

Online Appendix to “Inference for Treatment Effects Conditional on Generalized Principal Strata using Instrumental Variables”

Appendices C, D, E contain additional examples and details of the simulation study and the empirical application, and can be found at <https://github.com/edward-vytlacil/STRATA-ATE/blob/main/>.

A.4 Proof of Theorem 4.3

PROOF OF THEOREM 4.3. Suppose $P \in \mathbf{P}_0^w$. Then, it follows that $\theta_0 \in \Theta_0^w(P, \mathbf{Q}^w)$, so there exists a $Q \in \mathbf{Q}_0^w(P, \mathbf{Q}^w)$ such that $\theta_0 = \theta^w(Q)$ and $Q\{R \in \mathcal{R}', W \in \mathcal{W}'\} > 0$. Recall $Q \in \mathbf{Q}_0^w(P, \mathbf{Q}^w)$ if and only if $Q \in \mathbf{Q}^w$ and $P = Q(T^w)^{-1}$. Since $Q \in \mathbf{Q}^w$, it satisfies both Assumptions 4.2 and 2.2. Because $P = Q(T^w)^{-1}$, we can conclude that

$$Q\{Z = z, W = w\} = P\{Z = z, W = w\} \quad (49)$$

for all $(z, w) \in \mathcal{Z} \times \mathcal{W}$. Moreover, since $P\{Z = z, W = w\} > 0$ for all $(z, w) \in \mathcal{Z} \times \mathcal{W}$ for any $P \in \mathbf{P}_0^w$, Q satisfying Assumption 4.2 and $P = Q(T^w)^{-1}$ imply that

$$\begin{aligned} P\{Y = y, D = d|Z = z, W = w\} &= Q\{Y(d) = y, D(z) = d|Z = z, W = w\} \\ &= Q\{Y(d) = y, D(z) = d|W = w\}. \end{aligned} \quad (50)$$

Hence, results (49) and (50) together with $Q\{W = w\} = P\{W = w\}$ imply that

$$P\{Y = y, D = d|Z = z, W = w\} \times P\{W = w\} = Q\{Y(d) = y, D(z) = d, W = w\}. \quad (51)$$

Next, let q be a column vector with $|\mathcal{R}| \times |\mathcal{W}|$ elements indexed by $(r, w) \in \mathcal{R} \times \mathcal{W}$, where the (r, w) coordinate of q equals $q(r, w) = Q\{R = r, W = w\}$. Then note result (51) and the definitions of A^w and $\beta^w(P)$ imply that the first $|\mathcal{M}^w|$ rows of $A^w q$ equal the first $|\mathcal{M}^w|$ rows of $\beta^w(P)$. Moreover, result (49) and $Q\{R \in \mathcal{R}\} = 1$ by definition of $\mathbf{Q}_0^w(P, \mathbf{Q}^w)$ imply

$$\sum_{r \in \mathcal{R}} q(r, w) = \sum_{r \in \mathcal{R}} Q\{R = r, W = w\} = Q\{W = w\} = P\{W = w\}, \quad (52)$$

and therefore it follows that the $|\mathcal{M}^w| + 1$ through $|\mathcal{M}^w| + |\mathcal{W}|$ rows of $A^w q$ equal the $|\mathcal{M}^w| + 1$ through $|\mathcal{M}^w| + |\mathcal{W}|$ rows of $\beta^w(P)$. Finally, because $\theta_0 = \theta^w(Q)$ by hypothesis, the definition of the conditional expectation yields that

$$\theta_0 = \frac{E_Q[g(R, W)I\{R \in \mathcal{R}', W \in \mathcal{W}'\}]}{Q\{R \in \mathcal{R}', W \in \mathcal{W}'\}}, \quad (53)$$

where we recall $Q\{R \in \mathcal{R}', W \in \mathcal{W}'\}$ satisfies $Q\{R \in \mathcal{R}', W \in \mathcal{W}'\} > 0$. Rearranging the expression in (53) we obtain that $E_Q[(g(R, W) - \theta_0)I\{R \in \mathcal{R}', W \in \mathcal{W}'\}] = 0$ or, equivalently,

$$\sum_{r \in \mathcal{R}', w \in \mathcal{W}'} (g(r, w) - \theta_0)q(r, w) = 0 .$$

Thus, it follows from the definition of A^w that the $|\mathcal{M}^w| + |\mathcal{W}| + 1$ row of $A^w q$ equals the $|\mathcal{M}^w| + |\mathcal{W}| + 1$ row of $\beta^w(P)$. Hence, we can conclude that $\mathbf{P}_0^w \subseteq \tilde{\mathbf{P}}_0^w \equiv \{P \in \mathbf{P}^w : A^w x = \beta^w(P) \text{ for some } x \geq 0\}$, which establishes the first claim of the theorem.

Next, we establish the second claim by showing $\tilde{\mathbf{P}}_0^w \subseteq \text{cl}_{\mathbf{P}^w}(\mathbf{P}_0^w)$ whenever $\mathbf{P}_0^w \neq \emptyset$. To this end, fix any $\tilde{P} \in \tilde{\mathbf{P}}_0^w$ and note $A^w q = \beta^w(\tilde{P})$ for some $q \geq 0$ due to $\tilde{P} \in \tilde{\mathbf{P}}_0^w$. Since $\tilde{P} \in \mathbf{P}^w$ implies that $\tilde{P}\{Z = z, W = w\} > 0$ for all $(z, w) \in \mathcal{Z} \times \mathcal{W}$, it follows that we can define a finite measure $\tilde{Q}$ on the sample space $\mathcal{R} \times \mathcal{Z} \times \mathcal{W}$ according to

$$\tilde{Q}\{R = r, Z = z, W = w\} = q(r, w)\tilde{P}\{Z = z|W = w\} . \quad (54)$$

Next, note that the $|\mathcal{M}^w| + 1$ through $|\mathcal{M}^w| + |\mathcal{W}|$ rows of $A^w q$ being equal to the corresponding rows of $\beta^w(\tilde{P})$ further allow us to conclude that

$$\sum_{z \in \mathcal{Z}} \sum_{w \in \mathcal{W}} \sum_{r \in \mathcal{R}} \tilde{Q}\{R = r, Z = z, W = w\} = \sum_{z \in \mathcal{Z}} \sum_{w \in \mathcal{W}} \tilde{P}\{W = w\}\tilde{P}\{Z = z|W = w\} = 1,$$

which implies that $\tilde{Q}$ is a probability measure. Moreover, by construction, $R \perp\!\!\!\perp Z|W$ under $\tilde{Q}$ and $\tilde{Q}\{R \in \mathcal{R}\} = 1$, which implies that $\tilde{Q} \in \mathbf{Q}^w$. Similar arguments yield that $\tilde{Q}\{Z = z, W = w\} = \tilde{P}\{Z = z, W = w\}$ for any $(z, w) \in \mathcal{Z} \times \mathcal{W}$, while the first $|\mathcal{M}^w|$ rows of $A^w q$ being equal to the first $|\mathcal{M}^w|$ rows of $\beta^w(\tilde{P})$ imply that

$$\begin{aligned} \tilde{P}\{Y = y, D = d|Z = z, W = w\} &= \sum_{r \in \mathcal{R}} I\{y(d) = y, d(z) = d\} \frac{q(r, w)}{\tilde{P}\{W = w\}} \\ &= \sum_{r \in \mathcal{R}} I\{y(d) = y, d(z) = d\} \frac{\tilde{Q}\{R = r, Z = z, W = w\}}{\tilde{P}\{Z = z, W = w\}} \\ &= \tilde{Q}\{Y(d) = y, D(z) = d|Z = z, W = w\} , \end{aligned} \quad (55)$$

where in the second equality we used the definition of $\tilde{Q}$, and in the third that $\tilde{P}\{W = w, Z = z\} = \tilde{Q}\{W = w, Z = z\}$. Since (55) holds for any $(y, d, z, w) \in \mathcal{M}^w$ and $\tilde{P}\{(Y, D, Z, W) \in \mathcal{M}^w\} = 1$ due to $\tilde{P} \in \mathbf{P}^w$, it follows that $\tilde{P} = \tilde{Q}(T^w)^{-1}$. Furthermore, we also have

$$\begin{aligned} E_{\tilde{Q}}[(g(R, W) - \theta_0)I\{R \in \mathcal{R}', W \in \mathcal{W}'\}] \\ = \sum_{r \in \mathcal{R}'} \sum_{w \in \mathcal{W}'} (g(r, w) - \theta_0)I\{r \in \mathcal{R}', w \in \mathcal{W}'\}q(r, w) = 0, \end{aligned} \quad (56)$$

where the first equality follows by definition of $\tilde{Q}$, and the second from the final row of $A^w q$ being equal to zero.

On the other hand, because $\mathbf{P}_0^w \neq \emptyset$, there exists $Q^* \in \mathbf{Q}^w$, hence satisfying Assumptions 4.2 and 2.2, such that $Q^*\{R \in \mathcal{R}', W \in \mathcal{W}'\} > 0$ and $\theta^w(Q^*) = \theta_0$. Moreover, we have

$$Q^*\{R = r, Z = z, W = w\} = Q^*\{R = r, W = w\}Q^*\{Z = z|W = w\}$$

because Q^* satisfies Assumption 4.2 due to $Q^* \in \mathbf{Q}^w$. Finally, we define a measure Q by

$$Q\{R = r, Z = z, W = w\} = Q^*\{R = r, W = w\}\tilde{P}\{Z = z|W = w\},$$

and note that $Q \in \mathbf{Q}^w$ because $Q^* \in \mathbf{Q}^w$. Furthermore, $Q\{R = r, W = w\} = Q^*\{R = r, W = w\}$ for all $r \in \mathcal{R}$ and $w \in \mathcal{W}$, which implies $Q\{R \in \mathcal{R}', W \in \mathcal{W}'\} > 0$ and $\theta^w(Q) = \theta_0$. To conclude, we let $P = Q(T^w)^{-1}$, $P_n = n^{-1}P + (1 - n^{-1})\tilde{P}$, and $Q_n = n^{-1}Q + (1 - n^{-1})\tilde{Q}$. Since $P\{Z = z|W = w\} = \tilde{P}\{Z = z|W = w\}$ for any $(z, w) \in \mathcal{Z} \times \mathcal{W}$, it is straightforward to show that $P_n = Q_n(T^w)^{-1}$ and that Q_n satisfies Assumptions 2.2 and 4.2. Moreover, $Q\{R \in \mathcal{R}', W \in \mathcal{W}'\} > 0$, $\theta^w(Q) = \theta_0$, and result (56) yield that $\theta^w(Q_n) = \theta_0$. We thus obtain that $\theta_0 \in \Theta_0^w(P_n, \mathbf{Q}^w)$ for all n . Since $P_n \in \mathbf{P}^w$, it follows that $P_n \in \mathbf{P}_0^w$ and, because $P_n \rightarrow \tilde{P}$ as $n \rightarrow \infty$, that $\tilde{P} \in \text{cl}_{\mathbf{P}^w}(\mathbf{P}_0^w)$ as desired. ■

A.5 Proof of Theorem 5.1

In this appendix, we first establish Theorem 5.1 and subsequently show that the result continues to hold if we remove linearly dependent restrictions from A_L and $\beta_L(P)$.

Before proceeding with the proof, we first introduce some additional notation. Given the sets $\mathcal{R} \subseteq \mathcal{Y}^{|\mathcal{D}|} \times \mathcal{D}^{|\mathcal{Z}|}$ and $\mathcal{R}' \subseteq \mathcal{Y}^{|\mathcal{D}|} \times \mathcal{D}^{|\mathcal{Z}|}$ we define, for any $r_t \in \mathcal{D}^{|\mathcal{Z}|}$, the sets

$$\begin{aligned}\mathcal{R}_o(r_t) &\equiv \{r_o \in \mathcal{Y}^{|\mathcal{D}|} : (r_o, r_t) \in \mathcal{R}\} \\ \mathcal{R}'_o(r_t) &\equiv \{r_o \in \mathcal{Y}^{|\mathcal{D}|} : (r_o, r_t) \in \mathcal{R}'\}.\end{aligned}$$

Recall $B_{c_o} \equiv \prod_{d \in \mathcal{D}} B_{c_o(d), L} \subseteq \mathcal{Y}^{|\mathcal{D}|}$ for any $c_o \equiv (c_o(d) : d \in \mathcal{D}) \in \mathcal{L}^{|\mathcal{D}|}$ and $\mathcal{M}_L$ is defined as the set of realizations for $(\mathbb{L}, D, Z)$ that is compatible with the response types R belonging to $\mathcal{R}$ (see (24)). To define the matrix A_L we introduce a binary variable $\chi \in \{0, 1\}$ and for any $c_o = (c_o(d) : d \in \mathcal{D})$ and $r_t = (d(z) : z \in \mathcal{Z})$ we set

$$\mathcal{Y}(c_o, r_t, \chi) = \begin{cases} B_{c_o} \cap \mathcal{R}'_o(r_t) & \text{if } \chi = 1 \\ \mathcal{R}_o(r_t) \cap (B_{c_o} \setminus \mathcal{R}'_o(r_t)) & \text{if } \chi = 0. \end{cases}$$

Note that because $\mathcal{R}'_o(r_t) \subseteq \mathcal{R}_o(r_t)$, the set $\mathcal{Y}(c_o, r_t, 0) \cup \mathcal{Y}(c_o, r_t, 1)$ is non-empty if and only

if $B_{c_o} \cap \mathcal{R}_o(r_t) \neq \emptyset$. Using this notation we define the set of indices

$$\mathcal{I} \equiv \{(c_o, r_t, \chi) \in \mathcal{L}^{|\mathcal{D}|} \times \mathcal{D}^{|\mathcal{Z}|} \times \{0, 1\} : \mathcal{Y}(c_o, r_t, \chi) \neq \emptyset\}. \quad (57)$$

We can finally define A_L as a matrix with $|\mathcal{M}_L| + 3$ rows and $|\mathcal{I}| + 2$ columns. We index the first $|\mathcal{M}_L|$ rows of A_L by $(\ell, d, z) \in \mathcal{M}_L$, the first $|\mathcal{I}|$ columns by $(c_o, r_t, \chi) \in \mathcal{I}$, and set

$$\begin{aligned} A_{L,(\ell,d,z),(c_o,r_t,\chi)} &= I\{c_o(d) = \ell, d(z) = z\} \\ A_{L,|\mathcal{M}_L|+1,(c_o,r_t,\chi)} &= 1. \end{aligned}$$

We also set the $(|\mathcal{M}_L| + 2) \times (c_o, r_t, \chi)$ and $(|\mathcal{M}_L| + 3) \times (c_o, r_t, \chi)$ elements of A_L to be

$$\begin{aligned} A_{L,|\mathcal{M}_L|+2,(c_o,r_t,\chi)} &= \left(\sup_{r_o \in \mathcal{Y}(c_o,r_t,\chi)} g(r_o, r_t) - \theta_0 \right) \cdot \chi \\ A_{L,|\mathcal{M}_L|+3,(c_o,r_t,\chi)} &= \left(\inf_{r_o \in \mathcal{Y}(c_o,r_t,\chi)} g(r_o, r_t) - \theta_0 \right) \cdot \chi. \end{aligned}$$

Finally, the $(|\mathcal{M}_L| + 2) \times (|\mathcal{I}| + 1)$ element of A_L is given by -1 and the $(|\mathcal{M}_L| + 3) \times (|\mathcal{I}| + 2)$ element of A_L is given by 1 . All other elements of A_L are set to equal zero.

Given the introduced definitions, we next prove Theorem 5.1.

