# Supplemental Appendix for "Quantifying the Internal Validity of Weighted Estimands"

Alexandre Poirier* Tymon Słoczyński†

October 5, 2025

This document contains appendices that supplement the main text. Appendix B formalizes the connection between uniform causal representations and weakly causal estimands, defined in Blandhol, Bonney, Mogstad, and Torgovitsky (2022). Appendix C supplements the estimation and inference section (Section 6) and provides an algorithm for constructing confidence intervals. Appendices D–G provide proofs for results in the main text and for results in Appendices B and C. Appendix H contains additional calculations related to the weights for the TWFE estimand.

## B Weakly Causal Estimands and Uniform Causal Representations in $\mathcal{T}_{\text{all}}$

We now establish equivalence between *weakly causal* estimands as defined in Blandhol, Bonney, Mogstad, and Torgovitsky (2022) (henceforth, BBMT) and estimands that have uniform causal representations as in Theorem 3.1. As in BBMT, consider the case where $X$ has finite support and, as in this paper, assume the treatment is binary. We also abstract from choice groups denoted by $G$ in BBMT.

Since $X$ has finite support, let $\text{supp}(X \mid W_0 = 1) = \{x_1, \ldots, x_K\}$ and let $\tau \coloneqq (\tau(x_1), \ldots, \tau(x_K)) \in \mathbb{R}^K$ be the collection of CATEs. For $d \in \{0, 1\}$ let $\nu_d(x) \coloneqq \mathbb{E}[Y(d) \mid W_0 = 1, X = x]$ denote the *average structural function* (ASF) which also conditions on $W_0 = 1$, let $\nu_d \coloneqq (\nu_d(x_1), \ldots, \nu_d(x_K)) \in \mathbb{R}^K$, and let $\mathcal{M} \subseteq \mathbb{R}^{2K}$ be a set of possible ASFs such that $(\nu_0, \nu_1) \in \mathcal{M}$. We now state the definition of weakly causal estimands from BBMT (i.e., their Definition WC) in our setting which features binary treatments.

*Department of Economics, Georgetown University, alexandre.poirier@georgetown.edu
†Department of Economics, Brandeis University, tslocz@brandeis.edu

**Definition B.1.** The estimand $\beta$ is *weakly causal* if the following statements are true for all $(\nu_0, \nu_1) \in \mathcal{M}$:

1. If $\nu_1 - \nu_0 \geq \mathbf{0}_K$,[1] then $\beta \geq 0$.
2. If $\nu_1 - \nu_0 \leq \mathbf{0}_K$, then $\beta \leq 0$.

Thus, an estimand is weakly causal if all CATEs having the same sign implies the estimand also has that sign. Whether an estimand satisfies this condition also depends on $\mathcal{M}$, the set of allowed ASFs. To compare weak causality to our result on uniform causal representations, we consider $\mathcal{M}_{\text{all}} := \mathbb{R}^{2K}$, the unrestricted set of ASFs. The corresponding unrestricted set of CATE functions, which we denoted by $\mathcal{T}_{\text{all}}$, allows $\tau$ to be any vector in $\mathbb{R}^K$. We consider estimands characterized by our equation (1.1). We note that these estimands rule out "level dependence," i.e., that the estimand changes if potential outcomes $(Y(0), Y(1))$ are translated to $(Y(0) + c, Y(1) + c)$ for some constant $c \in \mathbb{R}$. For example, the IV estimand is generally level dependent when the propensity score $\mathbb{P}(Z = 1 \mid X)$ is nonlinear in $X$.[2] With these choices, we can show the two definitions are equivalent.

**Proposition B.1.** Let $\mu(a, \tau_0)$ be an estimand satisfying equation (1.1). Suppose Assumption 3.1 holds and that $a_{\max} < \infty$. Then $\mu(a, \tau_0)$ is weakly causal with $\mathcal{M} = \mathcal{M}_{\text{all}}$ if and only if it has a causal representation uniformly in $\mathcal{T}_{\text{all}}$.

The proof of this proposition hinges on the equivalence, under level independence, of weakly causal estimands and estimands with nonnegative weights, as in Proposition 4 of BBMT. Also, as shown in Theorem 3.1, estimands with nonnegative weights have a uniform causal representation in $\mathcal{T}_{\text{all}}$. Therefore, a weighted estimand has nonnegative weights if and only if it is weakly causal and if and only if it has a causal representation uniformly in $\mathcal{T}_{\text{all}}$. Thus, a weakly causal estimand admits a regular subpopulation $W^*$ such that the estimand measures the average effect of treatment over that subpopulation.

## C Details on Estimation and Inference

This appendix complements Section 6 in the main text. In it, we compute the limiting distribution of our estimated measure of internal validity and prove the validity of a nonstandard bootstrap algorithm for constructing confidence intervals around it. This

[1] Vector inequalities hold if they hold component-wise.
[2] See p. 17 in BBMT.

is done for the case where $\mathcal{T} = \mathcal{T}_{\text{all}}$. Estimation and inference for $\overline{P}(a, W_0; \{\tau_0\})$ is related to the question of estimation and inference in linear programs with estimated constraints. See Andrews, Roth, and Pakes (2023), Cox and Shi (2023), Fang, Santos, Shaikh, and Torgovitsky (2023), and Cho and Russell (2024) for recent advances on this topic.

As in Section 6, we consider the case where $X$ is discrete. We assume the existence of estimators for $a(\cdot)$ and $w_0(\cdot)$, but we do not assume knowledge of $\text{supp}(X \mid W_0 = 1)$. In this case, $a(x)$ and $w_0(x)$ are usually estimated "cell-by-cell" and their estimators are $\sqrt{n}$-consistent with a limiting Gaussian distribution. We will see that inference on $\overline{P}(a, W_0; \mathcal{T}_{\text{all}})$ is generally nonstandard and, as a result, most common bootstrap procedures fail.

We consider the following simple plug-in estimator:

$$\widehat{\overline{P}} = \frac{\frac{1}{n}\sum_{i=1}^{n} \widehat{a}(X_i)\widehat{w}_0(X_i)}{\frac{1}{n}\sum_{i=1}^{n} \widehat{w}_0(X_i) \cdot \max_{i:\widehat{w}_0(X_i)>c_n} \widehat{a}(X_i)},$$

where $c_n$ is a tuning parameter that converges to 0 as $n \to \infty$. Note that this tuning parameter is absent when $w_0$ is known, for example when $W_0 = 1$ almost surely. This estimator does not assume knowledge of the support of $X$ given $W_0 = 1$, but it can also be implemented by taking the maximum over $\text{supp}(X \mid W_0 = 1)$ when it is known.

