Supplemental Appendix: GMM with Multiple Missing Variables

Appendix B gives further details on the series estimators used in the paper. Appendix C collects proofs of all the theoretical results and claims. The notations are the same as in the original paper, and the equations are also numbered accordingly in the same order.

B Appendix: Series estimation of the nuisance parameters

For a variable U, consider $P_K(U) = (P_K^1(U), \dots, P_K^K(U))'$ that is a K-truncation of some approximating series such that the minimum eigenvalue of $E[P_K(U)P_K(U)']$ is bounded away from 0 uniformly in K. Let $R_K(U) := (E[P_K(U)P_K(U)'])^{-1/2}P_K(U)$ be a standardization for $P_K(U)$. See Newey (1997) for details. For estimation of $p(W)$ and $q(W;\beta)$ we take $U = W$, while for $q(Z_1, W;\beta)$ and $q(Z_2, W;\beta)$ we take $U = (Z_1', W')'$ and $U = (Z_2', W')'$ respectively.

$\widehat{p}(W) = (\widehat{p}_{00}(W), \widehat{p}_{10}(W), \widehat{p}_{01}(W), \widehat{p}_{11}(W))$ is estimated by multinomial series logit as:

$$\widehat{p}_{d_1d_2}(w) := L_{d_1d_2}(R_K(w), \widehat{\pi}_K) \text{ where } L_{d_1d_2}(R_K(w), \pi_K) = \frac{\exp[R_K(w)'\pi_{(d_1d_2)K}]}{\sum_{j,l=0,1}\exp[R_K(w)'\pi_{(jl)K}]}$$

for $d_1, d_2 = 0, 1$. $\widehat{p}_1(w) = \widehat{p}_{11}(w) + \widehat{p}_{10}(w)$ and $\widehat{p}_2(w) = \widehat{p}_{11}(w) + \widehat{p}_{01}(w)$. See Hirano et al. (2003) and Cattaneo (2010). For $\pi_K = [\pi'_{(00)K}, \pi'_{(10)K}, \pi'_{(01)K}, \pi'_{(11)K}]'$ the estimates are:

$$\widehat{\pi}_K \quad := \quad \arg \max_{\pi_K \in \mathbb{R}^{4d_K} | \pi_{(00)K} = 0} \sum_{i=1}^{N} \sum_{d_1, d_2 = 0,1} 1(D_{1i} = d_1, D_{2i} = d_2) \log L_{d_1d_2}(R_K(W_i), \pi_K).$$

Series estimators for the elements of $q(\beta)$ are obtained following Newey (1997) and Cattaneo (2010):

$$\begin{aligned} \widehat{q}(w;\beta) &= (I_{d_g} \otimes R_K(w)')\widehat{\gamma}_K(\beta), \\ \widehat{q}(z_j, w;\beta) &= (I_{d_g} \otimes R_K(u_j)')\widehat{\gamma}_K^{(j)}(\beta) \text{ where } U_j = (Z_j', W')' \text{ for } j = 1, 2. \end{aligned}$$

For $\gamma_K(\beta) = [\gamma_{K1}(\beta)', \dots, \gamma_{Kd_g}(\beta)']'$ and $\gamma_K^{(j)}(\beta) = [\gamma_{K1}^{(j)}(\beta)', \dots, \gamma_{Kd_g}^{(j)}(\beta)']'$ the estimates are:

$$\begin{aligned} \widehat{\gamma}_K(\beta) &:= \arg \min_{\gamma \in \mathbb{R}^{Kd_g}} \frac{1}{N}\sum_{i=1}^{N} D_{1i}D_{2i}\left(g(Z_i, W_i;\beta) - (I_{d_g} \otimes R_K(W_i)')\gamma\right)'\left(g(Z_i, W_i;\beta) - (I_{d_g} \otimes R_K(W_i)'\gamma)\right), \\ \widehat{\gamma}_K^{(j)}(\beta) &:= \arg \min_{\gamma \in \mathbb{R}^{Kd_g}} \frac{1}{N}\sum_{i=1}^{N} D_{1i}D_{2i}\left(g(Z_i, W_i;\beta) - (I_{d_g} \otimes R_K(U_{ji})')\gamma\right)'\left(g(Z_i, W_i;\beta) - (I_{d_g} \otimes R_K(U_{ji})'\gamma)\right), \end{aligned}$$

for $j = 1, 2$. Dividing $D_{1i}D_{2i}$ by $\widehat{p}_{11}(W_i)$ in the definitions for $\widehat{\gamma}_K(\beta)$ and $\widehat{\gamma}_K^{(j)}(\beta)$ often leads to improvement in the finite-sample properties of the AIPW estimators [see Hirano and Imbens (2001)]. Similarly, instead of using $\sum_{i=1}^N D_{1i}D_{2i}$ observations for all, improvement is possible in the P-case by taking advantage of the partition in (5) to estimate the elements of $\widehat{\gamma}_K(\beta)$ more precisely by using $\sum_{i=1}^N D_{1i}$ observations for the first Kd_{g_1} elements and $\sum_{i=1}^N D_{2i}$ observations for the last Kd_{g_2} elements. We follow this in Sections 4 and 5.

C Appendix: Proof of main results

Proof of Proposition 2.1:

Let f and F denote the density and distribution functions, with the concerned random variables specified inside parentheses. $L_0^2(F)$ denotes the space of mean-zero, square integrable functions with respect to F. The proof consists of three standard steps. (1) Obtain the tangent set for all regular parametric submodels satisfying the semiparametric assumptions on the observed data. (2) Conjecture the efficient influence function and then show pathwise differentiability of β^0 and verify that the efficient influence function lies in the tangent set. (3) Obtain the efficiency bound as the expectation of the outer product of the efficient influence function.

STEP - 1: Consider a regular parametric sub-model indexed by a finite-dimensional parameter θ for the joint distribution of the observed data $\mathcal{O} := (D_1, D_2, D_1 Z_1, D_2 Z_2, W)$. So the joint density $f_\theta(\mathcal{O})$ of the observed data can be expressed in terms of the full data (D_1, D_2, Z_1, Z_2, W) as

$$[p_{\theta,11}(W) f_\theta(Z_1, Z_2|W)]^{D_1 D_2} [p_{\theta,10}(W) f_\theta(Z_1|W)]^{D_1(1-D_2)} [p_{\theta,01}(W) f_\theta(Z_2|W)]^{(1-D_1)D_2} [p_{\theta,00}(W)]^{(1-D_1)(1-D_2)} f_\theta(W)$$

where (4) gives the factorization of the first three terms. The score with respect to θ is

$$S_\theta(\mathcal{O}) = D_1 D_2 s_\theta(Z_1, Z_2|W) + \sum_{j \neq k = 1,2} D_j(1 - D_k) s_\theta(Z_j|W) + s_\theta(W) + \sum_{d_1, d_2 = 0,1} \mathbf{1}(D_1 = d_1, D_2 = d_2) \frac{\dot{p}_{\theta,d_1 d_2}(W)}{p_{\theta,d_1 d_2}(W)}.$$

where $s_\theta(Z_1, Z_2|W) := \frac{\partial}{\partial\theta} \log f_\theta(Z_1, Z_2|W)$, $s_\theta(Z_1|W) := \frac{\partial}{\partial\theta} \log f_\theta(Z_1|W)$, $s_\theta(Z_2|W) := \frac{\partial}{\partial\theta} \log f_\theta(Z_2|W)$, $s_\theta(W) := \frac{\partial}{\partial\theta} \log f_\theta(W)$, and $\dot{p}_{\theta,d_1 d_2}(W) := \frac{\partial}{\partial\theta} p_{\theta,d_1 d_2}(W)$ for $d_1, d_2 = 0, 1$. Henceforth, we omit the subscript θ from quantities evaluated at $\theta = \theta_0$.

The tangent set for the model is characterized by functions of the form:

$$\mathcal{T} := D_1 D_2 a(Z_1, Z_2, W) + \sum_{j \neq k = 1,2} D_j(1 - D_k) a_j(Z_j, W) + a_0(W) + \sum_{d_1, d_2 = 0,1} \mathbf{1}(D_1 = d_1, D_2 = d_2) \frac{b_{d_1 d_2}(W)}{c_{d_1 d_2}(W)} \quad (25)$$

where $a(Z_1, Z_2, W) \in L_0^2(F(Z_1, Z_2|W))$, $a_1(Z_1, W) \in L_0^2(F(Z_1|W))$, $a_2(Z_2, W) \in L_0^2(F(Z_2|W))$, $a_0(W) \in L_0^2(F(W))$ and $\sum_{d_1,d_2=0,1} \mathbf{1}(D_1 = d_1, D_2 = d_2) \frac{b_{d_1 d_2}(W)}{c_{d_1 d_2}(W)} \in L_0^2(F(D_1, D_2|W))$ with the additional restriction that $\sum_{d_1,d_2=0,1} b_{d_1 d_2}(W) = 0$ and $\sum_{d_1,d_2=0,1} c_{d_1 d_2}(W) = 1$ for all W.