PROOF OF THEOREM 5.1. First note that $\mathcal{R}'_o(r_t) \subseteq \mathcal{R}_o(r_t)$ implies that $B_{c_o} \cap \mathcal{R}_o(r_t) \neq \emptyset$ if and only if $\mathcal{Y}(c_o, r_t, 0) \cup \mathcal{Y}(c_o, r_t, 1) \neq \emptyset$. Therefore, the definition of $\mathcal{I}$, $\mathcal{R} \subseteq \mathcal{Y}^{|\mathcal{D}|} \times \mathcal{D}^{|\mathcal{Z}|}$, and $\{B_{c_o}\}_{c_o \in \mathcal{L}^{|\mathcal{D}|}}$ being a partition of $\mathcal{Y}^{|\mathcal{D}|}$ by assumption allows us to conclude that

$$\bigcup_{(c_o,r_t,\chi) \in \mathcal{I}} (\mathcal{Y}(c_o, r_t, \chi) \times \{r_t\}) = \mathcal{R}. \quad (58)$$

Similarly, we also note that by the same arguments and the definition of $\mathcal{Y}(c_o, r_t, \chi)$ we have

$$\bigcup_{(c_o,r_t):(c_o,r_t,1) \in \mathcal{I}} (\mathcal{Y}(c_o, r_t, 1) \times \{r_t\}) = \mathcal{R}'. \quad (59)$$

To show $\mathbf{P}_0 \subseteq \mathbf{P}_{0,L}$ we first suppose $P \in \mathbf{P}_0$. Then, there exists a $Q \in \mathbf{Q}_0(P, \mathbf{Q})$ such that $\theta_0 = \theta(Q)$ and $Q\{R \in \mathcal{R}'\} > 0$. Letting $q(c_o, r_t, \chi) \equiv Q\{R_o \in \mathcal{Y}(c_o, r_t, \chi), R_t = r_t\}$ for any $(c_o, r_t, \chi) \in \mathcal{I}$ we then obtain from $Q \in \mathbf{Q}_0(P, \mathbf{Q})$, result (58), and $Q\{R \in \mathcal{R}\} = 1$ that

$$\begin{aligned} P\{Y \in B_{\ell,L}, D = d | Z = z\} &= Q\{Y(d) \in B_{\ell,L}, D(z) = d, R \in \mathcal{R}\} \\ &= \sum_{(c_o,r_t,\chi) \in \mathcal{I}} q(c_o, r_t, \chi) I\{c_o(d) = \ell, d(z) = z\}. \end{aligned} \quad (60)$$

Collect the probabilities $q(c_o, r_t, \chi)$ into a vector $q = (q(c_o, r_t, \chi) : (c_o, r_t, \chi) \in \mathcal{I})$ and define

$$\begin{aligned}\xi_1 &\equiv \sum_{(c_o, r_t) : (c_o, r_t, 1) \in \mathcal{I}} \left(\sup_{r_o \in \mathcal{Y}(c_o, r_t, 1)} g(r_o, r_t) - \theta_0 \right) q(c_o, r_t, 1) \\ \xi_2 &\equiv - \sum_{(c_o, r_t) : (c_o, r_t, 1) \in \mathcal{I}} \left(\inf_{r_o \in \mathcal{Y}(c_o, r_t, 1)} g(r_o, r_t) - \theta_0 \right) q(c_o, r_t, 1) .\end{aligned}$$

Next, note that since $\theta(Q) = \theta_0$ by hypothesis and $Q\{R \in \mathcal{R}'\} > 0$, it follows that Q satisfies $E_Q[(g(R) - \theta_0)I\{R \in \mathcal{R}'\}] = 0$. In particular, by result (59) we obtain that

$$-\xi_2 \leq E_Q[(g(R) - \theta_0)I\{R \in \mathcal{R}'\}] = 0 \leq \xi_1 .$$

Setting $x = (q', \xi_1, \xi_2)'$ we then note that $x \geq 0$. Moreover, it follows from (60) and the definition of A_L that the first $|\mathcal{M}_L|$ rows of $A_L x$ equal the first $|\mathcal{M}_L|$ rows of $\beta_L(P)$. Also note that result (58), the definition of q , and $Q\{R \in \mathcal{R}\} = 1$ allow us to conclude that

$$\sum_{(c_o, r_t, \chi) \in \mathcal{I}} q(c_o, r_t, \chi) = \sum_{(c_o, r_t, \chi) \in \mathcal{I}} Q\{R_o \in \mathcal{Y}(c_o, r_t, \chi), R_t = r_t\} = Q\{R \in \mathcal{R}\} = 1 .$$

In particular, it follows that the $|\mathcal{M}_L| + 1$ row of $A_L x$ equals the $|\mathcal{M}_L| + 1$ row of $\beta_L(P)$ by definition of A_L and $\beta_L(P)$. Finally, we note the $|\mathcal{M}_L| + 2$ and $|\mathcal{M}_L| + 3$ rows of $A_L x$ equal the $|\mathcal{M}_L| + 2$ and $|\mathcal{M}_L| + 3$ rows of $\beta_L(P)$ by definition of ξ_1, ξ_2, A_L and $\beta_L(P)$. Therefore, we have shown that $\mathbf{P}_0 \subseteq \mathbf{P}_{0,L} \equiv \{P \in \mathbf{P} : A_L x = \beta_L(P) \text{ for some } x \geq 0\}$.

To establish the second claim, we proceed by contradiction and assume there is a $P \in \limsup_{L \rightarrow \infty} \mathbf{P}_{0,L}$ with $P \notin \text{cl}_{\mathbf{P}}(\mathbf{P}_0)$. Since $P \notin \text{cl}_{\mathbf{P}}(\mathbf{P}_0)$, Lemma B.13 implies that

$$\int f dP = 1 , \quad \int f dP' \leq 0 \text{ for all } P' \in \mathbf{P}_0 \text{ satisfying } P'_Z = P_Z \quad (61)$$

for some continuous and bounded function $f : \mathcal{F} \rightarrow \mathbf{R}$ for $\mathcal{F} \equiv \mathcal{Y} \times \mathcal{D} \times \mathcal{Z}$ and P_Z denoting the marginal distribution of Z under P . Since the function $f : \mathcal{F} \rightarrow \mathbf{R}$ is continuous and $\mathcal{Y}$ is compact by Assumption 5.1(i), it follows that the function f is uniformly continuous on $\mathcal{F}$. For any $\eta > 0$ there therefore exists a $\delta > 0$ such that f satisfies

$$\max_{d \in \mathcal{D}, z \in \mathcal{Z}} \sup_{|y - y'| \leq 2\delta} |f(y, d, z) - f(y', d, z)| \leq \eta . \quad (62)$$

Moreover, by Assumption 5.1(iii) there is a $L_0 < \infty$ such that $\max_{1 \leq \ell \leq L} \text{diam}\{B_{\ell,L}\} < \delta$ for all $L \geq L_0$. Since in addition $P \in \limsup_{L \rightarrow \infty} \mathbf{P}_{0,L}$ implies that $P \in \mathbf{P}_{0,L}$ for some $L \geq L_0$, we can conclude that there is some $L < \infty$ such that

$$P \in \mathbf{P}_{0,L} \quad \text{and} \quad \max_{1 \leq \ell \leq L} \text{diam}\{B_{\ell,L}\} < \delta . \quad (63)$$

Next define λ to be a vector of dimension $|\mathcal{M}_L| + 3$ whose last three entries are zero, and each of its first $|\mathcal{M}_L|$ entries, which we index by $(\ell, d, z) \in \mathcal{M}_L$, is given by

$$\lambda(\ell, d, z) \equiv \sup_{y \in B_{\ell, L}} f(y, d, z) P\{Z = z\} .$$

Also observe that $P \in \mathbf{P}_{0, L}$ implies $\beta_L(P) = A_L x$ for some $x \geq 0$ belonging to $\mathbf{R}^{|\mathcal{I}|+2}$. Index the first $|\mathcal{I}|$ entries of x by $(c_o, r_t, \chi) \in \mathcal{I}$. For each $(c_o, r_t, \chi) \in \mathcal{I}$ we select an arbitrary $r_o^*(c_o, r_t, \chi) \in \mathcal{Y}(c_o, r_t, \chi)$ and define a finite measure Q'_R with support $\{(r_o^*(c_o, r_t, \chi), r_t)\}_{(c_o, r_t, \chi) \in \mathcal{I}}$ by setting $Q'_R\{(R_o, R_t) = (r_o^*(c_o, r_t, \chi), r_t)\} = x(c_o, r_t, \chi)$. Also note that by definition of Q'_R and the $|\mathcal{M}_L| + 1$ row of $A_L x$ being equal to one we have that

$$\sum_{(c_o, r_t, \chi) \in \mathcal{I}} Q'_R\{(R_o, R_t) = (r_o^*(c_o, r_t, \chi), r_t)\} = \sum_{(c_o, r_t, \chi) \in \mathcal{I}} x(c_o, r_t, \chi) = 1,$$

which implies Q'_R is a probability measure and, since $r_o^*(c_o, r_t, \chi) \in \mathcal{Y}(c_o, r_t, \chi)$, that

$$\sum_{(c_o, r_t, \chi) \in \mathcal{I}} Q'_R\{(R_o, R_t) \in (\mathcal{Y}(c_o, r_t, \chi) \times \{r_t\})\} = 1 . \quad (64)$$

Moreover, since the last two entries of $A_L x$ equal $(0, 0)$ because $A_L x = \beta_L(P)$, and in addition $Q'_R\{(R_o, R_t) \in (\mathcal{Y}(c_o, r_t, 1) \times \{r_t\})\} = x(c_o, r_t, 1)$ by definition of Q'_R , we obtain

$$\begin{aligned} \sum_{(c_o, r_t): (c_o, r_t, 1) \in \mathcal{I}} \left(\sup_{r_o \in \mathcal{Y}(c_o, r_t, 1)} g(r_o, r_t) - \theta_0 \right) \times Q'_R\{(R_o, R_t) \in (\mathcal{Y}(c_o, r_t, 1) \times \{r_t\})\} &\geq 0 \\ \sum_{(c_o, r_t): (c_o, r_t, 1) \in \mathcal{I}} \left(\inf_{r_o \in \mathcal{Y}(c_o, r_t, 1)} g(r_o, r_t) - \theta_0 \right) \times Q'_R\{(R_o, R_t) \in (\mathcal{Y}(c_o, r_t, 1) \times \{r_t\})\} &\leq 0 . \end{aligned} \quad (65)$$

Setting $Q' \equiv Q'_R \times P_Z$ and letting $P' = Q'T^{-1}$, then note $A_L x = \beta_L(P)$, the definition of Q'_R , and the manipulations in (60) (applied with P' and Q' in place of P and Q) imply $\beta_L(P) = \beta_L(P')$. Therefore, using that f satisfies condition (62), L was chosen to satisfy $\max_{1 \leq \ell \leq L} \text{diam}\{B_{\ell, L}\} < \delta$ in (63), the definition of λ , and $P\{(\mathbb{L}, D, Z) \in \mathcal{M}_L\} = 1$ gives

$$\begin{aligned} \int f dP &\leq \sum_{(\ell, d, z) \in \mathcal{M}_L} P\{Y \in B_{\ell, L}, D = d, Z = z\} \times \sup_{y \in B_{\ell, L}} f(y, d, z) \\ &= \lambda' \beta_L(P) = \lambda' \beta_L(P') \leq \int f dP' + \eta . \end{aligned} \quad (66)$$

On the other hand, Q'_R satisfying the restrictions in (64) and (65) implies by Lemma B.12 that there is a distribution $\tilde{Q}_R$ for R such that $\tilde{Q} \equiv \tilde{Q}_R \times P_Z$ satisfies $E_{\tilde{Q}}[(g(R) - \theta_0)I\{R \in \mathcal{R}'\}] = 0$, $\tilde{Q}\{R \in \mathcal{R}\} = 1$, and for which the induced distribution $\tilde{P} \equiv \tilde{Q}T^{-1}$ satisfies

$$\int f(dP' - d\tilde{P}) \leq 2\eta . \quad (67)$$

Finally, let $\mathbf{P}_0(P) \equiv \{\hat{P} \in \mathbf{P}_0 : \hat{P}_Z = P_Z\}$ and note that Lemma B.14 implies $\mathbf{P}_0(P)$ is not empty. Pick any $P_0 \in \mathbf{P}_0(P)$ and note that $P_0 = Q_0 T^{-1}$ for some $Q_0 \in \mathbf{Q}$ satisfying $Q_0 = Q_{0,R} \times P_Z$ for some distribution $Q_{0,R}$ for R . For any $\tau \in (0, 1]$, it then follows that $Q_\tau \equiv (1 - \tau)\tilde{Q} + \tau Q_0 = \{(1 - \tau)\tilde{Q}_R + \tau Q_{0,R}\} \times P_Z$ satisfies $E_{Q_\tau}[(g(R) - \theta_0)I\{R \in \mathcal{R}'\}] = 0$, $Q_\tau\{R \in \mathcal{R}\} = 1$, and $Q_\tau\{R \in \mathcal{R}'\} > 0$ because $Q_0\{R \in \mathcal{R}'\} > 0$. In particular, it follows that $P_\tau \equiv Q_\tau T^{-1} \in \mathbf{P}_0(P)$. Moreover, since f is bounded, we also have that $|\int f(d\tilde{P} - dP_\tau)| \leq 2\tau\|f\|_\infty$. Therefore, setting $\tau \leq \eta/(2\|f\|_\infty)$ allows us to conclude that

$$\int f d\tilde{P} \leq \int f dP_\tau + \eta \leq \eta, \quad (68)$$

where the final inequality holds by (61) and $P_\tau \in \mathbf{P}_0(P)$. Combining results (66), (67), and (68) finally yields $\int f dP \leq 4\eta$, which contradicts (61) for η sufficiently small. Hence, we conclude that there is no $P \in \limsup_{L \rightarrow \infty} \mathbf{P}_{0,L}$ satisfying $P \notin \text{cl}_{\mathbf{P}}(\mathbf{P}_0)$. ■

We next show that the conclusion of Theorem 5.1 also holds if we remove linearly redundant restrictions from A_L . To this end, recall that the first $|\mathcal{M}_L|$ rows of A_L and $\beta_L(P)$ are indexed by $(\ell, d, z) \in \mathcal{M}_L$. For each $z \in \mathcal{Z}$ we select an arbitrary $(\ell_z, d_z) \in \mathcal{M}_{L,z}$ (as defined in (23)), let A_L^s and $\beta_L^s(P)$ denote the matrix and vector obtained by removing the rows $((\ell_z, d_z, z) : z \in \mathcal{Z})$ from A_L and $\beta_L(P)$, and define the set of probabilities

$$\mathbf{P}_{0,L}^s \equiv \{P \in \mathbf{P} : A_L^s x = \beta_L^s(P) \text{ for some } x \geq 0\}.$$

Our next lemma establishes that $\mathbf{P}_{0,L} = \mathbf{P}_{0,L}^s$, which implies that the conclusion of Theorem 5.1 continues to hold if we replace A_L and $\beta_L(P)$ with A_L^s and β_L^s respectively.

Lemma A.3. *Let $\mathbf{P}$ denote the set of all distributions of (Y, D, Z) satisfying $P\{(\mathbb{L}, D, Z) \in \mathcal{M}_L\} = 1$ for all L and $P\{Z = z\} > 0$ for all $z \in \mathcal{Z}$. Then, it follows that $\mathbf{P}_{0,L} = \mathbf{P}_{0,L}^s$.*

PROOF OF LEMMA A.3. The proof follows similar arguments to those used in Lemma A.1. First note that any x satisfying $A_L x = \beta_L(P)$ also satisfies $A_L^s x = \beta_L^s(P)$, and therefore $\mathbf{P}_{0,L} \subseteq \mathbf{P}_{0,L}^s$. For the reverse inclusion select an arbitrary $(c_o, r_t, \chi) \in \mathcal{I}$ and $z \in \mathcal{Z}$. Since $\mathcal{Y}(c_o, r_t, \chi) \neq \emptyset$ for any $(c_o, r_t, \chi) \in \mathcal{I}$ and $\mathcal{R}' \subseteq \mathcal{R}$, it follows from the definition of $\mathcal{Y}(c_o, r_t, \chi)$ that there is a $r = (r_o, r_t) \in \mathcal{R}$ such that $r_o = (y(d) : d \in \mathcal{D})$ and $r_t = (d(z) : z \in \mathcal{Z})$ satisfy $r_o \in B_{c_o}$. By definition of $\mathcal{M}_{L,z}$ it follows that there exists an $(\ell, d) \in \mathcal{M}_{L,z}$ such that

$$1 = I\{y(d) \in B_{\ell,L}, d(z) = d\} = I\{c_o(d) = \ell, d(z) = d\}, \quad (69)$$

where the second equality follows from $r_o \in B_{c_o}$. Noting that, for any (c_o, r_t) , the indicator $I\{c_o(d) = \ell, d(z) = d\}$ can only equal one for a single (ℓ, d) , we obtain from (69) that

$$\sum_{(\ell,d) \in \mathcal{M}_{L,z}} A_{L,(\ell,d,z),(c_o,r_t,\chi)} = 1 \quad (70)$$

for any $(c_o, r_t, \chi) \in \mathcal{I}$ and $z \in \mathcal{Z}$. Hence, letting $A_{L,(\ell,d,z),\cdot}$ denote the (ℓ, d, z) row of the matrix A_L and using the definition of $A_{L,|\mathcal{M}_L|+1,\cdot}$ and that the final two columns of $A_{L,(\ell,d,z),\cdot}$ and $A_{L,|\mathcal{M}_L|+1,\cdot}$ equal zero, we obtain from result (70) that for any $z \in \mathcal{Z}$

$$\sum_{(\ell,d) \in \mathcal{M}_{L,z}} A_{L,(\ell,d,z),\cdot} = A_{L,|\mathcal{M}_L|+1,\cdot}. \quad (71)$$

Next select an arbitrary $P \in \mathbf{P}_{0,L}^s$ and note that by definition there exists an $x_0 \geq 0$ satisfying $A_L^s x_0 = \beta_L^s(P)$. Since the system $A_L^s x_0 = \beta_L^s(P)$ still contains the adding up constraint $A_{L,|\mathcal{M}_L|+1,\cdot} x_0 = 1$, result (71) allows us to conclude for any $z \in \mathcal{Z}$ that

$$1 = \sum_{(\ell,d) \in \mathcal{M}_{L,z}} A_{L,(\ell,d,z),\cdot} x_0 = \sum_{(\ell,d) \in \mathcal{M}_{L,z} \setminus \{(\ell_z, d_z)\}} \beta_{L,(\ell,d,z)}(P) + A_{L,(\ell_z, d_z, z),\cdot} x_0, \quad (72)$$

where in the second equality we used that $A_L^s x_0 = \beta_L^s(P)$ implies $A_{L,(\ell,d,z),\cdot} x_0 = \beta_{L,(\ell,d,z)}(P)$ for any $(\ell, d) \in \mathcal{M}_{L,z} \setminus \{(\ell_z, d_z)\}$ because those constraints remain in A_L^s . Moreover, since $P\{Z = z\} > 0$ for all $z \in \mathcal{Z}$ and $P\{(\mathbb{L}, D, Z) \in \mathcal{M}_L\} = 1$ due to $P \in \mathbf{P}$, we have

$$\sum_{(\ell,d) \in \mathcal{M}_{L,z}} \beta_{L,(\ell,d,z)}(P) = \sum_{(\ell,d) \in \mathcal{M}_{L,z}} P\{Y \in B_{\ell,L}, D = d | Z = z\} = 1. \quad (73)$$

Combining (72) and (73) we obtain that $A_{L,(\ell_z, d_z, z),\cdot} x_0 = \beta_{L,(\ell_z, d_z, z)}(P)$ for any $z \in \mathcal{Z}$. It follows that $A_L x_0 = \beta_L(P)$ and, since $x_0 \geq 0$, that $P \in \mathbf{P}_{0,L}$ as desired. ■

