Supplemental Appendix: "Treatment Effect Estimation with Noisy Conditioning Variables"

Kenichi Nagasawa

April 28, 2024

Contents

A Proofs

Notation Given a measure μ, $L^2(\mu)$ denotes the space of square μ-integrable functions. For a random variable A, F_A is the distribution of A. Equal signs involving random variables are understood as equality with probability one.

A.1 Proof of Theorem 1

Remark For the proof below, consider the linear operator $\Pi : L^2(F_{X^*}) \mapsto L^2(F_X)$, $\Pi(g)(x) = \mathbb{E}[g(X^*)|X = x]$. Assumption 5 implies that Π is compact (see e.g., Carrasco et al., 2007, p.5659). By Theorem 15.16 of Kress (2014), there exist non-negative singular values $\{\tau_j\}_{j\geq 1}$ and orthonormal sequences $\{\phi_j\}_{j\geq 1} \subset L^2(F_{X^*})$, $\{\varphi_j\}_{j\geq 1} \subset L^2(F_X)$ for the operator Π. Theorem 15.18 of Kress (2014) implies that for any function $g \in L^2(F_{X^*})$ in the orthogonal complement of the null space of Π, we have

$$g(x^*) = \sum_{j=1}^{\infty} \frac{\mathbb{E}[h(X)\varphi_j(X)]}{\tau_j}\phi_j(x^*), \qquad h(x) = \Pi(g)(x) \tag{S-1}$$

with $\sum_{j=1}^{\infty} \tau_j^{-2}\mathbb{E}[h(X)\varphi_j(X)]^2 < \infty$. □

By the law of iterated expectations and Assumption 2, for F_X-almost all x

$$\begin{aligned}
\frac{f_{X|DZ}(x|D,Z)}{f_X(x)} &= \int \frac{f_{X|X^*DZ}(x|x^*,D,Z)}{f_X(x)} f_{X^*|DZ}(x^*|D,Z)d\lambda_*(x^*) \\
&= \int \frac{f_{X|X^*}(x|x^*)}{f_X(x)} f_{X^*|DZ}(x^*|D,Z)d\lambda_*(x^*) \\
&= \Pi\left(\frac{f_{X^*|DZ}(\cdot|D,Z)}{f_{X^*}(\cdot)}\right)(x).
\end{aligned}$$

Since Assumptions 3 and 5 imply that $f_{X^*|DZ}(\cdot|d,z)/f_{X^*}(\cdot)$ lies in the orthogonal complement of the null space of Π, (S-1) implies

$$\begin{aligned}
\frac{f_{X^*|DZ}(x^*|D,Z)}{f_{X^*}(x^*)} &= \sum_{j=1}^{\infty} \tau_j^{-1} \int \frac{f_{X|DZ}(x|D,Z)}{f_X(x)} \varphi_j(x) f_X(x) d\lambda(x) \phi_j(x^*) \\
&= \sum_{j=1}^{\infty} \tau_j^{-1} \int f_{X|DZ}(x|D,Z)\varphi_j(x)d\lambda(x)\phi_j(x^*).
\end{aligned}$$

This equation implies, for any $y \in \mathbb{R}$,

$$
\begin{aligned}
&\Pr[Y(d) \le y | D, Z] \\
&= \int \Pr[Y(d) \le y | X^* = x^*, D, Z] f_{X^*|DZ}(x^*|D,Z) d\lambda_*(x^*) \\
&= \int \Pr[Y(d) \le y | X^* = x^*] f_{X^*|DZ}(x^*|D,Z) d\lambda_*(x^*) \\
&= \int \Pr[Y(d) \le y | X^* = x^*] \Big(\sum_{j=1}^{\infty} \tau_j^{-1} \int \varphi_j(x) f_{X|DZ}(x|D,Z) d\lambda(x) \phi_j(x^*) \Big) f_{X^*}(x^*) d\lambda_*(x^*)
\end{aligned}
$$

where the first equality is by the law of iterated expectations, and the second equality follows from Assumption 1. Lemma SA-1 below implies that the infinite sum and the outer integral are exchangeable, and thus,

$$
\begin{aligned}
\Pr[Y(d) \le y | D, Z] &= \sum_{j=1}^{\infty} \tau_j^{-1} \mathbb{E}[\Pr[Y(d) \le y | X^*] \phi_j(X^*)] \int \varphi_j(x) f_{X|DZ}(x|D,Z) d\lambda(x) \\
&\equiv \sum_{j=1}^{\infty} \frac{\varsigma_j(d)}{\tau_j} \int \varphi_j(x) \mathfrak{V}(x) d\lambda(x), \quad \varsigma_j(d) = \mathbb{E}[\Pr[Y(d) \le y | X^*] \phi_j(X^*)]
\end{aligned}
$$

where $\sum_{j=1}^{\infty} \tau_j^{-2} | \int \varphi_j(x) \mathfrak{V}(x) d\lambda(x)|^2 < \infty$ by (S-1). Then, $\Pr[Y(d) \le y | D, Z]$ is $\sigma(\mathfrak{V})$-measurable, and by $\sigma(\mathfrak{V}) \subseteq \sigma(D, Z)$, $\Pr[Y(d) \le y | D, Z] = \Pr[Y(d) \le y | \mathfrak{V}]$. Thus, $\Pr[Y(d) \le y | D, \mathfrak{V}] = \mathbb{E}[\Pr[Y(d) \le y | D, Z] | D, \mathfrak{V}] = \mathbb{E}[\Pr[Y(d) \le y | \mathfrak{V}] | D, \mathfrak{V}] = \Pr[Y(d) \le y | \mathfrak{V}]$, which is the desired conditional independence result.

To prove the second part of the theorem, first note $\mathbb{E}[Y | D = d, \mathfrak{V}] = \mathbb{E}[Y(d) | D = d, \mathfrak{V}] = \mathbb{E}[Y(d) | \mathfrak{V}]$ where the second equality holds by the conditional independence result above. Also, $\mathbb{E}[Y(d) | \mathfrak{V}] = \mathbb{E}[Y(d) | D, Z]$ by the argument above. Thus, $\mathbb{E}[Y | D = d, \mathfrak{V}] = \mathbb{E}[Y(d) | D, Z]$ and

$$
\mathbb{E}[Y(d)] = \mathbb{E}[\mathbb{E}[Y(d) | D, Z]] = \mathbb{E}[\mathbb{E}[Y | D = d, \mathfrak{V}]]
$$

provided that the last expectation is well-defined, which is equivalent to that the conditional support of $\mathfrak{V}$ given $D = d$ is equal to the marginal support of $\mathfrak{V}$.

Fix $v \in \text{supp}(\mathfrak{V})$, and by definition, for any open set $U \subset L^2(\lambda)$ containing v, $\mathbb{P}[\mathfrak{V} \in U] > 0$. By Bayes' rule,

$$
\mathbb{P}[\mathfrak{V} \in U | D = d] = \frac{\mathbb{P}[D = d | \mathfrak{V} \in U] \mathbb{P}[\mathfrak{V} \in U]}{\mathbb{P}[D = d]}
$$

and the right-hand side is strictly positive by $\mathbb{P}[\mathfrak{V} \in U] > 0$ and Assumption 4. Since U is an arbitrary open set, v belongs to the conditional support of $\mathfrak{V}$, and the desired result follows. $\square$

A.2 Proof of Theorem 2

Notations The underlying probability space is denoted by $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\Gamma(v) = \int G(x)f(x,v)d\lambda(x)$ and $\mathcal{V}_\psi$ be a probability one set on which $\frac{\partial}{\partial v^\intercal}\Gamma(v)$ is of column rank d_ψ. A diffeomorphism is a continuously differentiable bijective function whose inverse is also continuously differentiable. For a finite measure space $(\mathcal{X}, \mathcal{G}, \mathbb{P})$ where $\mathbb{P}\mathcal{X}$ is finite but may not equal one, I consider "conditional expectations" in the exactly same way as for probability spaces. That is, with $Y \in L_1$ and a sub sigma-field $\mathcal{H} \subset \mathcal{G}$, I define $\mathbb{E}[Y|\mathcal{H}]$ to be a $\mathcal{H}$-measurable random variable satisfying $\int_\mathcal{X} \mathbb{1}_H Y d\mathbb{P} = \int_\mathcal{X} \mathbb{1}_H \mathbb{E}[Y|\mathcal{H}]d\mathbb{P}$ for all $H \in \mathcal{H}$. $\square$

First note $\sigma(V) \subseteq \sigma(\mathfrak{V})$ since $V = \int G(x)\mathfrak{V}(x)d\lambda(x)$ is a measurable function of $\mathfrak{V}$.

Suppose that there is a partition $\{P_j \in \sigma(D,W,Z) : j \in \mathbb{N}\}$ of Ω such that $\mathfrak{V}$ is $\sigma(V)$-measurable on the measure space $(\Omega \cap P_j, \mathcal{F} \cap P_j, \mathbb{P}_j)$ where $\mathcal{F} \cap P_j = \{F \cap P_j : F \in \mathcal{F}\}$ and $\mathbb{P}_j(\cdot) = \mathbb{P}(\cdot \cap P_j)$. Then,

$$\mathbb{E}[Y|\mathfrak{V}, D, W] = \sum_{j\geq 1} \mathbb{E}[Y|\mathfrak{V}, D, W]\mathbb{1}_{P_j} = \sum_{j\geq 1} \mathbb{E}_j[Y|_{P_j}|\mathfrak{V}, D, W]\mathbb{1}_{P_j} = \sum_{j\geq 1} \mathbb{E}_j[Y|_{P_j}|V, D, W]\mathbb{1}_{P_j}$$

where the first equality follows from $1 = \sum_{j\geq 1} \mathbb{1}_{P_j}$ with at most one j having $\mathbb{1}_{P_j} > 0$, the second equality uses Lemma SA-2 below, and the third equality uses the fact that $\sigma(\mathfrak{V}) = \sigma(V)$ on $(\Omega \cap P_j, \mathcal{F} \cap P_j, \mathbb{P})$. Thus, to prove the theorem, it suffices to exhibit a partition $\{A_j \in \sigma(D,W,Z) : j \in \mathbb{N}\}$ such that $\mathfrak{V}$ is $\sigma(V)$-measurable on $(\Omega \cap A_j, \mathcal{F} \cap A_j, \mathbb{P})$.

Since $\frac{\partial}{\partial v^\intercal}\Gamma(v)$ is continuous on $\mathbb{R}^{d_\psi}$, for each $v \in \mathcal{V}_\psi$, there exists an open set containing v on which $\Gamma(v)$ is of rank d_ψ by Proposition 4.1 of Lee (2013). Let $I_v \subset \{1, \ldots, k\}$ be the set of indices whose corresponding rows of $\frac{\partial}{\partial v^\intercal}\Gamma(v)$ are linearly independent and $\pi_{I_v} : \mathbb{R}^k \to \mathbb{R}^{d_\psi}$ be the canonical projection of the corresponding elements. By the inverse function theorem, there is some open set $U_v \ni v$ on which $\pi_{I_v} \circ \Gamma$ is a diffeomorphism. The collection $\{U_v : v \in \mathcal{V}_\psi\}$ is an open cover of $\mathcal{V}_\psi$, and choose a countable open subcover $\{U_j : j \in \mathbb{N}\}$ of $\mathcal{V}_\psi$ to define $A_j = \{\omega \in \Omega : V_\psi(\omega) \in U_j \backslash \cup_{i=1}^{j-1} U_i\}$. Since $V_\psi = \psi(D,W,Z)$ and U_j is an open subset of $\mathbb{R}^{d_\psi}$, A_j is $\sigma(D,W,Z)$-measurable.

Now consider the measure space $(\Omega \cap A_j, \mathcal{F} \cap A_j, \mathbb{P}_j)$. As shown above, there exists a projection

π_j such that $\pi_j \circ \Gamma$ is a diffeomorphism on the open ball $U_j \subset \mathbb{R}^{d_\psi}$ and we have $V_\psi = (\pi_j \circ \Gamma)^{-1}(V)$ on A_j. Thus, V_ψ can be written as a measurable function of V, proving $\sigma(\mathfrak{V}) = \sigma(V)$ on $(\Omega \cap A_j, \mathcal{F} \cap A_j, \mathbb{P}_j)$. □

A.3 Auxiliary results

Lemma SA-1. *For a transformation $\mathscr{T}$, suppose $\mathbb{E}[\mathscr{T}(Y(d))^2] < \infty$. Under Assumptions 2, 3, and 5,*

$$\int \mathbb{E}[\mathscr{T}(Y(d))|X^* = x^*]\Big(\sum_{j=1}^{\infty} \tau_j^{-1} e_j(d,z)\phi_j(x^*)\Big) f_{X^*}(x^*) d\lambda_1(x^*) = \sum_{j=1}^{\infty} \tau_j^{-1} e_j(d,z)\varsigma_j(d)$$

where $e_j(t,z) = \int f_{X|DZ}(x|t,z)\varphi_j(x) d\lambda(x)$ and $\varsigma_j(d) = \mathbb{E}[\mathbb{E}[\mathscr{T}(Y(d))|X^*]\phi_j(X^*)]$.