STEP - 2: The moment conditions in (1) are equivalent to the requirement that for any $d_\beta \times d_g$ matrix A, the just-identified system of moment conditions $AE[g(Z, W; \beta^0)] = 0$ hold. Differentiating under the integral, and taking a full row rank A, we obtain by using (4) that

$$\frac{\partial \beta^0(\theta_0)}{\partial \theta'} = -(AG)^{-1} AE \left[g(Z, W; \beta^0) \frac{\partial \log f_{\theta_0}(Z, W)}{\partial \theta'} \right] = -(AG)^{-1} AE \left[g(Z, W; \beta^0) \{ s(W)' + s(Z_1, Z_2|W)' \} \right].$$

For an arbitrary A, pathwise differentiability follows if we can find $\psi(A, D_1, D_2, Z_1, Z_2, W; \beta^0) \in \mathcal{T}$ such that

$$E[\psi(A, D_1, D_2, Z_1, Z_2, W; \beta^0) S(\mathcal{O})'] = \frac{\partial \beta^0(\theta_0)}{\partial \theta'}. \quad (26)$$

Conjecture: $\psi(A, D_1, D_2, Z_1, Z_2, W; \beta^0) = -(AG)^{-1}A\varphi(D_1, D_2, Z_1, Z_2, W; \beta^0)$. Then verify (26) by showing

$$E[\varphi(D_1, D_2, Z_1, Z_2, W)S(\mathcal{O})'] = E\left[g(Z, W; \beta^0)\left\{s(W)' + s(Z_1, Z_2|W)'\right\}\right]. \quad (27)$$

We proceed term-by-term for the four terms in $\varphi(D_1, D_2, Z_1, Z_2, W; \beta^0)$. Dependence on β^0 is suppressed. Consider the first term. Taking expectation conditional on W and then using (4) we obtain:

$$E\left[\frac{D_1D_2}{p_{11}(W)}\left(g(Z, W) - q(W)\right)S(\mathcal{O})'\right] = E\left[\left(g(Z, W) - q(W)\right)s(Z_1, Z_2|W)'\right] = E\left[g(Z, W)s(Z_1, Z_2|W)'\right]$$

since $E[q(W)s(Z_1, Z_2|W)'] = 0$ by using $s(Z_1, Z_2|W) \in L^2_0(F(Z_1, Z_2|W))$. Now consider the second term. Taking expectation conditional on W and then using (4) we obtain: $E\left[q(W)S(\mathcal{O})'\right] = E\left[g(Z, W)s(W)'\right]$ since $s(Z_1, Z_2|W) \in L^2_0(F(Z_1, Z_2|W))$, $s(Z_1|W) \in L^2_0(F(Z_1|W))$, $s(Z_2|W) \in L^2_0(F(Z_2|W))$, and $\sum_{d_1,d_2=0,1}\dot{p}_{d_1d_2}(W) = 0$. Now consider the third term and note that similar arguments give

$$\begin{aligned}
& E\left[\frac{p_{10}(W)}{p_1(W)}\left(\frac{D_1(1-D_2)}{p_{10}(W)} - \frac{D_1D_2}{p_{11}(W)}\right)\left(q(Z_1, W) - q(W)\right)S(\mathcal{O})'\right] \\
= \; & E\left[\frac{p_{10}(W)}{p_1(W)}\left(\frac{D_1(1-D_2)}{p_{10}(W)} - \frac{D_1D_2}{p_{11}(W)}\right)\left(q(Z_1, W) - q(W)\right)\left\{D_1D_2s(Z_1, Z_2|W) + D_1(1-D_2)s(Z_1|W)\right\}'\right] \\
= \; & E\left[\frac{p_{10}(W)}{p_1(W)}\left(q(Z_1, W) - q(W)\right)\left\{s(Z_1|W) - s(Z_1, Z_2|W)\right\}'\right] \\
= \; & -E\left[\frac{p_{10}(W)}{p_1(W)}\left(q(Z_1, W) - q(W)\right)s(Z_2|Z_1, W)'\right] \; \left[\text{since } s(Z_1, Z_2|W) \stackrel{\text{def}}{=} s(Z_1|W) + s(Z_2|Z_1, W)\right] \\
= \; & 0
\end{aligned}$$

where the last line follows because, by definition of conditional score, $s(Z_2|Z_1, W) \in L^2_0(F(Z_2|Z_1, W))$. Similar arguments show that for the fourth term, $E\left[\frac{p_{01}(W)}{p_2(W)}\left(\frac{(1-D_1)D_2}{p_{01}(W)} - \frac{D_1D_2}{p_{11}(W)}\right)\left(q(Z_2, W) - q(W)\right)S(\mathcal{O})'\right] = 0$.

Hence (27) (and thereby (26)) is verified. To show that $\varphi(D_1, D_2, Z_1, Z_2, W; \beta^0)$ belongs to the tangent set $\mathcal{T}$ in (25), rearrange its terms suitably as follows:

$$\begin{aligned}
\varphi(D_1, D_2, Z_1, Z_2, W; \beta^0) \; = \; & \frac{D_1D_2}{p_{11}(W)}\left[\left(g(Z, W) - q(W)\right) - \frac{p_{10}(W)}{p_1(W)}\left(q(Z_1, W) - q(W)\right) - \frac{p_{01}(W)}{p_2(W)}\left(q(Z_2, W) - q(W)\right)\right] \\
& + \frac{D_1(1-D_2)}{p_1(W)}\left(q(Z_1, W) - q(W)\right) + \frac{(1-D_1)D_2}{p_2(W)}\left(q(Z_2, W) - q(W)\right) + q(W).
\end{aligned}$$

The first term of the RHS involves D_1D_2, Z_1, Z_2, W and belongs to $L^2_0(F(Z_1, Z_2|W))$ by (4). It corresponds to $D_1D_2a(Z_1, Z_2, W)$ of $\mathcal{T}$ in (25). The second term of the RHS $D_1(1-D_2), Z_1, W$ and belongs to $L^2_0(F(Z_1|W))$ by (4). It corresponds to $D_1(1-D_2)a_1(Z_1, W)$ of $\mathcal{T}$ in (25). Similarly, the third term corresponds to $(1-D_1)D_2a_2(Z_2, W)$ of $\mathcal{T}$ in (25). The fourth term involves W and belongs to $L^2_0(F(W))$. It corresponds $a_0(W)$ of $\mathcal{T}$ in (25). Remaining terms of $\mathcal{T}$ are corresponded identically by 0s.

STEP - 3: We verified that any regular estimator for β^0 is asymptotically linear with influence function of the form $-(AG)^{-1}Ag(Z, W; \beta^0)$. For a given A the projection of the above influence function on to the tangent set $\mathcal{T}$ is $\psi(A, D_1, D_2, Z_1, Z_2, W; \beta^0)$ which, therefore, is the efficient influence function given the A. The variance of $\psi(A, D_1, D_2, Z_1, Z_2, W; \beta^0)$ is $(AG)^{-1}A\ V\ A'(AG)^{-1'}$ where $V := Var\left(\varphi(D_1, D_2, Z_1, Z_2, W; \beta^0)\right)$. The efficient influence function involves the A that minimizes the above variance. Standard arguments give that the minimizer is $A_* = G'V^{-1}$. Hence the efficiency bound is $\Omega := (G'V^{-1}G)^{-1}$ and the efficient influence function with variance equal to the efficiency bound is

$$\psi(D_1, D_2, Z_1, Z_2, W) := \psi(A_*, D_1, D_2, Z_1, Z_2, W) = -\Omega^{-1}G'V^{-1}\varphi(D_1, D_2, Z_1, Z_2, W; \beta^0). \blacksquare$$

Remark: From the verification of (27) in Step 2 involving the first two terms of $\varphi(D_1, D_2, Z_1, Z_2, W)$, it follows naturally that the conventional form [see Chen et al. (2008)] based on the common complete sub-sample ($D_1 = D_2 = 1$):

$$\frac{D_1 D_2}{p_{11}(W)}\left[g(Z, W; \beta^0) - q(W; \beta^0)\right] + q(W; \beta^0)$$

is an influence function. However, in general it does not belong in $\mathcal{T}$ defined in (25) because that requires the parametric submodel to satisfy $s(Z_1|W) \equiv s(Z_2|W) \equiv 0$ but $s(Z_1, Z_2|W) \neq 0$. This is not possible except in the special case $s(Z_1, Z_2|W) \equiv s(Z_1|Z_2, W) \equiv s(Z_2|Z_1, W)$ which imposes the additional restrictions on $\mathcal{T}$ that $a(Z_1, Z_2, W) \in L_0^2(F(Z_1|Z_2, W))$ and $a(Z_1, Z_2, W) \in L_0^2(F(Z_2|Z_1, W))$.

Proof of Proposition 2.2:

STEP - 1: Same as that in the proof of Proposition 2.1 with W denoting the distinct collection of all elements of W_1 and W_2, and allowing for the possibility that $W = W_1$ and/or $W = W_2$.