Let $\text{supp}(X) = \{x_1, \ldots, x_K\}$, denote by $p_j = \mathbb{P}(X = x_j)$ the frequency of cell $j$, and let $\widehat{p}_j = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i = x_j)$ denote its sample frequency. Let $\widehat{\theta} = (\widehat{\mathbf{a}}, \widehat{\mathbf{w}}_0, \widehat{\mathbf{p}})$ where $\widehat{\mathbf{a}} = (\widehat{a}(x_1), \ldots, \widehat{a}(x_K))$, $\widehat{\mathbf{w}}_0 = (\widehat{w}_0(x_1), \ldots, \widehat{w}_0(x_K))$, and $\widehat{\mathbf{p}} = (\widehat{p}_1, \ldots, \widehat{p}_K)$. Let $\theta = (\mathbf{a}, \mathbf{w}_0, \mathbf{p})$ denote their population counterparts.

We make the following assumption on the behavior of the first-step estimators.

**Assumption C.1** (Preliminary estimators)**.** Let

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} \mathbb{Z}$$

where $\mathbb{Z} \coloneqq (\mathbb{Z}_{\mathbf{a}}, \mathbb{Z}_{\mathbf{w}_0}, \mathbb{Z}_X) \in \mathbb{R}^{3K}$ has a Gaussian distribution.

The above assumption is often satisfied when $X$ has finite support since estimators for $a(x_j)$ and $w_0(x_j)$ can be obtained using only the observations for which $X_i = x_j$. Note that the limiting distribution of $\mathbb{Z}_{\mathbf{w}_0}$ may be degenerate. For example, if $W_0 = 1$ almost surely, then $\widehat{w}_0(x) = w_0(x) = 1$ and thus $\mathbb{Z}_{\mathbf{w}_0} = \mathbf{0}_K$, a $K$-vector of zeros.

The next theorem shows the consistency of $\widehat{\overline{P}}$ and establishes the limiting distribution of this estimator. To simplify the exposition, we use $\overline{P}$ to denote $\overline{P}(a, W_0; \mathcal{T}_{\text{all}})$ in what follows.

**Theorem C.1** (Consistency and asymptotic distribution)**.** Suppose Assumption C.1 holds. Suppose $c_n = o(1)$ and $c_n\sqrt{n} \to \infty$ as $n \to \infty$. Suppose $\overline{P} \neq 0$. Then, $\widehat{\overline{P}}$ is consistent for $\overline{P}$ and

$$\sqrt{n}(\widehat{\overline{P}} - \overline{P}) \xrightarrow{d} \psi(\mathbb{Z}),$$

where $\psi$ is a continuous mapping defined by

$$\begin{aligned} \psi(\mathbb{Z}) \;=\; & \sum_{j=1}^{K} \frac{w_0(x_j)p_j}{\mathbb{P}(W_0=1)a_{\max}}\mathbb{Z}_{\mathbf{a}}(j) \;-\; \frac{\mathbb{E}[a(X) \mid W_0=1]}{a_{\max}^2} \max_{j\in\Psi_{\mathcal{X}+}} \mathbb{Z}_{\mathbf{a}}(j) \\ & + \sum_{j=1}^{K} \frac{(a(x_j)-\mathbb{E}[a(X) \mid W_0=1])p_j}{\mathbb{P}(W_0=1)a_{\max}}\mathbb{Z}_{\mathbf{w}_0}(j) \\ & + \sum_{j=1}^{K} \frac{(a(x_j)-\mathbb{E}[a(X) \mid W_0=1])w_0(x_j)}{\mathbb{P}(W_0=1)a_{\max}}\mathbb{Z}_{X}(j), \end{aligned} \tag{C.1}$$

where $\Psi_{\mathcal{X}+} = \{j \in \{1, \ldots, K\} : a(x_j) = a_{\max}\}$.

The mapping $\psi$ is linear if and only if $a(x)$ is maximized at a unique value $x \in \text{supp}(X \mid W_0 = 1)$, and nonlinear if multiple values maximize $a(x)$. The linearity of this mapping crucially affects the choice of the inference procedure. When $\psi$ is linear, the limiting distribution of $\widehat{\overline{P}}$ is Gaussian and common bootstrap procedures, such as the empirical bootstrap, are valid whenever they are valid for $\widehat{\theta}$.

However, when $a(x)$ is maximized at more than one value, the limiting distribution of $\widehat{\overline{P}}$ is nonlinear in $\mathbb{Z}$ and thus non-Gaussian. In this case, it can be shown (see Theorem 3.1 in Fang and Santos (2019)) that standard bootstrap approaches are invalid. However, the fact that the estimand $\overline{P}$ can be written as a Hadamard directionally differentiable mapping of $\theta$ implies that alternative bootstrap procedures, such as those proposed by Hong and Li (2018) and Fang and Santos (2019), can be applied to obtain valid inferences on $\overline{P}$.

We propose a bootstrap procedure that can be applied regardless of the linearity of $\psi$. In order to show its validity, we assume that the limiting distribution $\mathbb{Z}$ can be approximated via a bootstrap procedure.

**Assumption C.2** (Bootstrap for first-step estimators)**.** Let $\mathbb{Z}^* \coloneqq (\mathbb{Z}^*_{\mathbf{a}}, \mathbb{Z}^*_{\mathbf{w}_0}, \mathbb{Z}^*_X) \in \mathbb{R}^{3K}$ be a random vector such that $\mathbb{Z}^* \overset{p}{\rightsquigarrow} \mathbb{Z}$, where $\overset{p}{\rightsquigarrow}$ denotes convergence in probability conditioning on the data used to compute $\widehat{\theta}$.

This assumption is easily satisfied when $\widehat{\mathbf{p}}$ are sample frequencies, $(\widehat{\mathbf{a}}, \widehat{\mathbf{w}}_0)$ are cell-by-cell estimators that are asymptotically linear and Gaussian, and when $\mathbb{Z}^*$ is the distribution of these estimators under a standard bootstrap approach. For example, for the empirical bootstrap we can let $\mathbb{Z}^*_X(j) = \sqrt{n}(\widehat{p}^*_j - \widehat{p}_j)$ where $\widehat{p}^*_j = \frac{1}{n}\sum_{i=1}^{N} \mathbb{1}(X^*_i = x_j)$, where $(X^*_1, \ldots, X^*_n)$ are drawn from the empirical distribution of $(X_1, \ldots, X_n)$.