Proof. Note that the infinite sum $\sum_{j=1}^{\infty} \tau_j^{-1} e_j(d,z)\phi_j$ converges in $L^2(F_{X^*})$ as discussed around (S-1). Since $\mathbb{E}[\mathscr{T}(Y(d))|X^*] \in L^2(F_{X^*})$,

$$\begin{aligned}
&\int \mathbb{E}[\mathscr{T}(Y(d))|X^* = x^*]\Big(\sum_{j=1}^{\infty} \tau_j^{-1} e_j(d,z)\phi_j(x^*)\Big) f_{X^*}(x^*) d\lambda(x^*) \\
&= \lim_{N\to\infty} \int \mathbb{E}[\mathscr{T}(Y(d))|X^* = x^*]\Big(\sum_{j=1}^{N} \tau_j^{-1} e_j(d,z)\phi_j(x^*)\Big) f_{X^*}(x^*) d\lambda(x^*).
\end{aligned}$$

For each $N \in \mathbb{N}$,

$$\int \mathbb{E}[\mathscr{T}(Y(d))|X^* = x^*]\Big(\sum_{j=1}^{N} \tau_j^{-1} e_j(d,z)\phi_j(x^*)\Big) f_{X^*}(x^*) d\lambda(x^*) = \sum_{j=1}^{N} \tau_j^{-1} e_j(d,z)\varsigma_j(d)$$

and this infinite sum converges as $\sum_{j=1}^{\infty} \tau_j^{-2} e_j(d,z)^2 < \infty$ and $\sum_{j=1}^{\infty} \varsigma_j(d)^2 < \infty$. The latter inequality follows from $\mathbb{E}[\mathscr{T}(Y(d))|X^*] \in L^2(F_{X^*})$ and $\{\phi_j\}_{j\geq 1}$ being an orthonormal sequence in $L^2(F_{X^*})$. Then, the desired conclusion follows. □

Lemma SA-2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, Y be a random variable with a finite expectation, $\mathcal{G} \subset \mathcal{F}$ be a sub sigma field, and $\tilde{\Omega} \in \mathcal{G}$ be a subset of Ω. Also, $Y|_{\tilde{\Omega}}$ is the restriction of Y to $\tilde{\Omega}$, $\tilde{\mathcal{G}} = \{G \cap \tilde{\Omega} : G \in \mathcal{G}\}$, and $\tilde{\mathcal{F}} = \{F \cap \tilde{\Omega} : F \in \mathcal{F}\}$. Define $\mathbb{E}[Y|_{\tilde{\Omega}}|\tilde{\mathcal{G}}]$ to be a $\tilde{\mathcal{G}}$-measurable real-valued* *function satisfying* $\int_{\tilde{\Omega}} \mathbb{1}_{\tilde{G}} Y|_{\tilde{\Omega}} d\mathbb{P} = \int_{\tilde{\Omega}} \mathbb{1}_{\tilde{G}} \mathbb{E}[Y|_{\tilde{\Omega}}|\tilde{\mathcal{G}}] d\mathbb{P}$ *for all* $\tilde{G} \in \tilde{\mathcal{G}}$. *Then,*

$$\mathbb{E}[Y|\mathcal{G}] = \mathbb{E}[Y|_{\tilde{\Omega}}|\tilde{\mathcal{G}}]$$

except for a null set in $\tilde{\mathcal{G}}$.

Proof. Since $Y = Y|_{\tilde{\Omega}}$ on $\tilde{\Omega}$, for any $\tilde{G} \in \tilde{\mathcal{G}}$,

$$\int_{\tilde{\Omega}} \mathbb{1}_{\tilde{G}} Y|_{\tilde{\Omega}} d\mathbb{P} = \int_{\Omega} \mathbb{1}_{\tilde{G}} Y d\mathbb{P} = \int_{\Omega} \mathbb{1}_{\tilde{G}} \mathbb{E}[Y|\mathcal{G}] d\mathbb{P} = \int_{\tilde{\Omega}} \mathbb{1}_{\tilde{G}} \mathbb{E}[Y|\mathcal{G}] d\mathbb{P}$$

where the second equality uses $\tilde{G} \in \mathcal{G}$ and the definition of a conditional expectation and the last equality uses $\mathbb{1}_{\tilde{G}} \mathbb{E}[Y|\mathcal{G}] = 0$ almost surely outside $\tilde{\Omega}$. Since $\tilde{G}$ was arbitrary, the definition implies the desired result. □

Lemma SA-3. *Suppose that Assumptions 1, 2, 3, 5 hold. Then,* $\mathbb{E}[Y|D, Z]$ *is* $(\mathfrak{V}, D)$*-measurable. In particular,* $\mathbb{E}[\varepsilon|D, Z] = 0$ *for* $\varepsilon = Y - \mathbb{E}[Y|\mathfrak{V}, D]$.

Proof. Using $Y = Y(1)D + Y(0)(1 - D)$ and $Y(d) \perp\!\!\!\perp (D, Z)|X^*$,

$$\begin{aligned}
\mathbb{E}[Y|D, Z] &= \mathbb{E}[\mathbb{E}[Y(1)|X^*, D, Z]D + \mathbb{E}[Y(0)|X^*, D, Z](1 - D)|D, Z] \\
&= \mathbb{E}[\mathbb{E}[Y(1)|X^*]D + \mathbb{E}[Y(0)|X^*](1 - D)|D, Z] \\
&= \sum_{d \in \{0,1\}} \mathbb{1}\{D = d\} \int \mathbb{E}[Y(d)|X^* = x^*] f_{X^*|DZ}(x^*|d, Z) d\lambda_*(x^*)
\end{aligned}$$

Arguing as in the proof of Theorem 1,

$$\mathbb{E}[Y|D, Z] = \sum_{d \in \{0,1\}} \mathbb{1}\{D = d\} \sum_{j=1}^{\infty} \tau_j^{-1} \mathbb{E}[\mathbb{E}[Y(d)|X^*]\phi_j(X^*)] \int f_{X|DZ}(x|d, Z)\varphi_j(x) d\lambda(x).$$

Letting $\varsigma_{y,j}(d) = \mathbb{E}[\mathbb{E}[Y(d)|X^*]\phi_j(X^*)]$,

$$\mathbb{E}[Y|D, Z] = \sum_{j=1}^{\infty} \tau_j^{-1} \varsigma_{y,j}(D) \int \mathfrak{V}(x)\varphi_j(x) d\lambda(x),$$

which proves the desired result. □

The following is a slight modification of a known sufficient condition for bounded completeness.

Lemma SA-4. *Let $X = \chi(X^* + \eta)$ where $X^* \perp\!\!\!\perp \eta$ and χ is invertible. Suppose X^* and η are continuously distributed, and the characteristic function of η is non-zero everywhere. Then, the family of distributions $\{f_{X^*|X}(\cdot|x) : x \in \mathcal{X}\}$, where $\mathcal{X}$ is the support of X, is bounded complete.*

Proof. For any bounded function b, $\mathbb{E}[b(X^*)|X] = \mathbb{E}[b(X^*)|X^* + \eta]$ almost surely because χ is invertible, implying $\sigma(X) = \sigma(X^* + \eta)$. Thus, without loss of generality, assume χ is the identity function. Since $f_{X^*X}(y, x) = f_{X^*}(y)f_\eta(x - y)$,

$$\mathbb{E}[b(X^*)|X] = \int b(y)f_{X^*}(y)f_\eta(X - y)dy = \int \tilde{b}(y)f_\eta(y - X)dy$$

where $\tilde{b}(y) = b(-y)f_{X^*}(-y)$. By Theorem 2.1 of Mattner (1993), $\int \tilde{b}(y)f_\eta(y - X)dy = 0$ almost surely implies $\tilde{b}(y) = 0$ almost surely. This in turn implies $b(y) = 0$ for y such that $f_{X^*}(y) > 0$. Thus, bounded completeness holds. □

A.4 Proof of Theorem 3

The proof builds on the arguments of Chernozhukov et al. (2022) (henceforth CNS). Write $\beta_0(v) = \mu_0(1, v) - \mu_0(0, v)$ and $\widehat{\beta}(v) = \widehat{\mu}(1, v) - \widehat{\mu}(0, v)$. Let

$$\begin{aligned}\varphi(\Xi, \mu, \nu, \alpha) &= \alpha(D, \nu(D, Z))\big[Y - \mu(D, \nu(D, Z))\big] \\ &\quad + \left[\frac{\partial[\mu(1, v) - \mu(0, v)]}{\partial v^\intercal} - \alpha(D, \nu(D, Z))\frac{\partial\mu(D, v)}{\partial v^\intercal}\right]\Big|_{v=\nu(D,Z)}[G(X) - \nu(D, Z)].\end{aligned}$$

Note $\psi(\Xi, \mu, \nu, \alpha) = m(\Xi, \mu, \nu) + \varphi(\Xi, \mu, \nu, \alpha)$ and $m(\Xi, \mu, \nu) = \mu(1, \nu(D, Z)) - \mu(0, \mu(D, Z))$. I use the decomposition

$$\begin{aligned}\psi(\Xi, \widehat{\mu}_l, \widehat{\nu}_l, \widehat{\alpha}_l) = \psi(\Xi, \mu_0.\nu_0, \alpha_0) &\\ &+ \widehat{\beta}_l(\widehat{\nu}_l(D, Z)) - \beta_0(V) \\ &+ \varphi(\Xi, \widehat{\mu}_l, \widehat{\nu}_l, \alpha_0) - \varphi(\Xi, \mu_0, \nu_0, \alpha_0) \\ &+ \varphi(\Xi, \mu_0, \nu_0, \widehat{\alpha}_l) - \varphi(\Xi, \mu_0, \nu_0, \alpha_0) \\ &+ \varphi(\Xi, \widehat{\mu}_l, \widehat{\nu}_l, \widehat{\alpha}_l) - \varphi(\Xi, \widehat{\mu}_l, \widehat{\nu}_l, \alpha_0) - \varphi(\Xi, \mu_0, \nu_0, \widehat{\alpha}_l) + \varphi(\Xi, \mu_0, \nu_0, \alpha_0) \\ \equiv \psi(\Xi, \mu_0, \nu_0, \alpha_0) + R_{1l} + R_{2l} + R_{3l} + R_{4l}.&\end{aligned}$$

By the hypothesis, $\chi_{1,n}\omega_n^{2\zeta/(2\zeta+1)}\tilde{\kappa}_n + \chi_{n,2}\tilde{\omega}_n = o(1)$ and $\omega_n^{2\zeta/(2\zeta+1)}\tilde{\kappa}_n + (\chi_{0,n}\chi_{2,n})^{1/2}\tilde{\omega}_n = o(n^{-1/4})$. Then, Lemmas SA-5, SA-6, SA-7 imply $\sqrt{n}\mathbb{E}[R_{1l} + R_{2l}|I_l^c] = o_{\mathbb{P}}(1)$, $\mathbb{E}[R_{3l}|I_l^c] = 0$, $\sqrt{n}\mathbb{E}[R_{4l}|I_l^c] = o_{\mathbb{P}}(1)$, and $\mathbb{E}[R_{jl}^2|I_l^c] = o_{\mathbb{P}}(1)$ for $j = 1,2,3,4$ since n_l/n's are bounded away from zero. Then,

$$\frac{1}{\sqrt{n}}\sum_{l=1}^{L}\sum_{i\in I_l}\psi(\Xi_i, \widehat{\mu}_l, \widehat{\nu}_l, \widehat{\alpha}_l) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(\Xi_i, \mu_0, \nu_0, \alpha_0) + o_{\mathbb{P}}(1).$$

For consistency of the variance estimator, note

$$\begin{aligned}\big[\psi(\Xi, \widehat{\mu}_l, \widehat{\nu}_l, \widehat{\alpha}_l) - \widehat{\theta}_n\big]^2 = \big[\psi(\Xi, \mu_0, \nu_0, \alpha_0) - \theta_0\big]^2 + \big[\psi(\Xi, \widehat{\mu}_l, \widehat{\nu}_l, \widehat{\alpha}_l) - \psi(\Xi, \mu_0, \nu_0, \alpha_0)\big]^2 \\ + \big[\widehat{\theta}_n - \theta_0\big]^2 - 2\big[\psi(\Xi, \widehat{\mu}_l, \widehat{\nu}_l, \widehat{\alpha}_l) - \psi(\Xi, \mu_0, \nu_0, \alpha_0)\big]\big[\widehat{\theta}_n - \theta_0\big] \\ + 2\big[\psi(\Xi, \widehat{\mu}_l, \widehat{\nu}_l, \widehat{\alpha}_l) - \psi(\Xi, \mu_0, \nu_0, \alpha_0)\big]\big[\psi(\Xi, \mu_0, \nu_0, \alpha_0) - \theta_0\big] \\ - 2\big[\widehat{\theta}_n - \theta_0\big]\big[\psi(\Xi, \mu_0, \nu_0, \alpha_0) - \theta_0\big]\end{aligned}$$

and by the above argument and $\widehat{\theta}_n = \theta_0 + o_{\mathbb{P}}(1)$,

$$\widehat{\Psi}_n = \frac{1}{n}\sum_{i=1}^{n}\big[\psi(\Xi, \mu_0, \nu_0, \alpha_0) - \theta_0\big]^2 + o_{\mathbb{P}}(1) = \Psi_0 + o_{\mathbb{P}}(1).$$

□

A.4.1 Auxiliary lemmas

For reference later, I state a condition:

Condition SA-1. $\chi_{\ell,n} \le \chi_{\ell+1,n}$ *for* $\ell = 0,1$, $\log(K_n)/\log(n) = O(1)$, *and* $\chi_{1,n}(\kappa_n/\omega_n)\omega_n^{2\zeta/(2\zeta+1)} + \tilde{\omega}_n\chi_{2,n} + \chi_{0,n}/(n^{1/4}\log K_n) + \chi_{0,n}\omega_n^{(2\zeta-1)/(2\zeta+1)} = o(1)$.