STEP - 2: The moment conditions in (1) under (5) are equivalent to the requirement that for any $d_\beta \times d_g$ matrix $A = [A_1, A_2]$ where A_j is $d_\beta \times d_{g_j}$ for $j = 1, 2$, the following just-identified system of moment conditions

$$AE[g(Z, W; \beta^0)] \equiv [A_1, A_2]E\begin{bmatrix} g_1(Z_1, W_1; \beta_0^0, \beta_1^0) \\ g_2(Z_2, W_2; \beta_0^0, \beta_2^0) \end{bmatrix} = 0$$

hold. Differentiating under the integral, and taking a full row rank A, we obtain by using (4) that

$$\frac{\partial \beta^0(\theta_0)}{\partial \theta'} = -(A\widetilde{G})^{-1}AE\begin{bmatrix} g_1(Z_1, W_1; \beta_0^0, \beta_1^0)\left\{s(W_1)' + s(Z_1|W_1)'\right\} \\ g_2(Z_2, W_2; \beta_0^0, \beta_2^0)\left\{s(W_2)' + s(Z_2|W_2)'\right\} \end{bmatrix}.$$

Now as in STEP-2 in the proof of Proposition 2.1 we show that, for $j = 1, 2$:

$$E\left[\varphi_{Pj}(D_j, Z_j, W; \beta_0^0, \beta_j^0)S(\mathcal{O})'\right] = E\left[g_j(Z_j, W_j; \beta_0^0, \beta_j^0)\left\{s(W)' + s(Z_j|W)'\right\}\right]. \quad (28)$$

For $j = 1$, taking expectation conditional on W and then noting that $D_1 D_2 s(Z_1, Z_2|W) + D_1(1-D_2)s(Z_1|W) = D_1 s(Z_1|W) + D_1 D_2 s(Z_2|Z_1, W)$ in $S(D_1, D_2, D_1 Z_1, D_2 Z_2, W)$, it follows that

$$E\left[\frac{D_1}{p_1(W)}\left[g_1(Z_1, W_1) - q_1(W)\right]S(\mathcal{O})'\right] = E\left[\left[g_1(Z_1, W_1) - q_1(W)\right]s(Z_1|W)'\right] = E\left[g_1(Z_1, W_1)s(Z_1|W)'\right]$$

where the first equality follows from $s(Z_2|Z_1,W) \in L_0^2(F(Z_2|Z_1,W))$, and the second from $s(Z_1|W) \in L_0^2(F(Z_1|W))$. On the other hand, $E[q_j(W;\beta_0,\beta_j)S(\mathcal{O})'] = E[g_1(Z_1,W_1)s(W)']$ by using $s(Z_1,Z_2|W) \in L_0^2(F(Z_1,Z_2|W))$, $s(Z_1|W) \in L_0^2(F(Z_1|W))$, $s(Z_2|W) \in L_0^2(F(Z_2|W))$ and $\sum_{d_1 d_2} \dot{p}_{d_1 d_2}(W) = 0$. Therefore, (28) is verified for $j=1$. Similarly, it can be verified for $j=2$.

Now rearranging the terms in $\varphi_P(D_1,D_2,Z_1,Z_2,W;\beta^0)$ as follows:

$$\varphi_P(D_1,D_2,Z_1,Z_2,W) = \begin{bmatrix} \varphi_{P1}(D_1,Z_1,W_1) \\ \varphi_{P2}(D_2,Z_2,W) \end{bmatrix} = D_1 D_2 \begin{bmatrix} \frac{g_1(Z_1,W_1)-q_1(W)}{p_1(W)} \\ \frac{g_2(Z_2,W_2)-q_2(W)}{p_2(W)} \end{bmatrix} + D_1(1-D_2) \begin{bmatrix} \frac{g_1(Z_1,W_1)-q_1(W)}{p_1(W)} \\ 0 \end{bmatrix}$$
$$+(1-D_1)D_2 \begin{bmatrix} 0 \\ \frac{g_2(Z_2,W_2)-q_2(W)}{p_2(W)} \end{bmatrix} + \begin{bmatrix} q_1(W) \\ q_2(W) \end{bmatrix},$$

it is easy to see that when evaluated at $\beta = \beta^0$, the four terms on the RHS, by virtue of (4), are respectively in $L_0^2(F(Z_1,Z_2|W))$, $L_0^2(F(Z_1|W))$, $L_0^2(F(Z_2|W))$ and $L_0^2(F(W))$. The terms $D_1 D_2 a(Z_1,Z_2,W)$, $D_1(1-D_2)a_1(Z_1,W)$, $(1-D_1)D_2 a_2(Z_2,W)$ and $a_0(W)$ of $\mathcal{T}$ in (25) are corresponded by these four terms, whereas the remaining terms in $\mathcal{T}$ are corresponded identically by 0s. This completes step 2.

STEP - 3: Follows similarly as in the proof of Proposition 2.1. ■

Notations to be used in the rest of the Appendix: For any $a \times b$ matrix A (including $b=1$ or $a=b=1$), let $|A| := \sqrt{\text{Trace}(A'A)}$. For any $a \times b$ matrix $A(u,\beta)$ where the (i,j)-th element is a function $A_{ij}(u,\beta) : \mathcal{U} \times \Theta \to \mathbb{R}$, let $\|A(\beta)\|_\infty = \sup_{u\in\mathcal{U}} |A(u,\beta)|$ for any given $\beta \in \mathcal{B}$, and let $\|A\|_\infty = \sup_{\beta\in\mathcal{B}} \sup_{u\in\mathcal{U}} |A(u,\beta)|$.

Proof of Proposition-2.3:

Define the following quantities that will be used throughout the proof for notational convenience:

$$\begin{aligned} \omega_i &:= \frac{p_{10}(W_i)}{p_1(W_i)}, \text{ and } \widehat{\omega}_i := \frac{\widehat{p}_{10}(W_i)}{\widehat{p}_1(W_i)}, \\ \nu_i &:= \frac{D_{1i}(1-D_{2i})}{p_{10}(W_i)} - \frac{D_{1i}D_{2i}}{p_{11}(W_i)}, \text{ and } \widehat{\nu}_i := \frac{D_{1i}(1-D_{2i})}{\widehat{p}_{10}(W_i)} - \frac{D_{1i}D_{2i}}{\widehat{p}_{11}(W_i)}, \\ \tau_i(\beta) &:= q(Z_{1i},W_i;\beta) - q(W_i;\beta) \text{ and } \widehat{\tau}_i(\beta) := \widehat{q}(Z_{1i},W_i;\beta) - \widehat{q}(W_i;\beta). \end{aligned}$$

Under the conditions of the proposition,

$$\|\widehat{p} - p\|_\infty = o_p(N^{-1/4}), \tag{29}$$
$$\sup_{|\beta-\beta^0|<\delta} \|\widehat{q}(\beta) - q(\beta)\|_\infty = o_p(1) \text{ for some constant } \delta > 0 \text{ and,} \tag{30}$$
$$\|\widehat{q}(U;\beta^0) - q(U;\beta^0)\|_\infty = o_p(N^{-1/4}). \tag{31}$$

(29) is Cattaneo's condition (5.1), and holds by Theorem B-1 of Cattaneo (2010). (30) is Cattaneo's condition (5.2), and holds from the first result of Proposition A1(i) of Chen et al. (2005). (31) is shown in the proof of Theorem 8 (page 152) in Cattaneo (2010) [Theorem 4 of Newey (1997)].

Therefore, all that are left to be verified are two conditions: a condition similar to (5.3) of Cattaneo (2010) and a stochastic equicontinuity condition that, by virtue of the previous condition, gives (iii) in Theorem 3.3 in Pakes and Pollard (1989). Thanks to (a) symmetry in the terms in the second and third lines of (6), and (b) the proofs of Theorem 5 and 8 in Cattaneo (2010) this boils down to verifying:

$$o_p(1) \;=\; \sqrt{N}\left(\bar{\xi}_N(\beta^0, \widehat{p}, \widehat{q}(\beta^0)) - \bar{\xi}_N(\beta^0, p, \bar{q}(\beta^0))\right), \tag{32}$$

$$o_p(1) \;=\; \sup_{|\beta-\beta^0|\le\delta_N} \frac{\sqrt{N}\left|\bar{\xi}_N(\beta, \widehat{p}, \widehat{q}(\beta)) - E\left[\bar{\xi}_N(\beta, p, \bar{q}(\beta))\right] - \bar{\xi}_N(\beta^0, \widehat{p}, \widehat{q}(\beta^0))\right|}{1 + C\sqrt{N}|\beta - \beta^0|} \tag{33}$$

for all positive sequences $\delta_N = o(1)$ and a generic constant $C > 0$, where in terms of the notation above,

$$\bar{\xi}_N(\beta, \widehat{p}, \widehat{q}(\beta)) := \frac{1}{N}\sum_{i=1}^{N} \widehat{\omega}_i \widehat{\nu}_i \widehat{\tau}_i(\beta) \text{ and } \bar{\xi}_N(\beta, p, \bar{q}(\beta)) := \frac{1}{N}\sum_{i=1}^{N} \omega_i \nu_i \tau_i(\beta).$$

We start with the verification of (32). (Dependence on β is suppressed at $\beta = \beta^0$.) Note that (32)'s RHS is:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\widehat{\omega}_i - \omega_i)(\widehat{\nu}_i - \nu_i)(\widehat{\tau}_i - \tau_i) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_i(\widehat{\nu}_i - \nu_i)(\widehat{\tau}_i - \tau_i) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\widehat{\omega}_i - \omega_i)\nu_i(\widehat{\tau}_i - \tau_i)$$
$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\widehat{\omega}_i - \omega_i)(\widehat{\nu}_i - \nu_i)\tau_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_i\nu_i(\widehat{\tau}_i - \tau_i) + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_i(\widehat{\nu}_i - \nu_i)\tau_i + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\widehat{\omega}_i - \omega_i)\nu_i\tau_i.$$

Thanks to (29) and (31), and the fact that $\frac{1}{N}\sum_i |\omega_i|$, $\frac{1}{N}\sum_i |\nu_i|$ and $\frac{1}{N}\sum_i |\tau_i|$ are $O_p(1)$ under our assumptions, it is straightforward to show that the first four terms in the above expression are $o_p(1)$. We focus on the last three terms and show that each of them is $o_p(1)$. Since the convergence rates are faster for series estimators based on splines than on power series, it suffices to show the desired results for the power series case, i.e., with $\eta = 1$ in the statement of the proposition. We use Lemma A (below) with $s_p = s_q = s$ and $K_p = K_q = K$.