**Theorem C.2** (Bootstrap validity)**.** Suppose the assumptions of Theorem C.1 hold and that Assumption C.2 holds. Then,

$$\widehat{\psi}(\mathbb{Z}^*) \overset{p}{\rightsquigarrow} \psi(\mathbb{Z})$$

as $n \to \infty$, where $\widehat{\psi}$ is defined by

$$\begin{aligned}
\widehat{\psi}(\mathbb{Z}^*) \;=\; & \sum_{j=1}^{K} \frac{\widehat{w}_0(x_j)\widehat{p}_j}{\widehat{\mathbb{P}}(W_0 = 1)\widehat{a}_{\max}} \mathbb{Z}^*_{\mathbf{a}}(j) \;-\; \frac{\widehat{\mathbb{E}}[a(X) \mid W_0 = 1]}{\widehat{a}^2_{\max}} \max_{j \in \widehat{\Psi}_{\mathcal{X}^+}} \mathbb{Z}^*_{\mathbf{a}}(j) \\
& + \sum_{j=1}^{K} \frac{(\widehat{a}(x_j) - \widehat{\mathbb{E}}[a(X) \mid W_0 = 1])\widehat{p}_j}{\widehat{\mathbb{P}}(W_0 = 1)\widehat{a}_{\max}} \mathbb{Z}^*_{\mathbf{w}_0}(j) \\
& + \sum_{j=1}^{K} \frac{(\widehat{a}(x_j) - \widehat{\mathbb{E}}[a(X) \mid W_0 = 1])\widehat{w}_0(x_j)}{\widehat{\mathbb{P}}(W_0 = 1)\widehat{a}_{\max}} \mathbb{Z}^*_X(j),
\end{aligned} \tag{C.2}$$

where $\widehat{\mathbb{E}}[a(X) \mid W_0 = 1] = \frac{\frac{1}{n}\sum_{i=1}^{n} \widehat{a}(X_i)\widehat{w}_0(X_i)}{\frac{1}{n}\sum_{i=1}^{n} \widehat{w}_0(X_i)}$, $\widehat{\mathbb{P}}(W_0 = 1) = \frac{1}{n}\sum_{i=1}^{n} \widehat{w}_0(X_i)$, $\widehat{a}_{\max} = \max_{i:\widehat{w}_0(X_i) > c_n} \widehat{a}(X_i)$, and

$$\widehat{\Psi}_{\mathcal{X}^+} = \left\{ k \in \{1, \ldots, K\} : \widehat{a}(x_k) \geq \max_{i:\widehat{w}_0(X_i) > c_n} \widehat{a}(X_i) - \xi_n \right\},$$

where $\xi_n$ is a positive sequence satisfying $\xi_n = o(1)$ and $\xi_n\sqrt{n} \to \infty$ as $n \to \infty$.

The proof of Theorem C.2 shows that $\widehat{\Psi}_{\mathcal{X}^+}$ consistently estimates $\Psi_{\mathcal{X}^+}$ and thus satisfies the conditions of Theorem 3.2 in Fang and Santos (2019). The bootstrap procedure is valid whether the limiting distribution is Gaussian or not. If we assume $a(x)$ is maximized at a single value, standard bootstrap procedures can also be used to approximate the limiting distribution of $\widehat{\widehat{P}}$.

We propose the following bootstrap procedure to compute a one-sided $(1-\alpha)$ confidence interval for $\overline{P}$.

**Algorithm C.1** (One-sided confidence interval for $\overline{P}$)**.** We compute the confidence interval in three steps:

1. Compute $\widehat{\theta}$ and $\widehat{\overline{P}}$ using the random sample $\{(W_i, X_i)\}_{i=1}^n$;
2. For bootstrap samples $b = 1, \ldots, B$, compute $\widehat{\theta}^{*,b} = (\widehat{\mathbf{a}}^{*,b}, \widehat{\mathbf{w}}_0^{*,b}, \widehat{\mathbf{p}}^{*,b})$ and $\mathbb{Z}^{*,b} = \sqrt{n}(\widehat{\theta}^{*,b} - \widehat{\theta})$;
3. Compute $\widehat{q}_\alpha$, the $\alpha$ quantile of $\widehat{\psi}(\mathbb{Z}^{*,b})$, and report the interval $\left[0, \widehat{\overline{P}} - \widehat{q}_\alpha/\sqrt{n}\right]$.

We could also view these inferential problems through the lens of intersection or union bounds. For example, we can write

$$\overline{P}(a, W_0; \mathcal{T}_{\text{all}}) = \inf_{x \in \text{supp}(X|W_0=1)} \frac{\mathbb{E}[a(X)w_0(X)]}{\mathbb{E}[w_0(X)] \cdot a(x)} =: \inf_{x \in \text{supp}(X|W_0=1)} \overline{P}(x).$$

Computing a one-sided confidence interval for $\overline{P}(a, W_0; \mathcal{T}_{\text{all}})$ of the kind $[0, \widehat{\overline{P}}^+]$ can be cast as doing inference on intersection bounds. Chernozhukov, Lee, and Rosen (2013) offer methods for such problems. Equivalently, the computation of a one-sided confidence interval $[\widehat{\overline{P}}^-, 1]$ is related to inferential questions in union bounds: see Bei (2024). We leave all details for future work.

## D Proof of Theorem 4.2

We begin with a technical lemma that we use in the proof of Theorem 4.2.

**Lemma D.1.** Suppose Assumption 3.1 holds. Then,

1. The functions $\alpha \mapsto \mathbb{E}[T_\mu \mathbb{1}(T_\mu < \alpha) \mid W_0 = 1]$ and $\alpha \mapsto \mathbb{E}[T_\mu \mathbb{1}(T_\mu \geq \alpha) \mid W_0 = 1]$ are left-continuous.
2. The functions $\alpha \mapsto \mathbb{E}[T_\mu \mathbb{1}(T_\mu > \alpha) \mid W_0 = 1]$ and $\alpha \mapsto \mathbb{E}[T_\mu \mathbb{1}(T_\mu \leq \alpha) \mid W_0 = 1]$ are right-continuous.