Notations Write $\eta = G(X) - V$ where $V = \mathbb{E}[G(X)|D,Z]$. For sequences a_n, b_n, $a_n \lesssim b_n$ indicates that there exists a fixed constant $C > 0$ such that $a_n \le Cb_n$ where C does not dependent on n. For a function f, $\partial f(D,V) = \partial f(D,V)/\partial V^{\mathsf{T}}$. For $\ell = 1,2$, $\|\cdot\|_\ell$ is the ℓ-norm in Euclidean spaces. $\|\cdot\|_{L^2}$ is the $L^2(\mathbb{P})$-norm.

Using cross-fitting, I compute expectations conditional on observations not in I_l. Thus, I treat $(\widehat{\mu}_l, \widehat{\nu}_l, \widehat{\alpha}_l)$ as if fixed when computing expectations. To save space, I write $\mathbb{E}[\cdot|I_l^c] = \mathbb{E}_l[\cdot]$ and drop the subscript l from $\widehat{\mu}_l, \widehat{\nu}_l, \widehat{\alpha}_l, R_{jl}$ when the risk of confusion is small. $\square$

Lemma SA-5. *Under Assumptions 7-10 and Condition SA-1,*

$$\mathbb{E}_l[R_1 + R_2] = O_{\mathbb{P}}\left(\tilde{\omega}_n(\kappa_n/\omega_n)\omega_n^{2\zeta/(2\zeta+1)} + \chi_{2,n}\tilde{\omega}_n^2\right),\ \mathbb{E}_l[R_1^2] = o_{\mathbb{P}}(1),\ \textit{and}\ \ \mathbb{E}_l[R_2^2] = o_{\mathbb{P}}(1).$$

Proof. Since $R_{1l} = \widehat{\beta}_l(\widehat{\nu}_l(D, Z)) - \beta_0(V)$, Lemma SA-8 implies $\mathbb{E}_l[R_1^2] = o_{\mathbb{P}}(1)$. For R_2,

$$\begin{aligned}
R_2 &= \alpha_0(D, \widehat{V})[Y - \widehat{\mu}(D, \widehat{V})] - \alpha_0(D, V)\varepsilon \\
&\quad + [\partial\widehat{\beta}(\widehat{V}) - \alpha_0(D, \widehat{V})\partial\widehat{\mu}(D, \widehat{V})][G(X) - \widehat{V}] - [\partial\beta_0(V) - \alpha_0(D, V)\partial\mu_0(D, V)]\eta \\
&= [\alpha_0(D, \widehat{V}) - \alpha_0(D, V)]\varepsilon + \alpha_0(D, \widehat{V})[\mu_0(D, V) - \widehat{\mu}(D, \widehat{V})] \\
&\quad + [\partial\widehat{\beta}(\widehat{V}) - \partial\beta_0(V)]\eta - [\alpha_0(D, \widehat{V}) - \alpha_0(D, V)]\partial\widehat{\mu}(D, \widehat{V})\eta \\
&\quad + [\partial\widehat{\beta}(\widehat{V}) - \alpha_0(D, \widehat{V})\partial\widehat{\mu}(D, \widehat{V})][V - \widehat{V}] - \alpha_0(D, V)[\partial\widehat{\mu}(D, \widehat{V}) - \partial\mu_0(D, V)]\eta.
\end{aligned}$$

By the assumption, $\mathbb{E}[\varepsilon^2|D, Z]$, α_0, $\partial\mu_0$, $\partial\beta_0$ and $\mathbb{E}[\|\eta\|_2^2|D, Z]$ are all bounded and α_0 is Lipschitz continuous. Thus,

$$\begin{aligned}
\mathbb{E}_l[R_2^2] &\lesssim \|\widehat{\nu} - \nu_0\|_{L^2}^2 + \int |\widehat{\mu}(d, \widehat{\nu}(d, z)) - \mu_0(d, \nu_0(d, z))|^2 dF_{DZ}(d, z) \\
&\quad + \int \|\partial\{\widehat{\beta}(\widehat{\nu}(d, z)) - \beta_0(\nu_0(d, z))\}\|_2^2 dF_{DZ}(d, z) \\
&\quad + \chi_{1,n}^2\|\widehat{\nu} - \nu_0\|_{L^2}^2 + \int \|\partial\{\widehat{\mu}(d, \widehat{\nu}(d, z)) - \mu_0(d, \nu_0(d, z))\}\|_2^2 dF_{DZ}(d, z) = o_{\mathbb{P}}(1)
\end{aligned}$$

where I used Lemmas SA-8 and SA-18.

For $R_1 + R_2$, note $\mathbb{E}[\varepsilon|D, Z] = 0$ by Lemma SA-3. Then,

$$\begin{aligned}
\mathbb{E}_l[R_1 + R_2] &= \mathbb{E}_l[-\beta_0(V) + \alpha_0(D, V)\mu_0(D, V)] + \mathbb{E}_l[\widehat{\beta}(V) - \alpha_0(D, V)\widehat{\mu}(D, V)] && \text{(S-2)} \\
&\quad + \mathbb{E}_l[\{\alpha_0(D, \widehat{V}) - \alpha_0(D, V)\}\{\mu_0(D, V) - \widehat{\mu}(D, \widehat{V})\}] && \text{(S-3)} \\
&\quad - \mathbb{E}_l[\{\alpha_0(D, \widehat{V}) - \alpha_0(D, V)\}\partial\widehat{\mu}(D, \widehat{V})\{V - \widehat{V}\}] && \text{(S-4)} \\
&\quad + \mathbb{E}_l[\alpha_0(D, V)\{\widehat{\mu}(D, V) - \widehat{\mu}(D, \widehat{V}) - \partial\widehat{\mu}(D, \widehat{V})(V - \widehat{V})\}] && \text{(S-5)} \\
&\quad - \mathbb{E}_l[\widehat{\beta}(V) - \widehat{\beta}(\widehat{V}) - \partial\widehat{\beta}(\widehat{V})\{V - \widehat{V}\}]. && \text{(S-6)}
\end{aligned}$$

The right-hand side of (S-2) equals zero as $\mathbb{E}[\alpha_0(D,V)\mu(D,V)] = \mathbb{E}[m(\Xi,\mu,\nu_0)]$ for any $\mu \in L^2(F_{DV})$; $\mathbb{E}[D\mu(D,V)/\mathbb{P}[D=1|V]] = \mathbb{E}[\mu(1,V)]$ and $\mathbb{E}[(1-D)\mu(D,V)/\mathbb{P}[D=0|V]] = \mathbb{E}[\mu(0,V)]$.

For (S-3), the Cauchy-Schwartz inequality and Lemma SA-8 imply it is $O_{\mathbb{P}}(\tilde{\omega}_n(\kappa_n/\omega_n)\omega_n^{2\xi/(2\xi+1)})$.

For (S-4), Lemma SA-18 implies it is $O_{\mathbb{P}}(\chi_{1,n}\tilde{\omega}_n^2)$.

For (S-5), write $d\tilde{F}(t,z) = \alpha_0(t,\nu_0(t,z))dF_{DZ}(t,z)$. By Taylor expansion, for $\bar{s} \in [0,1]$ (dependent on (t,z)),

$$\text{(S-5)} = \int (\nu_0(t,z) - \widehat{\nu}(t,z))' \frac{\partial^2 \widehat{\rho}' q(t, \bar{s}\nu_0(t,z) + (1-\bar{s})\widehat{\nu}(t,z))}{\partial v \partial v^\mathsf{T}} (\nu_0(t,z) - \widehat{\nu}(t,z)) d\tilde{F}(t,z)$$

$$\lesssim \max_{1 \le l,l' \le \dim(V)} \sup_{t,v} |\widehat{\rho}' \partial^2 q(t,v)/\partial v_l \partial v_{l'}| \|\widehat{\nu} - \nu_0\|_{L^2}^2 = O_{\mathbb{P}}(\chi_{2,n}\tilde{\omega}_n^2).$$

An analogous argument shows (S-6) $= O_{\mathbb{P}}(\chi_{2,n}\tilde{\omega}_n^2)$. □

Lemma SA-6. *Under Assumptions 7-10 and Condition SA-1,* $\mathbb{E}_l[R_3] = 0$ *and* $\mathbb{E}_l[R_3^2] = o_{\mathbb{P}}(1)$.

Proof. Recall

$$R_3 = \widehat{\alpha}(D,V)\varepsilon + [\partial\beta_0(V) - \widehat{\alpha}(D,V)\partial\mu_0(D,V)]\eta$$
$$- \alpha_0(D,V)\varepsilon + [\partial\beta_0(V) - \alpha_0(D,V)\partial\mu_0(D,V)]\eta.$$

Using $\mathbb{E}[\varepsilon|D,V] = 0$ and $\mathbb{E}[\eta|D,V] = 0$, $\mathbb{E}_l[R_3] = 0$ almost surely. Also,

$$R_3^2 \lesssim |\widehat{\alpha}(D,V) - \alpha(D,V)|^2\varepsilon^2 + |\widehat{\alpha}(D,V) - \alpha_0(D,V)|^2 |\partial\mu_0(D,V)\eta|^2$$

and $\mathbb{E}_l[R_3^2] = o_{\mathbb{P}}(1)$ follows from $\|\widehat{\alpha} - \alpha_0\|_{L^2} = o_{\mathbb{P}}(1)$ and boundedness of $\mathbb{E}[\varepsilon^2|D,Z]$, $\partial\mu_0$, and $\mathbb{E}[\|\eta\|_2^2|D,Z]$. □

Lemma SA-7. *Under Assumptions 7-10 and Condition SA-1,*

$$\mathbb{E}_l[R_4] = O_{\mathbb{P}}(\chi_{0,n}\chi_{2,n}\tilde{\omega}_n^2 + (\kappa_n/\omega_n)^2\omega_n^{4\zeta/(2\zeta+1)}), \quad \textit{and} \quad \mathbb{E}_l[R_4^2] = o_{\mathbb{P}}(1).$$

Proof. Note

$$R_4 = [\widehat{\alpha}(D,\widehat{V}) - \alpha_0(D,\widehat{V})][Y - \widehat{\mu}(D,\widehat{V})] - [\widehat{\alpha}(D,\widehat{V}) - \alpha_0(D,\widehat{V})]\partial\widehat{\mu}(D,\widehat{V})[G(X) - \widehat{V}] - R_3.$$

By Lemma SA-18, $\sup_{d,v}|\widehat{\mu}(d,v)| + \sup_{d,v}|\widehat{\alpha}(d,v)| = O_{\mathbb{P}}(\chi_{0,n})$. Then,

$$\mathbb{E}_l[(R_4+R_3)^2] \lesssim \int |\widehat{\alpha}(t,\widehat{\nu}(t,z)) - \alpha_0(t,\nu_0(t,z))|^2 dF_{DZ}(t,z)[1+\chi_{0,n}+\chi_{1,n}]^2 + \|\widehat{\nu}-\nu_0\|_{L^2}^2[1+\chi_{0,n}+\chi_{1,n}]^2.$$

Thus, from Lemma SA-6, $\mathbb{E}_l[R_4^2] = o_{\mathbb{P}}(1)$ holds. For the expectation, note $\mathbb{E}_l[R_3]=0$ and

$$\begin{aligned}
\mathbb{E}_l[R_4] &= \mathbb{E}_l\big[\{\widehat{\alpha}(D,\widehat{V}) - \alpha_0(D,\widehat{V})\}\{\widehat{\mu}(D,V) - \widehat{\mu}(D,\widehat{V}) - \partial\widehat{\mu}(D,\widehat{V})(V-\widehat{V})\}\big] \\
&\quad + \mathbb{E}_l\big[\{\widehat{\alpha}(D,\widehat{V}) - \alpha_0(D,\widehat{V})\}\{\mu_0(D,V) - \widehat{\mu}(D,V)\}\big] \\
&\leq \big(\sup_{d,v}|\widehat{\alpha}(d,v)| + C\big)\mathbb{E}_l\big[|\widehat{\mu}(D,V) - \widehat{\mu}(D,\widehat{V}) - \partial\widehat{\mu}(D,\widehat{V})(V-\widehat{V})|\big] \\
&\quad + \sqrt{\mathbb{E}_l\big[\{\widehat{\alpha}(D,\widehat{V}) - \alpha_0(D,\widehat{V})\}^2\big]\mathbb{E}_l\big[\{\mu_0(D,V) - \widehat{\mu}(D,V)\}^2\big]}
\end{aligned}$$

where the first equality uses $\mathbb{E}[\varepsilon|D,Z]=0$ and $\mathbb{E}[\eta|D,Z]=0$. Using $\sup_{d,v}|\widehat{\alpha}(d,v)| = O_{\mathbb{P}}(\chi_{0,n})$, calculations similar to those in Lemma SA-5 imply $\mathbb{E}_l[R_4] = O_{\mathbb{P}}(\chi_{0,n}\chi_{2,n}\tilde{\omega}_n^2 + (\kappa_n/\omega_n)^2\omega_n^{4\zeta/(2\zeta+1)})$.

