\eta(x; \eta), \end{aligned} \tag{17}$$

where

$$s_\eta(y, x, d; \eta) = \nabla_\eta \log f(y, x, d; \eta),$$
$$s_\eta(y_1|x; \eta) = \nabla_\eta \log f(y_1|x; \eta), \quad t_\eta(x; \eta) = \nabla_\eta \log f(x; \eta).$$

By the usual mean zero property of (conditional) scores, we have

$$\mathbb{E}[s_\eta(Y_1|X)|X] = \mathbb{E}[t_\eta(X)] = 0, \tag{18}$$

where suppression of η in a function means that it is evaluated at its population value (e.g., $t_\eta(x) = t_\eta(x; \eta_0)$).

Condition (16) imposes additional restrictions on $s_\eta(Y_1|X)$ beyond conditional mean-zeroness. To see the structure of these restrictions, differentiate

(16) with respect to η through the integral and evaluate the result at $\eta = \eta_0$:

$$\begin{aligned}&\frac{\partial q_0(X)}{\partial \delta'}\frac{\partial \delta(\eta_0)}{\partial \eta'} + \frac{\partial q_0(X)}{\partial h'}\frac{\partial h(X_2; \eta_0)}{\partial \eta'}\\ &\quad = \mathbb{E}\big[\rho(Z, \delta_0, h_0(X_2); \beta_0) s_\eta(Y_1|X)'|X\big].\end{aligned}$$

The conditional covariance between $\rho(Z, \delta_0, h_0(X_2); \beta_0)$ and $s_\eta(Y_1|X)$ has a particular structure induced by the semiparametric restrictions on the form of $\mathbb{E}[\psi(Z, \beta)|x]$.

From (17), (18), and the above equality, the tangent set is evidently

$$\mathcal{T} = \big\{ds(y_1|x) + a(x)[d - p(x)] + t(x)\big\}, \tag{19}$$

where $a(x)$ is unrestricted, and $t(x)$ and $s(y_1|x)$ satisfy

$$\begin{aligned}&\mathbb{E}[t(X)] = 0,\\ &\mathbb{E}[s(Y_1|X)|X] = 0,\\ &\mathbb{E}\big[\rho(Z, \delta_0, h_0(X_2); \beta_0)s(Y_1|X)'|X\big]\\ &\quad = \left(\frac{\partial q_0(X)}{\partial \delta'}\right)c + \left(\frac{\partial q_0(X)}{\partial h'}\right)k(X_2),\end{aligned}$$

with c a constant matrix and $k(x_2)$ an unrestricted matrix-valued function of x_2.

STEP 2—Demonstration of Pathwise Differentiability: Under the parametric submodel, $\beta(\eta)$ is identified by the unconditional moment restriction

$$\mathbb{E}_\eta\big[\psi(Z; \beta(\eta))\big] = 0.$$

Differentiating under the integral and evaluating at $\eta = \eta_0$ gives

$$\frac{\partial \beta(\eta_0)}{\partial \eta'} = -\Gamma_0^{-1}\mathbb{E}\left[\psi(Z; \beta_0)\frac{\partial \log f(Y_1, X; \eta_0)'}{\partial \eta'}\right].$$

To demonstrate pathwise differentiability of β, we require $F(Y, X, D)$ such that

$$\frac{\partial \beta(\eta_0)}{\partial \eta'} = \mathbb{E}[F(Y, X, D)s_\eta(Y, X, D)'].$$

It is easy to verify that the function

$$\begin{aligned}F(Y, X, D) = -\Gamma_0^{-1}&\left\{\frac{D}{p_0(X)}\rho(Z, \delta_0, h_0(X_2); \beta_0)\right\}\\ &+ q(X, \delta_0, h_0(X_2); \beta_0)\end{aligned}$$

satisfies this condition (cf. Hahn (1998)).