*Proof of Lemma D.1.* The function $\alpha \mapsto \mathbb{E}[T_\mu \mathbb{1}(T_\mu < \alpha) \mid W_0 = 1]$ is left-continuous if for any strictly increasing sequence $\alpha_n \nearrow \alpha$ we have that $\mathbb{E}[T_\mu \mathbb{1}(T_\mu < \alpha_n) \mid W_0 = 1] \to \mathbb{E}[T_\mu \mathbb{1}(T_\mu < \alpha) \mid W_0 = 1]$. To see this holds, note that $f_n(t) \coloneqq t\mathbb{1}(t < \alpha_n) \to t\mathbb{1}(t < \alpha)$ since $t\mathbb{1}(t < \alpha_n) = 0$ for all $t \geq \alpha$, and $t\mathbb{1}(t < \alpha_n) = t$ whenever $t < \alpha$ for sufficiently large $n$. The random variable $|T_\mu \mathbb{1}(T_\mu < \alpha_n)|$ is dominated by $|T_\mu|$ and

$\mathbb{E}[|T_\mu| \mid W_0 = 1] < \infty$ by Assumption 3.1 and by $\mathbb{P}(W_0 = 1) > 0$. Therefore, by dominated convergence, $\mathbb{E}[T_\mu \mathbb{1}(T_\mu < \alpha_n) \mid W_0 = 1] \to \mathbb{E}[T_\mu \mathbb{1}(T_\mu < \alpha) \mid W_0 = 1]$ hence $\mathbb{E}[T_\mu \mathbb{1}(T_\mu < \alpha) \mid W_0 = 1]$ is left-continuous. The function $\alpha \mapsto \mathbb{E}[T_\mu \mathbb{1}(T_\mu \geq \alpha) \mid W_0 = 1]$ is also left-continuous because $\mathbb{E}[T_\mu \mathbb{1}(T_\mu \geq \alpha) \mid W_0 = 1] = \mathbb{E}[T_\mu \mid W_0 = 1] - \mathbb{E}[T_\mu \mathbb{1}(T_\mu < \alpha) \mid W_0 = 1]$. The lemma's second claim can be similarly shown by considering a sequence $\alpha_n \searrow \alpha$. $\square$

*Proof of Theorem 4.2.* We break down this proof into four cases.

**Case 1:** $\mu_0 \in \mathcal{S}(\tau_0; W_0)$ **and** $\mu_0 < E_0$

We want to maximize $\mathbb{P}(W^* = 1 \mid W_0 = 1)$ over the subpopulations $W^*$ in $\mathcal{W}(a; W_0, \{\tau_0\})$. Recall that $W^* \in \mathcal{W}(a; W_0, \{\tau_0\})$ if $\mu_0 = \mu(\underline{w}^*, \tau_0)$ and $W^* \in \text{SP}(W_0)$ hold, where $\underline{w}^*(X) = \mathbb{P}(W^* = 1 \mid X, W_0 = 1)$. Therefore,

$$\begin{aligned}
\overline{P}(a, W_0; \{\tau_0\}) &= \max_{W^* \in \mathcal{W}(a; W_0, \{\tau_0\})} \mathbb{P}(W^* = 1 \mid W_0 = 1) \\
&= \max_{W^* \in \{0,1\} : \mu_0 = \mu(\underline{w}^*, \tau_0), W^* \in \text{SP}(W_0)} \mathbb{E}[\underline{w}^*(X) \mid W_0 = 1] \\
&\leq \max_{W^* \in \{0,1\} : \mu_0 = \mu(\underline{w}^*, \tau_0)} \mathbb{E}[\underline{w}^*(X) \mid W_0 = 1] \\
&= \max_{\underline{w}^* : \mu_0 = \mu(\underline{w}^*, \tau_0), \underline{w}^*(X) \in [0,1]} \mathbb{E}[\underline{w}^*(X) \mid W_0 = 1].
\end{aligned}$$

We will first show a closed-form expression for an upper bound for $\overline{P}(a, W_0; \{\tau_0\})$. Then, we will show that this upper bound can be attained by a corresponding $W^+ \in \mathcal{W}(a; W_0, \{\tau_0\})$, and therefore it equals $\overline{P}(a, W_0; \{\tau_0\})$.

Before proceeding, let $\alpha^+ \coloneqq \inf\{\alpha \in \mathbb{R} : R(\alpha) \geq 0\}$ where $R(\alpha) \coloneqq \mathbb{E}[T_\mu \mathbb{1}(T_\mu \leq \alpha) \mid W_0 = 1]$. By construction, $\alpha^+ \geq 0$. By $\mu_0 < E_0$ we also have that $\alpha^+ < \infty$. By Lemma D.1, $R$ is a right-continuous function, and therefore $R(\alpha^+) = \lim_{\alpha \searrow \alpha^+} R(\alpha) \geq 0$. We now claim that $\alpha^+ > 0$. To show this claim, assume $\alpha^+ = 0$. Then, $0 \leq R(\alpha^+) = R(0) = \mathbb{E}[T_\mu \mathbb{1}(T_\mu \leq 0) \mid W_0 = 1] \leq 0$, which implies $\mathbb{P}(\tau_0(X) = \mu_0 \mid W_0 = 1) = 1$. This is ruled out by the assumption that $\mu_0 < E_0 = \mathbb{E}[\tau_0(X) \mid W_0 = 1] = \mu_0$. Therefore, $\alpha^+ > 0$.

Second, we show an upper bound for $\overline{P}(a, W_0; \{\tau_0\})$. For all $\underline{w}^*$ such that $\mu_0 =$

$\mu(\underline{w}^*, \tau_0)$ and $\underline{w}^*(X) \in [0, 1]$, we have that

$$
\begin{aligned}
&\mathbb{E}[\underline{w}^*(X) \mid W_0 = 1] \\
&= \frac{\mathbb{E}[\underline{w}^*(X)(\alpha^+ - T_\mu) \mid W_0 = 1]}{\alpha^+} + \frac{\mathbb{E}[\underline{w}^*(X) T_\mu \mid W_0 = 1]}{\alpha^+} \\
&= \frac{\mathbb{E}[\underline{w}^*(X)(\alpha^+ - T_\mu) \mid W_0 = 1]}{\alpha^+} \\
&= \frac{\mathbb{E}[\underline{w}^*(X)(\alpha^+ - T_\mu)\mathbb{1}(T_\mu \leq \alpha^+) \mid W_0 = 1]}{\alpha^+} + \frac{\mathbb{E}[\underline{w}^*(X)(\alpha^+ - T_\mu)\mathbb{1}(T_\mu > \alpha^+) \mid W_0 = 1]}{\alpha^+} \\
&\leq \frac{\mathbb{E}[1 \cdot (\alpha^+ - T_\mu)\mathbb{1}(T_\mu \leq \alpha^+) \mid W_0 = 1]}{\alpha^+} + \frac{\mathbb{E}[0 \cdot (\alpha^+ - T_\mu)\mathbb{1}(T_\mu > \alpha^+) \mid W_0 = 1]}{\alpha^+} \\
&= \mathbb{E}[\mathbb{1}(T_\mu \leq \alpha^+) \mid W_0 = 1] - \frac{\mathbb{E}[T_\mu \mathbb{1}(T_\mu \leq \alpha^+) \mid W_0 = 1]}{\alpha^+} \\
&=: P^+.
\end{aligned}
$$

The second equality follows from $\mu_0 = \mu(\underline{w}^*, \tau_0)$. The inequality follows from $\{0, 1\}$ being lower/upper bounds for $\underline{w}^*(X)$. Therefore, $\overline{P}(a, W_0; \{\tau_0\}) \leq P^+$.