□

Convergence rates of Lasso estimators

Lemma SA-8. *Suppose Assumptions 7-10 and SA-1 hold. For $s\in\{0,1\}$,*

$$\int \big|\widehat{\mu}(s,\widehat{\nu}(t,z)) - \mu_0(s,\nu_0(t,z))\big|^2 dF_{DZ}(t,z) = O_{\mathbb{P}}\big((\kappa_n/\omega_n)^2\omega_n^{4\zeta/(2\zeta+1)}\big),$$

$$\int \big|\widehat{\mu}(t,\widehat{\nu}(t,z)) - \mu_0(t,\nu_0(t,z))\big|^2 dF_{DZ}(t,z) = O_{\mathbb{P}}\big((\kappa_n/\omega_n)^2\omega_n^{4\zeta/(2\zeta+1)}\big),$$

$$\int \big|\widehat{\alpha}(t,\widehat{\nu}(t,z)) - \alpha_0(t,\nu_0(t,z))\big|^2 dF_{DZ}(t,z) = O_{\mathbb{P}}\big((\kappa_n/\omega_n)^2\omega_n^{4\zeta/(2\zeta+1)}\big).$$

$$\int \Big\|\frac{\partial}{\partial v^{\mathsf{T}}}\{\widehat{\mu}(s,\widehat{\nu}(t,z)) - \mu_0(s,\nu_0(t,z))\}\Big\|_2 dF_{DZ}(t,z) = o_{\mathbb{P}}(1).$$

□

Proof. First define some notations. Let $s_0 = \lfloor C\omega_n^{-2/(2\zeta+1)}\rfloor$ and let $\widetilde{\rho}$ be a vector as defined in the assumption 10 for $s = s_0$. Then,

$$\|\bar{\rho} - \widetilde{\rho}\|_2^2 \leq C s_0^{-2\zeta} \leq C \omega_n^{2\zeta/(2\zeta+1)}. \tag{S-7}$$

Let J_0 be the indices of nonzero elements of $\widetilde{\rho}$. Note $|J_0| = s_0$. Also, define

$$\rho_* \in \arg\min_{\rho} \left\{ (\rho - \bar{\rho})^{\mathsf{T}} \Omega (\rho - \bar{\rho}) + 2\omega_n \sum_{j \in \{1,\ldots,2K_n\} \setminus J_0} |\rho_j| \right\}$$

and J_* to be the set of indices of nonzero elements of ρ_*.

Below I prove the first display in the statement of the lemma. The second and the third displays follow from identical arguments. By adding and subtracting $\widehat{\mu}(s, \nu_0(t,z)) = q(s, \nu_0(t,z))^{\mathsf{T}} \widehat{\rho}$,

$$\begin{aligned} &\int \left| \widehat{\mu}(s, \widehat{\nu}(t,z)) - \mu_0(s, \nu_0(t,z)) \right|^2 dF_{DZ}(t,z) \\ &\leq 2 \int \left| \widehat{\mu}(s, \widehat{\nu}(t,z)) - \widehat{\mu}(s, \nu_0(t,z)) \right|^2 dF_{DZ}(t,z) + 2 \int \left| \widehat{\mu}(s,v) - \mu_0(s,v) \right|^2 dF_V(v) \end{aligned}$$

and for the first term after the inequality, Assumption 8 and Lemma SA-18 below imply

$$\begin{aligned} &\int \left(\{ q(s, \widehat{\nu}(t,z)) - q(s, \nu_0(t,z)) \}^{\mathsf{T}} \widehat{\rho} \right)^2 dF_{DZ}(t,z) \\ &\leq \sup_v \left\| \widehat{\rho}^{\mathsf{T}} \frac{\partial q(s,v)}{\partial v^{\mathsf{T}}} \right\|_2^2 \int \left\| \widehat{\nu}(d,z) - \nu_0(d,z) \right\|_2^2 dF_{DZ}(d,z) = O_{\mathbb{P}}\left(\omega_n^2\right). \end{aligned}$$

By the triangle inequality,

$$\|\widehat{\mu}(s,\cdot) - \mu_0(s,\cdot)\|_{L^2} \leq \|\widehat{\mu}(s,\cdot) - \rho_*^{\mathsf{T}} q(s,\cdot)\|_{L^2} + \|\rho_*^{\mathsf{T}} q(s,\cdot) - \bar{\rho}^{\mathsf{T}} q(s,\cdot)\|_{L^2} + \|\bar{\rho}^{\mathsf{T}} q(s,\cdot) - \mu_0(s,\cdot)\|_{L^2}$$

where $\|\bar{\rho}^{\mathsf{T}} q(s,\cdot) - \mu_0(s,\cdot)\|_{L^2} = o(n^{-1/4})$ by Assumption 10. Let

$$\tilde{\Omega}(s) = \begin{bmatrix} 1 & s \\ s & s \end{bmatrix} \otimes \mathbb{E}[p(V)p(V)^{\mathsf{T}}].$$

Since the largest eigenvalue of $\mathbb{E}[p(V)p(V)^{\top}]$ is uniformly bounded,

$$\|q(s,\cdot)^{\top}(\widehat{\rho}-\rho_*)\|_{L^2}^2 = (\widehat{\rho}-\rho_*)^{\top}\tilde{\Omega}(s)(\widehat{\rho}-\rho_*) \le C\|\widehat{\rho}-\rho_*\|_2^2 = O_{\mathbb{P}}\big((\kappa_n/\omega_n)^2\omega_n^{4\zeta/(2\zeta+1)}\big)$$

where the last equality uses Lemma SA-17 below. Also, Lemmas SA-12 and SA-17 below imply $\|q(s,\cdot)^{\top}(\rho_*-\bar{\rho})\|_{L^2}^2 \le C\omega_n^{2\zeta/(2\zeta+1)}$. Then,

$$\|\widehat{\mu}-\mu_0\|_{L^2} = O_{\mathbb{P}}\Big((\kappa_n/\omega_n)\omega_n^{2\zeta/(2\zeta+1)}\Big) + o\big(n^{-1/4}\big).$$

□

Lemmas for Lasso estimators I use some of lemmas in CNS and Bradic et al. (2022). For ease of reference, I collect them here.

Lemma SA-9 (Lemma A2 CNS)**.** *Under Assumptions 9-10,* $(\bar{\rho}-\rho_*)^{\top}\Omega(\bar{\rho}-\rho_*) \le C\omega_n^{4\zeta/(2\zeta+1)}$.

Lemma SA-10 (Lemma A3 CNS)**.** *Under Assumptions 9-10,* $\#|J_*| \le C\omega_n^{-2/(2\zeta+1)}$.

Lemma SA-11 (Lemma C1 Bradic et al. (2022))**.** *Consider* $a \in \mathbb{R}^p$ *such that* $\|a-b_s\|_2 \le Cs^{-r}$ *for any* $s \ge 0$*, where* $C, r > 0$ *are constants and* $b_s = \arg\min_{\|v\|_0 \le s}\|a-v\|_2$ *where* $\|\cdot\|_0$ *is the number of non-zero elements of the argument. If* $r > 1/2$ *and* $s \ge 2$*, then* $\|a-b_s\|_1 \le Ds^{1/2-r}$ *where* $D > 0$ *is a constant depending only on* C *and* r*.*

I establish additional lemmas to characterize the convergence rates of the Lasso estimators.

Lemma SA-12. *Under Assumptions 9-10,* $\|\rho_*-\widetilde{\rho}\|_2 \le C\omega_n^{2\zeta/(2\zeta+1)}$ *and* $\|\rho_*-\bar{\rho}\|_2 \le C\omega_n^{2\zeta/(2\zeta+1)}$.

Proof. Let $J_1 = J_0 \cup J_*$ and note that $\|(\rho_*)_{J_1}\|_2 = \|(\rho_*)\|_2$, $\|\bar{\rho}_{J_1}\|_2 = \|\bar{\rho}\|_2$. Then, $0 = \|(\rho_*-\widetilde{\rho})_{J_1^c}\|_2 \le 3\|(\rho_*-\widetilde{\rho})_{J_1}\|_2$ trivially holds. Lemma SA-10 implies $\#|J_1| \le C\omega_n^{-2/(2\zeta+1)}$. By Assumption 9,

$$\|\rho_*-\widetilde{\rho}\|_2^2 \le C(\rho_*-\widetilde{\rho})'\Omega(\rho_*-\widetilde{\rho}) \le 2C(\rho_*-\bar{\rho})'\Omega(\rho_*-\bar{\rho}) + 2C(\widetilde{\rho}-\bar{\rho})'\Omega(\widetilde{\rho}-\bar{\rho})$$

and the last two terms can be bounded by $C\omega_n^{4\zeta/(2\zeta+1)}$ by Lemma SA-9 and (S-7).

For the other conclusion, $\|\rho_*-\bar{\rho}\|_2 \le \|\rho_*-\widetilde{\rho}\|_2 + \|\widetilde{\rho}-\bar{\rho}\|_2 \le C\omega_n^{2\zeta/(2\zeta+1)}$. □

Lemma SA-13. *If Assumptions 9-10 hold, then* $\|\rho_*-\bar{\rho}\|_1 \le C\omega_n^{(2\zeta-1)/(2\zeta+1)}$.

Proof. By $\|\bar{\rho} - \widetilde{\rho}\|_2 \leq Cs^{-\zeta}$ and using the argument of Lemma SA-11,

$$\|\bar{\rho}_{J_0^c}\|_1 \leq Cs_0^{1/2-\zeta} \leq C\omega_n^{(2\zeta-1)/(2\zeta+1)}.$$

Let $J_1 = J_0 \cup J_*$ and note that $J_1^c \subset J_*^c$ and $J_1^c \subset J_0^c$. Then,

$$\|(\rho_*)_{J_1^c} - \bar{\rho}_{J_1^c}\|_1 = \|\bar{\rho}_{J_1^c}\|_1 \leq \|\bar{\rho}_{J_0^c}\|_1.$$

By Lemma SA-10, $\#|J_1| \leq \#|J_*| + \#|J_0| \leq C\omega_n^{-2/(2\zeta+1)} + s_0 \leq C\omega_n^{-2/(2\zeta+1)}$. Therefore,

$$\begin{aligned}
\|\rho_* - \bar{\rho}\|_1 &= \|(\rho_*)_{J_1} - \bar{\rho}_{J_1}\|_1 + \|(\rho_*)_{J_1^c} - \bar{\rho}_{J_1^c}\|_1 \leq \|(\rho_*)_{J_1} - \bar{\rho}_{J_1}\|_1 + \|\bar{\rho}_{J_0^c}\|_1 \\
&\leq \sqrt{\#|J_1|}\|(\rho_*)_{J_1} - \bar{\rho}_{J_1}\|_2 + C\omega_n^{(2\zeta-1)/(2\zeta+1)} \\
&\leq C\omega_n^{-1/(2\zeta+1)}\|\rho_* - \bar{\rho}\|_2 + C\omega_n^{(2\zeta-1)/(2\zeta+1)} \\
&\leq C\omega_n^{(2\zeta-1)/(2\zeta+1)}
\end{aligned}$$

where the last inequality follows from Lemma SA-12. □

Lemma SA-14. *If Assumptions 9-10 hold, $\|\bar{\rho}\|_1$ and $\|\rho_*\|_1$ are uniformly bounded.*

Proof. If $\max_{1\leq l\leq 2K_n} |\bar{\rho}_l| \leq C$ holds, then the conclusion follows by Lemma SA-11 since $\|\bar{\rho}\|_1 \leq \|\bar{\rho} - b_2\|_1 + \|b_2\|_1 \leq C$ where b_s is as defined in Lemma SA-11 and b_2 has only two non-zero elements which are the two largest elements of $\bar{\rho}$.

Now to show $\max_{1\leq l\leq 2K_n} |\bar{\rho}_l| \leq C$, note $\|(\rho_*)_{J_*}\|_2 = \|\rho_*\|_2$, and thus, $\|(\rho_*)_{J_*^c}\|_2 \leq 3\|(\rho_*)_{J_*}\|_2$. Using Assumption 10,

$$\rho_*^\mathsf{T}\rho_* \leq C\rho_*^\mathsf{T}\Omega\rho_* \leq C$$

where the last equality holds because $\mathbb{E}[|q(D,V)^\mathsf{T}(\bar{\rho} - \rho_*)|^2] = o(1)$ by Lemma SA-9, and by $\bar{\rho}$ being coefficients of least squares projection, $\mathbb{E}[|q(D,V)^\mathsf{T}\bar{\rho}|^2] \leq \mathbb{E}[\mu_0(D,V)^2] < \infty$. Since L^2 norm of ρ_* is uniformly bounded, $\max_{1\leq l\leq 2K_n} |\rho_{*l}| \leq C$. By Lemma SA-13, $\|\rho_* - \bar{\rho}\|_1 = o(1)$ and $\max_{1\leq l\leq 2K_n} |\bar{\rho}_l| \leq C$ follows.

By the triangle inequality, $\|\rho_*\|_1 \leq \|\rho_* - \bar{\rho}\|_1 + \|\bar{\rho}\|_1 \leq C$ also holds. □

Lemma SA-15. *Assumptions 7, 9, and 10 hold. If* $\chi_{0,n}\omega_n^{(2\zeta-1)/(2\zeta+1)} = o(1)$*, then* $|q(t,v)^\mathsf{T}\rho_*|$ *is uniformly bounded.*

Proof. By Lemma SA-13, $\sup_v |q(t,v)^\mathsf{T}\rho_* - q(t,v)^\mathsf{T}\bar{\rho}| \leq C\chi_{0,n}\omega_n^{(2\zeta-1)/(2\zeta+1)} = o(1)$, and by Assumption 10, $\sup_v |\mu_0(t,v) - q(t,v)^\mathsf{T}\bar{\rho}| = o(1)$. Since μ_0 is bounded by the hypothesis, the desired result holds. □

Given a vector $A \in \mathbb{R}^L$, let $\|A\|_\infty = \max_{1\leq l\leq L}|a_l|$. Also, define $M^\mu = \mathbb{E}[q(D,Z)Y]$.