STEP 3—Verification That the Conjectured Efficient Influence Function Equals the Required Projection: Inspection of the variance bounds associated with the conditional moment problem suggests the candidate efficient influence given by (7) in the main text. I first verify that $\phi_\beta^{\mathrm{f}}(Z,\eta_0,\beta_0)$ lies in the model tangent space. The last term in (7) plays the role of $t(x)$; a zero plays the role of $a(x)[d-p(x)]$. Finally, the first two terms in (7) play the role of $ds(y_1|x)$. To see this note that in addition to being both conditionally mean zero, we have

$$\begin{aligned}\mathbb{E}\Bigg[\rho(Z;\beta_0)\Bigg\{&H_0(X_2)Y_0^h(X_2)^{-1}\left(\frac{\partial q_0(X)}{\partial h'}\right)'\Sigma_0(X)^{-1}\rho(Z;\beta_0)\\&+\mathbb{E}[G_0(X)]\mathcal{I}_{\mathrm{m}}^{\mathrm{f}}(\delta_0)^{-1}G_0(X)'\Sigma(X)^{-1}\rho(Z;\beta_0)\Bigg\}'\Bigg|X\Bigg]\\=\left(\frac{\partial q_0(X)}{\partial\delta'}\right)&c+\left(\frac{\partial q_0(X)}{\partial h'}\right)k(X_2)\end{aligned}$$

with

$$\begin{aligned}c&=\mathcal{I}_{\mathrm{m}}^{\mathrm{f}}(\delta_0)^{-1}\mathbb{E}[G_0(X)]',\\k(X_2)&=Y_0^h(X_2)^{-1}\{H_0(X_2)'-Y_0^{h\delta}(X_2)c\}.\end{aligned}$$

The candidate efficient influence function therefore belongs to the model tangent space as required.

I next show that $\phi_\beta^{\mathrm{f}}(Z,\eta_0,\beta_0)$ is indeed the required projection by verifying that it satisfies

$$\mathbb{E}\big[\{F(Y,X,D)-\phi_\beta^{\mathrm{f}}(Z,\eta_0,\beta_0)\}\mathrm{t}'\big]=0\quad\text{for all }\mathrm{t}\in\mathcal{T}$$

(cf. equation (9) in Newey (1990, p. 106)). We have

$$\begin{aligned}&F(Y,X,D)-\phi_\beta^{\mathrm{f}}(Z,\eta_0,\beta_0)\\&\quad=\Gamma_0^{-1}D\Bigg\{\frac{1}{p_0(X)}-H_0(X_2)Y_0^h(X_2)^{-1}\left(\frac{\partial q_0(X)}{\partial h'}\right)'\Sigma(X)^{-1}\\&\qquad-\mathbb{E}[G_0(X)]\mathcal{I}_{\mathrm{m}}^{\mathrm{f}}(\delta_0)^{-1}G_0(X)'\Sigma(X)^{-1}\Bigg\}\rho(Z;\beta_0).\end{aligned}$$

By the conditional independence of Y_1 and D given X (Assumption 1.3), and by conditional mean-zeroness of $\rho(Z;\beta_0)$, it is easy to show that $F(Y,X,D)-$

$\phi^{\mathrm{f}}_{\beta}(Z, \eta_0, \beta_0)$ is orthogonal to any functions of the form $a(x)[d - p(x)]$ and $t(x)$. All that remains is to show orthogonality with $ds(y_1|x)$. We have

$$\begin{aligned}
&\mathbb{E}\big[\{F(Y, X, D) - \phi^{\mathrm{f}}_{\beta}(Z, \eta_0, \beta_0)\}Ds(Y_1|X)'\big] \\
&\quad = \mathbb{E}\left[\Gamma_0^{-1}\left\{I_K - H_0(X_2)\Upsilon_0^h(X_2)^{-1}\left(\frac{\partial q_0(X)}{\partial h'}\right)'\left(\frac{\Sigma(X)}{p(X)}\right)^{-1}\right.\right. \\
&\quad\quad \left. - \mathbb{E}[G_0(X)]\mathcal{I}^{\mathrm{f}}_{\mathrm{m}}(\delta_0)^{-1}G_0(X)'\left(\frac{\Sigma(X)}{p(X)}\right)^{-1}\right\} \\
&\quad\quad \left. \times \left\{\left(\frac{\partial q_0(X)}{\partial \delta'}\right)c + \left(\frac{\partial q_0(X)}{\partial h'}\right)k(X_2)\right\}\right],
\end{aligned}$$

where I have made use of the special structure of the conditional covariance $\mathbb{E}[\rho(Z; \beta_0)s_{\eta}(Y_1|X)'|X]$. Multiplying out terms yields