Let us start with the fifth term:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_i\nu_i(\widehat{\tau}_i - \tau_i) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_i\nu_i(\widehat{q}(Z_{1i}, W_i) - q(Z_{1i}, W_i)) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_i\nu_i(\widehat{q}(W_i) - q(W_i)).$$

We use Lemma A to show the second term on the RHS is $o_p(1)$. Modifying (B4)-(B6) in Lemma A accordingly, and then following the same steps as below will give the first term on the RHS is $o_p(1)$. Hence this is omitted for brevity. Note that $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_i\nu_i(\widehat{q}(Z_{1i}, W_i) - q(Z_{1i}, W_i)) = T_{5aN} + T_{5bN}$ where

$$T_{5aN} := \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_i\nu_i(\widehat{q}(Z_{1i}, W_i) - (I_{d_g} \otimes R_K'(W_i))\gamma_K^*) \text{ and } T_{5bN} := \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_i\nu_i((I_{d_g} \otimes R_K'(W_i))\gamma_K^* - q(W_i)).$$

Consider T_{5aN} and note that $|T_{5aN}| \le \left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_i\nu_i(I_{d_g} \otimes R_K'(W_i))\right| |\widehat{\gamma}_K - \gamma_K^*| = O_p\left(KN^{-1/2} + K^{1/2-s/d_w}\right)$, because under our assumptions, $E[\omega_i\nu_i(I_{d_g} \otimes R_K'(W_i))] = 0$, $Var(\omega_i\nu_i(I_{d_g} \otimes R_K'(W_i))) = O(K)$ and hence $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_i\nu_i(I_{d_g} \otimes R_K'(W_i)) = O_p(K^{1/2})$; while Lemma A (B4) gives $|\widehat{\gamma}_K - \gamma_K^*| = O_p\left(K^{1/2}N^{-1/2} + K^{-s/d_w}\right)$. Since taking $\eta = 1$ means $v < 1/6$ and $s/d_w > 3$ under our assumptions, noting $K = N^v$ gives $|T_{5aN}| = o_p(1)$.

Our assumptions give $E[\omega_i\nu_i((I_{d_g}\otimes R_K'(W_i))\gamma_K^* - q(Z_{1i},W_i))] = 0$ and $Var(\omega_i\nu_i((I_{d_g}\otimes R_K'(W_i))\gamma_K^* - q(Z_{1i},W_i)) = O\left(\sup_w \|((I_{d_g}\otimes R_K'(w))\gamma_K^* - q(w))\|^2\right)$. So $|T_{5bN}| = O_p(K^{-s/d_w}) = o_p(1)$ by Lemma A (B4).

Now consider the sixth term: $\frac{1}{\sqrt{N}}\sum_{i=1}^N \omega_i(\widehat{\nu}_i - \nu_i)\tau_i = T_{6aN} + T_{6bN}$ where

$$\begin{aligned} T_{6aN} &:= \frac{1}{\sqrt{N}}\sum_{i=1}^N \omega_i \frac{D_{1i}(1-D_{2i})}{\widehat{p}_{10}(W_i)p_{10}(W_i)}(p_{10}(W_i) - \widehat{p}_{10}(W_i))\tau_i, \\ T_{6bN} &:= \frac{1}{\sqrt{N}}\sum_{i=1}^N \omega_i \frac{D_{1i}D_{2i}}{\widehat{p}_{11}(W_i)p_{11}(W_i)}(\widehat{p}_{11}(W_i) - p_{11}(W_i))\tau_i. \end{aligned}$$

We will show $T_{6aN} = o_p(1)$. Define $\mathbf{1}_N^p := \mathbf{1}\left(\inf_{w\in\mathcal{W}}\widehat{p}_{10}(w)\geq\kappa\right)$. $\mathbf{1}_N^p \xrightarrow{P} 1$ under Assumption M(2) since $\|\widehat{p} - p\|_\infty = o_p(1)$. (4) gives $E[\mathbf{1}_N^p T_{6aN}|W_1,\ldots,W_N] = 0$ and hence $E[\mathbf{1}_N^p T_{6aN}] = 0$. Similarly, (4) also gives

$$E\left[\mathbf{1}_N^p\omega_i\omega_j\frac{D_{1i}(1-D_{2i})}{\widehat{p}_{10}(W_i)p_{10}(W_i)}(p_{10}(W_i)-\widehat{p}_{10}(W_i))\frac{D_{1j}(1-D_{2j})}{\widehat{p}_{10}(W_j)p_{10}(W_j)}(p_{10}(W_j)-\widehat{p}_{10}(W_j))\tau_i\tau_j|W_1,\ldots,W_N\right] = 0$$

for all $i\neq j$. Hence $Var(\mathbf{1}_N^p T_{6aN}) = E\left[Var(\mathbf{1}_N^p T_{6aN}|W_1,\ldots,W_N)\right] = E\left[O_p(\|\widehat{p}-p\|_\infty^2)\right]$ under our assumptions. Now (29) gives $Var(\mathbf{1}_N^p T_{6aN}) = o_p(1)$ and hence $|T_{6aN}| = o_p(1)$. Similar steps give $T_{6bN} = o_p(1)$.

Finally consider the seventh (last) term: $\frac{1}{\sqrt{N}}\sum_{i=1}^N(\widehat{\omega}_i - \omega_i)\nu_i\tau_i$ and note that steps similar to that for the sixth term also show that this last term is $o_p(1)$. Hence (32) is verified.

Now we verify (33), i.e., for all positive $\delta_N = o_p(1)$ and a generic positive constant C,

$$o_p(1) = \sup_{|\beta-\beta^0|\leq\delta_N}\frac{\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N\widehat{\omega}_i\widehat{\nu}_i\left(\widehat{\tau}_i(\beta) - \widehat{\tau}_i(\beta^0)\right)\right|}{1 + C\sqrt{N}|\beta-\beta^0|}$$

using that $E\left[\bar{\xi}_N(\beta,p,\bar{q}(\beta))\right] = 0$ under (4). Define $\zeta_{1i}(\beta) = \tau_i(\beta) - E[\tau_i(\beta)]$ and $\zeta_{2i}(\beta) = \widehat{\tau}_i(\beta) - \tau_i(\beta)$. Therefore, the RHS of the above is

$$\sup\nolimits_{|\beta-\beta^0|\leq\delta_N}\frac{\sqrt{N}\,|A_{1N}(\beta) + A_{2N}(\beta)|}{1 + C\sqrt{N}|\beta-\beta^0|}, \text{ where}$$
$$A_{1N}(\beta) = \tfrac{1}{N}\textstyle\sum_{i=1}^N\widehat{\omega}_i\widehat{\nu}_i\left(\zeta_{1i}(\beta) - \zeta_{1i}(\beta^0)\right), \text{ and } A_{2N}(\beta) = \tfrac{1}{N}\textstyle\sum_{i=1}^N\widehat{\omega}_i\widehat{\nu}_i\left(\zeta_{2i}(\beta) - \zeta_{2i}(\beta^0)\right)$$

since $E[\tau_i(\beta)] = 0$ for all β by (4). Now, the verification of (33) follows directly by following *exactly the same steps* as that for the corresponding terms $R_{2N}(\beta)$ (for $A_{1N}(\beta)$) and $R_{3N}(\beta)$ (for $A_{2N}(\beta)$) in the proof of Proposition -2.4 below. (Details are available from the authors.) ∎

The series estimators in Proposition-2.4 are based on power series. Lemma A summarizes some well known results for such estimators. The presentation omits an important intermediate step concerning the maximizer and minimizer of the limiting objective functions for the coefficients for the power series that are treated carefully in Hirano et al. (2003) and Imbens et al. (2009). Instead we directly consider the approximation error (B2) and (B5) for the intermediate target quantities defined below in (B1) and (B4) respectively. (B4)-(B6) can be modified to accommodate for the nuisance parameters $q(Z_1,W;\beta)$ and $q(Z_2,W;\beta)$.

Lemma A: The following results hold under the conditions of Proposition-2.4:

(B1) For a fixed K_p there exists a $\pi^*_{K_p} \in \mathbb{R}^{K_p}$ such that $\|p_{d_1d_2} - L_{d_1d_2}(R_{K_p}, \pi^*_{K_p})\|_\infty = O(K_p^{-s_p/d_w})$.

(B2) $|\widehat{\pi}_{K_p} - \pi^*_{K_p}| = O_p\left(K_p^{1/2}N^{-1/2} + K_p^{1/2}K_p^{-s_p/d_w}\right)$ as $N \to \infty$.

(B3) $\|\widehat{p} - p\|_\infty = O_p\left(K_p[K_p^{1/2}N^{-1/2} + K_p^{1/2}K_p^{-s_p/d_w}]\right)$ as $N \to \infty$.

(B4) For a fixed K_q there exists a $\gamma^*_{K_q}(\beta^0) \in \mathbb{R}^{K_q}$ such that $\|q(\beta^0) - (I_{d_g} \otimes R'_{K_q})\gamma^*_{K_q}(\beta^0)\|_\infty = O(K_q^{-s_q/d_w})$.