Lemma SA-16. *Suppose Assumptions 7-10 hold. If* $\log(K_n)/\log(n) = O(1)$, $\chi_{0,n} \leq \chi_{1,n}$*, and* $\chi_{0,n}/(n^{1/4}\log K_n) + \chi_{0,n}\omega_n^{(2\zeta-1)/(2\zeta+1)} = o(1)$*, then*

$$\|\widehat{\Omega}\rho_* - \Omega\rho_*\|_\infty = O_\mathbb{P}(\omega_n), \qquad \|\widehat{M}_n^\mu - M^\mu\|_\infty = O_\mathbb{P}(\omega_n).$$

Proof. I first prove $\|\widehat{M}_n^\mu - M^\mu\|_\infty = O_\mathbb{P}(\omega_n)$. Let $\bar{M}_n^\mu = n^{-1}\sum_{i=1}^n q(D_i,V_i)Y_i$. With $X = q(D,V)$, $Y = Y$, and $s = 4$, Lemma SA-19 implies

$$\|\bar{M}_n^\mu - M^\mu\|_\infty = O_\mathbb{P}\left(\frac{\chi_{0,n}\log(K_n)}{\sqrt{n}}\right).$$

It remains to bound $\|\widehat{M}_n^\mu - \bar{M}_n^\mu\|_\infty$.

$$\begin{aligned}\|\widehat{M}_n^\mu - \bar{M}_n^\mu\|_\infty &= \max_{1\leq l\leq 2K_n}\left|\frac{1}{n}\sum_{i=1}^n \big[q(D_i,\widehat{V}_i) - q(D_i,V_i)\big]Y_i\right| \\ &\leq C\max_{1\leq l\leq \dim(V)}\sup_v\left\|\frac{\partial p(v)}{\partial v_l}\right\|_\infty \frac{1}{n}\sum_{i=1}^n \|\widehat{V}_i - V_i\|_1|Y_i|.\end{aligned}$$

and

$$\mathbb{E}\Big[\|\widehat{M}_n^\mu - \bar{M}_n^\mu\|_\infty\Big] \leq C\chi_{1,n}\left|\int \|\widehat{\nu}_n(t,z) - \nu_0(t,z)\|_2^2 dF_{DZ}(t,z)\mathbb{E}[Y^2]\right|^{1/2} = O_\mathbb{P}\big(\chi_{1,n}\tilde{\omega}_n\big)$$

where the expectation is conditional on observations used to estimate $\widehat{\nu}_n$, which are independent of the observations used to form the sum.

Now consider $\widehat{\Omega}_n\rho_*$. Let $\bar{\Omega}_n = n^{-1}\sum_{i=1}^n q(D_i,V_i)q(D_i,V_i)^\mathsf{T}$. By Lemma SA-15, $q(D,V)^\mathsf{T}\rho_*$ is uniformly bounded, and thus, with $X = q(D, V)$ and $Y = q(D, V)^{\mathsf{T}}\rho_*$, Lemma SA-19 implies

$$\|\bar{\Omega}_n\rho_* - \Omega\rho_*\|_\infty = O_{\mathbb{P}}\left(\frac{\chi_{0,n}\log(K_n)}{\sqrt{n}}\right).$$

Now,

$$\begin{aligned}\|\widehat{\Omega}_n\rho_* - \bar{\Omega}_n\rho_*\|_\infty &\leq \max_{1\leq l\leq 2K_n}\left|\frac{1}{n}\sum_{i=1}^n\big[q_l(D_i,\widehat{V}_i) - q_l(D_i,V_i)\big]q(D_i,\widehat{V}_i)^{\mathsf{T}}\rho_*\right| \\ &\quad + \max_{1\leq l\leq 2K_n}\left|\frac{1}{n}\sum_{i=1}^n q_l(D_i,V_i)\big[q(D_i,\widehat{V}_i) - q(D_i,V_i)\big]^{\mathsf{T}}\rho_*\right| \\ &\leq C\max_{1\leq \ell\leq \dim(V)}\sup_v\left\|\frac{\partial p(v)}{\partial v_\ell}\right\|_\infty\frac{1}{n}\sum_{i=1}^n\|\widehat{V}_i - V_i\|_1\big|q(D_i,\widehat{V}_i)^{\mathsf{T}}\rho_*\big| \\ &\quad + C\max_{1\leq \ell\leq \dim(V)}\sup_v\left|\frac{\partial p(v)}{\partial v_\ell}^{\mathsf{T}}\rho_*\right|\max_{1\leq l\leq 2K_n}\left|\frac{1}{n}\sum_{i=1}^n\|\widehat{V}_i - V_i\|_1|q_l(D_i,V_i)|\right|.\end{aligned}$$

The first term of the right-hand side is $O_{\mathbb{P}}(\omega_n)$ using the argument for $\widehat{M}_n^\mu$. For the other term, $\|\rho_*\|_1 \leq C$ by Lemma SA-14 and

$$\begin{aligned}\max_{1\leq l\leq 2K_n}\left|\frac{1}{n}\sum_{i=1}^n\|\widehat{V}_i - V_i\|_1|q_l(D_i,V_i)|\right| &\leq \max_{1\leq l\leq 2K_n}\mathbb{E}[\|\widehat{V}_i - V_i\|_1|q_l(D_i,V_i)|] \\ &\quad + \max_{1\leq l\leq 2K_n}\left|\frac{1}{n}\sum_{i=1}^n\|\widehat{V}_i - V_i\|_1|q_l(D_i,V_i)| - \mathbb{E}[\|\widehat{V}_i - V_i\|_1|q_l(D_i,V_i)|]\right|\end{aligned}$$

where the expectations are conditional on the observations used to construct $\widehat{\nu}$, and the result follows from Lemma SA-20 and $\mathbb{E}[\|\widehat{V}_i - V_i\|_1|q_l(D_i,V_i)|] \leq C\sqrt{\mathbb{E}[\|\widehat{V}_i - V_i\|_1^2]} = O_{\mathbb{P}}(\tilde{\omega}_n)$. □

Lemma SA-17. *Suppose Assumptions 7-10,* $\log(K_n)/\log(n) = O(1)$*,* $\chi_{0,n} \leq \chi_{1,n}$*, and* $\chi_{0,n}/(n^{1/4}\log K_n) + \chi_{0,n}\omega_n^{(2\zeta-1)/(2\zeta+1)} = o(1)$*. Then,*

$$\|\widehat{\rho} - \rho_*\|_1 = O_{\mathbb{P}}\left((\kappa_n/\omega_n)\omega_n^{(2\zeta-1)/(2\zeta+1)}\right), \qquad \|\widehat{\rho} - \rho_*\|_2 = O_{\mathbb{P}}\left((\kappa_n/\omega_n)\omega_n^{2\zeta/(2\zeta+1)}\right).$$

Proof. The result follows from the arguments of Lemma A6 of CNS and Lemma B7 of Bradic et al. (2022). □

Lemma SA-18. *Suppose Assumptions 7-10,* $\log(K_n)/\log(n) = O(1)$*,* $\chi_{0,n} \leq \chi_{1,n}$*, and* $\chi_{0,n}/(n^{1/4}\log K_n)+$

$\chi_{0,n}\omega_n^{(2\zeta-1)/(2\zeta+1)} = o(1)$. *Then,*

$$\|\widehat{\rho}\|_1 = O_{\mathbb{P}}(1),$$

and for $\hat{f} \in \{\widehat{\mu}, \widehat{\alpha}\}$,

$$\sup_{d,v} |\hat{f}(d,v)| = O_{\mathbb{P}}(\chi_{0,n}), \quad \sup_{d,v} \|\partial \hat{f}(d,v)/\partial v^{\mathsf{T}}\|_1 = O_{\mathbb{P}}(\chi_{1,n}).$$

Proof. The first result follows from Lemmas SA-14 and SA-17 since

$$\|\widehat{\rho}\|_1 \leq \|\widehat{\rho} - \rho_*\|_1 + \|\rho_*\|_1.$$

The second result follows from the first since $\partial\widehat{\mu}(d,v)/\partial v^{\mathsf{T}} = \widehat{\rho}^{\mathsf{T}} \partial q(d,v)/\partial v^{\mathsf{T}}$ and $\|\partial q(d,v)/\partial v^{\mathsf{T}}\|_1 = O(\chi_{1,n})$. □

The following uniform convergence rate result builds on Lemma B.1 of Cattaneo et al. (2013).

Lemma SA-19. *Let* $(X_{i1}, \ldots, X_{ip}, Y_i)$ *be i.i.d. across* i *and I write* $X_i = (X_{i1}, \ldots, X_{ip})^{\mathsf{T}}$. *For some positive sequence* ς_n, $\max_{1\leq i\leq n, 1\leq j\leq p} |X_{ij}| \leq \varsigma_n$, $\mathbb{E}[Y^2|X] \leq C$, $\mathbb{E}[Y^s] < \infty$ *for some* $s \geq 2$, $\max_{1\leq j\leq p} \mathbb{E}[X_{ij}^2] \leq C$, *and* $\varsigma_n/n^{1/2-1/s}(\log p)^{-1/2} = o(1)$. *Letting* $Z_{ij} = X_{ij}Y_i - \mathbb{E}[X_{ij}Y_i]$,

$$\max_{1\leq j\leq p} \left|\frac{1}{n}\sum_{i=1}^n Z_{ij}\right| = O_{\mathbb{P}}\left(\sqrt{\frac{\log(p)}{n}} \max\left\{1, \varsigma_n\sqrt{\frac{\log(p)}{n^{1-2/s}}}\right\}\right) + o\big(\varsigma_n n^{-(s-1)/s}\big).$$

Proof. Let $\tau_n = n^{1/s}$ and $Y_{in} = Y_i \mathbb{1}\{|Y_i| \leq \tau_n\}$.

$$\begin{aligned} \Pr\left[\sum_{i=1}^n X_{ij}Y_i \neq \sum_{i=1}^n X_{ij}Y_{in} \text{ for some } j\right] &\leq \Pr\Big[\max_{1\leq i\leq n} |Y_i| > \tau_n\Big] \\ &\leq \sum_{i=1}^n \Pr\big[|Y_i| > \tau_n\big] \\ &\leq n\mathbb{E}\big[|Y|^s \mathbb{1}\{|Y| > \tau_n\}\big]\tau_n^{-1/s} = o(1). \end{aligned}$$

For the difference in the expectations,

$$\big|\mathbb{E}[X_{ij}Y_i] - \mathbb{E}[X_{ij}Y_{in}]\big| \leq \varsigma_n \mathbb{E}[|Y|\mathbb{1}\{|Y| > \tau_n\}] \leq \varsigma_n \tau_n^{-(s-1)} \mathbb{E}[|Y|^s \mathbb{1}\{|Y| > \tau_n\}] = o\big(\varsigma_n n^{-(s-1)/s}\big).$$

Now let $Z_{ijn} = X_{ij}Y_{in} - \mathbb{E}[X_{ij}Y_{in}]$ and for any $c > 0$,

$$\begin{aligned}
\Pr\Big[\max_{1\le j\le p}\Big|\frac{1}{n}\sum_{i=1}^{n} Z_{ij}\Big| > c\Big] &\le \Pr\Big[\max_{1\le j\le p}\Big|\frac{1}{n}\sum_{i=1}^{n} Z_{ijn}\Big| > c\Big] + \Pr\Big[\sum_{i=1}^{n} X_{ij}Y_i \ne \sum_{i=1}^{n} X_{ij}Y_{in} \text{ for some } j\Big] \\
&\quad + \mathbb{1}\Big\{\max_{1\le j\le p}\big|\mathbb{E}[X_jY\mathbb{1}\{|Y| > \tau_n\}]\big| > c\Big\} \\
&\le \Pr\Big[\max_{1\le j\le p}\Big|\frac{1}{n}\sum_{i=1}^{n} Z_{ijn}\Big| > c\Big] + o(1) \\
&\quad + \mathbb{1}\Big\{\varsigma_n n^{-(s-1)/s}\mathbb{E}[|Y|^s\mathbb{1}\{|Y| > \tau_n\}] > c\Big\}.
\end{aligned}$$

Thus, the desired result follows if $\max_{1\le j\le p}|\frac{1}{n}\sum_{i=1}^{n} Z_{ijn}| = O_{\mathbb{P}}(\varrho_n)$ with

$$\varrho_n = \sqrt{\frac{\log(p)}{n}}\max\Big\{1, \varsigma_n\sqrt{\frac{\log(p)}{n^{1-2/s}}}\Big\}.$$

Note $\mathbb{E}[Z_{ijn}^2] \le 2\mathbb{E}[X_{ij}^2Y_i^2] = 2\mathbb{E}[X_{ij}^2\mathbb{E}[Y_i^2|X_i]] \le C\mathbb{E}[X_{ij}^2] \le C$. Since $\max_{1\le i\le n}|Z_{ijn}| \le 2\tau_n\varsigma_n$, Bernstein's inequality implies that for any $M > 0$,

$$\begin{aligned}
\Pr\Big(\max_{1\le j\le p}\Big|\frac{1}{n}\sum_{i=1}^{n} Z_{ijn}\Big| > \varrho_n M\Big) &\le \sum_{j=1}^{p}\Pr\Big(\Big|\sum_{i=1}^{n} Z_{ijn}\Big| > n\varrho_n M\Big) \\
&\le 2\exp\Big(\log(p) - \frac{n(\varrho_n M)^2/16}{C + \tau_n\varsigma_n M\varrho_n/6}\Big).
\end{aligned}$$

Now, suppose $\limsup_n \frac{\varsigma_n^2\log(p)}{n^{1-2/s}} < \infty$. Then, $\varrho_n = O(\sqrt{\log(p)/n})$, and

$$\begin{aligned}
\exp\Big(-\frac{n(\varrho_n M)^2/16}{C + \tau_n\varsigma_n\varrho_n M/6}\Big) &\le \exp\Big(-\frac{M\log(p)/16}{CM^{-1} + \tau_n\varsigma_n\sqrt{\log(p)/n}/6}\Big) \\
&\le \exp\Big(-M\log(p)/(C+1)\Big)
\end{aligned}$$

where the last inequality uses $\tau_n\varsigma_n\sqrt{\log(p)/n} \le C$.