A.6 Proof of Corollary 5.1

PROOF OF COROLLARY 5.1. Setting $g = 1$, $\theta_0 = 1$, and $\mathcal{R}' = \mathcal{R}$, we have

$$\Theta_0(P, \mathbf{Q}) = \begin{cases} \{1\} & \text{if } \mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset \\ \emptyset & \text{otherwise} \end{cases}.$$

Therefore, $\mathbf{P}_0 = \{P \in \mathbf{P} : 1 \in \Theta_0(P, \mathbf{Q})\} = \{P \in \mathbf{P} : \mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset\}$. For this choice of g , θ_0 , and $\mathcal{R}'$, because the last two rows of the equation $A_L x = \beta_L(P)$ are always satisfied, we have $\tilde{\mathbf{P}}_{0,L} = \mathbf{P}_{0,L}$. Theorem 5.1 then implies $\{P \in \mathbf{P} : \mathbf{Q}_0(P, \mathbf{Q}) \neq \emptyset\} \subseteq \tilde{\mathbf{P}}_{0,L}$ as well as the last statement of the corollary. ■

Appendix B Auxiliary Lemmas

Lemma B.1. *If Assumptions 4.1(i)-(iii) hold and $\log^2(|\mathcal{M}|)|\mathcal{M}|/n = o(1)$ then Assumptions 4.1, 4.2, 4.3, 4.4(v), and A.1 of Fang et al. (2023) are satisfied for the test defined in (18) with $a_n = \log(|\mathcal{M}|)\sqrt{|\mathcal{M}|/n}$ and $M_{3,\Psi} = \sqrt{|\mathcal{M}|}$.*

PROOF. Assumption 4.1(i) in Fang et al. (2023) is equivalent to our Assumption 4.1(i), while their Assumption 4.1(ii) holds with $a_n = \log(|\mathcal{M}|)\sqrt{|\mathcal{Z}|/n}$ by Lemma B.3 and Assumption 4.1(ii). Next note that Assumptions 4.2(i)(ii) in Fang et al. (2023) hold by Lemma B.7(i)(ii), their Assumption 4.2(iii) holds with $M_{3,\Psi} = \sqrt{|\mathcal{M}|}$ by Lemma B.7(iv), and their Assumption 4.2(iv) holds due to $\Omega(P)$ being diagonal. Further observe their Assumption 4.3(i) holds due to the first $|\mathcal{M}|$ diagonal entries of $\Omega(P)$ being non-zero and the final two rows of $\psi(V, P)$ being zero. In turn, Assumptions 4.3(ii) and 4.4(v) in Fang et al. (2023) hold by Lemma B.7(v) and Assumption 4.1(ii). Finally, we note that Assumption A.1 in Fang et al. (2023) holds with $a_n = \log(|\mathcal{M}|)\sqrt{|\mathcal{M}|/n}$ by Lemma B.5. ■

Lemma B.2. *Suppose Assumptions 4.1(i) and 4.1(iii) hold, $\log(|\mathcal{M}|)|\mathcal{M}|/n = o(1)$, and for $\psi(V, P)$ as in (41) let $\Sigma(P) \equiv E_P[\psi(V, P)\psi(V, P)']$. Then, it follows that*

(i) *There exists a Gaussian vector $\mathbb{G}(P) \sim N(0, \Sigma(P))$ satisfying uniformly in $P \in \mathbf{P}$*

$$\left\| \Omega(P)^\dagger \left(\sqrt{n}(\hat{\beta}_n - \beta(P)) - \mathbb{G}(P) \right) \right\|_\infty = O_P \left(\frac{\log^{3/2}(n)\sqrt{|\mathcal{M}|}}{\sqrt{n}} \right). \quad (74)$$

(ii) *There exists a Gaussian vector $\mathbb{G}^*(P) \sim N(0, \Sigma(P))$ that is independent of the sample $\{V_i\}_{i=1}^n$ and satisfying uniformly in $P \in \mathbf{P}$*

$$\left\| \Omega(P)^\dagger \left(\sqrt{n}(\hat{\beta}_n^* - \hat{\beta}_n) - \mathbb{G}^*(P) \right) \right\|_\infty = O_P \left(\frac{\log^{3/2}(n)\sqrt{|\mathcal{M}|}}{\sqrt{n}} \right). \quad (75)$$

Proof. First note that Lemma B.3 allows us to conclude, uniformly in $P \in \mathbf{P}$, that we have

$$\left\| \Omega(P)^\dagger \left\{ \sqrt{n}(\hat{\beta}_n - \beta(P)) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(V_i, P) \right\} \right\|_\infty = O_P \left(\frac{\log(|\mathcal{M}|)\sqrt{|\mathcal{Z}|}}{\sqrt{n}} \right). \quad (76)$$

To establish the lemma, we next apply Theorem 1 in Massart (1989) to couple the vector $\sum_i \psi(V_i, P)/\sqrt{n}$ to a Gaussian variable $\mathbb{G}(P)$. To this end, we let $\{U_i\}_{i=1}^n$ be an i.i.d. sample with $U_i \sim \text{Unif}[0, 1]$. We also divide $[0, 1]$ into $|\mathcal{Z}|$ disjoint intervals, denoted $C_z(P)$, satisfying $P(U_i \in C_z(P)) = P_z$. We further subdivide each interval C_z into disjoint subintervals, denoted $C_{ydz}(P)$, satisfying $P(U_i \in C_{ydz}(P)) = P_{ydz}$ for any (y, d) such that $(y, d, z) \in \mathcal{M}$. In what follows, we will assume that (Y_i, D_i, Z_i) are generated from U_i according to

$$I\{Y_i = y, D_i = d, Z_i = z\} = I\{U_i \in C_{ydz}(P)\}, \quad (77)$$

which we note is without loss of generality in that the distribution of (Y_i, D_i, Z_i) is still P in this probability space. Next, let $\mathcal{S}$ denote the class of all intervals contained in $[0, 1]$. Because all intervals are convex, $\mathcal{S}$ satisfies the uniform Minkowski condition in Definition

2 in [Massart \(1989\)](#). Moreover, since $\mathcal{S}$ is a VC-class, it follows that $\mathcal{S}$ satisfies assumption $H(\zeta)$ in [Massart \(1989\)](#) with $\zeta = 0$. Defining $\mathbb{G}_n \in \ell^\infty(\mathcal{S})$ by

$$\mathbb{G}_n(S) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n (I\{U_i \in S\} - P(U \in S))$$

for any $S \in \mathcal{S}$, we then obtain from Theorem 1 in [Massart \(1989\)](#) that there is a Brownian bridge $\mathbb{W}$ satisfying

$$\text{Cov}[\mathbb{G}_n(S), \mathbb{G}_n(S')] = \text{Cov}[\mathbb{W}(S), \mathbb{W}(S')] \quad (78)$$

for all $S, S' \in \mathcal{S}$ and such that for some constants K, Λ , and θ depending only on $\mathcal{S}$ we have

$$P \left\{ \sup_{S \in \mathcal{S}} |\mathbb{G}_n(S) - \mathbb{W}(S)| > \frac{\sqrt{\log(n)}}{\sqrt{n}} (t + K \log(n)) \right\} \leq \Lambda \exp(-\theta t) \quad (79)$$

for all $t > 0$. In particular, by setting t to be large enough in (79) we obtain that

$$\sup_{S \in \mathcal{S}} |\mathbb{G}_n(S) - \mathbb{W}(S)| = O_P \left(\frac{\log^{3/2}(n)}{\sqrt{n}} \right). \quad (80)$$

To conclude, define a Gaussian vector $\mathbb{G}(P)$ of dimension $|\mathcal{M}| + 2$ by letting its last two coordinates equal zero and its first $|\mathcal{M}|$ coordinates, indexed by $(y, d, z) \in \mathcal{M}$, be

$$\mathbb{G}_{ydz}(P) \equiv \frac{1}{P_z} \mathbb{W}(C_{ydz}(P)) - \frac{P_{ydz}}{P_z^2} \mathbb{W}(C_z(P)). \quad (81)$$

Similarly, by (77), definition of $\psi(V, P)$, and the construction of $C_{ydz}(P)$ and $C_z(P)$ we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{ydz}(V_i, P) = \frac{1}{P_z} \mathbb{G}_n(C_{ydz}(P)) - \frac{P_{ydz}}{P_z^2} \mathbb{G}_n(C_z(P)). \quad (82)$$

In particular, results (78) and (82) imply that $E[\mathbb{G}(P)\mathbb{G}(P)'] = \Sigma(P)$ as desired. Moreover, by definition of $\Omega(P)$, Assumption 4.1(iii), and (81) and (82) we also have that

$$\begin{aligned} & \left\| (\Omega(P))^\dagger \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(V_i, P) - \mathbb{G}(P) \right\} \right\|_\infty \\ &= \max_{(y,d,z) \in \mathcal{M}} \frac{P_z^{1/2}}{P_{ydz}^{1/2} (1 - P_{ydz})^{1/2}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_{ydz}(V_i, P) - \mathbb{G}_{ydz}(P) \right| \\ &\lesssim \max_{(y,d,z) \in \mathcal{M}} \frac{P_z^{1/2}}{P_{ydz}^{1/2}} \left(\frac{1}{P_z} + \frac{P_{ydz}}{P_z^2} \right) \times \sup_{S \in \mathcal{S}} |\mathbb{G}_n(S) - \mathbb{W}(S)| \\ &= O_P \left(\frac{\log^{3/2}(n) \sqrt{|\mathcal{M}|}}{\sqrt{n}} \right), \end{aligned} \quad (83)$$

where the final result holds uniformly in $P \in \mathbf{P}$ by Assumption 4.1(iii) and result (80). The first claim of the lemma therefore follows from (76) and (83).

In order to establish the second claim of the lemma, we first define the vector $\hat{\mathbb{G}}_n$ by

$$\hat{\mathbb{G}}_n \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n (\psi(V_i^*, P) - \frac{1}{n} \sum_{j=1}^n \psi(V_j, P)) \quad (84)$$

Given this notation, then note that Lemma B.4 implies, uniformly in $P \in \mathbf{P}$, that

$$\|(\Omega(P))^\dagger \{\sqrt{n}\{\hat{\beta}_n^* - \hat{\beta}_n\} - \hat{\mathbb{G}}_n\}\|_\infty = O_P \left(\frac{\log(|\mathcal{M}|) \sqrt{|\mathcal{M}|}}{\sqrt{n}} \right). \quad (85)$$

Applying the same construction based on Massart (1989) conditionally on $\{V_i\}_{i=1}^n$ implies that there is $\mathbb{G}(\hat{P})$ satisfying $\mathbb{G}(\hat{P}) \sim N(0, \text{Var}_{\hat{P}}\{\psi(V, P)\})$ conditionally on $\{V_i\}_{i=1}^n$ and

$$P \left\{ \|\hat{\mathbb{G}}_n - \mathbb{G}(\hat{P})\|_\infty > C_1 \frac{|\mathcal{Z}| \sqrt{\log(n)}}{\sqrt{n}} (t + K \log(n)) \middle| \{V_i\}_{i=1}^n \right\} \leq \Lambda \exp\{-\theta t\} \quad (86)$$

for some $C_1 < \infty$ not depending on $\{V_i\}_{i=1}^n$. Further let $\mathcal{B}$ denote the Borel σ -field, and for any $B \in \mathcal{B}$ denote its ϵ -enlargement under $\|\cdot\|_\infty$ by $B^\epsilon \equiv \{\tilde{b} : \inf_{b \in B} \|b - \tilde{b}\|_\infty \leq \epsilon\}$. Defining $\delta_n = 2C_1 K \log^{3/2}(n) |\mathcal{Z}| / \sqrt{n}$ and setting $t = C_2$, we then obtain from result (86) and Strassen's Theorem (see, e.g., Theorem 10.3 in Pollard (2002)), that

$$\sup_{P \in \mathbf{P}} E_P \left[\sup_{B \in \mathcal{B}} \left\{ P \left\{ \hat{\mathbb{G}}_n \in B \middle| \{V_i\}_{i=1}^n \right\} - P \left\{ \mathbb{G}(\hat{P}) \in B^{K\delta_n} \middle| \{V_i\}_{i=1}^n \right\} \right\} \right] \leq \Lambda \exp\{-\theta C_2\} \quad (87)$$

for n sufficiently large. Since the right hand side of (87) can be made arbitrarily small by setting C_2 sufficiently large, we obtain from Theorem 4 in Monrad and Philipp (1991) that there exists a random variable $\bar{\mathbb{G}}(P)$ with distribution $N(0, \text{Var}_{\hat{P}}\{\psi(V, P)\})$ conditionally on $\{V_i\}_{i=1}^n$ and such that uniformly in $P \in \mathbf{P}$ it satisfies

$$\|\hat{\mathbb{G}}_n - \bar{\mathbb{G}}(P)\|_\infty = O_P \left(\frac{\log^{3/2}(n) |\mathcal{Z}|}{\sqrt{n}} \right). \quad (88)$$

Moreover, result (88), the definition of $\Omega(P)$, and Assumption 4.1(iii) yield that

$$\begin{aligned} & \|(\Omega(P))^\dagger \{\hat{\mathbb{G}}_n - \bar{\mathbb{G}}(P)\}\|_\infty \\ & \leq \sup_{(y, d, z) \in \mathcal{M}} \left(\frac{P_z}{P_{y|d|z}(1 - P_{y|d|z})} \right)^{1/2} \|\hat{\mathbb{G}}_n - \bar{\mathbb{G}}(P)\|_\infty = O_P \left(\frac{\log^{3/2}(n) \sqrt{|\mathcal{M}|}}{\sqrt{n}} \right) \end{aligned} \quad (89)$$

uniformly in $P \in \mathbf{P}$. However, since $(\Omega(P))^\dagger \bar{\mathbb{G}}(P) \sim N(0, (\Omega(P))^\dagger \text{Var}_{\hat{P}}\{\psi(V, P)\} (\Omega(P))^\dagger)$ conditionally on the data, we may apply the arguments in Lemma A.7 in Fang et al. (2023)

(see in particular their equation (A.57)) and Lemma B.6 to conclude that there is a $\bar{\mathbb{G}}^*(P) \sim N(0, (\Omega(P))^\dagger \text{Var}_P\{\psi(V, P)\}(\Omega(P))^\dagger)$ that is independent of the data, and satisfies

$$\|(\Omega(P))^\dagger \bar{\mathbb{G}}(P) - \bar{\mathbb{G}}^*(P)\|_\infty = O_P \left(\frac{\log(|\mathcal{M}|) \sqrt{|\mathcal{M}|}}{\sqrt{n}} \right) \quad (90)$$

uniformly in $P \in \mathbf{P}$. Finally, let $\mathbb{G}^*(P) = \Omega(P) \bar{\mathbb{G}}^*(P)$ and note that $\mathbb{G}^*(P)$ is independent of the data, because $\bar{\mathbb{G}}^*(P)$ is, and $\mathbb{G}^*(P) \sim N(0, \text{Var}_P\{\psi(V, P)\})$ because the columns of $\text{Var}_P\{\psi(V, P)\}$ are in the range of $\Omega(P)$. The second claim of the lemma then follows from results (85), (89), and (90). ■

Lemma B.3. *Let Assumptions 4.1(i) and 4.1(iii) hold and $\psi(V, P)$ be as defined in (41). If $\log(|\mathcal{M}|)|\mathcal{M}|/n = o(1)$, then it follows that, uniformly in $P \in \mathbf{P}$, we have*

$$\|(\Omega(P))^\dagger \{\sqrt{n}\{\hat{\beta}_n - \beta(P)\} - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(V_i, P)\}\|_\infty = O_P \left(\frac{\log(|\mathcal{M}|) \sqrt{|\mathcal{Z}|}}{\sqrt{n}} \right).$$

PROOF. We first note that by Lemma B.10(iv), $\min_{z \in \mathcal{Z}} \hat{P}_z > 0$ with probability tending to one uniformly in $P \in \mathbf{P}$. Therefore, the first $|\mathcal{M}|$ coordinates of $\hat{\beta}_n - \beta(P)$ have the following structure for some $(y, d, z) \in \mathcal{M}$ with probability tending to one

$$\frac{\hat{P}_{ydz}}{\hat{P}_z} - \frac{P_{ydz}}{P_z} = \frac{1}{P_z} (\hat{P}_{ydz} - P_{ydz}) - \frac{P_{ydz}}{P_z^2} (\hat{P}_z - P_z) + \hat{\gamma}_{ydz},$$

where

$$\hat{\gamma}_{ydz} = (\hat{P}_{ydz} - P_{ydz}) \left(\frac{1}{\hat{P}_z} - \frac{1}{P_z} \right) - \frac{P_{ydz}}{P_z} \left(\frac{1}{\hat{P}_z} - \frac{1}{P_z} \right) (\hat{P}_z - P_z).$$