(B5) $|\widehat{\gamma}_{K_q}(\beta^0) - \gamma_{K_q}(\beta^0)| = O_p\left(K_q^{1/2}N^{-1/2} + K_q^{-s_q/d_w}\right)$ as $N \to \infty$.

(B6) For $\widehat{q}(\beta^0) = (I_{d_g} \otimes R'_{K_q})\widehat{\gamma}_{K_q}(\beta^0)$, $\|\widehat{q}(\beta^0) - q(\beta^0)\|_\infty = O_p\left(K_q[K_q^{1/2}N^{-1/2} + K_q^{-s_q/d_w}]\right)$ as $N \to \infty$.

Proof of Lemma A:

See Theorem B-1 of Cattaneo (2010) for (B1)-(B3). See Lemma 1 of Newey (1994) for (B4). See Theorem 1 (including the proof) and Theorem 4 of Newey (1997) for (B5) and (B6). ■

Proof of Proposition-2.4:

Since the idea is the same, for notational simplicity let us present the proof for the case with only two-level missingness. Accordingly, *only in this proof* let $D := D_1D_2$, $p(W) := p_{11}(W)$, and let it be known that $D_1(1 - D_2) \equiv (1 - D_1)D_2 \equiv 0$ and $p_{10}(W) \equiv p_{01}(W) \equiv 0$. Without loss of generality, let $d_g = 1$. Define $L(u) := \exp(u)/[1 + \exp(u)]$ for some scalar u (to replace the general formula $L_{d_1d_2}(.)$).

The proof is similar to that of Theorem 5 in Cattaneo (2010). The main difference is that we will not require his condition (5.1), i.e., $\|\widehat{p} - p\|_\infty = o_p(N^{-1/4})$. His condition (5.2), i.e.,

$$\sup_{|\beta-\beta^0|<\delta} \|\widehat{q}(\beta) - q(\beta)\|_\infty = o_p(1) \tag{34}$$

is still satisfied in the same way from the first result of Proposition A1(i) of Chen et al. (2005). His condition (5.3) also holds under our setup as is shown in Lemma B below (note the contrast with the proof of Theorem 8 in Cattaneo (2010)). Hence, similar to (33) in the proof of Proposition 2.3, we only need to verify that for any positive $\delta_N = o(1)$ and some generic positive constant C:

$$\sup_{|\beta-\beta^0|\le\delta_N} \frac{\sqrt{N}\ |R_N(\beta)|}{1 + C\sqrt{N}|\beta - \beta^0|} = o_p(1), \tag{35}$$

$$\text{where } R_N(\beta) = \bar{m}_N(\beta, \widehat{p}, \widehat{q}(\beta)) - E\left[\bar{m}_N(\beta, p, q(\beta))\right] - \bar{m}_N(\beta^0, \widehat{p}, \widehat{q}(\beta^0))$$

and $\bar{m}(\beta, p^*, q^*) := \frac{1}{N}\sum_{i=1}^N \left\{\frac{D_i}{p^*(W_i)}\left[g(Z_i, W_i; \beta) - q^*(W_i; \beta)\right] + q^*(W_i; \beta)\right\}$ for some generic p^* and q^*.

Since $E\left[\bar{m}_N(\beta, p, q(\beta))\right] = E[g(Z_i, W_i; \beta)] = E[q(W_i; \beta)]$ and $E[g(Z_i, W_i; \beta^0)] = E[q(W_i; \beta^0)] = 0$ by (1), we obtain: $\sqrt{N}R_N(\beta) = R_{1N}(\beta) + R_{2N}(\beta) + R_{3N}(\beta)$ where

$$\begin{aligned} R_{1N}(\beta) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{D_i}{\widehat{p}(W_i)}[v_1(Z_i,W_i;\beta) - v_1(Z_i,W_i;\beta^0)], \\ R_{2N}(\beta) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(1-\frac{D_i}{\widehat{p}(W_i)}\right)[v_2(W_i;\beta) - v_2(W_i;\beta^0)], \\ R_{3N}(\beta) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(1-\frac{D_i}{\widehat{p}(W_i)}\right)[v_3(W_i;\beta) - v_3(W_i;\beta^0)], \end{aligned}$$

and the individual components are $v_1(Z_i,W_i;\beta) := g(Z_i,W_i;\beta) - E[g(Z_i,W_i;\beta)]$, $v_2(W_i;\beta) := \widehat{q}(W_i;\beta) - q(W_i;\beta)$ and $v_3(W_i;\beta) := q(W_i;\beta) - E[q(W_i;\beta)]$. Now we verify (35) by working through the terms $R_{1N}(\beta)$, $R_{2N}(\beta)$ and $R_{3N}(\beta)$ respectively. First choose δ_N converging to zero slowly enough to ensure that $\mathbf{1}_N^q := \mathbf{1}\left(\sup_{\beta\in\mathcal{N}_{\delta_N}}\|\widehat{q}(\beta) - q(\beta)\|_\infty \le \delta_N\right) \xrightarrow{P} 1$ by appealing to (34). Also define $\mathbf{1}_N^p := \mathbf{1}\left(\inf_{w\in\mathcal{W}}\widehat{p}(w) \ge \kappa\right)$. Lemma (B3) gives $\|\widehat{p} - p\|_\infty = o_p(1)$ because $v_p < \frac{1}{3}$ and $\frac{s_p}{d_w} > \frac{1}{2}$ by (14). So $\mathbf{1}_N^p \xrightarrow{P} 1$ by Assumption M(2).

Consider $R_{1N}(\beta)$. Using Assumption M(2), Assumptions g(2), g(3) and g(4), arguments along the line of Theorem 4 in Cattaneo (2010) imply that the class of functions $\left\{\mathbf{1}_N^p\frac{D}{\widehat{p}(.)}v_1(.;\beta) : \beta\in\mathcal{N}_{\delta_N}\right\}$ is Donsker with finite integrable envelope and L_2 continuous. Recalling that $\mathbf{1}_N^p \xrightarrow{P} 1$ (not depending on β), it follows that $\sup_{\beta\in\mathcal{N}_{\delta_N}}|R_{1N}(\beta)|/[1 + C\sqrt{N}|\beta-\beta^0|] = o_p(1)$.

Now consider $R_{2N}(\beta)$. First by a mean-value expansion (with the mean-value being subsumed by the $\sup_{\beta\in\mathcal{N}_{\delta_N}}$ clause) and then by appealing to (34) to use Assumption q(2b) we obtain

$$\sup_{\beta\in\mathcal{N}_{\delta_N}}\mathbf{1}_N^q\mathbf{1}_N^p|R_{2N}(\beta)| \le \sup_{\beta\in\mathcal{N}_{\delta_N},\|\tilde{q}-q\|_\infty\le\delta_N}\sqrt{N}|\beta-\beta^0|\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_N^q\mathbf{1}_N^p\left|1-\frac{D_i}{\widehat{p}(W_i)}\right|\left|\frac{\partial}{\partial\beta'}[\tilde{q}(W_i;\beta) - q(W_i;\beta)]\right|.$$

Since $\mathbf{1}_N^q\mathbf{1}_N^p\left|1-\frac{D_i}{\widehat{p}(W_i)}\right| \le \max\left(1,\left|1-\frac{1}{\kappa}\right|\right)$ is bounded, using (4) and Assumption q(2b) we obtain

$$\sup_{\beta\in\mathcal{N}_{\delta_N}}\mathbf{1}_N^q\mathbf{1}_N^p\frac{|R_{2N}(\beta)|}{1+C\sqrt{N}|\beta-\beta^0|} \le C_1\delta_N^\epsilon\left[\frac{1}{N}\sum_{i=1}^{N}b(W_i)\right]$$

for some generic positive constant C_1, some non-negative measurable function $b(w)$ with $E[b(W)] < \infty$, and some $\epsilon > 0$. Letting $\delta_N \to 0$ and recalling that $\mathbf{1}_N^q \xrightarrow{P} 1$ and $\mathbf{1}_N^p \xrightarrow{P} 1$ (not depending on β) give $\sup_{\beta\in\mathcal{N}_{\delta_N}}|R_{2N}(\beta)|/[1 + C\sqrt{N}|\beta-\beta^0|] = o_p(1)$.