Next, if $\liminf_n \frac{\varsigma_n^2\log(p)}{n^{1-2/s}} > 0$, then $\varrho_n = O(\varsigma_n\log(p)/n^{1-1/s})$ and

$$\begin{aligned}
\exp\Big(-\frac{n(\varrho_n M)^2/16}{C + \tau_n\varsigma_n\varrho_n M/6}\Big) &\le \exp\Big(-\frac{M\varsigma_n^2\log(p)^2/16n^{1-2/s}}{CM^{-1} + \varsigma_n^2\log(p)/6n^{1-2/s}}\Big) \\
&= \exp\Big(-\frac{M\log(p)/16}{Cn^{1-2/s}/(M\log(p)\varsigma_n^2) + 1/6}\Big) \\
&\le \exp\Big(-M\log(p)/(C+1)\Big)
\end{aligned}$$

where the last inequality uses $\limsup_n \frac{n^{1-2/s}}{\varsigma_n^2 \log(p)} = (\liminf_n \frac{\varsigma_n^2 \log(p)}{n^{1-2/s}})^{-1} < \infty$. Thus, for both cases, by taking M large enough,

$$\Pr\Big(\max_{1\le j\le p}\Big|\frac{1}{n}\sum_{i=1}^n Z_{ijn}\Big| > \varrho_n M\Big) = o(1).$$

For general cases, arguing along subsequences and using the above two cases lead to the desired result. □

Lemma SA-20. *Suppose Assumptions 7-8 hold. If $\tilde{\omega}_n(\log(K_n)+\chi_{0,n}) = O(1)$, then*

$$\max_{1\le l\le 2K_n}\left|\frac{1}{n}\sum_{i=1}^n \|\widehat{V}_i - V_i\|_1|q_l(D_i,V_i)| - \mathbb{E}[\|\widehat{V}_i - V_i\|_1|q_l(D_i,V_i)|]\right| = O_{\mathbb{P}}\Bigg(\sqrt{\frac{\log(K_n)}{n}} + \chi_{0,n}\tilde{\omega}_n\Bigg)$$

where the expectations are computed conditional on observations used to estimate $\widehat{\nu}_n$, which are independent of the observations used to form the sample mean.

Proof. The argument is analogous to the one for Lemma SA-19. In the notation of the lemma, $Y_i = \|\widehat{V}_i - V_i\|_1$, $X_{ij} = |q_j(D_i,V_i)|$. Let $\tau_n = Cn^{1/2}\tilde{\omega}_n$ for some large $C>0$, and $n\mathbb{E}[\|\widehat{V}_i - V_i\|_1^2]\tau_n^{-2} = C^{-2}O_{\mathbb{P}}(1)$. Also, $\mathbb{E}[|q_j(D_i,V_i)|\|\widehat{V}_i - V_i\|_1\mathbb{1}\{\|\widehat{V}_i - V_i\|_1 > \tau_n\}] \le \chi_{0,n}\tilde{\omega}_n$. Letting $Z_{ijn} = X_{ij}Y_i\mathbb{1}\{|Y_i|\le \tau_n\} - \mathbb{E}[X_{ij}Y_i\mathbb{1}\{|Y_i|\le\tau_n\}]$, $\mathbb{E}[Z_{ijn}^2] \le C\chi_{0,n}^2 O_{\mathbb{P}}(\tilde{\omega}_n^2)$. Then, with C large enough, $\mathbb{E}[Z_{ijn}^2] \le C\chi_{0,n}^2\tilde{\omega}_n^2$ occurs with probability close to one. On this event,

$$\begin{aligned}\Pr\Bigg(\max_{1\le j\le 2K_n}\Big|\frac{1}{n}\sum_{i=1}^n Z_{ijn}\Big| > M\sqrt{\frac{\log(K_n)}{n}}\Bigg) &\le 4\exp\Bigg(\log(K_n) - \frac{M^2\log(K_n)/16}{C\chi_{0,n}^2\tilde{\omega}_n^2 + \tau_n M\sqrt{\log(K_n)/n}/6}\Bigg)\\ &\le 4\exp\Big(-\log(K_n)\Big) = o(1)\end{aligned}$$

where the last equality holds if M and n are sufficiently large. □

A.5 Flexible parametric estimation

Recall the setup

$$\theta(d) = \mathbb{E}[\mathbb{E}[Y|D=d,V]]$$

$$\mathbb{E}[Y|D,V] = \Lambda\big(p(D,V)'\gamma_0\big)$$

$$\mathbb{E}[G(X)|D,Z] = Q(D,Z)\delta_0.$$

Define the regression residuals $\varepsilon = Y - \mathbb{E}[Y|D,V], \eta = G(X) - \mathbb{E}[G(X)|D,Z]$. I impose the following conditions.

Condition SA-2. *Let* $\beta = (\delta, \gamma)$ *and* $\mathcal{N} = \{\beta : \|\delta - \delta_0\| \vee \|\gamma - \gamma_0\| \le \eta\}$ *for some* $\eta > 0$.

(i) $\{Y(d) : d \in \mathcal{D}\} \perp\!\!\!\perp (D,Z)|X^*$.

(ii) Λ *is strictly monotone and twice continuously differentiable with bounded derivatives, and* p *is continuously differentiable in* V*.*

(iii) The matrices $\Gamma_2 = \mathbb{E}[\dot{\Lambda}(p(D,V)'\gamma_0)p(D,V)p(D,V)']$, $\Gamma_1 = \mathbb{E}[Q(D,Z)'Q(D,Z)]$ *are non-singular where* $\dot{\Lambda}$ *denotes the first derivative of* Λ*.*

(iv) $\mathbb{E}[\sup_{\beta\in\mathcal{N}} \|\{Y - \Lambda(p(D,Q(D,Z)\delta)'\gamma)\}p(D,Q(D,Z)\delta)\|^2]$, $\mathbb{E}[\sup_{\beta\in\mathcal{N}} \|Q(D,Z)'(X - Q(D,Z)\delta)\|^2]$, $\mathbb{E}[\sup_{\beta\in\mathcal{N}} |\Lambda(p(D,Q(D,Z)\delta)'\gamma)|^2]$, $\mathbb{E}[\sup_{\beta\in\mathcal{N}} \|p(D,Q(D,Z)\delta)p(D,Q(D,Z)\delta)'\|]$, *and* $\mathbb{E}[\sup_{\beta\in\mathcal{N}} |Y - \Lambda(p(D,Q(D,Z)\delta)'\gamma)|\{\|p(D,Q(D,Z)\delta)\| + 1\}\|\frac{\partial p(D,Q(D,Z)\delta)}{\partial V'}Q(D,Z)\|]$ *are all finite.*

Theorem SA-1. *Under Assumption SA-2, the flexible parametric estimator satisfies*

$$\sqrt{n}\big(\widehat{\theta}_n(d) - \theta(d)\big) \rightsquigarrow \text{Normal}(0, \Psi(d))$$

where

$$\Psi(d) = \text{Var}\big[\Lambda(p(d,V)'\gamma_0) - \theta(d) + c_1(d)\Gamma_2^{-1}P\dot{\Lambda}(P'\gamma_0)\varepsilon + \{c_1(d)\Gamma_2^{-1}\Gamma_3 + c_2(d)\}\Gamma_1^{-1}Q'\zeta\big],$$

$P = p(D,V)$, $Q = Q(D,Z)$, *and*

$$\begin{aligned} c_1(d) &= \mathbb{E}[\dot{\Lambda}(p(d,V)'\gamma_0)p(d,V)'] \\ c_2(d) &= \mathbb{E}\Big[\dot{\Lambda}(p(d,V)'\gamma_0)\gamma_0'\frac{\partial p(d,V)}{\partial V'}Q\Big] \\ \Gamma_3 &= -\mathbb{E}\Big[|\dot{\Lambda}(P'\gamma_0)|^2 P\gamma_0'\frac{\partial p(D,V)}{\partial V'}Q\Big]. \end{aligned}$$

In addition, the variance estimator in the main paper is consistent for $\Psi(d)$*.*

A.5.1 Proof

In the sequel, I refer to Newey and McFadden (1994) as NM. Note that $\widehat{\delta}_n \to_{\mathbb{P}} \delta_0$ follows from standard arguments. Using Lemma 2.4 of NM, one can show $\widehat{\gamma}_n \to_{\mathbb{P}} \gamma_0$ and $\widehat{\theta}_n(d) \to_{\mathbb{P}} \theta(d)$. Then, asymptotic normality follows from Theorem 6.1 of NM where the moment function is

$$\begin{bmatrix} Q(D,Z)'(X - Q(D,Z)\delta) \\ \{Y - \Lambda(p(D,Q(D,Z)\delta)'\gamma)\}\dot{\Lambda}(p(D,Q(D,Z)\delta)'\gamma)p(D,Q(D,Z)\delta) \\ \Lambda(p(d,Q(D,Z)\delta)'\gamma) - \theta \end{bmatrix}.$$

The derivative of this moment condition with respect to (δ, γ, β) is

$$\begin{bmatrix} -Q(D,Z)'Q(D,Z) & 0 & 0 \\ \varphi_1(Y,D,Z,\delta,\gamma) & \varphi_2(Y,D,Z,\delta,\gamma) & 0 \\ \dot{\Lambda}(p(d,Q(D,Z)\delta)'\gamma)\gamma'\frac{\partial p(d,Q(D,Z)\delta)}{\partial V'}Q(D,Z) & \dot{\Lambda}(p(d,Q(D,Z)\delta)'\gamma)p(d,Q(D,Z)\delta)' & -1 \end{bmatrix}$$

where

$$\begin{aligned}\varphi_1(Y,D,Z,\delta,\gamma) =& \big[\{Y - \Lambda(p(D,Q(D,Z)\delta)'\gamma)\}\ddot{\Lambda}(p(D,Q(D,Z)\delta)'\gamma) - \big|\dot{\Lambda}(p(D,Q(D,Z)\delta)'\gamma)\big|^2\big] \\ & \times p(D,Q(D,Z)\delta)\gamma'\frac{\partial p(D,Q(D,Z)\delta)}{\partial V'}Q(D,Z) \\ & + \{Y - \Lambda(p(D,Q(D,Z)\delta)'\gamma)\}\dot{\Lambda}(p(D,Q(D,Z)\delta)'\gamma)\frac{\partial p(D,Q(D,Z)\delta)}{\partial V'}Q(D,Z),\end{aligned}$$

$$\begin{aligned}\varphi_2(Y,D,Z,\delta,\gamma) =& \big[\{Y - \Lambda(p(D,Q(D,Z)\delta)'\gamma)\}\ddot{\Lambda}(p(D,Q(D,Z)\delta)'\gamma) - \big|\dot{\Lambda}(p(D,Q(D,Z)\delta)'\gamma)\big|^2\big] \\ & \times p(D,Q(D,Z)\delta)p(D,Q(D,Z)\delta)',\end{aligned}$$

and $\ddot{\Lambda}$ is the second derivatives of Λ. By the formula given in NM, the asymptotic linear representation of $\widehat{\theta}_n(d)$ is

$$\begin{aligned}\sqrt{n}\big(\widehat{\theta}_n(d) - \theta(d)\big) =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\Lambda(p(d,V)'\gamma_0) - \theta(d)\} \\ & - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} G_{12}G_{11}^{-1}\begin{bmatrix} Q(D_i,Z_i)'\zeta_i \\ p(D_i,V_i)\dot{\Lambda}(p(D_i,V_i)'\gamma_0)\varepsilon_i \end{bmatrix} + o_{\mathbb{P}}(1)\end{aligned}$$

where

$$G_{12} = \begin{bmatrix} \mathbb{E}\big[\dot{\Lambda}(p(d,V)'\gamma_0)\gamma_0'\frac{\partial p(d,V)}{\partial V'}Q(D,Z)\big] & \mathbb{E}\big[\dot{\Lambda}(p(d,V)'\gamma_0)p(d,V)'\big] \end{bmatrix}$$

and

$$G_{11} = \begin{bmatrix} -\mathbb{E}[Q(D,Z)'Q(D,Z)] & 0 \\ \mathbb{E}[\varphi_1(Y,D,Z,\delta_0,\gamma_0)] & \mathbb{E}[\varphi_2(Y,D,Z,\delta_0,\gamma_0)] \end{bmatrix}.$$