Since the final two rows of $\{\hat{\beta}_n - \beta(P)\}$ are identically zero, the definition of $\Omega(P)$ implies

$$\begin{aligned} & \left\| (\Omega(P))^\dagger \left\{ \sqrt{n}\{\hat{\beta}_n - \beta(P)\} - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(V_i, P) \right\} \right\|_\infty \\ &= \max_{(y,d,z) \in \mathcal{M}} \left| \left(\frac{P_z}{P_{ydz}(1 - P_{ydz})} \right)^{1/2} \sqrt{n} \hat{\gamma}_{ydz} \right| \lesssim \left(\frac{|\mathcal{M}|}{|\mathcal{Z}|^2} \right)^{1/2} \times \max_{(y,d,z) \in \mathcal{M}} \sqrt{n} |\hat{\gamma}_{ydz}|, \quad (91) \end{aligned}$$

where the inequality follows from Assumption 4.1(iii). Next note Lemma B.10 implies that

$$\begin{aligned} & \max_{(y,d,z) \in \mathcal{M}} \left| (\hat{P}_{ydz} - P_{ydz}) \left(\frac{1}{\hat{P}_z} - \frac{1}{P_z} \right) \right| \\ & \leq \max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz} - P_{ydz}| \times \max_{z \in \mathcal{Z}} |\hat{P}_z - P_z| \times \max_{z \in \mathcal{Z}} \frac{1}{P_z^2} \times O_P(1) = O_P \left(\frac{\log(|\mathcal{M}|) |\mathcal{Z}|^2}{n \sqrt{|\mathcal{M}|} |\mathcal{Z}|} \right) \quad (92) \end{aligned}$$

uniformly in $P \in \mathbf{P}$. Similarly, Lemma B.10 and Assumption 4.1(iii) also yield that

$$\begin{aligned} \max_{(y,d,z) \in \mathcal{M}} \left| \frac{P_{ydz}}{P_z} \left(\frac{1}{\hat{P}_z} - \frac{1}{P_z} \right) (\hat{P}_z - P_z) \right| \\ \leq \max_{(y,d,z) \in \mathcal{M}} \frac{P_{ydz}}{P_z^3} \times \max_{z \in \mathcal{Z}} |\hat{P}_z - P_z|^2 \times O_P(1) = O_P \left(\frac{|\mathcal{Z}|^2 \log(|\mathcal{Z}|)}{|\mathcal{M}|n} \right) \end{aligned} \quad (93)$$

uniformly in $P \in \mathbf{P}$. The claim of the lemma then follows from combining results (91), (92), and (93). ■

Lemma B.4. *Let Assumptions 4.1(i) and 4.1(iii) hold and $\psi(V, P)$ be as defined in (41). If $\log(|\mathcal{M}|)|\mathcal{M}|/n = o(1)$, then it follows that, uniformly in $P \in \mathbf{P}$, we have*

$$\|(\Omega(P))^\dagger \{\sqrt{n}\{\hat{\beta}_n^* - \hat{\beta}_n\} - \frac{1}{\sqrt{n}} \sum_{i=1}^n (\psi(V_i^*, P) - \frac{1}{n} \sum_{j=1}^n \psi(V_j, P))\}\|_\infty = O_P \left(\frac{\log(|\mathcal{M}|)\sqrt{|\mathcal{M}|}}{\sqrt{n}} \right).$$

PROOF. The proof follows similar arguments to those employed in the proof of Lemma B.3. We first note that by Lemmas B.10(iv) and B.11(iv), $\min_{z \in \mathcal{Z}} \hat{P}_z^* \wedge \hat{P}_z > 0$ with probability tending to one uniformly in $P \in \mathbf{P}$. Therefore, the first $|\mathcal{M}|$ coordinates of $\hat{\beta}_n^* - \hat{\beta}_n$ have the following structure for some $(y, d, z) \in \mathcal{M}$ with probability tending to one

$$\frac{\hat{P}_{ydz}^*}{\hat{P}_z^*} - \frac{\hat{P}_{ydz}}{\hat{P}_z} = \frac{1}{\hat{P}_z} (\hat{P}_{ydz}^* - \hat{P}_{ydz}) - \frac{\hat{P}_{ydz}}{\hat{P}_z^2} (\hat{P}_z^* - \hat{P}_z) + \hat{\gamma}_{ydz}^*, \quad (94)$$

where

$$\hat{\gamma}_{ydz}^* = (\hat{P}_{ydz}^* - \hat{P}_{ydz}) \left(\frac{1}{\hat{P}_z^*} - \frac{1}{\hat{P}_z} \right) - \frac{\hat{P}_{ydz}}{\hat{P}_z} \left(\frac{1}{\hat{P}_z^*} - \frac{1}{\hat{P}_z} \right) (\hat{P}_z^* - \hat{P}_z). \quad (95)$$

Since the final two rows of $\{\hat{\beta}_n^* - \hat{\beta}_n\}$ are identically zero, the definition of $\Omega(P)$ implies

$$\begin{aligned} \|(\Omega(P))^\dagger \{\sqrt{n}\{\hat{\beta}_n^* - \hat{\beta}_n\} - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(V_i^*, \hat{P})\}\|_\infty \\ = \max_{(y,d,z) \in \mathcal{M}} \left| \left(\frac{P_z}{P_{ydz}(1 - P_{ydz})} \right)^{1/2} \sqrt{n} \hat{\gamma}_{ydz}^* \right| \lesssim \left(\frac{|\mathcal{M}|}{|\mathcal{Z}|^2} \right)^{1/2} \times \max_{(y,d,z) \in \mathcal{M}} \sqrt{n} |\hat{\gamma}_{ydz}^*|, \end{aligned} \quad (96)$$

where the inequality follows from Assumption 4.1(iii). Next note that Lemma B.11(iv) yields

$$\begin{aligned} \max_{(y,d,z) \in \mathcal{M}} \left| (\hat{P}_{ydz}^* - \hat{P}_{ydz}) \left(\frac{1}{\hat{P}_z^*} - \frac{1}{\hat{P}_z} \right) \right| \\ \leq \max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz}^* - \hat{P}_{ydz}| \times \max_{z \in \mathcal{Z}} |\hat{P}_z^* - \hat{P}_z| \times \max_{z \in \mathcal{Z}} \frac{1}{\hat{P}_z^2} \times O_P(1) = O_P \left(\frac{\log(|\mathcal{M}|)|\mathcal{Z}|^2}{n\sqrt{|\mathcal{M}||\mathcal{Z}|}} \right) \end{aligned} \quad (97)$$

uniformly in $P \in \mathbf{P}$, where in the final result we used Lemmas B.11(i)(ii), Lemma B.10(iv),

and Assumption 4.1(iii). Similarly, applying Lemmas B.10(iii)(iv) and Lemma B.11(iv) yields, uniformly in $P \in \mathbf{P}$, that

$$\begin{aligned} \max_{(y,d,z) \in \mathcal{M}} \left| \frac{\hat{P}_{ydz}}{\hat{P}_z} \left(\frac{1}{\hat{P}_z^*} - \frac{1}{\hat{P}_z} \right) (\hat{P}_z^* - \hat{P}_z) \right| \\ \leq \max_{(y,d,z) \in \mathcal{M}} \frac{P_{ydz}}{P_z^3} \times \max_{z \in \mathcal{Z}} |\hat{P}_z^* - \hat{P}_z|^2 \times O_P(1) = O_P \left(\frac{|\mathcal{Z}|^2 \log(|\mathcal{Z}|)}{|\mathcal{M}|n} \right) \end{aligned} \quad (98)$$

where the final equality used Lemma B.11(ii) and Assumption 4.1(iii). Hence, we obtain

$$\|(\Omega(P))^\dagger \{ \sqrt{n} \{ \hat{\beta}_n^* - \hat{\beta}_n \} - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(V_i^*, \hat{P}) \} \|_\infty = O_P \left(\frac{\log(|\mathcal{M}|) \sqrt{|\mathcal{Z}|}}{\sqrt{n}} \right) \quad (99)$$

uniformly in $P \in \mathbf{P}$, by results (96), (97), and (98). Next, note that for any $(y, d, z) \in \mathcal{M}$

$$\psi_{ydz}(V_i^*, \hat{P}) = \frac{1}{\hat{P}_z} (I\{Y_i^* = y, D_i^* = d, Z_i^* = z\} - \hat{P}_{ydz}) - \frac{\hat{P}_{ydz}}{\hat{P}_z^2} (I\{Z_i^* = z\} - \hat{P}_z), \quad (100)$$

and therefore

$$\begin{aligned} \psi_{ydz}(V_i^*, \hat{P}) - (\psi_{ydz}(V_i^*, P) - \frac{1}{n} \sum_{j=1}^n \psi_{ydz}(V_j, P)) \\ = \left(\frac{1}{\hat{P}_z} - \frac{1}{P_z} \right) (I\{Y_i^* = y, D_i^* = d, Z_i^* = z\} - \hat{P}_{ydz}) - \left(\frac{\hat{P}_{ydz}}{\hat{P}_z^2} - \frac{P_{ydz}}{P_z^2} \right) (I\{Z_i^* = z\} - \hat{P}_z). \end{aligned}$$

Therefore, the definition of $\Omega(P)$ and Assumption 4.1(iii) together allow us to conclude that

$$\begin{aligned} \|(\Omega(P))^\dagger \{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(V_i^*, \hat{P}) - \frac{1}{\sqrt{n}} \sum_{i=1}^n (\psi(V_i^*, P) - \frac{1}{n} \sum_{j=1}^n \psi(V_j, P)) \} \|_\infty \\ \lesssim \max_{(y,d,z) \in \mathcal{M}} \left| \left(\frac{P_z}{P_{ydz}} \right)^{1/2} \sqrt{n} \left\{ \left(\frac{1}{\hat{P}_z} - \frac{1}{P_z} \right) (\hat{P}_{ydz}^* - \hat{P}_{ydz}) - \left(\frac{\hat{P}_{ydz}}{\hat{P}_z^2} - \frac{P_{ydz}}{P_z^2} \right) (\hat{P}_z^* - \hat{P}_z) \right\} \right|. \end{aligned} \quad (101)$$

Next, observe that Assumption 4.1(iii) and Lemmas B.10(ii)(iv) and B.11(i) imply that

$$\begin{aligned} \max_{(y,d,z) \in \mathcal{M}} \left| \left(\frac{P_z}{P_{ydz}} \right)^{1/2} \sqrt{n} \left(\frac{1}{\hat{P}_z} - \frac{1}{P_z} \right) (\hat{P}_{ydz}^* - \hat{P}_{ydz}) \right| \\ \lesssim \frac{\sqrt{n|\mathcal{M}|}}{|\mathcal{Z}|} \times \max_{z \in \mathcal{Z}} \left| \frac{\hat{P}_z - P_z}{\hat{P}_z P_z} \right| \times \max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz}^* - \hat{P}_{ydz}| = O_P \left(\frac{\log(|\mathcal{M}|) \sqrt{|\mathcal{Z}|}}{\sqrt{n}} \right) \end{aligned} \quad (102)$$

uniformly in $P \in \mathbf{P}$. Moreover, Lemmas B.10(ii)(v) and Assumption 4.1(iii) also imply that

$$\begin{aligned} & \max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{yd|z} P_z - P_{yd|z} \hat{P}_z| \\ & \leq \max_{(y,d,z) \in \mathcal{M}} \left\{ |\hat{P}_{yd|z} - P_{yd|z}| P_z + P_{yd|z} |\hat{P}_z - P_z| \right\} = O_P \left(\frac{\sqrt{\log(|\mathcal{M}|)}}{\sqrt{n|\mathcal{Z}|}} \right) \end{aligned} \quad (103)$$

uniformly in $P \in \mathbf{P}$. Therefore, result (103), Assumption 4.1(iii), and Lemmas B.10(iv) and B.11(ii) yield

$$\begin{aligned} & \max_{(y,d,z) \in \mathcal{M}} \left| \left(\frac{P_z}{P_{yd|z}} \right)^{1/2} \sqrt{n} \left(\frac{\hat{P}_{ydz}}{\hat{P}_z^2} - \frac{P_{ydz}}{P_z^2} \right) (\hat{P}_z^* - \hat{P}_z) \right| \\ & \lesssim \frac{\sqrt{n|\mathcal{M}|}}{|\mathcal{Z}|} \times \max_{(y,d,z) \in \mathcal{M}} \left| \frac{\hat{P}_{ydz} P_z - P_{yd|z} \hat{P}_z}{\hat{P}_z P_z} \right| \times \max_{z \in \mathcal{Z}} |\hat{P}_z^* - \hat{P}_z| = O_P \left(\frac{\log(|\mathcal{M}|) \sqrt{|\mathcal{M}|}}{\sqrt{n}} \right) \end{aligned} \quad (104)$$

uniformly in $P \in \mathbf{P}$. The claim of the lemma then follows from (99), (101), (102), and (104). $\blacksquare$

Lemma B.5. *Let Assumptions 4.1(i) and 4.1(iii) hold, and let $\|C\|_{o,\infty} = \sup_{\|a\|_\infty \leq 1} \|Ca\|_\infty$ for any $p \times p$ matrix C . If $\log(|\mathcal{M}|)|\mathcal{M}|/n = o(1)$, then it follows that $\|\Omega(P)^\dagger(\hat{\Omega}_n - \Omega(P))\|_{o,\infty} = O_P(\sqrt{\log(|\mathcal{M}|)|\mathcal{M}|/n})$ uniformly in $P \in \mathbf{P}$.*

PROOF. For any constant $c \geq 0$, we have $|c - 1| = |\sqrt{c} - 1| |\sqrt{c} + 1|$ and therefore that $|\sqrt{c} - 1| \leq |c - 1|$. Therefore, employing the definitions of $\Omega(P)$ and $\hat{\Omega}_n$ yields that

$$\begin{aligned} \|\Omega(P)^\dagger(\hat{\Omega}_n - \Omega(P))\|_{o,\infty} &= \max_{(y,d,z) \in \mathcal{M}} \left| \frac{\hat{P}_{ydz}^{1/2} (1 - \hat{P}_{yd|z})^{1/2}}{\hat{P}_z^{1/2}} \times \frac{P_z^{1/2}}{P_{yd|z}^{1/2} (1 - P_{yd|z})^{1/2}} - 1 \right| \\ &\leq \max_{(y,d,z) \in \mathcal{M}} \left| \frac{\hat{P}_{ydz} (1 - \hat{P}_{yd|z})}{\hat{P}_z^2} \times \frac{P_z^2}{P_{ydz} (1 - P_{yd|z})} - 1 \right|. \end{aligned} \quad (105)$$

Next, note that $(a^2 - b^2) = (a - b)(a + b)$, Lemmas B.10(ii)(iv) and Assumption 4.1(iii) imply

$$\max_{z \in \mathcal{Z}} \left| \frac{P_z^2}{\hat{P}_z^2} - 1 \right| \leq \max_{z \in \mathcal{Z}} \frac{1}{\hat{P}_z^2} \times \max_{z \in \mathcal{Z}} |\hat{P}_z - P_z| \times \max_{z \in \mathcal{Z}} |\hat{P}_z + P_z| = O_P \left(\frac{\sqrt{\log(|\mathcal{Z}|)|\mathcal{Z}|}}{\sqrt{n}} \right) \quad (106)$$

uniformly in $P \in \mathbf{P}$. Similarly relying on Assumption 4.1(iii) and Lemma B.10(v) yields

$$\max_{(y,d,z) \in \mathcal{M}} \left| \frac{(1 - \hat{P}_{yd|z})}{(1 - P_{yd|z})} - 1 \right| = \max_{(y,d,z) \in \mathcal{M}} \left| \frac{(P_{yd|z} - \hat{P}_{yd|z})}{(1 - P_{yd|z})} \right| = O_P \left(\frac{\sqrt{\log(|\mathcal{M}|)|\mathcal{Z}|}}{\sqrt{n}} \right) \quad (107)$$

uniformly in $P \in \mathbf{P}$. The lemma then follows from (105), Lemma B.10(iii), and results (106)

and (107). ■

Lemma B.6. *Let Assumptions 4.1(i) and 4.1(iii) hold, $\psi(V, P)$ be as defined in (41) and for any $k \times k$ symmetric matrix M let $\|M\|_o$ denote its largest absolute eigenvalue. If $\delta_n^2 \equiv \log(|\mathcal{M}|)|\mathcal{M}|/n$ satisfies $\delta_n = o(1)$, then uniformly in $P \in \mathbf{P}$ it follows that*

$$\|(\Omega(P))^\dagger(\text{Var}_{\hat{P}}\{\psi(V, P)\} - \text{Var}_P\{\psi(V, P)\})(\Omega(P))^\dagger\|_o = O_P(\delta_n) \quad (108)$$

$$\|(\Omega(P)^\dagger \text{Var}_{\hat{P}}\{\psi(V, P)\} \Omega(P)^\dagger)^{1/2} - (\Omega(P)^\dagger \text{Var}_P\{\psi(V, P)\} \Omega(P)^\dagger)^{1/2}\|_o = O_P(\delta_n). \quad (109)$$

PROOF. First define the vector $\bar{\psi}_n(P) \equiv \sum_i \psi(V_i, P)/n$ for notational simplicity. Next, note that for any unit-length vector v (under $\|\cdot\|_2$), the Cauchy-Schwarz inequality implies

$$\|\bar{\psi}_n(P) \bar{\psi}_n(P)' v\|_2 = \|\bar{\psi}_n(P)\|_2 |\bar{\psi}_n(P)' v| \leq \|\bar{\psi}_n(P)\|_2^2 = \sum_{(y,d,z) \in \mathcal{M}} \left(\frac{1}{n} \sum_{i=1}^n \psi_{ydz}(V_i, P) \right)^2,$$

where the final equality follows from the definition of $\psi(V, P)$. Therefore, since Lemma B.7(i) implies that $E_P[\psi(V, P)] = 0$ for all $P \in \mathbf{P}$ and the sample $\{V_i\}_{i=1}^n$ is i.i.d., Assumption 4.1(iii) and the definitions of $\Omega(P)$ and $\psi(V, P)$ yield the bound

$$\begin{aligned} \sup_{P \in \mathbf{P}} E_P \left[\|(\Omega(P))^\dagger \bar{\psi}_n \bar{\psi}_n' (\Omega(P))^\dagger\|_o \right] &\leq \sup_{P \in \mathbf{P}} \left\{ \|(\Omega(P))^\dagger\|_o^2 \times \frac{1}{n} \sum_{(y,d,z) \in \mathcal{M}} E_P \left[\psi_{ydz}^2(V, P) \right] \right\} \\ &\leq \sup_{P \in \mathbf{P}} \left\{ \max_{(y,d,z) \in \mathcal{M}} \frac{P_z}{P_{ydz}(1 - P_{ydz})} \times \frac{1}{n} \sum_{(y,d,z) \in \mathcal{M}} \left(\frac{P_{ydz}}{P_z^2} + \frac{P_{ydz}^2}{P_z^3} \right) \right\} \lesssim \frac{|\mathcal{M}|}{n}. \end{aligned} \quad (110)$$