Third, and finally, we show this inequality is binding by finding $W^+ \in \mathcal{W}(a; W_0, \{\tau_0\})$ such that $\mathbb{P}(W^+ = 1 \mid W_0 = 1) = P^+$.

We start by defining

$$
\underline{w}^+(X) = \mathbb{1}(T_\mu < \alpha^+) + \left(1 - \frac{R(\alpha^+)\mathbb{1}(\mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1) \neq 0)}{\alpha^+ \mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1)}\right) \mathbb{1}(T_\mu = \alpha^+).
$$

This function is bounded above by 1 because $R(\alpha^+) \geq 0$ and $\alpha^+ > 0$. To show $\underline{w}^+$ is bounded below by 0, consider cases where $\mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1)$ or $R(\alpha^+)$ equal and differ from 0. If $\mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1) = 0$ or $R(\alpha^+) = 0$, then $\underline{w}^+(X) \in \{0, 1\}$ and it is therefore bounded below by 0. If $\mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1) > 0$ and $R(\alpha^+) > 0$, then

$$
\begin{aligned}
1 - \frac{R(\alpha^+)}{\alpha^+ \mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1)} &= \frac{\alpha^+ \mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1) - \mathbb{E}[T_\mu \mathbb{1}(T_\mu \leq \alpha^+) \mid W_0 = 1]}{\alpha^+ \mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1)} \\
&= \frac{\mathbb{E}[T_\mu \mathbb{1}(T_\mu = \alpha^+) \mid W_0 = 1] - \mathbb{E}[T_\mu \mathbb{1}(T_\mu \leq \alpha^+) \mid W_0 = 1]}{\alpha^+ \mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1)} \\
&= \frac{-\mathbb{E}[T_\mu \mathbb{1}(T_\mu < \alpha^+) \mid W_0 = 1]}{\alpha^+ \mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1)}.
\end{aligned}
$$

By the definition of $\alpha^+$ as an infimum, we must have that $R(\alpha^+ - \varepsilon) < 0$ for all $\varepsilon > 0$, implying that $R(\alpha)$ is discontinuous at $\alpha^+$. By Lemma D.1, $\alpha \mapsto \mathbb{E}[T_\mu \mathbb{1}(T_\mu < \alpha) \mid$

$W_0 = 1]$ is left-continuous and satisfies $\mathbb{E}[T_\mu \mathbb{1}(T_\mu < \alpha) \mid W_0 = 1] \leq R(\alpha)$. Therefore, since $R(\alpha^+ - \varepsilon) < 0$ for all $\varepsilon > 0$, we must have that $\mathbb{E}[T_\mu \mathbb{1}(T_\mu < \alpha^+ - \varepsilon) \mid W_0 = 1] < 0$ for all $\varepsilon > 0$. Letting $\varepsilon \searrow 0$ yields that $\mathbb{E}[T_\mu \mathbb{1}(T_\mu < \alpha^+) \mid W_0 = 1] \leq 0$. Therefore $-\mathbb{E}[T_\mu \mathbb{1}(T_\mu < \alpha^+) \mid W_0 = 1]/(\alpha^+ \mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1)) \geq 0$ and $\underline{w}^+(X) \geq 0$.

Next, we compute

$$
\begin{aligned}
&\mathbb{E}[\underline{w}^+(X) \mid W_0 = 1] \\
&= \mathbb{P}(T_\mu < \alpha^+ \mid W_0 = 1) + \left(1 - \frac{R(\alpha^+)\mathbb{1}(\mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1) \neq 0)}{\alpha^+ \mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1)}\right) \mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1) \\
&= \mathbb{P}(T_\mu \leq \alpha^+ \mid W_0 = 1) - \frac{\mathbb{E}[T_\mu \mathbb{1}(T_\mu \leq \alpha^+) \mid W_0 = 1]}{\alpha^+} \mathbb{1}(\mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1) \neq 0) \\
&= \mathbb{P}(T_\mu \leq \alpha^+ \mid W_0 = 1) - \frac{\mathbb{E}[T_\mu \mathbb{1}(T_\mu \leq \alpha^+) \mid W_0 = 1]}{\alpha^+} \\
&= P^+.
\end{aligned}
$$

The indicator function disappears in the third equality because $\mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1) = 0$ implies $\mathbb{E}[T_\mu \mathbb{1}(T_\mu \leq \alpha^+) \mid W_0 = 1] = 0$ as shown above.

We next verify $\mu(\underline{w}^*, \tau_0) = \mu_0$. This condition is equivalent to $\mathbb{E}[\underline{w}^+(X) T_\mu \mid W_0 = 1] = 0$, which we verify here:

$$
\begin{aligned}
\mathbb{E}[\underline{w}^+(X) T_\mu \mid W_0 = 1] &= \mathbb{E}[T_\mu \mathbb{1}(T_\mu \leq \alpha^+) \mid W_0 = 1] \\
&\quad - \frac{R(\alpha^+)\mathbb{1}(\mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1) \neq 0)}{\alpha^+ \mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1)} \alpha^+ \mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1) \\
&= R(\alpha^+) - R(\alpha^+)\mathbb{1}(\mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1) \neq 0) \\
&= R(\alpha^+)\mathbb{1}(\mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1) = 0).
\end{aligned}
$$

Therefore, $\mathbb{E}[\underline{w}^+(X) T_\mu \mid W_0 = 1]$ equals 0 when $R(\alpha^+) = 0$. When $R(\alpha^+) > 0$, we have that $\mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1) > 0$ as shown earlier. So $\mathbb{E}[\underline{w}^+(X) T_\mu \mid W_0 = 1]$ is also equal to 0 in this case.