Using the block matrix inverse formula,

$$\begin{aligned}
&- G_{12}G_{11}^{-1} \begin{bmatrix} Q(D_i,Z_i)'\zeta_i \\ p(D_i,V_i)\dot{\Lambda}(p(D_i,V_i)'\gamma_0)\varepsilon_i \end{bmatrix} \\
&= \mathbb{E}\Big[\dot{\Lambda}(p(d,V)'\gamma_0)\gamma_0'\frac{\partial p(d,V)}{\partial V'}Q(D,Z)\Big]\mathbb{E}[Q(D,Z)'Q(D,Z)]^{-1}Q(Z_i,D_i)'\zeta_i \\
&\quad + \mathbb{E}[\dot{\Lambda}(p(d,V)'\gamma_0)p(d,V)']\mathbb{E}[|\dot{\Lambda}(p(D,V)'\gamma_0)|^2 p(D,V)p(D,V)']^{-1}p(D_i,V_i)\dot{\Lambda}(p(D_i,V_i)'\gamma_0)\varepsilon_i \\
&\quad + \mathbb{E}[\dot{\Lambda}(p(d,V)'\gamma_0)p(d,V)']\mathbb{E}[|\dot{\Lambda}(p(D,V)'\gamma_0)|^2 p(D,V)p(D,V)']^{-1}\mathbb{E}[\varphi_1(Y,D,Z,\delta_0,\gamma_0)] \\
&\qquad \times \mathbb{E}[Q(D,Z)'Q(D,Z)]^{-1}Q(Z_i,D_i)'\zeta_i.
\end{aligned}$$

Consistency of the variance estimator follows from repeated applications of Lemma 4.3 of NM. □

B Multi-valued treatment variables

In the main paper, I use estimation of grade retention effects as a running example, in which the treatment variable of interest is binary. Here, I discuss how the assumptions for Theorem 1 change with multi-valued treatment.

Replacing $d \in \{0,1\}$ with $d \in \mathcal{D} = \mathrm{supp}(D)$, Assumptions 1, 2, 3, and 5 remain unchanged. For Assumption 4, the condition is modified as:

Condition SA-3. *Let* $\mathcal{X} = \mathrm{supp}(X)$ *and* $\mathfrak{V} = \{f_{X|DZ}(x|D,Z) : x \in \mathcal{X}\}$*. The conditional support of* $\mathfrak{V}$ *given* $D = d$*,* $\mathrm{supp}(\mathfrak{V}|D=d)$*, equals the marginal support* $\mathrm{supp}(\mathfrak{V})$ *for almost all* $d \in \mathcal{D}$*.*

Recall that the support of a random element in $L^2(\lambda)$, say $\mathfrak{U}$, can be defined as

$$\mathrm{supp}(\mathfrak{U}) = \big\{x \in L^2(\lambda) : \mathbb{P}[\mathfrak{U} \in O_x] > 0 \quad \text{for any open } O_x \text{ containing } x\big\}$$

where $L^2(\lambda)$ is equipped with the norm topology.

The formal assumption above may not be intuitive, but it is the same type of restriction as to the common support condition in the control function literature (e.g., Imbens and Newey, 2009, Assumption 2). Similar to the binary treatment case discussed in the main paper, Assumption SA-3 essentially imposes that the conditional distribution of the proxy given (D, Z) satisfy some index restrictions: there exist fixed functions f, ψ such that

$$f_{X|DZ}(X|D,Z) = f(X, \psi(D,Z)) \quad a.s.$$

Then, Assumption SA-3 holds if for each $d \in \text{supp}(D)$, $\text{supp}(\psi(D,Z)|D = d) = \text{supp}(\psi(D,Z))$.

To give an example where Assumption SA-3 holds, consider the following random coefficient model

$$D = \eta_1 Z_1 + \eta_2 Z_2$$

where the excluded variable $Z = (Z_1, Z_2)^\top$ is two-dimensional and $\eta = (\eta_1, \eta_2)^\top$ may be correlated with X^*, causing the endogeneity issue. Assume $Z \perp\!\!\!\perp (X^*, \eta)$ i.e., Z is an instrument for D.[1] Then, the conditional distribution of X^* given $(D, Z) = (d, z)$ is determined by the index $\psi(d, z) = (d/z_1, z_2/z_1)$ (for simplicity, $Z_1 = 0$ is a probability zero event). Then, under Assumption 2, the index restriction holds. Now, Assumption SA-3 holds if for each $\tilde{d} \in \text{supp}(D)$ and $(d, z) \in \text{supp}(D, Z)$, there exists $\tilde{z} \in \text{supp}(Z|D = \tilde{d})$ such that $(d/z_1, z_2/z_1) = (\tilde{d}/\tilde{z}_1, \tilde{z}_2/\tilde{z}_1)$.

Two points should be noted. First, the above example again requires the large support of Z, analogous to the existing literature of control function methods. Secondly, the above random coefficient model cannot be handled using the existing control function methods, such as Imbens and Newey (2009), because the first-stage equation is not invertible in the unobserved heterogeneity.

C Comparison with proximal control/proxy control approaches

My identifying assumptions are closely related to those used by Deaner (2018); Miao et al. (2018), and some comparison may be warranted. I focus on their results using the outcome equation, and I refer to them as integral equation approaches, to contrast with my control function approach. I

[1] Even with the availability of valid instruments, causal effects might not be identified with existing approaches if the outcome of interest is a non-separable function of X^*.

should note that Deaner (2018) also developed a method based on re-weighting of the outcome variable, which is related to the integral equation approach but does not exactly fit into the discussion that follows.

Both my method and the integral equation approach impose Assumptions 1, 2, and 3. The main difference is Assumption SA-3. In the integral equation approach, instead of Assumption SA-3, the identifying assumptions are:

Condition SA-4. *Let* $\mathcal{D} = \operatorname{supp}(D)$*. Given* $d \in \mathcal{D}$*, define the operator*

$$\Pi_d : L^2(F_{X|D=d}) \mapsto L^2(F_{Z|D=d}), \quad \Pi_d(h)(z) = \mathbb{E}[h(X)|D=d, Z=z]$$

Assume this linear operator Π_d *is compact, and denote the singular values by* $\{\pi_{j,d}\}_{j\geq 1}$ *and associated singular-value functions* $\{v_{j,d}\}_{j\geq 1} \subset L^2(F_{X|D=d})$*,* $\{u_{j,d}\}_{j\geq 1} \subset L^2(F_{Z|D=d})$*. See Theorem 15.16 of Kress (2014). Write* $\pi_j(d) \equiv \pi_{j,d}$*,* $v_j(d,x) \equiv v_{j,d}(x)$*, and* $u_j(d,z) \equiv u_{j,d}(z)$*. Assume*

$$\sum_{j=1}^{\infty} \frac{1}{\pi_j(d)^2} \mathbb{E}\big[\mathbb{E}[Y|D,Z]u_j(D,Z)\big|D=d\big]^2 < \infty.$$

Condition SA-5. *For any* $b \in L^2(F_{X^*})$ *and* $d \in \mathcal{D}$*,* $\mathbb{P}[\mathbb{E}[b(X^*)|D,Z] = 0|D=d] = 1$ *implies* $\mathbb{E}[b(X^*)|D] = 0$ *with probability one.*

Assumption SA-5 is a weak version of a completeness condition. For reference, a standard completeness conditional on D is

Condition SA-6. *For any* $b \in L^2(F_{X^*})$ *and* $d \in \mathcal{D}$*,* $\Pr[\mathbb{E}[b(X^*)|D,Z] = 0|D=d] = 1$ *implies* $b(X^*) = 0$ *with probability one.*

Since this completeness implies Assumption SA-5, I refer to the latter as a weak version of completeness.

Deaner (2018) used a standard completeness condition rather than Assumption SA-5. Miao et al. (2018) similarly imposed a standard completeness condition on X given (D,Z). Yet, for the integral equation approach, the weak version of Assumption SA-5 suffices for the identification of the average structural functions (and its conditional version given D).[2]

[2]An anonymous referee pointed out this sufficiency of the weak version of completeness conditions. They also proved and shared a version of Lemma SA-21 in a report. I thank them for very helpful feedback.

The following result shows that Assumption SA-5 follows from the assumptions imposed by the control function method.

Lemma SA-21. *Suppose Assumptions 2-3, the index restriction (3) in the main paper, and Assumption SA-3 hold. Suppose also that there is a measurable selection from the set* $\{z \in \text{supp}(Z|D = d) : \psi(d, z) = v\}$ *for each* $(d, v) \in \text{supp}(D, \psi(D, Z))$*. Then, Assumption SA-5 holds.*

Proof. As shown in the proof of Theorem 1,

$$\frac{f_{X^*|DZ}(x^*|D, Z)}{f_{X^*}(x^*)} = \Pi^\dagger\left(\frac{f_{X|DZ}(\cdot|D, Z)}{f_X(\cdot)}\right)(x^*)$$

for some fixed mapping $\Pi^\dagger$. By the hypothesis, for each $\tilde{d} \in \mathcal{D}$, there exists a measurable mapping $s : \mathcal{D} \times \text{supp}(\psi(D, Z)) \to \text{supp}(Z|D = \tilde{d})$ such that

$$\frac{f_{X|DZ}(x|D, Z)}{f_X(x)} = \frac{f(x, \psi(\tilde{d}, s(\psi(D, Z), \tilde{d})))}{f_X(x)} \qquad \forall x \text{ s.t. } f_X(x) > 0$$

holds almost surely. Then,

$$f_{X^*|DZ}(x^*|D, Z) = f_{X^*|DZ}(x^*|\tilde{d}, s(\psi(D, Z), \tilde{d})) \qquad \forall x^* \text{ s.t. } f_{X^*}(x^*) > 0$$

with probability one. Now, suppose that with $F_{Z|D=t}$ probability one,

$$0 = \mathbb{E}[b(X^*)|D = t, Z] = \int b(x^*) f_{X^*|DZ}(x^*|t, Z) d\lambda_*(x^*). \tag{S-8}$$

Then, $\int b(x^*) f_{X^*|DZ}(x^*|D, Z) dx^* = 0$ holds with probability one; otherwise,

$$0 \neq \int b(x^*) f_{X^*|DZ}(x^*|D, Z) dx^* = \int b(x^*) f_{X^*|DZ}(x^*|t, s(\psi(D, Z), t)) d\lambda_*(x^*)$$

with some positive probability. That is, there exists a positive probability set $\mathcal{A} \subseteq \text{supp}(Z|D = t)$ such that

$$\int b(x^*) f_{X^*|DZ}(x^*|t, z) d\lambda_*(x^*) \neq 0 \qquad \forall z \in \mathcal{A}$$

but this contradicts the hypothesis (S-8). Therefore, if $\int b(x^*) f_{X^*|DZ}(x^*|t, Z) dx^* = 0$ holds with $F_{Z|D=t}$ probability one, then $\mathbb{E}[b(X^*)|D] = 0$ almost surely follows. □

The hypothesis of the existence of a measurable selection may be deemed as a mild regularity condition as smoothness of $(d,z) \mapsto \psi(d,z)$ implies such condition.

Given the above result, the main difference in the control function approach and the integral equation approach is Assumption 4 versus Assumption SA-4. This is indeed the key distinction because of the two results in the next section. First, I show that one can achieve the identification by replacing Assumption 4 with Assumption SA-5 and a high-level condition that is analogous to Assumption SA-4. Second, I show that in some model, Assumption SA-4 fails while Assumptions 1-4 hold.

C.1 Replacing the common support condition

As in Lemma SA-3, under Assumptions 1-5 and $\{Y(d) : d \in \mathcal{D}\} \perp\!\!\!\perp (D, Z)|X^*$,

$$\mathbb{E}[Y|D = d, Z] = \sum_{j=1}^{\infty} \tau_j^{-1} \mathbb{E}[\mathbb{E}[Y(d)|X^*]\phi_j(X^*)] \int f_{X|DZ}(x|d, Z)\varphi_j(x)d\lambda(x).$$

Suppose we impose the following condition

$$\sum_{j=1}^{\infty} \frac{1}{\tau_j^2} \mathbb{E}[\mathbb{E}[Y(d)|X^*]\phi_j(X^*)]^2 < \infty, \tag{S-9}$$

which is analogous to Assumption SA-4. With this condition imposed, the infinite sum and the integral with respect to x in the above display can be interchanged to obtain

$$\begin{aligned}\mathbb{E}[Y|D = d, Z] &= \int \left(\sum_{j=1}^{\infty} \tau_j^{-1} \mathbb{E}[\mathbb{E}[Y(d)|X^*]\phi_j(X^*)]\varphi_j(x) \right) f_{X|DZ}(x|d, Z)d\lambda(x) \\ &\equiv \mathbb{E}[\mathfrak{B}(d, X)|D = d, Z]\end{aligned}$$

where the function $\mathfrak{B}$ is what Miao et al. (2018) calls a bridge function. With Assumption SA-5, the ASF can be identified via $\mathbb{E}[\mathfrak{B}(d, X)]$. Note that the above identification argument did not use Assumption 4 and thus (S-9) can replace Assumption 4 (along with Assumption SA-5) to achieve the identification. Yet, as discussed in the main paper, verifying a high-level condition like (S-9) seems difficult in practice.