Next, set $M_i(P) \equiv (\Omega(P))^\dagger(\psi(V_i, P)\psi(V_i, P)' - E_P[\psi(V, P)\psi(V, P)']) (\Omega(P))^\dagger/n$, and note result (110), the triangle inequality, and Markov's inequality yield, uniformly in $P \in \mathbf{P}$, that

$$\|(\Omega(P))^\dagger \{\text{Var}_{\hat{P}}\{\psi(V, P)\} - \text{Var}_P\{\psi(V, P)\}\} (\Omega(P))^\dagger\|_o \leq \left\| \sum_{i=1}^n M_i(P) \right\|_o + O_P \left(\frac{|\mathcal{M}|}{n} \right). \quad (111)$$

Moreover, for any unit length vector v (under $\|\cdot\|_2$) the Cauchy-Schwarz inequality implies

$$\begin{aligned} \|\psi(V, P)\psi(V, P)' v\|_2 &\leq \|\psi(V, P)\|_2^2 \\ &= \sum_{(y,d,z) \in \mathcal{M}} \left(\frac{1}{P_z} I\{Y = y, D = d, Z = z\} - \frac{P_{ydz}}{P_z^2} I\{Z = z\} \right)^2 \\ &\leq 2 \max_{z \in \mathcal{Z}} \sum_{(y,d):(y,d,z) \in \mathcal{M}} \left(\frac{1}{P_z^2} I\{Y = y, D = d\} + \frac{P_{ydz}^2}{P_z^4} \right) \\ &\leq 2 \max_{z \in \mathcal{Z}} \frac{1}{P_z^2} + 2 \max_{z \in \mathcal{Z}} \sum_{(y,d):(y,d,z) \in \mathcal{M}} \frac{P_{ydz}^2}{P_z^4} \\ &\lesssim |\mathcal{Z}|^2, \end{aligned} \quad (112)$$

where the first equality follows by definition of $\psi(V, P)$ and the final inequality from Assumption 4.1(iii) and $\sum_{(y,d): (y,d,z) \in \mathcal{M}} P_{yd|z}^2 \leq 1$. Hence, the definitions of $\Omega(P)$ and $M_i(P)$, the triangle inequality, result (112), and Lemma B.7(ii) imply

$$\begin{aligned} \|M_i(P)\|_o &\leq \frac{1}{n} \{ \|(\Omega(P))^\dagger\|_o^2 \|\psi(V, P)\psi(V, P)'\|_o + \|(\Omega(P))^\dagger E_P[\psi(V, P)\psi(V, P)'](\Omega(P))^\dagger\|_o \} \\ &\lesssim \frac{1}{n} \left(\sup_{(y,d,z) \in \mathcal{M}} \frac{P_z}{P_{yd|z}(1 - P_{yd|z})} \times |\mathcal{Z}|^2 + O(1) \right) = O\left(\frac{|\mathcal{M}|}{n}\right). \end{aligned} \quad (113)$$

Next, observe that the definition of $\Omega(P)$, result (112), and Assumption 4.1(iii) yield that

$$\|(\Omega(P))^\dagger \psi(V, P)\psi(V, P)'(\Omega(P))^\dagger\|_o \leq \max_{(y,d,z) \in \mathcal{M}} \frac{P_z}{P_{yd|z}(1 - P_{yd|z})} \times \|\psi(V, P)\|_2^2 \lesssim |\mathcal{M}|. \quad (114)$$

Hence, the triangle inequality, $\{M_i(P)\}_{i=1}^n$ being i.i.d., Lemma B.7(ii), and result (114) imply

$$\left\| \sum_{i=1}^n E_P[M_i^2(P)] \right\|_o \leq \frac{1}{n} \left\{ \|E_P[(\Omega(P))^\dagger \psi(V, P)\psi(V, P)'(\Omega(P))^\dagger]^2\|_o + O(1) \right\} \lesssim \frac{|\mathcal{M}|}{n}. \quad (115)$$

Results (113) and (115) allow us to apply Bernstein's inequality for matrices (see, e.g., Theorem 1.4 in Tropp, 2012) to conclude there is a $C < \infty$ such that for all $t \geq 0$ we have

$$P \left\{ \left\| \sum_{i=1}^n M_i(P) \right\|_o > t \right\} \leq |\mathcal{M}| \exp \left\{ -C \frac{nt^2}{|\mathcal{M}|(1+t)} \right\}. \quad (116)$$

Evaluating the bound in (116) at $t = K\sqrt{|\mathcal{M}| \log(|\mathcal{M}|)}/\sqrt{n}$ for K sufficiently large implies

$$\left\| \sum_{i=1}^n M_i(P) \right\|_o = O_P \left(\frac{\sqrt{\log(|\mathcal{M}|) |\mathcal{M}|}}{\sqrt{n}} \right) \quad (117)$$

uniformly in $P \in \mathbf{P}$. Result (108) then follows from (111) and (117).

In order to establish (109), let $N(P)$ and $N(\hat{P})$ denote the null spaces of the matrices $\Omega(P)^\dagger \text{Var}_P\{\psi(V, P)\}\Omega(P)^\dagger$ and $\Omega(P)^\dagger \text{Var}_{\hat{P}}\{\psi(V, P)\}\Omega(P)^\dagger$ respectively, and note that the null space of $\text{Var}_P\{\psi(V, P)\}$ being a subset of the null space of $\text{Var}_{\hat{P}}\{\psi(V, P)\}$ with probability one implies $N(P) \subseteq N(\hat{P})$ with probability one. On the other hand, Weyl's Perturbation Theorem (see, e.g., Corollary III.2.6 in Bhatia (2013)), the positive eigenvalues of $\Omega(P)^\dagger \text{Var}_P\{\psi(V, P)\}\Omega(P)^\dagger$ being (weakly) larger than one for all $P \in \mathbf{P}$ by Lemma B.7(iii), and result (108) imply that the dimension of $N(\hat{P})$ is (weakly) smaller than the dimension of $N(P)$ with probability tending to one uniformly in $P \in \mathbf{P}$. Since, as argued, $N(P) \subseteq N(\hat{P})$ with probability one, we can therefore conclude that

$$\liminf_{n \rightarrow \infty} \inf_{P \in \mathbf{P}} P\{N(P) = N(\hat{P})\} = 1. \quad (118)$$

Next set $\underline{\lambda}^+(P)$ and $\underline{\lambda}^+(\hat{P})$ to be the smallest strictly positive eigenvalues of the matrices $\Omega(P)^\dagger \text{Var}_P\{\psi(V, P)\} \Omega(P)^\dagger$ and $\Omega(P)^\dagger \text{Var}_{\hat{P}}\{\psi(V, P)\} \Omega(P)^\dagger$ respectively. A second application of Weyl's Perturbation Theorem, Lemma B.7(iii), and results (108) and (118) then imply

$$\liminf_{n \rightarrow \infty} \inf_{P \in \mathbf{P}} P\{\underline{\lambda}^+(\hat{P}) > 1/2\} = 1. \quad (119)$$

In particular, results (118) and (119) imply the event $E_0(P) \equiv \{N(P) = N(\hat{P}) \text{ and } \underline{\lambda}^+(\hat{P}) > 1/2\}$ occurs with probability tending to one uniformly in $P \in \mathbf{P}$. Under $E_0(P)$ we can apply Theorem X.3.8 in Bhatia (2013) (on the subspace orthogonal to $N(P)$) to conclude

$$\begin{aligned} & \|(\Omega(P)^\dagger \text{Var}_{\hat{P}}\{\psi(V, P)\} \Omega(P)^\dagger)^{1/2} - (\Omega(P)^\dagger \text{Var}_P\{\psi(V, P)\} \Omega(P)^\dagger)^{1/2}\|_o \\ & \leq \frac{1}{\sqrt{2}} \|(\Omega(P))^\dagger (\text{Var}_{\hat{P}}\{\psi(V, P)\} - \text{Var}_P\{\psi(V, P)\}) (\Omega(P))^\dagger\|_o, \end{aligned} \quad (120)$$

and therefore (109) holds due to (120) holding with probability tending to one and (108). ■

Lemma B.7. *Let Assumptions 4.1(i) and 4.1(iii) hold, $\log(|\mathcal{M}|)|\mathcal{M}|/n = o(1)$, $\psi(V, P)$ be as in (41), and $\Sigma(P) \equiv E_P[\psi(V, P)\psi(V, P)']$. Then, it follows that*

- (i) $\psi(V, P)$ satisfies $E_P[\psi(V, P)] = 0$;
- (ii) The eigenvalues of $\Omega(P)^\dagger \Sigma(P) \Omega(P)^\dagger$ are uniformly bounded in $P \in \mathbf{P}$;
- (iii) The nonzero eigenvalues of $\Omega(P)^\dagger \Sigma(P) \Omega(P)^\dagger$ are weakly bigger than one for all $P \in \mathbf{P}$;
- (iv) $\Psi(V, P) \equiv \|\Omega(P)^\dagger \psi(V, P)\|_\infty$ satisfies $\sup_{P \in \mathbf{P}} E_P[|\Psi(V, P)|^3] \lesssim |\mathcal{M}|^{3/2}$;
- (v) $(\hat{\beta}_n - \beta(P)), (\hat{\beta}_n^* - \hat{\beta}_n) \in \text{range}(\Sigma(P))$ with probability tending to one uniformly in $P \in \mathbf{P}$.
- (vi) The null space of $\Sigma(P)$ has dimension $|\mathcal{Z}| + 2$ for all $P \in \mathbf{P}$.

PROOF. In order to establish the first claim, recall that the first $|\mathcal{M}|$ entries of $\psi(V, P)$ are indexed by $(y, d, z) \in \mathcal{M}$ and have the following structure

$$\psi_{ydz}(V, P) = \frac{1}{P_z} I\{Y = y, D = d, Z = z\} - \frac{P_{ydz}}{P_z^2} I\{Z = z\}.$$

By direct calculation, $E_P[\psi_{ydz}(V, P)] = P_{yd|z} - P_{yd|z} = 0$ and, since the final two coordinates of $\psi(V, P)$ are identically equal to zero, the first claim of the lemma follows.

To establish the second claim, we first characterize the covariance matrix $\Sigma(P)$ by noting

$$\text{Cov}[\psi_{ydz}(V, P), \psi_{\tilde{y}\tilde{d}\tilde{z}}(V, P)] = \begin{cases} 0 & \text{if } z \neq \tilde{z} \\ -\frac{P_{yd|z}P_{\tilde{y}\tilde{d}|\tilde{z}}}{P_z} & \text{if } z = \tilde{z} \text{ and } (y, d) \neq (\tilde{y}, \tilde{d}) \\ \frac{P_{yd|z}(1-P_{yd|z})}{P_z} & \text{if } (y, d, z) = (\tilde{y}, \tilde{d}, \tilde{z}) \end{cases}. \quad (121)$$

Also observe that the last two rows and columns of $\Sigma(P)$ are zero because the final two rows of $\psi(V, P)$ equal zero. Since $\Omega(P)$ is a diagonal matrix in which the first $|\mathcal{M}|$ entries

have the structure $(P_{ydz}(1 - P_{ydz})/P_z)^{1/2}$ for some $(y, d, z) \in \mathcal{M}$ and the final two entries equal to zero, result (121) implies that $\Omega(P)^\dagger \Sigma(P) \Omega(P)^\dagger$ is block diagonal and the final two columns and rows equal zero. In particular, there are $|\mathcal{Z}|$ blocks, which we denote by $\Gamma_z(P)$ for $z \in \mathcal{Z}$. Letting $a_z(P)$ be a vector with coordinates $\{P_{ydz}^{1/2} : (y, d, z) \in \mathcal{M}\}$, $b_z(P)$ a vector with coordinates $\{(1 - P_{ydz})^{-1/2} : (y, d, z) \in \mathcal{M}\}$ and $\text{diag}\{b_z(P)\}$ denote a diagonal matrix with diagonal $b_z(P)$, it then follows by direct calculation that each $\Gamma_z(P)$ satisfies

$$\Gamma_z(P) = \text{diag}\{b_z(P)\}(I - a_z(P)a_z(P)')\text{diag}\{b_z(P)\}, \quad (122)$$

where I is the identity matrix. Since $\|b_z(P)\|_\infty$ is uniformly bounded due to P_{ydz} being uniformly bounded away from one by Assumption 4.1(iii), the second claim of the lemma follows from (122) and $\Omega(P)^\dagger \Sigma(P) \Omega(P)^\dagger$ being block diagonal with blocks $\Gamma_z(P)$.

For the third claim, set $M_z(P) \equiv (I - a_z(P)a_z(P)')$ and let $\lambda_0 > 0$ be any strictly positive eigenvalue of $\Gamma_z(P)$ and x_0 denote its corresponding unit norm eigenvector. By result (122) and $M_z(P)M_z(P) = M_z(P)$ due to $M_z(P)$ being a projection matrix, we then obtain

$$\text{diag}\{b_z(P)\}M_z(P)M_z(P)\text{diag}\{b_z(P)\}x_0 = \lambda_0 x_0. \quad (123)$$

Next set $v_0 \equiv M_z(P)\text{diag}\{b_z(P)\}x_0$ and note that result (123) implies that $\|v_0\|_2^2 = \lambda_0$ because $\|x_0\|_2 = 1$. Moreover, by direct calculation we can also conclude that

$$M_z(P)(\text{diag}\{b_z(P)\})^2 M_z(P) \frac{v_0}{\sqrt{\lambda_0}} = M_z(P)\text{diag}\{b_z(P)\}\Gamma_z(P) \frac{x_0}{\sqrt{\lambda_0}} = \lambda_0 \frac{v_0}{\sqrt{\lambda_0}}, \quad (124)$$

which implies λ_0 is an eigenvalue of the matrix $M_z(P)(\text{diag}\{b_z(P)\})^2 M_z(P)$. Also note

$$\begin{aligned} \min_{\|x\|_2=1, \langle x, a_z(P) \rangle=0} x' M_z(P)(\text{diag}\{b_z(P)\})^2 M_z(P) x \\ = \min_{\|x\|_2=1, \langle x, a_z(P) \rangle=0} x' (\text{diag}\{b_z(P)\})^2 x \geq \min_{(y,d,z) \in \mathcal{M}} \frac{1}{1 - P_{ydz}} \geq 1 \end{aligned} \quad (125)$$

where the equality follows from the definition of $M_z(P)$ and the inequality from the definition of $b_z(P)$. In particular, since $a_z(P)$ spans the null space of $M_z(P)(\text{diag}\{b_z(P)\})^2 M_z(P)$, result (125) implies that the positive eigenvalues of $M_z(P)(\text{diag}\{b_z(P)\})^2 M_z(P)$, and hence of $\Gamma_z(P)$, are bounded below by one. Since $\Omega(P)^\dagger \Sigma(P) \Omega(P)^\dagger$ is block diagonal with blocks $\Gamma_z(P)$, the third claim of the lemma follows.

For the fourth claim of the lemma, use the definitions of $\psi(V, P)$ and $\Omega(P)$ to obtain

$$\begin{aligned} \|\Omega(P)^\dagger \psi(V, P)\|_\infty \\ = \max_{(y,d,z) \in \mathcal{M}} \left(\frac{\sqrt{P_z}}{\sqrt{P_{ydz}(1 - P_{ydz})}} \times \left| \frac{1}{P_z} I\{Y = y, D = d, Z = z\} - \frac{P_{ydz}}{P_z^2} I\{Z = z\} \right| \right) \end{aligned}$$

$$\begin{aligned}
&\leq \max_{(y,d,z) \in \mathcal{M}} \frac{\sqrt{P_z}}{\sqrt{P_{yd|z}(1-P_{yd|z})}} \times \max_{(y,d,z) \in \mathcal{M}} \left| \frac{1}{P_z} + \frac{P_{yd|z}}{P_z} \right| \\
&\lesssim \sqrt{|\mathcal{M}|},
\end{aligned}$$

where the final inequality holds uniformly in $P \in \mathbf{P}$ by Assumption 4.1(iii).