Finally consider $R_{3N}(\beta)$. By a mean-value expansion and then using Assumption q(2a) that allows for interchanging the order of integration and differentiation we obtain

$$\begin{aligned} \sup_{\beta\in\mathcal{N}_{\delta_N}}\mathbf{1}_N^p|R_{3N}(\beta)| &\le \sup_{\beta\in\mathcal{N}_{\delta_N}}\sqrt{N}|\beta-\beta^0|\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_N^p\left|1-\frac{D_i}{\widehat{p}(W_i)}\right|\left|\frac{\partial}{\partial\beta'}[q(W_i;\beta) - E[q(W_i;\beta)]]\right| \\ \Rightarrow \sup_{\beta\in\mathcal{N}_{\delta_N}}\mathbf{1}_N^p\frac{|R_{3N}(\beta)|}{1+C\sqrt{N}|\beta-\beta^0|} &\le C_1\sup_{\beta\in\mathcal{N}_{\delta_N}}\frac{1}{N}\sum_{i=1}^{N}\left|\frac{\partial}{\partial\beta'}q(W_i;\beta) - E\left[\frac{\partial}{\partial\beta'}q(W_i;\beta)\right]\right| \end{aligned}$$

for some generic positive constant C_1. The dominating integrable function in Assumption q(2a) also ensures that $\frac{1}{N}\sum_{i=1}^{N}\left|\frac{\partial}{\partial\beta'}q(W_i;\beta)-E\left[\frac{\partial}{\partial\beta'}q(W_i;\beta)\right]\right|\xrightarrow{P}0$ uniformly in $\beta\in\mathcal{N}_{\delta_N}$. Recalling that $\mathbf{1}_N^p\xrightarrow{P}1$ (not depending on β) gives $\sup_{\beta\in\mathcal{N}_{\delta_N}}|R_{3N}(\beta)|/[1+C\sqrt{N}|\beta-\beta^0|]=o_p(1)$. ■

Lemma B: The following result holds under the conditions of Proposition-2.4 and its proof:

$$\bar{m}_N(\beta^0,\widehat{p},\widehat{q}(\beta^0))=\bar{m}_N(\beta^0,p,q(\beta^0))+o_p(N^{-1/2}).$$

where, $\bar{m}_N(\beta,p^*,q^*):=\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{D_i}{p^*(W_i)}\left[g(Z_i,W_i;\beta)-q^*(W_i;\beta)\right]+q^*(W_i;\beta)\right\}$ for some generic p^* and q^*.

Proof of Lemma B:

Note that $\sqrt{N}\left[\bar{m}_N(\beta^0,\widehat{p},\widehat{q}(\beta^0))-\bar{m}_N(\beta^0,p_1,q(\beta^0))\right]=A_{1N}+A_{2N}+A_{3N}$ where

$$\begin{aligned}
A_{1N} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{D_i}{p(W_i)\widehat{p}(W_i)}\left[g(Z_i,W_i;\beta^0)-q(W_i;\beta^0)\right]\left[p(W_i)-\widehat{p}(W_i)\right],\\
A_{2N} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{D_i}{p(W_i)}-1\right)\left[q(W_i;\beta^0)-\widehat{q}(W_i;\beta^0)\right],\\
A_{3N} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{D_i}{p(W_i)\widehat{p}(W_i)}\left[q(W_i;\beta^0)-\widehat{q}(W_i;\beta^0)\right]\left[p(W_i)-\widehat{p}(W_i)\right].
\end{aligned}$$

This is an exact relation in contrast to the proof of Theorem 8 in Cattaneo (2010). We will show one by one that $A_{1N}=o_p(1)$, $A_{2N}=o_p(1)$ and $A_{3N}=o_p(1)$.

We start with A_{1N}. Now, by (4), $E[\mathbf{1}_N^p A_{1N}|W_1,\ldots,W_N]=E[\mathbf{1}_N^p A_{1N}]=0$, and $Var(\mathbf{1}_N^p A_{1N}|W_1,\ldots,W_N)=O_p\left(\|\widehat{p}-p\|_\infty^2\right)$ further using Assumption T, and because for each $i\neq j$, we can condition on W_i,W_j to obtain

$$E\left[\frac{D_iD_j\left[p(W_i)-\widehat{p}(W_i)\right]\left[p(W_j)-\widehat{p}(W_j)\right]}{p(W_i)\widehat{p}(W_i)p(W_j)\widehat{p}(W_j)}\left[g(Z_i,W_i;\beta)-q(W_i;\beta)\right]\left[g(Z_j,W_j;\beta)-q(W_j;\beta)\right]'\right]=0$$

by (4). Therefore, Assumption T(2) gives $Var(\mathbf{1}_N^p A_{1N})=O\left(\|\widehat{p}-p\|_\infty^2\right)$ and hence $\mathbf{1}_N^p A_{1N}=O_p\left(\|\widehat{p}-p\|_\infty\right)$. Recalling that $\mathbf{1}_N^p\xrightarrow{P}1$, we obtain $A_{1N}=O_p\left(\|\widehat{p}-p\|_\infty\right)$ which is $o_p(1)$ by Lemma (B3) under (14).

Now consider $A_{2N}=B_{1N}+B_{2N}$ where

$$\begin{aligned}
B_{1N} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{D_i}{p(W_i)}-1\right)\left[q(W_i;\beta^0)-R_{K_q}(W_i)'\gamma_{K_q}^*\right],\\
B_{2N} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{D_i}{p(W_i)}-1\right)\left[R_{K_q}(W_i)'\gamma_{K_q}^*-\widehat{q}(W_i;\beta^0)\right].
\end{aligned}$$

Now, by (4), $E[B_{1N}|W_1,\ldots,W_N]=E[B_{1N}]=0$, and $Var(B_{1N}|W_1,\ldots,W_N)=O_p\left(\|q^0-R_{K_q}'\gamma_{K_q}^*\|_\infty^2\right)$. Therefore, $Var(B_{1N})=O_p\left(\|q^0-R_{K_q}'\gamma_{K_q}^*\|_\infty^2\right)$ by Assumption T(2) and using the same arguments as for A_{1N}. Therefore, $B_{1N}=O_p\left(\|q^0-R_{K_q}'\gamma_{K_q}^*\|_\infty\right)$ which is $o_p(1)$ by Lemma (B1) since $\frac{s_q}{d_w}>0$ under (14). On the other hand,

$$|B_{2N}| = \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{D_i}{p(W_i)} - 1 \right) R_{K_q}(W_i)' \left(\gamma^*_{K_q} - \widehat{\gamma}^0_{K_q} \right) \right| \leq |\widehat{\gamma}^0_{K_q} - \gamma^*_{K_q}| \left| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{D_i}{p(W_i)} - 1 \right) R_{K_q}(W_i)' \right|.$$

$E\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{D_i}{p(W_i)}-1\right)R_{K_q}(W_i)'\right] = 0$ and $Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{D_i}{p(W_i)}-1\right)R_{K_q}(W_i)'\right) = O\left(E|R_{K_q}(W)|^2\right) = O(K_q)$ by definition of $R_{K_q}(W)$ and using the same arguments as for A_{1N}. Hence $|B_{2N}| = O_p\left(|\widehat{\gamma}^0_{K_q} - \gamma^*_{K_q}|K_q^{1/2}\right)$, which is $o_p(1)$ by Lemma (B5) because (14) requires $v_q < \frac{1}{2}$ and $\frac{s_q}{d_w} > \frac{1}{2}$. Therefore, we obtain $A_{2N} = B_{1N} + B_{2N} = o_p(1)$.

Finally consider $A_{3N} = B_{3N} + B_{4N} + B_{5N} + B_{6N}$ where

$$\begin{aligned}
B_{3N} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{D_i}{p(W_i)\widehat{p}(W_i)} \left[q(W_i;\beta^0) - R_{K_q}(W_i)'\gamma^*_{K_q} \right] \left[p(W_i) - L(R_{K_p}(W_i)'\pi^*_{K_p}) \right], \\
B_{4N} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{D_i}{p(W_i)\widehat{p}(W_i)} \left[q(W_i;\beta^0) - R_{K_q}(W_i)'\gamma^*_{K_q} \right] \left[L(R_{K_p}(W_i)'\pi^*_{K_p}) - \widehat{p}(W_i) \right], \\
B_{5N} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{D_i}{p(W_i)\widehat{p}(W_i)} \left[R_{K_q}(W_i)'\gamma^*_{K_q} - \widehat{q}(W_i;\beta^0) \right] \left[p(W_i) - L(R_{K_p}(W_i)'\pi^*_{K_p}) \right], \\
B_{6N} &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{D_i}{p(W_i)\widehat{p}(W_i)} \left[R_{K_q}(W_i)'\gamma^*_{K_q} - \widehat{q}(W_i;\beta^0) \right] \left[L(R_{K_p}(W_i)'\pi^*_{K_p}) - \widehat{p}(W_i) \right].
\end{aligned}$$

Note that $|B_{3N}| \leq \left[\frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{p(W_i)\widehat{p}(W_i)}\right]\sqrt{N}\|q^0 - R'_{K_q}\gamma^*_{K_q}\|_\infty \|p - L(R'_{K_p}\pi^*_{K_p})\|_\infty$. Also, $\mathbf{1}^p_N \frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{p(W_i)\widehat{p}(W_i)} = O_p(1)$. Hence $\frac{1}{N}\sum_{i=1}^{N}\frac{D_i}{p(W_i)\widehat{p}(W_i)} = O_p(1)$ recalling that $\mathbf{1}^p_N \xrightarrow{P} 1$. Thus, $|B_{3N}| \leq \sqrt{N}O_p(K_q^{-s_q/d_w})O(K_p^{-s_p/d_w})$ by Lemmas (B1) and (B4). Since $v_p\frac{s_p}{d_w} + v_q\frac{s_q}{d_w} > \frac{1}{2}$ by (14), it follows that $|B_{3N}| = o_p(1)$.