$$\begin{aligned}
&\mathbb{E}\big[\{F(Y, X, D) - \phi^{\mathrm{f}}_{\beta}(Z, \eta_0, \beta_0)\}Ds(Y_1|X)\big] \\
&\quad = \Gamma_0^{-1}\mathbb{E}\left[\frac{\partial q_0(X)}{\partial \delta'}c + H_0(X_2)k(X_2)\right. \\
&\quad\quad - H_0(X_2)\Upsilon_0^h(X_2)^{-1}\Upsilon_0^{h\delta}(X_2)c - H_0(X_2)k(X_2) \\
&\quad\quad - \mathbb{E}[G_0(X)]\mathcal{I}^{\mathrm{f}}_{\mathrm{m}}(\delta_0)^{-1}\Upsilon_0^{\delta}(X_2)c \\
&\quad\quad + \mathbb{E}[G_0(X)]\mathcal{I}^{\mathrm{f}}_{\mathrm{m}}(\delta_0)^{-1}\Upsilon_0^{h\delta}(X_2)'\Upsilon_0^h(X_2)^{-1}\Upsilon_0^{h\delta}(X_2)c \\
&\quad\quad - \mathbb{E}[G_0(X)]\mathcal{I}^{\mathrm{f}}_{\mathrm{m}}(\delta_0)^{-1}\Upsilon_0^{h\delta}(X_2)'k(X_2) \\
&\quad\quad \left. + \mathbb{E}[G_0(X)]\mathcal{I}^{\mathrm{f}}_{\mathrm{m}}(\delta_0)^{-1}\Upsilon_0^{h\delta}(X_2)'k(X_2)\right] \\
&\quad = \Gamma_0^{-1}\big\{\mathbb{E}[G_0(X)]c - \mathbb{E}[G_0(X)]c\big\} = 0,
\end{aligned}$$

where

$$\Upsilon_0^{\delta}(X_2) = \mathbb{E}\left[D\left(\frac{\partial_{q_0}(X)}{\partial \delta'}\right)'\Sigma_0(X)^{-1}\left(\frac{\partial_{q_0}(X)}{\partial \delta'}\right)\Big| X_2\right]$$

and the first equality follows from iterated expectations and the second equality follows from the definitions of $G_0(X)$ and $\mathcal{I}^{\mathrm{f}}_{\mathrm{m}}(\delta_0)$ in the main text.

The result then follows from an application of Theorem 3.1 in Newey (1990).

D. PROOF OF PROPOSITION 3.1

The difference in the variance bounds is given by

$$\mathcal{I}_{\mathrm{m}}(\beta_0)^{-1} - \mathcal{I}^{\mathrm{f}}_{\mathrm{m}}(\beta_0)^{-1} = \Gamma_0^{-1}(\Lambda_0 - \Xi_0)\Gamma_0^{-1\prime},$$

with Λ_0 and Ξ_0 as defined in the main text.

First observe that $\mathbb{E}[G_0(X)]$ has the covariance representation

$$\begin{aligned}\mathbb{E}[G_0(X)] &= \mathbb{E}\left[\frac{\partial q_0(X)}{\partial \delta'} - \left(\frac{\partial q_0(X)}{\partial h'}\right)\Upsilon_0^h(X_2)^{-1}\Upsilon_0^{h\delta}(X_2)\right] \\ &= \mathbb{C}(\xi_1, \xi_2'),\end{aligned}$$

with ξ_1 and ξ_2 as defined in the main text. This follows since

$$\begin{aligned}&\mathbb{E}\left[\frac{D}{p_0(X)}\rho(Z;\beta_0)\rho(Z;\beta_0)'\Sigma_0(X)^{-1}\right. \\ &\qquad \left.\times\left\{\frac{\partial q_0(X)}{\partial \delta'} - \left(\frac{\partial q_0(X)}{\partial h'}\right)\Upsilon_0^h(X_2)^{-1}\Upsilon_0^{h\delta}(X_2)\right\}\right] \\ &\quad = \mathbb{E}\left[\frac{\partial q_0(X)}{\partial \delta'} - \left(\frac{\partial q_0(X)}{\partial h'}\right)\Upsilon_0^h(X_2)^{-1}\Upsilon_0^{h\delta}(X_2)\right]\end{aligned}$$

and also

$$\begin{aligned}&\mathbb{E}\left[DH_0(X_2)\Upsilon_0^h(X_2)^{-1}\left(\frac{\partial q_0(X)}{\partial h'}\right)'\Sigma_0(X)^{-1}\rho(Z;\beta_0)\rho(Z;\beta_0)'\Sigma_0(X)^{-1}\right. \\ &\quad \left.\times\left\{\frac{\partial q_0(X)}{\partial \delta'} - \left(\frac{\partial q_0(X)}{\partial h'}\right)\Upsilon_0^h(X_2)^{-1}\Upsilon_0^{h\delta}(X_2)\right\}\right] = 0.\end{aligned}$$