We conclude by showing that $\underline{w}^+(X)$ corresponds to $\mathbb{P}(W^+ = 1 \mid X, W_0 = 1)$ for some $W^+ \in \mathrm{SP}(W_0)$. Let $U \sim \mathrm{Unif}(0,1)$ satisfy $U \perp\!\!\!\perp (Y(1), Y(0), X, W_0)$ and define

$$
W^+ = \left(\mathbb{1}(T_\mu < \alpha^+) + \mathbb{1}\left(T_\mu = \alpha^+, U \leq 1 - \frac{R(\alpha^+)\mathbb{1}(\mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1) \neq 0)}{\alpha^+ \mathbb{P}(T_\mu = \alpha^+ \mid W_0 = 1)}\right)\right) \cdot W_0.
$$

By construction, $W^+ \in \{0,1\}$, $\mathbb{P}(W^+ = 1 \mid X, W_0 = 1) = \underline{w}^+(X)$, $\mathbb{P}(W_0 = 1 \mid$

$W^+ = 1) = 1$, and $W^+ \perp\!\!\!\perp (Y(1), Y(0)) \mid X, W_0 = 1$. Also, since $\mu_0 \in \mathcal{S}(\tau_0; W_0)$, $\mathbb{P}(T_\mu \leq 0 \mid W_0 = 1) = \mathbb{P}(\tau_0(X) \leq \mu_0 \mid W_0 = 1) > 0$. Since $\alpha^+ > 0$ we have that $\mathbb{P}(W^+ = 1 \mid W_0 = 1) \geq \mathbb{P}(T_\mu < \alpha^+ \mid W_0 = 1) \geq \mathbb{P}(T_\mu \leq 0 \mid W_0 = 1) > 0$. Therefore $W^+$ is a regular subpopulation of $W_0$ for which $\mathbb{P}(W^+ = 1 \mid W_0 = 1) = P^+$, hence $P^+$ is the maximum.

**Case 2:** $\mu_0 \in \mathcal{S}(\tau_0; W_0)$ **and** $\mu_0 > E_0$

As in Case 1, $\overline{P}(a, W_0; \{\tau_0\}) \leq \max_{\underline{w}^*:\mu(\underline{w}^*,\tau_0)=\mu_0,\underline{w}^*(X)\in[0,1]} \mathbb{E}[\underline{w}^*(X) \mid W_0 = 1]$.

Let $\alpha^- = \sup\{\alpha \in \mathbb{R} : L(\alpha) \leq 0\}$ where $L(\alpha) = \mathbb{E}[T_\mu \mathbb{1}(T_\mu \geq \alpha) \mid W_0 = 1]$. By construction, $\alpha^- \leq 0$ and by $\mu_0 > E_0$ we have that $\alpha^- > -\infty$. By Lemma D.1, $L$ is a left-continuous function, and therefore $L(\alpha^-) = \lim_{\alpha \nearrow \alpha^-} L(\alpha) \leq 0$. Similarly to Case 1, we can show that $\alpha^- < 0$.

We now show an upper bound for $\overline{P}(a, W_0; \{\tau_0\})$. For all $\underline{w}^*$ such that $\mu(\underline{w}^*, \tau_0) = \mu_0$ and $\underline{w}^*(X) \in [0, 1]$, we have that

$$
\begin{aligned}
&\mathbb{E}[\underline{w}^*(X) \mid W_0 = 1] \\
&\leq \frac{\mathbb{E}[1 \cdot (\alpha^- - T_\mu)\mathbb{1}(T_\mu \geq \alpha^-) \mid W_0 = 1]}{\alpha^-} + \frac{\mathbb{E}[0 \cdot (\alpha^- - T_\mu)\mathbb{1}(T_\mu < \alpha^-) \mid W_0 = 1]}{\alpha^-} \\
&= \mathbb{E}[\mathbb{1}(T_\mu \geq \alpha^-) \mid W_0 = 1] - \frac{\mathbb{E}[T_\mu \mathbb{1}(T_\mu \geq \alpha^-) \mid W_0 = 1]}{\alpha^-} \\
&=: P^-,
\end{aligned}
$$

which follows a similar argument as above. This implies $\overline{P}(a, W_0; \{\tau_0\}) \leq P^-$. We now show that this inequality is an equality by finding $W^- \in \mathcal{W}(a; W_0, \{\tau_0\})$ such that $\mathbb{P}(W^- = 1 \mid W_0 = 1) = P^-$. Let

$$
\underline{w}^-(X) = \mathbb{1}(T_\mu > \alpha^-) + \left(1 - \frac{L(\alpha^-)\mathbb{1}(\mathbb{P}(T_\mu = \alpha^- \mid W_0 = 1) \neq 0)}{\alpha^- \mathbb{P}(T_\mu = \alpha^- \mid W_0 = 1)}\right) \mathbb{1}(T_\mu = \alpha^-).
$$

The rest of the proof symmetrically follows the one for the previous case.

**Case 3:** $\mu_0 = E_0 \in \mathcal{S}(\tau_0; W_0)$

Since $W^* = W_0 \in \mathrm{SP}(W_0)$, we have that $\mu_0 = \mathbb{E}[Y(1) - Y(0) \mid W_0 = 1]$ and thus $\mathbb{P}(W^* = 1 \mid W_0 = 1)$ is maximized at 1.

**Case 4:** $\mu_0 \notin \mathcal{S}(\tau_0; W_0)$

By Theorem 3.2, there does not exist a regular subpopulation $W^*$ satisfying $\mu_0 = \mathbb{E}[Y(1) - Y(0) \mid W^* = 1]$ and therefore the supremum equals 0 by its definition. $\square$

## E Proofs for Section 5

*Proof of Proposition 5.3.* We begin by noting that

$$\beta_{\text{TWFE}} = \frac{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\ddot{D}_t Y_t]}{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\ddot{D}_t^2]} = \frac{\sum_{t=1}^{T}\mathbb{E}[\ddot{D}_t Y_t \mid P=t]\mathbb{P}(P=t)}{\sum_{t=1}^{T}\mathbb{E}[\ddot{D}_t^2 \mid P=t]\mathbb{P}(P=t)} = \frac{\mathbb{E}[\ddot{D}Y]}{\mathbb{E}[\ddot{D}^2]},$$

where the second equality follows from the uniform distribution of $P$ which is independent from $(\ddot{D}_t, Y_t)$ for all $t \in \{1, \dots, T\}$. The third equality follows from defining $\ddot{D} \coloneqq \ddot{D}_P$. We also note that

$$\begin{aligned}
\ddot{D} &= D_P - \frac{1}{T}\sum_{s=1}^{T} D_s - \sum_{t=1}^{T}\mathbb{E}[D_t]\mathbb{1}(P=t) + \sum_{s=1}^{T}\mathbb{E}[D_s]\mathbb{E}[\mathbb{1}(P=s)] \\
&= D - \frac{1}{T}\sum_{s=1}^{T}\mathbb{1}(G \le s) - \mathbb{E}[D \mid P] + \mathbb{E}[D] \\
&= D - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D].
\end{aligned}$$

The third equality follows from $\mathbb{E}[D \mid G] = \mathbb{E}[\mathbb{1}(G \le P) \mid G] = \frac{1}{T}\sum_{s=1}^{T}\mathbb{1}(G \le s) = \frac{1}{T}\sum_{s=1}^{T} D_s$. We break down the rest of this proof into four steps.