Turning to the fifth and sixth claims, note that by Lemma B.10(iv), the event $E_0 \equiv \{\min_{z \in \mathcal{Z}} \hat{P}_z > 0\}$ has probability tending to one uniformly in $P \in \mathbf{P}$. We next argue that $(\hat{\beta}_n - \beta(P)) \in \text{range}(\Sigma(P))$ whenever the event E_0 occurs. To this end, note that by result (121), $\Sigma(P)$ is block diagonal with $|\mathcal{Z}|$ blocks we denote by $\Lambda_z(P)$, and has the final two rows and columns equal to zero. To examine the null space of $\Lambda_z(P)$ note that a vector $a_z \equiv (a_{yd} : (y, d, z) \in \mathcal{M})$ satisfies $\Lambda_z(P)a_z = 0$ if and only if it satisfies the equality

$$0 = \text{Var}_P \left(\sum_{(y,d):(y,d,z) \in \mathcal{M}} a_{yd} \left(\frac{1}{P_z} I\{Y=y, D=d, Z=z\} - \frac{P_{ydz}}{P_z^2} I\{Z=z\} \right) \right). \quad (126)$$

Defining a function f of (Y, D) by $f(y, d) = a_{yd}$, condition (126) may be equivalently expressed as f satisfying $(f(Y, D) - E_P[f(Y, D)|Z=z])I\{Z=z\} = 0$. However, by Assumption 4.1(iii) such an equality can only hold if f , and hence a , is constant. Given the structure of $\Sigma(P)$, we thus obtain that the null space of $\Sigma(P)$ has dimension $|\mathcal{Z}| + 2$ and its basis is given by: (a) $|\mathcal{Z}|$ vectors whose final two coordinates equal zero and first $|\mathcal{M}|$ coordinates equal a vector $a \equiv (a_{ydz} : (y, d, z) \in \mathcal{M})$ satisfying $a_{ydz} = I\{z = z_0\}$ for some z_0 ; and (b) The coordinate vectors for the last two coordinates. Because the final two coordinates of $(\hat{\beta}_n - \beta(P))$ equal zero and in addition $\sum_{(y,d)} \hat{P}_{yd|z} = \sum_{(y,d)} P_{yd|z} = 1$ whenever the event E_0 occurs, it follows that $(\hat{\beta}_n - \beta(P))$ is orthogonal to the null space of $\Sigma(P)$ and hence belongs to its range because $\Sigma(P)$ is symmetric. Identical arguments further imply that $(\hat{\beta}_n^* - \hat{\beta}_n) \in \text{range}(\Sigma(P))$ with probability tending to one uniformly in $P \in \mathbf{P}$. ■

Lemma B.8. *Let Assumptions 4.1(i) and 4.1(iii) hold, $e_j \in \mathbf{R}^{|\mathcal{M}|+2}$ has its j^{th} coordinate equal to one and all other coordinates equal to zero, and $v_z \in \mathbf{R}^{|\mathcal{M}|+2}$ has its (y, d, z) coordinate equal to one for all $(y, d, z) \in \mathcal{M}$ and all other coordinates equal to zero. Then:*

- (i) *If $\log(|\mathcal{M}|)|\mathcal{M}|/n = o(1)$ and the null space of A' is a subset of the span of the vectors $\{v_z : z \in \mathcal{Z}\} \cup \{e_{|\mathcal{M}|+1}, e_{|\mathcal{M}|+2}\}$, then it follows that Assumption 4.1(ii) holds.*
- (ii) *If $|\mathcal{M}|$ is fixed with n and Assumption 4.1(ii) holds, then it follows that the null space of A' is a subset of the span of the vectors $\{v_z : z \in \mathcal{Z}\} \cup \{e_{|\mathcal{M}|+1}, e_{|\mathcal{M}|+2}\}$.*

PROOF. For notational simplicity we set V to equal the span of $\{v_z : z \in \mathcal{Z}\} \cup \{e_{|\mathcal{M}|+1}, e_{|\mathcal{M}|+2}\}$, $\text{null}(A')$ to denote the null space of A' , and for any vector space H we let $H^\perp$ denote its orthogonal complement (i.e., $H^\perp \equiv \{h' : \langle h, h' \rangle = 0 \text{ for all } h \in H\}$).

For the first claim, first note that set inclusion implies that $V^\perp \subseteq [\text{null}(A')]^\perp$, and Theorem 6.6.3 in [Luenberger \(1997\)](#) therefore allows us to conclude that

$$V^\perp \subseteq \text{range}(A). \quad (127)$$

Next define the event $E_0 \equiv \{\hat{P}_z > 0 \text{ for all } z \in \mathcal{Z}\}$ and note that under E_0 we have $\langle v_z, \hat{\beta}_n \rangle = 1$ for any $z \in \mathcal{Z}$. Since $\langle v_z, \beta(P) \rangle = 1$ for any $z \in \mathcal{Z}$ and $\langle \hat{\beta}_n - \beta(P), e_j \rangle = 0$ for $j \in \{|\mathcal{M}| + 1, |\mathcal{M}| + 2\}$ by construction, it follows that under the event E_0 we have $(\hat{\beta}_n - \beta(P)) \in V^\perp$. Therefore, result (127) and E_0 occurring with probability tending to one uniformly in $P \in \mathbf{P}$ by Lemma B.10(iv) imply that Assumption 4.1(ii) holds when $\text{null}(A') \subseteq V$.

For the second claim, fix any $P \in \mathbf{P}$ and let $\psi(V, P)$ be defined in (41). Then note that $\Omega(P)\Omega(P)^\dagger(\hat{\beta}_n - \beta(P)) = (\hat{\beta}_n - \beta(P))$ and $\Omega(P)\Omega(P)^\dagger\psi(V, P) = \psi(V, P)$ due to the last two coordinates of $(\hat{\beta}_n - \beta(P))$ and $\psi(V, P)$ being zero, which imply that

$$\begin{aligned} & \|\sqrt{n}\{\hat{\beta}_n - \beta(P)\} - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(V_i, P)\|_2 \\ & \leq \|\Omega(P)\|_o \|\Omega(P)^\dagger\{\sqrt{n}\{\hat{\beta}_n - \beta(P)\} - \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(V_i, P)\}\|_2 = o_P(1) \end{aligned} \quad (128)$$

where the final result follows from Lemma B.3 and the dimension $|\mathcal{M}|$ being fixed. Setting $\Sigma(P) \equiv E_P[\psi(V, P)\psi(V, P)']$ and $\mathbb{G}(P) \sim N(0, \Sigma(P))$, it then follows that $\sqrt{n}\{\hat{\beta}_n - \beta(P)\} \xrightarrow{d} \mathbb{G}(P)$ by the central limit theorem and result (128). We therefore obtain that

$$1 = \limsup_{n \rightarrow \infty} P \left\{ \sqrt{n}(\hat{\beta}_n - \beta(P)) \in \text{range}(A) \right\} \leq P \left\{ \mathbb{G}(P) \in \text{range}(A) \right\} \quad (129)$$

where the equality follows from Assumption 4.1(ii) being satisfied by hypothesis, and the inequality by the Portmanteau Theorem (see, e.g., Theorem 1.3.4(iii) in [van der Vaart and Wellner \(1996\)](#)). Next note that since $\mathbb{G}(P) \sim N(0, \Sigma(P))$, Theorem 3.6.1 in [Bogachev \(1998\)](#) and result (129) imply that $\text{range}(\Sigma(P)) \subseteq \text{range}(A)$ or, equivalently, that $[\text{null}(\Sigma(P))]^\perp \subseteq [\text{null}(A')]^\perp$ by Theorem 6.6.3 in [Luenberger \(1997\)](#). Therefore, by set inclusion we can conclude that if Assumption 4.1(ii) holds, then we must have

$$\text{null}(A') \subseteq \text{null}(\Sigma(P)). \quad (130)$$

Finally, note that $\psi(V, P)'v = 0$ for any $v \in V$ by definition of $\psi(V, P)$. Hence, $\Sigma(P)v = 0$ for all $v \in V$ and, since the dimension of V is $|\mathcal{Z}| + 2$, Lemma B.7(vi) implies that $\text{null}(\Sigma(P)) = V$. The second claim of the lemma therefore follows from result (130). ■

Lemma B.9. *For each $z \in \mathcal{Z}$ select a $(y_z, d_z) \in \mathcal{M}_z$, let $\psi(V, P)$ be as defined in (41), $\psi^s(V, P)$ equal the vector obtained from removing the rows $((y_z, d_z, z) : z \in \mathcal{Z})$ from $\psi(V, P)$,*

$\Sigma^s(P) \equiv E_P[\psi^s(V, P)\psi^s(V, P)']$, and $\Omega^s(P)$ be a diagonal matrix with (j, j) entry $\Omega_{j,j}^s(P)$ equal to $(\Sigma_{j,j}(P))^{1/2}$. If Assumptions 4.1(i) and 4.1(iii) hold, then it follows that

- (i) The eigenvalues of $\Omega^s(P)^\dagger \Sigma^s(P) \Omega^s(P)^\dagger$ are uniformly bounded in $P \in \mathbf{P}$.
- (ii) The nonzero eigenvalues of $\Omega^s(P)^\dagger \Sigma^s(P) \Omega^s(P)^\dagger$ are bounded from below by $\pi(|\mathcal{Z}|/|\mathcal{M}|)$ for all $P \in \mathbf{P}$ and some constant $\pi > 0$ not depending on $|\mathcal{M}|$ or $|\mathcal{Z}|$.
- (iii) The range of $\Omega^s(P)^\dagger \Sigma^s(P) \Omega^s(P)^\dagger$ equals all vectors with the final two coordinates equal to zero for any $P \in \mathbf{P}$.

PROOF. For the first part of the lemma note that $\Omega(P)^\dagger \Sigma(P) \Omega(P)^\dagger$ is a $(|\mathcal{M}| + 2) \times (|\mathcal{M}| + 2)$ matrix whose first $|\mathcal{M}|$ rows and columns can be indexed by $(y, d, z) \in \mathcal{M}$. Since $\Omega^s(P)^\dagger \Sigma^s(P) \Omega^s(P)^\dagger$ equals the matrix obtained from removing every row and column indexed by $((y_z, d_z, z) : z \in \mathcal{Z})$, part (i) of the lemma follows from Lemma B.7(ii) and Cauchy's Interlacing Theorem (see, e.g., Corollary III.1.5 in Bhatia (2013)).

For the second part, note that $\Omega^s(P)^\dagger \Sigma^s(P) \Omega^s(P)^\dagger$ is a block diagonal matrix containing $|\mathcal{Z}|$ blocks and whose final two rows and columns equal zero. Letting $b_z^s(P)$ denote a vector with coordinates $((1 - P_{yd|z})^{-1} : (y, d) \in \mathcal{M}_z \setminus \{(y_z, d_z)\})$, $a_z^s(P)$ denote a vector with coordinates $(P_{yd|z}^{1/2}/(1 - P_{yd|z})^{1/2} : (y, d) \in \mathcal{M}_z \setminus \{(y_z, d_z)\})$, and $\text{diag}\{b_z^s(P)\}$ denote the diagonal matrix with entries $b_z^s(P)$, it then follows from direct calculation that

$$\Gamma_z^s(P) = \text{diag}\{b_z^s(P)\} - a_z^s(P)a_z^s(P)'. \quad (131)$$

Let $\pi \equiv c_1/c_2$ for c_1 and c_2 the constants in Assumption 4.1(iii) and define $c^* \equiv \pi(|\mathcal{Z}|/|\mathcal{M}|)$. Then note that, because of the block diagonal structure of $\Omega^s(P)^\dagger \Sigma^s(P) \Omega^s(P)^\dagger$, in order to establish part (ii) of the lemma it suffices to show that for any x we have

$$x'(\text{diag}\{b_z^s(P)\} - c^*I - a_z^s(P)a_z^s(P)')x \geq 0, \quad (132)$$

where I denotes the $(|\mathcal{M}_z| - 1) \times (|\mathcal{M}_z| - 1)$ identity matrix. Also note that all coordinates of $b_z^s(P)$ are larger than one, which implies that the matrix $\text{diag}\{b_z^s(P)\} - c^*I$ is invertible because $c^* < 1$. Hence, using the Cauchy-Schwarz inequality allows us to conclude that

$$\begin{aligned} (x'a_z^s(P))^2 &= (x'(\text{diag}\{b_z^s(P)\} - c^*I)^{1/2}(\text{diag}\{b_z^s(P)\} - c^*I)^{-1/2}a_z^s(P))^2 \\ &\leq (x'(\text{diag}\{b_z^s(P)\} - c^*I)x)(a_z^s(P)'(\text{diag}\{b_z^s(P)\} - c^*I)^{-1}a_z^s(P)). \end{aligned} \quad (133)$$

Moreover, by direct calculation and using that $(1 - P_{yd|z}) \leq 1$ and $c^* < 1$ we obtain

$$\begin{aligned} a_z^s(P)'(\text{diag}\{b_z^s(P)\} - c^*I)^{-1}a_z^s(P) &= \sum_{(y,d) \in \mathcal{M}_z \setminus \{(y_z, d_z)\}} \frac{P_{yd|z}}{1 - c^*(1 - P_{yd|z})} \\ &\leq \sum_{(y,d) \in \mathcal{M}_z \setminus \{(y_z, d_z)\}} \frac{P_{yd|z}}{1 - c^*} = \frac{1 - P_{y_z d_z | z}}{1 - c^*} \leq 1, \end{aligned} \quad (134)$$

where the final inequality follows from $c^* \leq P_{ydz}$ for all $(y, d, z) \in \mathcal{M}$ and $P \in \mathbf{P}$ by Assumption 4.1(iii). The second part of the lemma then follows by noting that results (133) and (134) imply that (132) is satisfied.

The third claim of the lemma is immediate from the block diagonal structure of the matrix $\Omega^s(P)^\dagger \Sigma^s(P) \Omega^s(P)^\dagger$ and each block $\Gamma_z^s(P)$ being full rank by result (132) and $c^* > 0$. ■

Lemma B.10. *Let Assumptions 4.1(i) and 4.1(iii) be satisfied. If in addition the condition $\log(|\mathcal{M}|)|\mathcal{M}|/n = o(1)$ holds, then it follows that uniformly in $P \in \mathbf{P}$:*

- (i) $\max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz} - P_{ydz}| = O_P(\sqrt{\log(|\mathcal{M}|)}/\sqrt{|\mathcal{M}|n});$
- (ii) $\max_{z \in \mathcal{Z}} |\hat{P}_z - P_z| = O_P(\sqrt{\log(|\mathcal{Z}|)}/\sqrt{|\mathcal{Z}|n});$
- (iii) $\max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz}/P_{ydz} - 1| = O_P(\sqrt{\log(|\mathcal{M}|)|\mathcal{M}|}/\sqrt{n});$
- (iv) $\max_{z \in \mathcal{Z}} |\hat{P}_z/P_z - 1| = O_P(\sqrt{\log(|\mathcal{Z}|)|\mathcal{Z}|}/\sqrt{n});$
- (v) $\max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz} - P_{ydz}| = O_P(\sqrt{\log(|\mathcal{M}|)|\mathcal{Z}|/n}).$

PROOF. First note Bernstein's inequality (see, for example, Lemma 2.2.9 in [van der Vaart and Wellner, 1996](#)) allows us to conclude that

$$P \left\{ |\hat{P}_{ydz} - P_{ydz}| > x \right\} \leq 2 \exp \left(-\frac{1}{2} \frac{x^2}{\frac{P_{ydz}(1-P_{ydz})}{n} + \frac{x}{3n}} \right),$$

where the inequality holds for all $(y, d, z) \in \mathcal{M}$ and $P \in \mathbf{P}$. Therefore, Lemma 2.2.10 in [van der Vaart and Wellner \(1996\)](#) and the norm inequality $\|\cdot\|_1 \leq \|\cdot\|_{\psi_1}$ yield

$$E \left[\max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz} - P_{ydz}| \right] \lesssim \frac{\log(|\mathcal{M}|)}{n} + \max_{(y,d,z) \in \mathcal{M}} (P_{ydz}(1-P_{ydz}))^{1/2} \frac{\sqrt{\log(|\mathcal{M}|)}}{\sqrt{n}}. \quad (135)$$

Next observe $P_{ydz}(1-P_{ydz}) \lesssim 1/|\mathcal{M}|$ uniformly in $(y, d, z) \in \mathcal{M}$ and $P \in \mathbf{P}$ by Assumption 4.1(iii). Therefore, result (135), Markov's inequality, and $\log(|\mathcal{M}|)|\mathcal{M}|/n = o(1)$ imply

$$\max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz} - P_{ydz}| = O_P \left(\frac{\sqrt{\log(|\mathcal{M}|)}}{\sqrt{|\mathcal{M}|n}} \right) \quad (136)$$

uniformly in $P \in \mathbf{P}$, which establishes the first claim. Moreover, by identical arguments

$$\max_{z \in \mathcal{Z}} |\hat{P}_z - P_z| = O_P \left(\frac{\sqrt{\log(|\mathcal{Z}|)}}{\sqrt{|\mathcal{Z}|n}} \right) \quad (137)$$

uniformly in $P \in \mathbf{P}$, which establishes the second claim. For the third claim, note that

$$\max_{(y,d,z) \in \mathcal{M}} \left| \frac{\hat{P}_{ydz}}{P_{ydz}} - 1 \right| \leq \max_{(y,d,z) \in \mathcal{M}} \frac{1}{P_{ydz}} \times \max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz} - P_{ydz}| = O_P \left(\frac{\sqrt{\log(|\mathcal{M}|)|\mathcal{M}|}}{\sqrt{n}} \right), \quad (138)$$

where the equality holds uniformly in $P \in \mathbf{P}$ by Assumption 4.1(iii) and result (136). The fourth claim of the lemma follows by identical arguments to those employed in (138) by relying on (137) instead of (136). To establish the final claim of the Lemma, note that

$$\begin{aligned} \max_{(y,d,z) \in \mathcal{M}} \left| \frac{\hat{P}_{ydz}}{\hat{P}_z} - \frac{P_{ydz}}{P_z} \right| &\leq \max_{z \in \mathcal{Z}} \frac{1}{\hat{P}_z} \times \max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz} - P_{ydz}| + \max_{(y,d,z) \in \mathcal{M}} P_{ydz} \times \max_{z \in \mathcal{Z}} \frac{|\hat{P}_z - P_z|}{\hat{P}_z P_z} \\ &= O_P \left(|\mathcal{Z}| \frac{\sqrt{\log(|\mathcal{M}|)}}{\sqrt{|\mathcal{M}|n}} \right) + O_P \left(\frac{|\mathcal{Z}|^2 \sqrt{\log(|\mathcal{Z}|)}}{|\mathcal{M}| \sqrt{|\mathcal{Z}|n}} \right) = O_P \left(\frac{\sqrt{\log(|\mathcal{M}|)} |\mathcal{Z}|}{\sqrt{n}} \right) \end{aligned} \quad (139)$$

where the equalities hold uniformly in $P \in \mathbf{P}$ by parts (i), (ii), and (iv) of this Lemma and Assumption 4.1(iii). ■