C.2 Counterexample

I show that the high-level condition of the integral equation (Assumption SA-4) fails to hold in the following simple example.

$$Y = \mathbb{1}\{\beta D \geq X^*\}, \qquad X = X^* + U_x, \qquad Z = X^* + U_z$$

where

$$\begin{bmatrix} X^* \\ D \end{bmatrix} \sim \text{Normal}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma_x\sigma_d\rho \\ \sigma_x\sigma_d\rho & \sigma_d^2 \end{bmatrix} \right),$$

$(X^*, D) \perp\!\!\!\perp (U_x, U_z)$, $U_x \perp\!\!\!\perp U_z$, and U_x, U_z each follows a standard normal distribution. We have

$$(X, Z)|D \sim \text{Normal}\left(\begin{bmatrix} \frac{\sigma_x}{\sigma_d}\rho D \\ \frac{\sigma_x}{\sigma_d}\rho D \end{bmatrix}, \begin{bmatrix} (1-\rho^2)\sigma_x^2 + 1 & (1-\rho^2)\sigma_x^2 \\ (1-\rho^2)\sigma_x^2 & (1-\rho^2)\sigma_x^2 + 1 \end{bmatrix} \right).$$

In the notation of Assumption SA-4, we can take (see e.g., Carrasco et al., 2007)

$$u_j(d, z) = \frac{1}{\sqrt{j!}} He_j\left(\frac{z - \frac{\sigma_x}{\sigma_d}\rho d}{(1-\rho^2)\sigma_x^2 + 1} \right), \qquad \pi_j(d) = \left(\frac{(1-\rho^2)\sigma_x^2}{(1-\rho^2)\sigma_x^2 + 1} \right)^j$$

where $He_j(\cdot)$ is the jth order Hermite polynomial. For reference, let $\varsigma = \frac{(1-\rho^2)\sigma_x^2}{(1-\rho^2)\sigma_x^2+1}$ so $\pi_j(d) = \varsigma^j$.

For the outcome, note

$$\begin{bmatrix} X^* \\ D \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} X^* \\ D \\ U_z \end{bmatrix} \sim \text{Normal}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma_x\sigma_d\rho & \sigma_x^2 \\ \sigma_x\sigma_d\rho & \sigma_d^2 & \sigma_x\sigma_d\rho \\ \sigma_x^2 & \sigma_x\sigma_d\rho & \sigma_x^2 + 1 \end{bmatrix} \right)$$

and $X^*|D, Z \sim \text{Normal}(\frac{\varsigma\rho}{(1-\rho^2)\sigma_x\sigma_d} D + \varsigma Z, \varsigma)$. Thus,

$$\mathbb{E}[Y|D, Z] = \Phi\left(\frac{1}{\sqrt{\varsigma}} \Big[\beta - \frac{\sigma_x\rho/\sigma_d}{(1-\rho^2)\sigma_x^2 + 1} \Big] D - \sqrt{\varsigma} Z \right) \equiv \Phi\big(\bar{\beta} D - \sqrt{\varsigma} Z \big).$$

Computing inner products

$$\int \Phi(a + bz) He_j\left(\frac{z - \mu}{\sigma} \right) \sigma^{-1} \phi\left(\frac{z - \mu}{\sigma} \right) dz = \int \Phi(a + b\mu + b\sigma t) He_j(t) \phi(t) dt.$$

For $j \geq 1$, using integration by part,

$$\int \Phi(\tilde{a}+\tilde{b}t)He_j(t)\phi(t)dt = -\Phi(\tilde{a}+\tilde{b}t)He_{j-1}(t)\phi(t)\Big|_{-\infty}^{\infty} + \tilde{b}\int \phi(\tilde{a}+\tilde{b}t)He_{j-1}(t)\phi(t)dt$$
$$= \frac{\tilde{b}}{2\pi}\exp\Big(\frac{-\tilde{a}^2}{2(\tilde{b}^2+1)}\Big)\int \exp\Big(-\frac{(\tilde{b}^2+1)}{2}\Big(t+\frac{\tilde{a}\tilde{b}}{\tilde{b}^2+1}\Big)^2\Big)He_{j-1}(t)dt$$

where I used that $\frac{d}{dx}(He_{j-1}(x)\exp(-x^2/2)) = (-1)He_j(x)\exp(-x^2/2)$.

Now for some constants $\gamma > 1, \delta$,

$$\int \exp\Big(-\frac{\gamma}{2}(t+\delta)^2\Big)He_j(t)dt$$
$$= \frac{1}{\sqrt{\gamma}}\int \exp\Big(-\frac{s^2}{2}\Big)He_j\Big(\frac{s}{\sqrt{\gamma}}-\delta\Big)ds$$
$$= \frac{1}{\sqrt{\gamma}}\int \exp\Big(-\frac{s^2}{2}\Big)\sum_{k=0}^{j}\binom{j}{k}(-\delta)^{j-k}He_k\Big(\frac{s}{\sqrt{\gamma}}\Big)ds$$
$$= \frac{1}{\sqrt{\gamma}}\int \exp\Big(-\frac{s^2}{2}\Big)\sum_{k=0}^{j}\binom{j}{k}(-\delta)^{j-k}\sum_{i=0}^{\lfloor k/2\rfloor}\gamma^{-k/2+i}(\gamma^{-1}-1)^i\binom{k}{2i}\frac{(2i)!}{i!}2^{-i}He_{k-2i}(s)ds$$
$$= \sqrt{\frac{2\pi}{\gamma}}\sum_{k=0}^{\lfloor j/2\rfloor}\binom{j}{2k}(-\delta)^{j-2k}(\gamma^{-1}-1)^k\frac{(2k)!}{k!}2^{-k}$$
$$= \sqrt{\frac{2\pi}{\gamma}}\sum_{k=0}^{\lfloor j/2\rfloor}\binom{j}{2k}(-1)^{j-k}\delta^{j-2k}\left(\frac{\gamma}{\gamma-1}\right)^{-k}\frac{(2k)!}{k!}2^{-k}$$
$$= \sqrt{\frac{2\pi}{\gamma}}(-1)^j\left(\frac{\gamma-1}{\gamma}\right)^{j/2}j!\sum_{k=0}^{\lfloor j/2\rfloor}\frac{(-1)^k}{k!(j-2k)!}\delta^{j-2k}\left(\frac{\gamma}{\gamma-1}\right)^{j/2-k}2^{-k}$$
$$= \sqrt{\frac{2\pi}{\gamma}}(-1)^j\left(\frac{\gamma-1}{\gamma}\right)^{j/2}He_j\left(\delta\sqrt{\frac{\gamma}{\gamma-1}}\right).$$

Putting all together, with $\mu_d = \frac{\sigma_x}{\sigma_d}\rho d$, $s^2 = (1-\rho^2)\sigma_x^2 + 1$

$$\mathbb{E}\big[\mathbb{E}[Y|D,Z]u_{j+1}(D,Z)\big|D=d\big] = \frac{1}{\sqrt{(j+1)!}}\int \Phi(\bar{\beta}d - \sqrt{\varsigma}z)He_{j+1}\Big(\frac{z-\mu_d}{s}\Big)s^{-1}\phi\left(\frac{z-\mu_d}{s}\right)dz$$
$$= \frac{\tilde{b}(-1)^j}{\sqrt{2\pi\gamma(j+1)!}}\exp\left(\frac{-\tilde{a}^2}{2(\tilde{b}^2+1)}\right)\left(\frac{\gamma-1}{\gamma}\right)^{j/2}He_j\left(\delta\sqrt{\frac{\gamma}{\gamma-1}}\right).$$

where $\tilde{a} = (\bar{\beta} - \sqrt{\varsigma}\frac{\sigma_x}{\sigma_d}\rho)d$, $\tilde{b} = -\sqrt{\varsigma(1-\rho^2)\sigma_x^2 + \varsigma}$,

$$\gamma = \varsigma[(1-\rho^2)\sigma_x^2 + 1] + 1 = (1-\rho^2)\sigma_x^2 + 1, \quad \text{and} \quad \delta = -\frac{d[\bar{\beta} - \sqrt{\varsigma}\sigma_x\rho/\sigma_d]\sqrt{\varsigma}}{\sqrt{(1-\rho^2)\sigma_x^2 + 1}}.$$

Asymptotics Using (8.22.8) in Szegő (1975, p.200),

$$\frac{\Gamma(\frac{j}{2}+1)}{\Gamma(j+1)} e^{-x^2/4} 2^{j/2} He_j(x) = \cos\left(x\sqrt{j+1/2} - \frac{j\pi}{2}\right) + \frac{x^3}{12\sqrt{2}}(2j+1)^{-1/2}\sin\left(x\sqrt{j+1/2} - \frac{j\pi}{2}\right) + O(j^{-1})$$

for $j \in \mathbb{N}$. Then, for infinitely many $n \in \mathbb{N}$, for some constants $c_1, c_2 > 0$,

$$\begin{aligned}
\mathbb{E}\big[\mathbb{E}[Y|D,Z]u_{2n+1}(D,Z)\big|D=d\big] &\geq c_1 \frac{(2n)!}{n!2^n} \frac{1}{\sqrt{(2n+1)!}} \left(\frac{\gamma-1}{\gamma}\right)^n \\
&= c_1(2n+1)^{-1/2} \frac{(2n-1)!!}{\sqrt{(2n)!}} \left(\frac{\gamma-1}{\gamma}\right)^n \\
&= c_1(2n+1)^{-1/2} \sqrt{\frac{(2n-1)!!}{(2n)!!}} \left(\frac{\gamma-1}{\gamma}\right)^n \\
&\geq c_2(2n+1)^{-1/2} n^{-1/4} \left(\frac{\gamma-1}{\gamma}\right)^n
\end{aligned}$$

where $n!!$ denotes the double factorial of n, the second and third equality use $(2n)! = (2n)!!(2n-1)!! = n!2^n(2n-1)!!$, and the last inequality follows from results on Wallis' integrals. Then,

$$\pi_{2n+1}(d)^{-2}\mathbb{E}\big[\mathbb{E}[Y|D,Z]u_{2n+1}(D,Z)\big|D=d\big]^2 \geq cn^{-3/2}\frac{1}{\varsigma^{4n}}\left(\frac{\gamma-1}{\gamma}\right)^{2n}$$

and if $(\gamma-1)/\gamma > \varsigma^2$, Assumption SA-4 fails. Since $(\gamma-1)/\gamma = \varsigma$ and by $\varsigma < 1$, Assumption SA-4 fails.

C.2.1 Verifying conditions for the control function method

I verify Assumptions 1-6 for the above model. Note $Y(d) = \mathbb{1}\{\beta d \geq X^*\}$ so Assumption 1 $Y(d) \perp\!\!\!\perp (D,Z)|X^*$ trivially holds. Assumption 2 holds as $U_x \perp\!\!\!\perp (X^*, D, U_z)$. Assumption 3 follows from the completeness property of exponential families (see e.g., Newey and Powell, 2003). Assumption 5 can be verified using the normality assumption. For Assumption 6, letting $\beta = \frac{\varsigma\rho}{(1-\rho^2)\sigma_x\sigma_d}$ and

$\varsigma = \frac{(1-\rho^2)\sigma_x^2}{(1-\rho^2)\sigma_x^2+1}$, $X|D,Z \sim \text{Normal}(\beta D + \varsigma Z, \varsigma + 1)$, and $\psi(d,z) = \beta d + \varsigma z$ satisfies (3) in the main paper. Choosing $G(x) = x$ will do for Assumption 6. Now, Assumption SA-3 follows from the joint normality of (Z, D) combined with $f_{X|DZ}(x|d,z) = \phi((x - \psi(d,z))^2/\sqrt{\varsigma+1})/\sqrt{\varsigma+1}$ where $\phi(\cdot)$ is the density of standard normal.

D Bibliography

Bradic, J., V. Chernozhukov, W. K. Newey, and Y. Zhu (2022): "Minimax Semiparametric Learning with Approximate Sparsity," Working paper.

Carrasco, M., J.-P. Florens, and E. Renault (2007): "Econometrics Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, vol. 6, 5633 – 5751.

Cattaneo, M. D., R. K. Crump, and M. Jansson (2013): "Generalized Jackknife Estimators of Weighted Average Derivatives," *Journal of the American Statistical Association*, 108, 1243–1256.

Chernozhukov, V., W. Newey, and R. Singh (2022): "Automatic Debiased Machine Learning of Causal and Structural Effects," *Econometrica*, 90, 967–1027.

Deaner, B. (2018): "Proxy Controls and Panel Data," Working Paper.

Imbens, G. W. and W. K. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations models without Additivity," *Econometrica*, 77, 1481–1512.

Kress, R. (2014): *Linear Integral Equations*, New York, NY: Springer.

Lee, J. M. (2013): *Introduction to Smooth Manifolds*, Graduate Texts in Mathematics, Springer.

Mattner, L. (1993): "Some Incomplete But Boundedly Complete Location Families," *Annals of Statistics*, 21, 2158–2162.

Miao, W., Z. Geng, and E. J. Tchetgen Tchetgen (2018): "Identifying Causal Effects with Proxy Variables of an Unmeasured Confounder," *Biometrika*, 105, 987–993.

Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Elsevier, vol. 4, 2111 – 2245.

Newey, W. K. and J. L. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Szegő, G. (1975): *Orthogonal polynomials*, vol. 23, American Mathematical Soc.