**Step 1: Splitting the Numerator in Two**
We have that

$$\begin{aligned}
\mathbb{E}[\ddot{D}Y] &= \mathbb{E}[\ddot{D}(Y(0) + D(Y(1) - Y(0)))] \\
&= \mathbb{E}[\ddot{D}\mathbb{E}[Y(0) \mid G, P]] + \mathbb{E}[\ddot{D}D\mathbb{E}[Y(1) - Y(0) \mid G, P]].
\end{aligned}$$

The first equality follows from $Y = Y(0) + D(Y(1) - Y(0))$ and the second from iterated expectations and $\mathbb{E}[D \mid G, P] = D$.

**Step 2: First Numerator Term**
We have that

$$\mathbb{E}[\ddot{D}\mathbb{E}[Y(0) \mid G, P]] = \mathbb{E}[\ddot{D}\theta(G, P)] = \mathbb{E}[\ddot{D}\ddot{\theta}(G, P)],$$

where $\theta(G, P) = \mathbb{E}[Y(0) \mid G, P]$. The second equality follows by properties of projections and from defining $\ddot{\theta}(G,P)$ as follows:

$$\begin{aligned}\ddot{\theta}(G,P) &:= \theta(G,P) - \mathbb{E}[\theta(G,P) \mid G] - \mathbb{E}[\theta(G,P) \mid P] + \mathbb{E}[\theta(G,P)] \\ &= \mathbb{E}[Y(0) \mid G,P] - \mathbb{E}[Y(0) \mid G] - \mathbb{E}[Y(0) \mid P] + \mathbb{E}[Y(0)].\end{aligned}$$

Then, we note that

$$\begin{aligned}\ddot{\theta}(g',t') &= \mathbb{E}[Y(0) \mid G=g', P=t'] - \mathbb{E}[Y(0) \mid G=g'] - \mathbb{E}[Y(0) \mid P=t'] + \mathbb{E}[Y(0)] \\ &= \mathbb{E}[Y_{t'}(0) \mid G=g'] - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_t(0) \mid G=g'] \\ &- \sum_{g\in\mathcal{G}}\left(\mathbb{E}[Y_{t'}(0) \mid G=g] - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_t(0) \mid G=g]\right)\mathbb{P}(G=g) \\ &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_{t'}(0) - Y_t(0) \mid G=g'] - \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_{t'}(0) - Y_t(0) \mid G]\right].\end{aligned}$$

Assumption 5.3.2 implies that for any pair $t, t' \in \{1, \ldots, T\}$ and any $g' \in \mathcal{G}$

$$\mathbb{E}[Y_{t'}(0) - Y_t(0) \mid G=g'] = \mathbb{E}[Y_{t'}(0) - Y_t(0)].$$

This can be shown for $t' > t$ by writing $\mathbb{E}[Y_{t'}(0) - Y_t(0) \mid G=g'] = \sum_{s=t+1}^{t'}\mathbb{E}[Y_s(0) - Y_{s-1}(0) \mid G=g'] = \sum_{s=t+1}^{t'}\mathbb{E}[Y_s(0) - Y_{s-1}(0)] = \mathbb{E}[Y_{t'}(0) - Y_t(0)]$. Similar derivations show this holds for $t' < t$. The case where $t' = t$ is trivial. Therefore,

$$\begin{aligned}\theta(g',t') &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_{t'}(0) - Y_t(0) \mid G=g'] - \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_{t'}(0) - Y_t(0) \mid G]\right] \\ &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_{t'}(0) - Y_t(0)] - \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[Y_{t'}(0) - Y_t(0)]\right] \\ &= 0\end{aligned}$$

for all $(g',t') \in \mathcal{G} \times \{1, \ldots, T\}$, which implies $\mathbb{E}[\ddot{D}\mathbb{E}[Y(0) \mid G,P]] = 0$.

**Step 3: Second Numerator Term**

We can write

$$\begin{aligned}\ddot{D}D &= (D - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D])D \\ &= (1 - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D])\mathbb{P}(D = 1 \mid G, P) \end{aligned} \tag{E.1}$$

by $D^2 = D = \mathbb{E}[D \mid G, P]$. Thus,

$$\begin{aligned}&\mathbb{E}[\ddot{D}D\mathbb{E}[Y(1) - Y(0) \mid G, P]] \\ &= \mathbb{E}[(1 - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D])\mathbb{E}[Y(1) - Y(0) \mid G, P, D = 1]\mathbb{P}(D = 1 \mid G, P)].\end{aligned}$$

**Step 4: Denominator**
In this step, we show that

$$\mathbb{E}[\ddot{D}^2] = \mathbb{E}[\ddot{D}D] = \mathbb{E}[(1 - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D])\mathbb{P}(D = 1 \mid G, P)].$$

The first equality is obtained from properties of linear projections and the second follows from equation (E.1).

We conclude the proof by noting the equivalence of integrating over the distribution of $P$ and averages over time periods, which shows the equivalence between $\beta_{\text{TWFE}}$, the expression in Proposition 5.3, and

$$\frac{\mathbb{E}[(1 - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D]) \cdot \mathbb{E}[Y(1) - Y(0) \mid G, P, D = 1] \cdot \mathbb{P}(D = 1 \mid G, P)]}{\mathbb{E}[(1 - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D]) \cdot \mathbb{P}(D = 1 \mid G, P)]}.$$

□

*Proof of Proposition 5.4.* Proposition 5.3 and $\mathbb{P}(D = 1 \mid G, P) = D$ yields

$$\beta_{\text{TWFE}} = \frac{\mathbb{E}[(1 - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D]) \cdot D \cdot \mathbb{E}[Y(1) - Y(0) \mid G, P, D = 1]]}{\mathbb{E}[(1 - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D]) \cdot D]}.$$

Since $\mathbb{E}[Y(1) - Y(0) \mid G, P, D = 1] = \mathbb{E}[Y(1) - Y(0) \mid G, D = 1]$ by assumption, we can use the law of iterated expectations to obtain

$$\beta_{\text{TWFE}} = \frac{\mathbb{E}[\mathbb{E}[(1 - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D]) \cdot D \mid G] \cdot \mathbb{E}[Y(1) - Y(0) \mid G, D = 1]]}{\mathbb{E}[\mathbb{E}[(1 - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D]) \cdot D \mid G]]}.$$

We now calculate the conditional expectation $\mathbb{E}[(1 - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D]) \cdot$