Similar calculations yield the variance representations

$$\begin{aligned}\mathbb{V}(\xi_1) &= \mathbb{E}\left[\frac{\Sigma_0(X)}{p_0(X)} - H_0(X_2)\Upsilon_0^h(X_2)^{-1}H_0(X_2)'\right], \\ \mathbb{V}(\xi_2) &= \mathbb{E}[DG_0(X)'\Sigma_0(X)^{-1}G_0(X)],\end{aligned}$$

with the result directly following.

E. PROOF OF PROPOSITION 3.2

Part (i) follows from Theorem 3.2. For part (ii), condition (a) implies the equality

$$\begin{aligned}&\mathbb{E}\left[DH_0(X_2)\Upsilon_0^h(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)'\Sigma_0(X)^{-1}\rho_*(Z;\beta_0)\middle| X_2\right] \\ &\quad = H_0(X_2)\mathbb{E}\left[\left(\frac{\partial q_*(X)}{\partial h'}\right)'\Sigma_0(X)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\middle| X_2\right]^{-1} \\ &\qquad \times \mathbb{E}\left[\left(\frac{\partial q_*(X)}{\partial h'}\right)'\Sigma_0(X)^{-1}\rho_*(Z;\beta_0)\middle| X_2\right].\end{aligned}$$

Condition (b) implies that $\Sigma_0(X) = \Phi_0(X_2)$. Let $L(X_2)L(X_2)' = \Phi_0(X_2)$ be the Cholesky decomposition of $\Phi_0(X_2)$. This implies that the term to the right of the last equality equals

$$H_0(X_2)\mathbb{E}\left[\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}'\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}\Big|X_2\right]^{-1}$$
$$\times\mathbb{E}\left[\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}'L(X_2)^{-1}\rho_*(Z;\beta_0)\Big|X_2\right].$$

Since all expectations in the above expression condition on X_2, $L(X_2)$ may be treated as nonstochastic so that

$$L(X_2)^{-1}H_0(X_2) = \mathbb{E}\left[L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\Big|X_2\right].$$

Recall that a linear predictor passes through the mean of the outcome variable at the means of the predictor variables (when a constant is included). Condition (c) implies that each row of $\partial q_*(X)/\partial h'$ includes such a constant and hence that

$$\begin{aligned}
&L(X_2)^{-1}\mathbb{E}[\rho_*(Z;\beta_0)|X_2]\\
&\quad= L(X_2)^{-1}H_0(X_2)\\
&\qquad\times\mathbb{E}\left[\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}'\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}\Big|X_2\right]^{-1}\\
&\qquad\times\mathbb{E}\left[\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}'\{L(X_2)^{-1}\rho(Z;\beta_0)\}\Big|X_2\right]
\end{aligned}$$

and, therefore, that

$$\begin{aligned}
&\mathbb{E}\left[DH_0(X_2)Y_0^h(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)'\Sigma_0(X)^{-1}\rho_*(Z;\beta_0)\Big|X_2\right]\\
&\quad= \mathbb{E}[\rho_*(Z;\beta_0)] = -\mathbb{E}[q_*(X)].
\end{aligned}$$

This implies that the first part of $\phi_\beta^{\mathrm{f}}(Z,\eta_*,\beta_0)$ has mean $-\mathbb{E}[q_*(X)]$.