Lemma B.11. *Let Assumptions 4.1(i) and 4.1(iii) be satisfied. If in addition the condition $\log(|\mathcal{M}|)|\mathcal{M}|/n = o(1)$ holds, then it follows that uniformly in $P \in \mathbf{P}$:*

- (i) $\max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz}^* - \hat{P}_{ydz}| = O_P(\sqrt{\log(|\mathcal{M}|)}/\sqrt{|\mathcal{M}|n});$
- (ii) $\max_{z \in \mathcal{Z}} |\hat{P}_z^* - \hat{P}_z| = O_P(\sqrt{\log(|\mathcal{Z}|)}/\sqrt{|\mathcal{Z}|n});$
- (iii) $\max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz}^*/\hat{P}_{ydz} - 1| = O_P(\sqrt{\log(|\mathcal{M}|)}|\mathcal{M}|/\sqrt{n});$
- (iv) $\max_{z \in \mathcal{Z}} |\hat{P}_z^*/\hat{P}_z - 1| = O_P(\sqrt{\log(|\mathcal{Z}|)}|\mathcal{Z}|/\sqrt{n}).$

PROOF. We follow similar arguments to those used in the proof of Lemma B.10. First define

$$A_n \equiv \left\{ \max_{(y,d,z) \in \mathcal{M}} \left| \frac{\hat{P}_{ydz}}{P_{ydz}} - 1 \right| \leq \frac{1}{2} \text{ and } \max_{z \in \mathcal{Z}} \left| \frac{\hat{P}_z}{P_z} - 1 \right| \leq \frac{1}{2} \right\},$$

which we note is a function of $\{V_i\}_{i=1}^n$. Therefore, the law of iterated expectations yields

$$\begin{aligned} P \left\{ \max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz}^* - \hat{P}_{ydz}| > x \right\} \\ \leq E_P \left[P \left\{ \max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz}^* - \hat{P}_{ydz}| > x \middle| \{V_i\}_{i=1}^n \right\} I\{A_n\} \right] + P\{A_n^c\} \end{aligned} \quad (140)$$

for any $x > 0$. Next, note that Bernstein's inequality (see, for example, Lemma 2.2.9 in van der Vaart and Wellner (1996)) allows us to conclude that

$$P \left\{ |\hat{P}_{ydz}^* - \hat{P}_{ydz}| > x \middle| \{V_i\}_{i=1}^n \right\} \leq 2 \exp \left(-\frac{1}{2} \frac{x^2}{\frac{\hat{P}_{ydz}(1-\hat{P}_{ydz})}{n} + \frac{x}{3n}} \right),$$

where the inequality holds for all $(y, d, z) \in \mathcal{M}$. Therefore, Lemma 2.2.10 in van der Vaart and Wellner (1996) and the norm inequality $\|\cdot\|_1 \leq \|\cdot\|_{\psi_1}$ yield that

$$E_P \left[\max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz}^* - \hat{P}_{ydz}| \middle| \{V_i\}_{i=1}^n \right] I\{A_n\} \lesssim \frac{\log(|\mathcal{M}|)}{n} + \max_{(y,d,z) \in \mathcal{M}} P_{ydz}^{1/2} \frac{\sqrt{\log(|\mathcal{M}|)}}{\sqrt{n}}. \quad (141)$$

Next observe that $P_{ydz} \lesssim 1/|\mathcal{M}|$ uniformly in $(y, d, z) \in \mathcal{M}$ and $P \in \mathbf{P}$ by Assumption 4.1(iii). Therefore, result (141) together with Markov's inequality and the condition $\log(|\mathcal{M}|)|\mathcal{M}|/n = o(1)$ yield for any constant $K > 0$ that

$$E_P \left[P \left\{ \max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz}^* - \hat{P}_{ydz}| > K \frac{\sqrt{\log(|\mathcal{M}|)}}{\sqrt{|\mathcal{M}|n}} \left| \{V_i\}_{i=1}^n \right\} I\{A_n\} \right\} \right] \lesssim \frac{1}{K}. \quad (142)$$

Hence, since $\sup_{P \in \mathbf{P}} P\{A_n^c\} = o(1)$ by Lemmas B.10(iii),(iv), results (140) and (142) yield

$$\max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz}^* - \hat{P}_{ydz}| = O_P \left(\frac{\sqrt{\log(|\mathcal{M}|)}}{\sqrt{|\mathcal{M}|n}} \right) \quad (143)$$

uniformly in $P \in \mathbf{P}$, which establishes the first claim. Moreover, by identical arguments,

$$\max_{z \in \mathcal{Z}} |\hat{P}_z^* - \hat{P}_z| = O_P \left(\frac{\sqrt{\log(|\mathcal{Z}|)}}{\sqrt{|\mathcal{Z}|n}} \right) \quad (144)$$

uniformly in $P \in \mathbf{P}$, which establishes the second claim. For the third claim, note that

$$\max_{(y,d,z) \in \mathcal{M}} \left| \frac{\hat{P}_{ydz}^*}{\hat{P}_{ydz}} - 1 \right| \leq \max_{(y,d,z) \in \mathcal{M}} \frac{1}{\hat{P}_{ydz}} \times \max_{(y,d,z) \in \mathcal{M}} |\hat{P}_{ydz}^* - \hat{P}_{ydz}| = O_P \left(\frac{\sqrt{\log(|\mathcal{M}|)|\mathcal{M}|}}{\sqrt{n}} \right), \quad (145)$$

where the equality holds uniformly in $P \in \mathbf{P}$ by Lemma B.10(iii), Assumption 4.1(iii), and (143). The fourth claim follows by identical arguments to those employed in (145) but relying on (144) instead of (143) and Lemma B.10(iv) instead of Lemma B.10(iii). ■

Lemma B.12. *Let Assumption 5.1 hold, $\text{diam}\{B_{\ell,L}\} < \delta$ for some $\delta > 0$ and all $1 \leq \ell \leq L$, $P\{(Y, D, Z) \in \mathcal{Y} \times \mathcal{D} \times \mathcal{Z}\} = 1$, and $f : \mathcal{Y} \times \mathcal{D} \times \mathcal{Z} \rightarrow \mathbf{R}$ be a function satisfying*

$$\max_{d \in \mathcal{D}, z \in \mathcal{Z}} \sup_{|y-y'| \leq 2\delta} |f(y, d, z) - f(y', d, z)| \leq \eta \quad (146)$$

for some $\eta > 0$. Define $\mathcal{I} \equiv \{(c_o, r_t, \chi) \in \mathcal{L}^{|\mathcal{D}|} \times \mathcal{R}_t \times \{0, 1\} : \mathcal{Y}(c_o, r_t, \chi) \neq \emptyset\}$ and suppose $P = QT^{-1}$ where $Q = Q_R \times P_Z$ for P_Z the marginal distribution of Z and Q_R satisfying

$$\sum_{(c_o, r_t, \chi) \in \mathcal{I}} Q_R\{(R_o, R_t) \in (\mathcal{Y}(c_o, r_t, \chi) \times \{r_t\})\} = 1 \quad (147)$$

$$\sum_{(c_o, r_t) : (c_o, r_t, 1) \in \mathcal{I}} \left(\sup_{r_o \in \mathcal{Y}(c_o, r_t, 1)} g(r_o, r_t) - \theta_0 \right) \times Q_R\{(R_o, R_t) \in (\mathcal{Y}(c_o, r_t, 1) \times \{r_t\})\} \geq 0 \quad (148)$$

$$\sum_{(c_o, r_t) : (c_o, r_t, 1) \in \mathcal{I}} \left(\inf_{r_o \in \mathcal{Y}(c_o, r_t, 1)} g(r_o, r_t) - \theta_0 \right) \times Q_R\{(R_o, R_t) \in (\mathcal{Y}(c_o, r_t, 1) \times \{r_t\})\} \leq 0. \quad (149)$$

Then, there is a distribution $\tilde{Q}_R$ of R satisfying $\tilde{Q}_R\{R \in \mathcal{R}\} = 1$, $E_{\tilde{Q}_R}[(g(R) - \theta_0)I\{R \in \mathcal{R}'\}] = 0$, and such that $\tilde{P} = \tilde{Q}T^{-1}$ with $\tilde{Q} = \tilde{Q}_R \times P_Z$ satisfies $\int f(dP - d\tilde{P}) \leq 2\eta$.

PROOF. The proof proceeds by constructing a discrete measure $\tilde{Q}_R$ satisfying the desired properties. We distinguish multiple cases depending on whether the inequalities in (148) and (149) hold strictly or with equality. In what follows, it will be helpful to note that $P = QT^{-1}$ and R being independent of Z under Q allow us to conclude that

$$\begin{aligned} P\{Y \in B_{\ell,L}, D = d, Z = z\} &= \sum_{c_o \in \mathcal{L}^{|\mathcal{D}|}, r_t \in \mathcal{R}_t} Q_R\{R_o \in B_{c_o}, R_t = r_t\} I\{c_o(d) = \ell, d(z) = d\} P\{Z = z\} \\ &= \sum_{(c_o, r_t, \chi) \in \mathcal{I}} Q_R\{R_o \in \mathcal{Y}(c_o, r_t, \chi), R_t = r_t\} I\{c_o(d) = \ell, d(z) = d\} P\{Z = z\}, \end{aligned} \quad (150)$$

where the final equality follows from the definition of $\mathcal{Y}(c_o, r_t, \chi)$ and Q_R satisfying (147). Moreover, the same equality holds if we replace P and Q with $\tilde{P}$ and $\tilde{Q}$ because of $\tilde{Q}_R\{R \in \mathcal{R}\} = 1$ and result (58).

Case I. First, suppose that the inequalities in (148) and (149) both hold strictly and define

$$\mathbf{V} \equiv \bigotimes_{(c_o, r_t) : (c_o, r_t, 1) \in \mathcal{I}} \mathcal{Y}(c_o, r_t, 1).$$

Note that the set $\mathbf{V}$ is connected because $\mathcal{Y}(c_o, r_t, 1) = B_{c_o} \cap \mathcal{R}'_o(r_t)$ is connected for all (c_o, r_t) such that $(c_o, r_t, 1) \in \mathcal{I}$ by Assumption 5.1(iv). Further write any element $v \in \mathbf{V}$ as $v = (v(c_o, r_t) \in \mathcal{Y}(c_o, r_t, 1) : (c_o, r_t, 1) \in \mathcal{I})$ and define the function

$$F(v) \equiv \sum_{(c_o, r_t) : (c_o, r_t, 1) \in \mathcal{I}} (g(v(c_o, r_t), r_t) - \theta_0) \times Q\{(R_o, R_t) \in (\mathcal{Y}(c_o, r_t, 1) \times \{r_t\})\}.$$

Note that $F : \mathbf{V} \rightarrow \mathbf{R}$ and that inequalities (148) and (149) holding strictly imply there are $\underline{v}, \bar{v} \in \mathbf{V}$ satisfying $F(\underline{v}) < 0 < F(\bar{v})$. Since $F : \mathbf{V} \rightarrow \mathbf{R}$ is continuous and $\mathbf{V}$ is connected, it follows that $F(\mathbf{V})$ is connected as well; see, e.g., Theorem 23.5 in Munkres (2000). Therefore, we conclude that there is a $v^* \in \mathbf{V}$ such that $F(v^*) = 0$. Finally, define a discrete measure $\tilde{Q}_R$ to have support points $(s(c_o, r_t, \chi) : (c_o, r_t, \chi) \in \mathcal{I})$, where

$$s(c_o, r_t, \chi) = \begin{cases} (v^*(c_o, r_t), r_t) & \text{if } \chi = 1 \\ \text{any } (r_o, r_t) \in (\mathcal{Y}(c_o, r_t, \chi) \times \{r_t\}) & \text{if } \chi = 0. \end{cases}$$

Note that such a definition is possible because $\mathcal{Y}(c_o, r_t, \chi) \neq \emptyset$ for any $(c_o, r_t, \chi) \in \mathcal{I}$. We assign probabilities $\tilde{Q}_R\{(R_o, R_t) = s(c_o, r_t, \chi)\} = Q_R\{(R_o, R_t) \in \mathcal{Y}(c_o, r_t, \chi) \times \{r_t\}\}$ for any $(c_o, r_t, \chi) \in \mathcal{I}$. Since $s(c_o, r_t, \chi) \in (\mathcal{Y}(c_o, r_t, \chi) \times \{r_t\})$ and $\mathcal{Y}(c_o, r_t, \chi) \times \{r_t\} \subseteq \mathcal{R}$ for all $(c_o, r_t, \chi) \in \mathcal{I}$ by definition of $\mathcal{Y}(c_o, r_t, \chi)$, it follows from result (147) that $\tilde{Q}_R\{R \in \mathcal{R}\} = 1$.

Further note that $F(v^*) = 0$ implies that $E_{\tilde{Q}_R}[(g(R) - \theta_0)I\{R \in \mathcal{R}'\}] = 0$. Next, use result (150) and Q_R satisfying the restriction (147) to conclude that

$$\begin{aligned} \int f dP &\leq \sum_{(\ell, d, z) \in \mathcal{L} \times \mathcal{D} \times \mathcal{Z}} P\{Y \in B_{\ell, L}, D = d, Z = z\} \times \sup_{y \in B_{\ell, L}} f(y, d, z) \\ &= \sum_{(\ell, d, z) \in \mathcal{L} \times \mathcal{D} \times \mathcal{Z}} \sum_{(c_o, r_t, \chi) \in \mathcal{I}} \left\{ Q_R\{R_o \in \mathcal{Y}(c_o, r_t, \chi), R_t = r_t\} I\{c_o(d) = \ell, d(z) = d\} \right. \\ &\quad \left. \times P\{Z = z\} \sup_{y \in B_{\ell, L}} f(y, d, z) \right\}. \end{aligned} \quad (151)$$

Similarly, $\tilde{Q}_R\{(R_o, R_t) = s(c_o, r_t, \chi)\} = Q_R\{(R_o, R_t) \in \mathcal{Y}(c_o, r_t, \chi) \times \{r_t\}\}$ for $(c_o, r_t, \chi) \in \mathcal{I}$ by definition of $\tilde{Q}_R$ and $\tilde{P}_Z = P_Z$ allow us to obtain that

$$\begin{aligned} \int f d\tilde{P} &\geq \sum_{(\ell, d, z) \in \mathcal{L} \times \mathcal{D} \times \mathcal{Z}} \tilde{P}\{Y \in B_{\ell, L}, D = d, Z = z\} \times \inf_{y \in B_{\ell, L}} f(y, d, z) \\ &= \sum_{(\ell, d, z) \in \mathcal{L} \times \mathcal{D} \times \mathcal{Z}} \sum_{(c_o, r_t, \chi) \in \mathcal{I}} \left\{ Q_R\{R_o \in \mathcal{Y}(c_o, r_t, \chi), R_t = r_t\} I\{c_o(d) = \ell, d(z) = d\} \right. \\ &\quad \left. \times P\{Z = z\} \inf_{y \in B_{\ell, L}} f(y, d, z) \right\}. \end{aligned} \quad (152)$$

Results (151) and (152) together with $\text{diam}\{B_{\ell, L}\} < \delta$ and the function f satisfying restriction (146) then imply $\int f(dP - d\tilde{P}) \leq \eta$.