Now denoting $\dot{L}(u) := \frac{\partial}{\partial u}L(u)$, a mean-value expansion gives for some mean-value $\tilde{\pi}$

$$B_{4N} = -\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \frac{D_i}{p(W_i)\widehat{p}(W_i)} \left[q(W_i;\beta^0) - R_{K_q}(W_i)'\gamma^*_{K_q} \right] \dot{L}(R_{K_p}(W_i)'\tilde{\pi}) R_{K_p}(W_i)' \left(\widehat{\pi}_{K_p} - \pi^*_{K_p} \right).$$

Noting that $\dot{L}(u) = L(u)[1 - L(u)] \in (0,1)$ for all u, we obtain

$$\begin{aligned}
E|\mathbf{1}^p_N B_{4N}| &\leq \|q^0 - R'_{K_q}\gamma^*_{K_q}\|_\infty |\widehat{\pi}_{K_p} - \pi^*_{K_p}| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} E\left[\frac{\mathbf{1}^p_N D_i \dot{L}(R_{K_p}(W_i)'\overline{\pi})}{p(W_i)\widehat{p}(W_i)} |R_{K_p}(W_i)'| \right] \\
&\leq \|q^0 - R'_{K_q}\gamma^*_{K_q}\|_\infty |\widehat{\pi}_{K_p} - \pi^*_{K_p}| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sqrt{E\left[\frac{\mathbf{1}^p_N D_i}{p(W_i)\widehat{p}(W_i)} \right]^2} \sqrt{E\left[|R_{K_p}(W_i)'|^2\right]}.
\end{aligned}$$

But $E\left[|R_{K_p}(W_i)'|^2\right] = K_p$ by definition of $R_{K_p}(W)$. Therefore, by Lemmas (B4), (B2), and (4) and Assumption M(2) respectively,

$$E|\mathbf{1}^p_N B_{4N}| \leq O(K_q^{-s_q/d_w}) O_p\left(K_p^{1/2} N^{-1/2} + K_p^{1/2} K_p^{-s_p/d_w} \right) \sqrt{N} O(1) \sqrt{K_p}.$$

Hence, $E|\mathbf{1}^p_N B_{4N}| \leq O_p\left(N^{v_p - v_q\frac{s_q}{d_w}} + N^{\frac{1}{2} + v_p - \left(v_p\frac{s_p}{d_w} + v_q\frac{s_q}{d_w}\right)} \right) = o_p(1)$ since $v_q\frac{s_q}{d_w} > v_p$ and $v_p\frac{s_p}{d_w} + v_q\frac{s_q}{d_w} > \frac{1}{2} + v_p$ by (14). This gives $|\mathbf{1}^p_N B_{4N}| = o_p(1)$. Recalling that $\mathbf{1}^p_N \xrightarrow{P} 1$, we obtain that $B_{4N} = o_p(1)$.

Following steps similar for B_{4N} we obtain

$$\begin{aligned} B_{5N} &= -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{D_i}{p(W_i)\widehat{p}(W_i)}R_{K_q}(W_i)'\left(\widehat{\gamma}^0_{K_q}-\gamma^*_{K_q}\right)\left[p(W_i)-L(R_{K_p}(W_i)'\pi^*_{K_p})\right], \text{ and hence} \\ E|\mathbf{1}^p_N B_{5N}| &\leq \|p-L(R'_{K_q}\gamma^*_{K_p})\|_\infty|\widehat{\gamma}^0_{K_q}-\gamma^*_{K_q}|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}E\left[\frac{\mathbf{1}^p_N D_i}{p(W_i)\widehat{p}(W_i)}|R_{K_q}(W_i)'|\right] \\ &= O(K_p^{-s_p/d_w})O_p\left(K_q^{1/2}N^{-1/2}+K_q^{-s_q/d_w}\right)\sqrt{N}O(1)\sqrt{K_q} \end{aligned}$$

by Lemmas (B1), (B5), and (4) and Assumption M(2), and the definition of $R_{K_q}(W)$ respectively. Therefore, $E|\mathbf{1}^p_N B_{5N}| \leq O_p\left(N^{v_q - v_p\frac{s_p}{d_w}} + N^{\frac{1}{2}+\frac{v_q}{2}-\left(v_p\frac{s_p}{d_w}+v_q\frac{s_q}{d_w}\right)}\right) = o_p(1)$ since $v_p\frac{s_p}{d_w} > v_q$ and $v_p\frac{s_p}{d_w}+v_q\frac{s_q}{d_w} > \frac{1}{2}+\frac{v_q}{2}$ by (14). Hence, as before, it follows that $B_{5N} = o_p(1)$.

Finally, again denoting $\dot{L}(u) := \frac{\partial}{\partial u}L(u)$, a mean-value expansion gives for some mean-value $\tilde{\pi}$,

$$B_{6N} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{D_i}{p(W_i)\widehat{p}(W_i)}R_{K_q}(W_i)'\left(\widehat{\gamma}^0_{K_q}-\gamma^*_{K_q}\right)\dot{L}(R_{K_p}(W_i)'\tilde{\pi})R_{K_p}(W_i)'\left(\widehat{\pi}_{K_p}-\pi^*_{K_p}\right).$$

Using $\dot{L}(u) \in (0,1)$, note that (4) and Assumption M(2) give

$$\begin{aligned} |\mathbf{1}^p_N B_{6N}| &\leq |\widehat{\gamma}^0_{K_q}-\gamma^*_{K_q}||\widehat{\pi}_{K_p}-\pi^*_{K_p}|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\frac{\mathbf{1}^p_N D_i\dot{L}(R_{K_p}(W_i)'\tilde{\pi})}{p(W_i)\widehat{p}(W_i)}|R_{K_q}(W_i)R_{K_p}(W_i)'| \\ &\leq |\widehat{\gamma}^0_{K_q}-\gamma^*_{K_q}||\widehat{\pi}_{K_p}-\pi^*_{K_p}|\frac{\sqrt{N}}{\kappa^2}\frac{1}{N}\sum_{i=1}^{N}|R_{K_q}(W_i)R_{K_p}(W_i)'|. \end{aligned}$$

$E\left[|R_{K_q}(W)R_{K_p}(W)'|\right] \leq E\left[|R_{K_q}(W)||R_{K_p}(W)'|\right] \leq \sqrt{E\left[|R_{K_q}(W)|^2\right]E\left[|R_{K_p}(W)'|^2\right]} = \sqrt{K_qK_p}$. Hence $\frac{1}{N}\sum_{i=1}^{N}|R_{K_q}(W_i)R_{K_p}(W_i)'| \leq O_p(\sqrt{K_qK_p})$. Therefore, by Lemmas (B5) and (B2),

$$|\mathbf{1}^p_N B_{6N}| \leq O_p\left(K_q^{1/2}N^{-1/2}+K_q^{-s_q/d_w}\right)O_p\left(K_p^{1/2}N^{-1/2}+K_p^{1/2}K_p^{-s_p/d_w}\right)\sqrt{N}\sqrt{K_pK_q},$$

giving $|\mathbf{1}^p_N B_{6N}| \leq O_p\left(N^{-\frac{1}{2}+v_p+v_q}+N^{v_p+v_q\left(\frac{1}{2}-\frac{s_q}{d_w}\right)}+N^{v_q+v_p\left(1-\frac{s_p}{d_w}\right)}+N^{\frac{1}{2}+v_p+\frac{v_q}{2}-\left(v_p\frac{s_p}{d_w}+v_q\frac{s_q}{d_w}\right)}\right) = o_p(1)$ since $v_p+v_q < \frac{1}{2}$, $v_q\frac{s_q}{d_w} > v_p+\frac{v_q}{2}$, $v_p\frac{s_p}{d_w} > v_p+v_q$ and $v_p\frac{s_p}{d_w}+v_q\frac{s_q}{d_w} > v_p+\frac{v_q}{2}+\frac{1}{2}$ by (14). $\mathbf{1}^p_N \xrightarrow{P} 1$ implies $B_{6N} = o_p(1)$. Hence $A_{3N} = o_p(1)$. ■

Verification of the comment in Section 2 that the asymptotic variance of IPW-GMM equals the efficiency bound based on a smaller set of moment restrictions:

Toward the end of Section 2 we noted that: The asymptotic variance of the semiparametric IPW-GMM estimator in the general case equals the efficiency bound for estimation of β by combining the moment restrictions in (15) and (18) and instead considering a modified restriction:

$$E\left[\text{Proj}_W\left(\phi(D_1,D_2,Z,W;\beta)|\phi_0\right)\right] = 0 \text{ for } \beta \in \mathcal{B} \subset \mathbb{R}^{d_\beta} \text{ if and only if } \beta = \beta^0.$$

We verify this by obtaining the concerned efficiency bound. The idea is same as Theorem 2.1 and the discussion following it in Graham (2011). To convert the conditional (on W) restrictions into unconditional ones, we consider W with a known finite support $\mathcal{W} = \{w_1, w_2, \ldots, w_L\}$. This gets rid of the infinite dimensional nuisance parameters $p(W)$ that arises with an infinite support of W, and instead introduces a finite number of unknown nuisance parameters $\rho = (\rho'_{10}, \rho'_{01}, \rho'_{11})'$ where $\rho_{jk} = (\rho_{jk}(1) := P(D_1 = j, D_2 = k | W = w_1), \ldots, \rho_{jk}(L) := P(D_1 = j, D_2 = k | W = w_L))'$ for $j, k = 0, 1$. In Lemma C we obtain the Fisher information bound for β^0 in this model treating ρ as an unknown (finite dimensional) nuisance parameter. Since the bound does not depend on the multinomial assumption for W, the same arguments as in Graham (2011) (page 442) establish that this is the semiparametric efficiency bound β^0 under the moment restrictions (15) and (18).