Using conditions (a)–(c) and arguments analogous to those given immediately above, we have

$$\begin{aligned}
&L(X_2)^{-1}G_0(X)\\
&\quad= L(X_2)^{-1}\frac{\partial q_0(X)}{\partial\delta'}\\
&\qquad- L(X_2)^{-1}\left(\frac{\partial q_0(X)}{\partial h'}\right)
\end{aligned}$$

$$\times\,\mathbb{E}\left[\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}'\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}\Big|X_2\right]^{-1}$$
$$\times\,\mathbb{E}\left[\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}'\left\{L(X_2)^{-1}\frac{\partial q_0(X)}{\partial \delta'}\right\}\Big|X_2\right],$$

so that $\mathbb{E}[L(X_2)^{-1}G_0(X)|X_2] = L(X_2)^{-1}\mathbb{E}[G_0(X)|X_2] = 0$. The law of iterated expectations then gives $\mathbb{E}[G_0(X)] = 0$. This implies that the second part of $\phi^{\mathrm{f}}_{\beta}(Z, \eta_*, \beta_0)$ is mean zero. The third part of $\phi^{\mathrm{f}}_{\beta}(Z, \eta_*, \beta_0)$ has mean $\mathbb{E}[q_*(X)]$. The result follows as claimed.

F. PROOF OF COROLLARY 3.1

From the proof to Proposition 3.2, we have $\mathbb{E}[G_0(X)] = 0$. So the result follows if

$$\mathbb{E}\left[\frac{\Sigma_0(X)}{p_0(X)} - H_0(X_2)\Upsilon^h_0(X_2)^{-1}H_0(X_2)'\right] = 0.$$

Under conditions (a) and (b) of Proposition 3.2(ii), we have

$$\begin{aligned}
&\mathbb{E}\left[\frac{\Sigma_0(X)}{p_0(X)} - H_0(X_2)\Upsilon^h_0(X_2)^{-1}H_0(X_2)'\right]\\
&= \mathbb{E}\Bigg[\frac{\Phi_0(X_2)}{e_0(X_2)} - \frac{1}{e_0(X_2)}H_0(X_2)\\
&\quad\times \mathbb{E}\left[\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}'\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}\Big|X_2\right]^{-1}\\
&\quad\times H_0(X_2)'\Bigg]\\
&= \mathbb{E}\Bigg[\frac{\Phi_0(X_2)}{e_0(X_2)} - \frac{L(X_2)}{e_0(X_2)}\mathbb{E}\left[L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\Big|X_2\right]\\
&\quad\times \mathbb{E}\left[\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}'\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}\Big|X_2\right]^{-1}\\
&\quad\times \mathbb{E}\left[L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\Big|X_2\right]'L(X_2)'\Bigg],
\end{aligned}$$

where $L(X_2)L(X_2)' = \Phi_0(X_2)$ as above. Observe that

$$\mathbb{E}\left[L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\Big|X_2\right]$$
$$\times\,\mathbb{E}\left[\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}'\left\{L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\right\}\Big|X_2\right]^{-1}$$
$$\times\,\mathbb{E}\left[L(X_2)^{-1}\left(\frac{\partial q_*(X)}{\partial h'}\right)\Big|X_2\right]'$$

is equal to the multivariate conditional linear predictor of the $K \times K$ identity matrix given $L(X_2)^{-1}(\frac{\partial q_*(X)}{\partial h'})$ evaluated at $\mathbb{E}[L(X_2)^{-1}(\frac{\partial q_*(X)}{\partial h'})|X_2]$; therefore this object equals I_K and we have

$$\mathbb{E}\left[\frac{\Sigma_0(X)}{p_0(X)} - H_0(X_2)Y_0^h(X_2)^{-1}H_0(X_2)'\right]$$
$$= \mathbb{E}\left[\frac{\Phi_0(X_2)}{e_0(X_2)} - \frac{L(X_2)L(X_2)'}{e_0(X_2)}\right] = 0,$$

as required.

REFERENCES

BICKEL, P. J., C. A. J. KLAASSEN, Y. RITOV, AND J. A. WELLNER (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. New York: Springer-Verlag. [9]

CHAMBERLAIN, G. (1987): "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334. [2]

——— (1992): "Comment: Sequential Moment Restrictions in Panel Data," *Journal of Business & Economic Statistics*, 10, 20–26. [1]

HAHN, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66, 315–331. [12]

NEWEY, W. K. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99–135. [9,12,13]

Dept. of Economics, New York University, 19 West 4th Street 6FL, New York, NY 10012, U.S.A. and National Bureau of Economic Research; bryan.graham@nyu.edu.

Manuscript received August, 2007; final revision received June, 2010.