Case II. Second, suppose that (148) holds with equality. Let $\bar{B}_{c_o}$ denote the closure of B_{c_o} and then note that $\bar{B}_{c_o} \cap \mathcal{R}'_o(r_t)$ is compact since $\mathcal{R}'_o(r_t)$ is closed by Assumption 5.1(iv) and $B_{\ell, L}$ is bounded. Then note that the definition of $\mathcal{Y}(c_o, r_t, \chi)$ implies that $\mathcal{Y}(c_o, r_t, 1) = B_{c_o} \cap \mathcal{R}'_o(r_t) \neq \emptyset$ for any (c_o, r_t) such that $(c_o, r_t, 1) \in \mathcal{I}$. Since g is continuous, it follows that for each (c_o, r_t) such that $(c_o, r_t, 1) \in \mathcal{I}$, there is a $v^*(c_o, r_t) \in \bar{B}_{c_o} \cap \mathcal{R}'_o(r_t)$ satisfying

$$\sup_{r_o \in \mathcal{Y}(c_o, r_t, 1)} g(r_o, r_t) = \sup_{r_o \in B_{c_o} \cap \mathcal{R}'_o(r_t)} g(r_o, r_t) = g(v^*(c_o, r_t), r_t). \quad (153)$$

Note that $v^*(c_o, r_t)$ does not necessarily belong to $B_{c_o} \cap \mathcal{R}'_o(r_t)$. However, since $\{B_{\tilde{c}_o}\}_{\tilde{c}_o \in \mathcal{L}^{|\mathcal{D}|}}$ is a partition of $\mathcal{Y}^{|\mathcal{D}|}$ and $\mathcal{Y}^{|\mathcal{D}|}$ is closed because $\mathcal{Y}$ is closed by Assumption 5.1(i), it follows that $v^*(c_o, r_t) \in B_{\tilde{c}_o}$ for some $\tilde{c}_o$ and, since $v^*(c_o, r_t) \in \mathcal{R}'_o(r_t)$, we also have

$$v^*(c_o, r_t) \in B_{\tilde{c}_o} \cap \mathcal{R}'_o(r_t) = \mathcal{Y}(\tilde{c}_o, r_t, 1) \quad (154)$$

by definition of $\mathcal{Y}(\tilde{c}_o, r_t, \chi)$. Moreover, for any $d \in \mathcal{D}$ and $v_d^*(c_o, r_t)$ denoting the d^{th} coordinate of $v^*(c_o, r_t) \in \mathcal{Y}^{|\mathcal{D}|}$ we can conclude from the triangle inequality that

$$I\{v^*(c_o, r_t) \in \mathcal{Y}(\tilde{c}_o, r_t, 1)\} \times \sup_{y \in B_{c_o(d), L}} \sup_{y' \in B_{\tilde{c}_o(d), L}} |y - y'|$$

$$\begin{aligned}
&\leq I\{v^*(c_o, r_t) \in \mathcal{Y}(\tilde{c}_o, r_t, 1)\} \times \left\{ \sup_{y \in B_{c_o(d), L}} |y - v_d^*(c_o, r_t)| + \sup_{y' \in B_{\tilde{c}_o(d), L}} |v_d^*(c_o, r_t) - y'| \right\} \\
&\leq 2\delta,
\end{aligned} \tag{155}$$

where the final inequality follows from $v_d^*(c_o, r_t)$ belonging to the closure of $B_{c_o(d), L}$ because $v^*(c_o, r_t) \in \bar{B}_{c_o}$, $v_d^*(c_o, r_t) \in B_{\tilde{c}_o(d), L}$, and $\text{diam}\{B_{\ell, L}\} < \delta$. Finally, we let $\tilde{Q}_R$ be a discrete measure with support points $(s(c_o, r_t, \chi) : (c_o, r_t, \chi) \in \mathcal{I})$ given by

$$s(c_o, r_t, \chi) = \begin{cases} (v^*(c_o, r_t), r_t) & \text{if } \chi = 1 \\ \text{any } (r_o, r_t) \in (\mathcal{Y}(c_o, r_t, \chi) \times \{r_t\}) & \text{if } \chi = 0, \end{cases}$$

and assign them probabilities $\tilde{Q}_R\{(R_o, R_t) = s(c_o, r_t, \chi)\} = \sum_{\tilde{c}_o: (\tilde{c}_o, r_t, \chi) \in \mathcal{I}} I\{s(\tilde{c}_o, r_t, \chi) = s(c_o, r_t, \chi)\} Q_R\{(R_o, R_t) \in \mathcal{Y}(\tilde{c}_o, r_t, \chi) \times \{r_t\}\}$, where we use the fact that $s(\tilde{c}_o, \tilde{r}_t, \tilde{\chi}) = s(c_o, r_t, \chi)$ implies $(\tilde{r}_t, \tilde{\chi}) = (r_t, \chi)$. Since $(v^*(c_o, r_t), r_t) \in \mathcal{R}' \subseteq \mathcal{R}$ for any (c_o, r_t) such that $(c_o, r_t, 1) \in \mathcal{I}$ and in addition $(\mathcal{Y}(c_o, r_t, 0) \times \{r_t\}) \subseteq \mathcal{R}$ for any $(c_o, r_t, 0) \in \mathcal{I}$, it follows that $\tilde{Q}_R\{R \in \mathcal{R}\} = 1$ due to Q_R satisfying (147). Moreover, $\tilde{Q}_R$ also satisfies $E_{\tilde{Q}_R}[(g(R) - \theta_0)I\{R \in \mathcal{R}'\}] = 0$ because of result (153) and restriction (148) holding with equality by hypothesis. Further note that

$$\begin{aligned}
\int f d\tilde{P} &\geq \sum_{(\ell, d, z) \in \mathcal{L} \times \mathcal{D} \times \mathcal{Z}} \sum_{(c_o, r_t, \chi) \in \mathcal{I}} \left\{ \tilde{Q}_R\{R_o \in \mathcal{Y}(c_o, r_t, \chi), R_t = r_t\} I\{c_o(d) = \ell, d(z) = d\} \right. \\
&\quad \left. \times P\{Z = z\} \inf_{y \in B_{\ell, L}} f(y, d, z) \right\} \tag{156}
\end{aligned}$$

by result (150) applied with $\tilde{P}$ and $\tilde{Q}$ in place of P and Q . In addition, note that $s(\tilde{c}_o, \tilde{r}_t, \tilde{\chi}) \in (\mathcal{Y}(c_o, r_t, \chi) \times \{r_t\})$ implies that $(\tilde{r}_t, \tilde{\chi}) = (r_t, \chi)$ and therefore by definition of $\tilde{Q}$ we obtain

$$\begin{aligned}
&\tilde{Q}_R\{R_o \in \mathcal{Y}(c_o, r_t, \chi), R_t = r_t\} \\
&= \sum_{\tilde{c}_o: (\tilde{c}_o, r_t, \chi) \in \mathcal{I}} I\{s(\tilde{c}_o, r_t, \chi) \in (\mathcal{Y}(c_o, r_t, \chi) \times \{r_t\})\} Q_R\{R_o \in \mathcal{Y}(\tilde{c}_o, r_t, \chi), R_t = r_t\}. \tag{157}
\end{aligned}$$

Furthermore, noting that $s(\tilde{c}_o, r_t, 0) \in (\mathcal{Y}(c_o, r_t, 0) \times \{r_t\})$ if and only if $\tilde{c}_o = c_o$, and using result (155) and that the function $f : \mathcal{Y} \times \mathcal{D} \times \mathcal{Z} \rightarrow \mathbf{R}$ satisfies restriction (146) we obtain that for any $c_o, \tilde{c}_o$ such that $(c_o, r_t, \chi), (\tilde{c}_o, r_t, \chi) \in \mathcal{I}$ the following inequality holds

$$\begin{aligned}
&\sum_{\ell=1}^L I\{c_o(d) = \ell, d(z) = d\} I\{s(\tilde{c}_o, r_t, \chi) \in (\mathcal{Y}(c_o, r_t, \chi) \times \{r_t\})\} \inf_{y \in B_{\ell, L}} f(y, d, z) \\
&\geq \sum_{\ell=1}^L I\{\tilde{c}_o(d) = \ell, d(z) = d\} I\{s(\tilde{c}_o, r_t, \chi) \in (\mathcal{Y}(c_o, r_t, \chi) \times \{r_t\})\} \left(\inf_{y \in B_{\ell, L}} f(y, d, z) - \eta \right). \tag{158}
\end{aligned}$$

Since in addition $\sum_{c_o: (c_o, r_t, \chi) \in \mathcal{I}} I\{s(\tilde{c}_o, r_t, \chi) \in (\mathcal{Y}(c_o, r_t, \chi) \times \{r_t\})\} = 1$ for any $\tilde{c}_o$ satisfying $(\tilde{c}_o, r_t, \chi) \in \mathcal{I}$, results (156)–(158) allow us to conclude that

$$\begin{aligned} \int f d\tilde{P} \geq & \sum_{(\ell, d, z) \in \mathcal{L} \times \mathcal{D} \times \mathcal{Z}} \sum_{(\tilde{c}_o, r_t, \chi) \in \mathcal{I}} \left\{ Q\{R_o \in \mathcal{Y}(\tilde{c}_o, r_t, \chi), R_t = r_t\} I\{\tilde{c}_o(d) = \ell, r_t(z) = d\} \right. \\ & \left. \times P\{Z = z\} \times \left(\inf_{y \in B_{\ell, L}} f(y, d, z) - \eta \right) \right\}. \end{aligned}$$

Therefore, the upper bound for $\int f dP$ obtained in (151), the function $f : \mathcal{Y} \times \mathcal{D} \times \mathcal{Z} \rightarrow \mathbf{R}$ satisfying (146), and $\text{diam}\{B_{\ell, L}\} < \delta$ for all $1 \leq \ell \leq L$ imply that $\int f(dP - d\tilde{P}) \leq 2\eta$.

Case III. The third case with (149) holding with equality follows from the same arguments as Case II. ■

Lemma B.13. *For any $P \in \mathbf{P}$ let $P_{YD|Z}$ and P_Z denote the conditional distribution of (Y, D) given Z and the marginal distribution of Z under P . Suppose $\mathbf{P}_0 \neq \emptyset$, $\mathbf{P}$ is convex, $P_{YD|Z}\tilde{P}_Z \in \mathbf{P}$ for any $P, \tilde{P} \in \mathbf{P}$, $P\{(Y, D, Z) \in \mathcal{Y} \times \mathcal{D} \times \mathcal{Z}\} = 1$ for all $P \in \mathbf{P}$, and let $\mathbf{Q}$ be the set of distributions for (R_o, R_t, Z) satisfying Assumptions 2.1–2.2 and $\text{cl}_{\mathbf{P}}(\mathbf{P}_0)$ denote the closure of $\mathbf{P}_0$ under weak convergence (in $\mathbf{P}$). If $P \in \mathbf{P} \setminus \text{cl}_{\mathbf{P}}(\mathbf{P}_0)$, then there is a continuous and bounded function $f : \mathcal{Y} \times \mathcal{D} \times \mathcal{Z} \rightarrow \mathbf{R}$ satisfying $\int f dP = 1$ and $\int f d\tilde{P} \leq 0$ for all $\tilde{P} \in \mathbf{P}_0$ satisfying $\tilde{P}_Z = P_Z$.*

PROOF. Let $\mathbf{V}$ denote the linear span of $\mathbf{P}$, which forms a vector space, set $\mathcal{F} \equiv \mathcal{Y} \times \mathcal{D} \times \mathcal{Z}$ noting that $P\{(Y, D, Z) \in \mathcal{F}\} = 1$ for any $P \in \mathbf{P}$ by assumption, and define

$$C_b(\mathcal{F}) \equiv \{f : \mathcal{F} \rightarrow \mathbf{R} \text{ s.t. } f \text{ is continuous and bounded}\}$$

We equip $\mathbf{V}$ with the weakest topology that makes the linear maps $\tilde{P} \mapsto \int f d\tilde{P}$ continuous for each $f \in C_b(\mathcal{F})$, which we denote by $\sigma(\mathbf{V}, C_b(\mathcal{F}))$. By Theorem 5.73 and Lemmas 2.52 and 2.53 in Aliprantis and Border (2006), $(\mathbf{V}, \sigma(\mathbf{V}, C_b(\mathcal{F})))$ is a locally convex topological vector space and the relative topology on $\mathbf{P}$ induced by $\sigma(\mathbf{V}, C_b(\mathcal{F}))$ equals the topology of weak convergence on $\mathbf{P}$. Next set $\mathbf{P}_0(P) \equiv \{\tilde{P} \in \mathbf{P}_0 : \tilde{P}_Z = P_Z\}$, let $\text{cl}_{\mathbf{P}}(\mathbf{P}_0(P))$ denote the closure of $\mathbf{P}_0(P)$ under weak convergence (in $\mathbf{P}$), and note that $\mathbf{P}_0(P) \subseteq \mathbf{P}_0$ implies $\text{cl}_{\mathbf{P}}(\mathbf{P}_0(P)) \subseteq \text{cl}_{\mathbf{P}}(\mathbf{P}_0)$ and hence $P \notin \text{cl}_{\mathbf{P}}(\mathbf{P}_0(P))$ because $P \notin \text{cl}_{\mathbf{P}}(\mathbf{P}_0)$ by hypothesis. Moreover, setting $\text{cl}_{\mathbf{V}}(\mathbf{P}_0(P))$ to denote the $\sigma(\mathbf{V}, C_b(\mathcal{F}))$ -closure of $\mathbf{P}_0(P)$ in $\mathbf{V}$, it follows from Theorem 17.4 in Munkres (2000) that $\text{cl}_{\mathbf{P}}(\mathbf{P}_0(P)) = \text{cl}_{\mathbf{V}}(\mathbf{P}_0(P)) \cap \mathbf{P}$. Therefore, since $P \in \mathbf{P} \setminus \text{cl}_{\mathbf{P}}(\mathbf{P}_0(P))$, we can conclude that $P \notin \text{cl}_{\mathbf{V}}(\mathbf{P}_0(P))$. Also note that since $\mathbf{P}_0(P)$ is nonempty and convex by Lemma B.14, Lemma 5.27(6) in Aliprantis and Border (2006) implies $\text{cl}_{\mathbf{V}}(\mathbf{P}_0(P))$ is nonempty and convex as well.

We have so far shown that $P \notin \text{cl}_{\mathbf{V}}(\mathbf{P}_0(P))$ and that $\text{cl}_{\mathbf{V}}(\mathbf{P}_0(P))$ is a nonempty closed convex subset of the locally convex topological vector space $(\mathbf{V}, \sigma(\mathbf{V}, C_b(\mathcal{F})))$. By Corollary

5.80 and Theorem 5.93 in Aliprantis and Border (2006), there therefore exists a $\tilde{f} \in C_b(\mathcal{F})$ satisfying $\int \tilde{f} d\tilde{P} \leq 0$ for all $\tilde{P} \in \text{cl}_{\mathbf{V}}(\mathbf{P}_0(P))$ and $\int \tilde{f} dP > 0$. The lemma then follows by setting $f = \tilde{f} / \int \tilde{f} dP$ and using that $\mathbf{P}_0(P) \subseteq \text{cl}_{\mathbf{V}}(\mathbf{P}_0(P))$. ■

Lemma B.14. *For any $P \in \mathbf{P}$ let $P_{YD|Z}$ and P_Z denote the conditional distribution of (Y, D) given Z and the marginal distribution of Z under P . Suppose $\mathbf{P}$ is convex, that for any $P, \tilde{P} \in \mathbf{P}$ we have $P_{YD|Z}\tilde{P}_Z \in \mathbf{P}$, and define $\mathbf{P}_0(P) \equiv \{\tilde{P} \in \mathbf{P}_0 : \tilde{P}_Z = P_Z\}$. If $\mathbf{Q}$ is the set of all distributions for (R_o, R_t, Z) satisfying Assumptions 2.1–2.2 and $\mathbf{P}_0$ is nonempty, then it follows that $\mathbf{P}_0(P)$ is nonempty and convex for any $P \in \mathbf{P}$.*

PROOF. We first show $\mathbf{P}_0(P)$ is nonempty. Because $\mathbf{P}_0$ is nonempty, there exists $\hat{P} \in \mathbf{P}_0$ such that $\hat{P} = QT^{-1}$ for some Q that satisfies Assumptions 2.1–2.2, $Q\{R \in \mathcal{R}'\} > 0$, and $\theta(Q) = \theta_0$. By Assumption 2.1, $R \perp\!\!\!\perp Z$ under Q and therefore $Q = Q_R \times Q_Z$ where Q_R and Q_Z denote the marginal distributions of R and Z under Q . Next, define $\tilde{Q} = Q_R \times P_Z$ and note that $\tilde{Q} \in \mathbf{Q}$ because $Q \in \mathbf{Q}$, while $\theta(\tilde{Q}) = \theta(Q) = \theta_0$ and $\tilde{Q}\{R \in \mathcal{R}'\} = Q\{R \in \mathcal{R}'\} > 0$ because $Q_R = \tilde{Q}_R$. Set $\tilde{P} = \tilde{Q}T^{-1}$ and note that $\tilde{P} = \hat{P}_{YD|Z}P_Z$ and therefore $\tilde{P} \in \mathbf{P}$ by our supposition on $\mathbf{P}$ because $\hat{P}, P \in \mathbf{P}$. Hence, we obtain that $\tilde{P} \in \mathbf{P}_0(P)$, which shows that $\mathbf{P}_0(P)$ is nonempty.

Next, we show that $\mathbf{P}_0(P)$ is convex. To this end, fix arbitrary distributions $P^1, P^2 \in \mathbf{P}_0(P)$ and $\lambda \in [0, 1]$. By definition of $\mathbf{P}_0(P)$, there then exist Q^1 and Q^2 that satisfy Assumptions 2.1–2.2 and $P^j = Q^jT^{-1}$, $Q^j\{R \in \mathcal{R}'\} > 0$, $\theta(Q^j) = \theta_0$, and $Q_Z^j = P_Z$ for $j \in \{1, 2\}$. Defining $Q = \lambda Q^1 + (1 - \lambda)Q^2$, it then follows that

$$Q = \lambda Q_R^1 \times P_Z + (1 - \lambda)Q_R^2 \times P_Z = (\lambda Q_R^1 + (1 - \lambda)Q_R^2) \times P_Z$$

and hence that Q satisfies Assumption 2.1 with $Q_Z = P_Z$. In addition, for any set $\mathcal{A}$ we have $Q\{R \in \mathcal{A}\} = \lambda Q^1\{R \in \mathcal{A}\} + (1 - \lambda)Q^2\{R \in \mathcal{A}\}$, which implies $Q\{R \in \mathcal{R}\} = 1$ and $Q\{R \in \mathcal{R}'\} > 0$. Moreover, because $\theta(Q^1) = \theta(Q^2) = \theta_0$, we can also conclude that

$$\begin{aligned} \theta(Q) &= \frac{\lambda E_{Q^1}[g(R)I\{R \in \mathcal{R}'\}] + (1 - \lambda)E_{Q^2}[g(R)I\{R \in \mathcal{R}'\}]}{\lambda Q^1\{R \in \mathcal{R}'\} + (1 - \lambda)Q^2\{R \in \mathcal{R}'\}} \\ &= \frac{\lambda \theta_0 Q^1\{R \in \mathcal{R}'\} + (1 - \lambda)\theta_0 Q^2\{R \in \mathcal{R}'\}}{\lambda Q^1\{R \in \mathcal{R}'\} + (1 - \lambda)Q^2\{R \in \mathcal{R}'\}} = \theta_0. \end{aligned}$$

Since $\lambda P^1 + (1 - \lambda)P^2 = QT^{-1}$ and $\mathbf{P}$ is convex, it follows that $\lambda P^1 + (1 - \lambda)P^2 \in \mathbf{P}_0(P)$, and therefore that $\mathbf{P}_0(P)$ is convex. ■

Appendices C, D, E contain additional examples and details of the simulation study and the empirical application, and can be found at <https://github.com/edward-vytlacil/STRATA-ATE/blob/main/>.