Lemma C: Suppose that (i) the distribution of W has a known, finite support $\mathcal{W} = \{w_1, \ldots, w_L\}$, (ii) there is some $\beta^0 \in \mathcal{B} \subset \mathbb{R}^{d_\beta}$ and $\rho^0 = (\rho^{0'}_{10}, \rho^{0'}_{01}, \rho^{0'}_{11})' \in R_1 \times \ldots \times R_L$ such that (15) and (18) hold. (iii) For each $l = 1, \ldots, L$ the space $R_l := \{(r_l(1), r_l(2), r_l(3)) :$ such that $r_l(1), r_l(2), r_l(3), 1 - (r_l(1) + r_l(2) + r_l(3)) \geq \kappa \in (0, 1)\}$ satisfies Assumption M(2). (iv) Other assumptions in Theorem 2.1 of Graham (2011) hold. Then the Fisher information bound for β^0 is $\left(G'[V + \Delta]^{-1} G\right)^{-1}$ [see (8)].

Proof of Lemma C: To simplify notations let β and ρ denote β^0 and ρ^0 unless explicitly stated otherwise. The result follows from the same three steps in the proof of Theorem 2.1 in Graham (2011).

Step 1: Let C be an $L \times 1$ vector with 1 in the l-th row if $W = w_l$ and 0 elsewhere, and $\tau_l := P(W = w_l)$. Exactly following Graham (2011), it can be established that the restrictions (15) and (18) are, in the multinomial case, equivalent to a finite number $(d_g + 3L)$ of unconditional moment restrictions:

$$E[m(D_1, D_2, Z_1, Z_2, W; \beta, \rho)] = E \begin{bmatrix} m_1(D_1, D_2, Z_1, Z_2, W; \beta, \rho) \\ m_2(D_1, D_2, W; \rho) \end{bmatrix} = 0,$$

$$\text{where} \qquad m_1(.; \beta, \rho) = \frac{D_1 D_2}{C' \rho_{11}} g(Z_1, Z_2, W; \beta) \text{ and } m_2(.; \rho) = C \otimes \begin{bmatrix} D_1(1 - D_2) - C'\rho_{10} \\ (1 - D_1) D_2 - C'\rho_{01} \\ D_1 D_2 - C'\rho_{11} \end{bmatrix}.$$

Step 2: Following Graham (2011) it can be shown that the variance bound for β under the sole restriction $E[m(D_1, D_2, Z_1, Z_2, W; \beta, \rho)] = 0$ is the upper (north-west) $d_\beta \times d_\beta$ block of the matrix $(M' \bar{V}^{-1} M)^{-1}$ where

$$\begin{aligned} M &= \left[M_\beta := E\left[\frac{\partial}{\partial \beta'} m(.; \beta, \rho)\right] = G, M_\rho := E\left[\frac{\partial}{\partial \rho'} m(.; \beta, \rho)\right]\right] \\ \bar{V} &= \begin{bmatrix} \bar{V}_{11} := E[m_1(.; \beta, \rho) m_1(.; \beta, \rho)'] & \bar{V}_{12} := E[m_1(.; \beta, \rho) m_2(.; \rho)'] \\ \bar{V}_{21} := E[m_2(.; \rho) m_1(.; \beta, \rho)'] & \bar{V}_{22} := E[m_2(.; \rho) m_2(.; \rho)'] \end{bmatrix} \end{aligned}$$

Since $m_2(.; \beta)$ does not involve β (meaning, $M_\beta = [G', 0]'$, i.e., the bottom $3L$ rows of M_β are identically 0), it follows after some algebra (shown below) that this bound is equal to

$$\left(G' \left(\bar{V}_{11} - \bar{V}_{12} \bar{V}_{22}^{-1} \bar{V}_{21}\right)^{-1} G\right)^{-1}. \tag{36}$$

This holds because the $d_\beta \times 3L$ block in the north-east of $(M'\bar{V}^{-1}M)^{-1}$ is a zero-block (and same for the $3L \times d_\beta$ block in the south-west). Under assumptions M(2)-(4), we show this below by equivalently showing that the $d_\beta \times 3L$ block in the north-east of $M'\bar{V}^{-1}M$ is a zero-block. This is equivalent to showing that the $d_\beta \times 3L$ matrix $G'\left(\bar{V}_{11} - \bar{V}_{12}\bar{V}_{22}^{-1}\bar{V}_{21}\right)^{-1} M_{1\rho} - G'\left(\bar{V}_{11} - \bar{V}_{12}\bar{V}_{22}^{-1}\bar{V}_{21}\right)^{-1} \bar{V}_{12}\bar{V}_{22}^{-1}M_{2\rho}$ is zero, where $M_{1\rho}$ and $M_{2\rho}$ respectively denote the first d_g and the last $3L$ rows of M_ρ. A sufficient condition for this is $M_{1\rho} = \bar{V}_{12}\bar{V}_{22}^{-1}M_{2\rho}$, and in the rest of Step 2 we verify that it holds. Define $A := \left[\frac{\tau_1}{\rho_{11}(1)}q(1), \ldots, \frac{\tau_L}{\rho_{11}(L)}q(L)\right]$ where $q(l) := E[g(Z, W; \beta^0)|W = w_l]$. Hence $M_{1\rho} = -[0, 0, A]$. On the other hand, $M_{2\rho} = -[\tau_1 B(1)', \ldots, \tau_L B(L)']'$ where, for $l = 1, \ldots, L$, $B(l) := [e(l), e(L + l), e(2L + l)]'$ and $e(k)$ is a $3L \times 1$ unit vector with 1 in the k-th element and zeros elsewhere. Define $E[\phi(.;\beta^0)\phi_0(.)'|W] = H(W)'$ and $(E[\phi_0(.)\phi_0(.)'|W])^{-1} = K(W)$ where

$$
\begin{aligned}
H(W)' &:= -\left[p_{10}(W), p_{01}(W), p_{11}(W) - 1\right] q(W), \\
K(W) &:= J(W)^{-1} = \begin{bmatrix} \frac{p_{00}(W)+p_{10}(W)}{p_{00}(W)p_{10}(W)} & \frac{1}{p_{00}(W)} & \frac{1}{p_{00}(W)} \\ \frac{1}{p_{00}(W)} & \frac{p_{00}(W)+p_{01}(W)}{p_{00}(W)p_{01}(W)} & \frac{1}{p_{00}(W)} \\ \frac{1}{p_{00}(W)} & \frac{1}{p_{00}(W)} & \frac{p_{00}(W)+p_{11}(W)}{p_{00}(W)p_{11}(W)} \end{bmatrix} \text{ and} \\
J(W) &:= \begin{bmatrix} p_{10}(W)(1-p_{10}(W)) & -p_{10}(W)p_{01}(W) & -p_{10}(W)p_{11}(W) \\ -p_{01}(W)p_{10}(W) & p_{01}(W)(1-p_{01}(W)) & -p_{01}(W)p_{11}(W) \\ -p_{11}(W)p_{10}(W) & -p_{11}(W)p_{01}(W) & p_{11}(W)(1-p_{11}(W)) \end{bmatrix},
\end{aligned}
$$

and for $W = w_l$ ($l = 1, \ldots, L$) denote them by $H(l)$, $K(l)$ and $J(l)$. Therefore, some algebra gives

$$
\begin{aligned}
\bar{V}_{12} &= [\tau_1 H(1)', \tau_2 H(2)', \ldots, \tau_L H(L)'] = \bar{V}_{21}', \\
\bar{V}_{22} &= \text{diag}\{\tau_1 J(1), \tau_2 J(2), \ldots, \tau_L J(L)\}, \\
\text{and } \bar{V}_{22}^{-1} &= \text{diag}\left\{\frac{1}{\tau_1}K(1), \frac{1}{\tau_2}K(2), \ldots, \frac{1}{\tau_L}K(L)\right\}, \\
\text{and hence } \bar{V}_{12}\bar{V}_{22}^{-1} &= [H(1)'K(1), H(2)'K(2), \ldots, H(L)'K(L)].
\end{aligned}
$$

Letting $T_j(l)$ denote the j-th column of the $d_g \times 3$ matrix $H(l)'K(l)$ for $j = 1, 2, 3$ and $l = 1, \ldots, L$, we obtain

$$
\bar{V}_{12}\bar{V}_{22}^{-1}M_{2\rho} = -\left[\{\tau_1 T_1(1), \ldots, \tau_L T_1(L)\}, \{\tau_1 T_2(1), \ldots, \tau_L T_2(L)\}, \{\tau_1 T_3(1), \ldots, \tau_L T_3(L)\}\right] = M_{1\rho}
$$

by distributing the columns according to the selection elements in the matrices $B(1), \ldots, B(L)$. Therefore, the sufficient condition is verified and hence (36) gives the variance bound.[16]

Step 3: Since $\bar{V}_{11} = E[\phi(.;\beta^0)\phi(.;\beta^0)']$ and $\bar{V}_{12}\bar{V}_{22}^{-1}\bar{V}_{21} = \sum_{l=1}^{L} \tau_l H(l)'K(l)H(l) = E\left[E\left[\phi\phi_0'|W\right]\left(E\left[\phi_0\phi_0'|W\right]\right)^{-1} E\left[\phi_0\phi'|W\right]\right]$, simple algebra gives $\bar{V}_{11} - \bar{V}_{12}\bar{V}_{22}^{-1}\bar{V}_{21} = V + \Delta$ [see (8)] and hence completes the proof . ■

[16] Recall that this sufficient condition has the important implication that knowing the nuisance parameters ρ^0 does not lead to more efficient estimator of β^0 under the current setup.