$D \mid G = g]$ for $g \in \mathcal{G}$. If $g = +\infty$, then this conditional expectation is 0 by construction, so we focus on the case where $g \in \{2, \dots, T\}$. For these derivations, we let $F_G(p) := \mathbb{P}(G \leq p)$ denote the cdf of $G$ at $p$. We have that:

$$
\begin{aligned}
&\mathbb{E}[(1 - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D])D \mid G = g] \\
&= \mathbb{E}[D \mid G = g]\left(1 - \mathbb{E}[D \mid G = g] - \frac{\mathbb{E}[\mathbb{E}[D \mid P]D \mid G = g]}{\mathbb{E}[D \mid G = g]} + \mathbb{E}[D]\right) \\
&= \mathbb{E}[D \mid G = g]\Bigg(1 - \mathbb{E}[\mathbb{1}(G \leq P) \mid G = g] - \frac{\mathbb{E}[F_G(P)\mathbb{1}(G \leq P) \mid G = g]}{\mathbb{E}[\mathbb{1}(G \leq P) \mid G = g]} \\
&\quad + \mathbb{E}[\mathbb{E}[\mathbb{1}(G \leq P) \mid P]]\Bigg) \\
&= \mathbb{E}[D \mid G = g]\left(1 - \mathbb{E}[\mathbb{1}(g \leq P)] - \frac{\mathbb{E}[F_G(P)\mathbb{1}(g \leq P)]}{\mathbb{E}[\mathbb{1}(g \leq P)]} + \mathbb{E}[F_G(P)]\right) \\
&= \mathbb{E}[D \mid G = g]\,(1 - \mathbb{E}[\mathbb{1}(g \leq P)] - \mathbb{E}[F_G(P) \mid g \leq P] \\
&\quad + \mathbb{E}[F_G(P) \mid g \leq P]\mathbb{E}[\mathbb{1}(g \leq P)] + \mathbb{E}[F_G(P) \mid g > P]\mathbb{E}[\mathbb{1}(g > P)]) \\
&= \mathbb{E}[D \mid G = g]\mathbb{E}[\mathbb{1}(g > P)](1 + \mathbb{E}[F_G(P) \mid g > P] - \mathbb{E}[F_G(P) \mid g \leq P]) \\
&= \mathbb{E}[D \mid G = g](1 - \mathbb{E}[D \mid G = g])(1 + \mathbb{E}[D \mid g > P] - \mathbb{E}[D \mid g \leq P]) \\
&= \mathbb{P}(D = 1 \mid G = g) \cdot \mathbb{P}(D = 0 \mid G = g) \cdot (\mathbb{P}(D = 1 \mid P < g) + \mathbb{P}(D = 0 \mid P \geq g)).
\end{aligned}
$$

The first equality follows from $\mathbb{E}[D \mid G = g] > 0$ for $g \in \{2, \dots, T\}$, the second from $D = \mathbb{1}(G \leq P)$ and the law of iterated expectations, the third from $G \perp\!\!\!\perp P$, the fourth from the law of iterated expectations, the fifth from combining terms, the sixth from the law of iterated expectations again, and the last line is obtained by the fact that $D \in \{0, 1\}$. The representation in Proposition 5.4 follows. □

## F Proofs for Appendix B

*Proof of Proposition B.1.* By Theorem 3.1, $\mu(a, \tau_0)$ has a uniform causal representation in $\mathcal{T}_{\text{all}}$ if and only if $a(x_k) \geq 0$ for $k \in \{1, \dots, K\}$. Therefore, it is sufficient to show the equivalence between weakly causal estimands and estimands with nonnegative weights. A similar result was shown in Proposition 4 of BBMT, but we nevertheless provide a proof here to account for the slight differences in notation.

Suppose $\mu(a, \tau_0)$ is weakly causal. Let $\nu_1 = (\mathbb{1}(a(x_1) < 0), \dots, \mathbb{1}(a(x_K) < 0))$ and $\nu_0 = \mathbf{0}_K$. Trivially, $(\nu_1, \nu_0) \in \mathcal{M}_{\text{all}}$ and $\tau^- := \nu_1 - \nu_0 \in \mathcal{T}_{\text{all}}$, where $\tau^- \geq \mathbf{0}_K$. Since

$\mu(a, \tau_0)$ is weakly causal, $\mu(a, \tau^-) \geq 0$ where

$$\begin{aligned}\mu(a, \tau^-) &= \frac{\mathbb{E}[a(X)\tau^-(X) \mid W_0 = 1]}{\mathbb{E}[a(X) \mid W_0 = 1]} \\ &= \frac{1}{\mathbb{E}[a(X) \mid W_0 = 1]} \sum_{k=1}^{K} a(x_k)\mathbb{1}(a(x_k) < 0)\mathbb{P}(X = x_k \mid W_0 = 1) \geq 0.\end{aligned}$$

This implies $a(x_k) \geq 0$ for all $k \in \{1, \ldots, K\}$. Thus, $\mu(a, \tau_0)$ has a uniform causal representation in $\mathcal{T}_{\text{all}}$.

Now suppose $\mu(a, \tau_0)$ has a uniform causal representation in $\mathcal{T}_{\text{all}}$, or that $a(x_k) \geq 0$ for $k \in \{1, \ldots, K\}$. Then, for any $(\nu_0, \nu_1) \in \mathcal{M}_{\text{all}}$ such that $\tau \coloneqq \nu_1 - \nu_0 \geq \mathbf{0}_K$,

$$\mu(a, \tau) = \frac{1}{\mathbb{E}[a(X) \mid W_0 = 1]} \sum_{k=1}^{K} a(x_k)\tau(x_k)\mathbb{P}(X = x_k \mid W_0 = 1) \geq 0.$$

The inequality holds because $a(x_k)$ and $\tau(x_k)$ are nonnegative for all $k \in \{1, \ldots, K\}$. This last inequality is reversed if we instead assume that $\tau \leq \mathbf{0}_K$. Thus, $\mu(a, \tau_0)$ is weakly causal. $\square$

## G Proofs for Appendix C

We use the following lemma in the proof of Theorem C.1.

**Lemma G.1.** Let $\theta = (\theta(1), \ldots, \theta(K)) \in \mathbb{R}^K$ and define the mapping $\phi : \mathbb{R}^K \to \mathbb{R}$ by $\phi(\theta) = \max_{j \in \{1,\ldots,K\}} \theta(j)$. Then, $\phi$ is Hadamard directionally differentiable for all $\theta \in \mathbb{R}^K$ tangentially to $\mathbb{R}^K$ with directional derivative at $\theta$ in direction $h \in \mathbb{R}^K$

$$\phi'_\theta(h) = \max_{j \in \arg\max_{k \in \{1,\ldots,K\}} \theta(k)} h(j).$$

*Proof of Lemma G.1.* Let $h_n \to h \in \mathbb{R}^K$ and $t_n \searrow 0$ as $n \to \infty$. Then,

$$t_n^{-1}(\phi(\theta + t_n h_n) - \phi(\theta)) = t_n^{-1}\left(\max_{k \in \{1,\ldots,K\}}(\theta(k) + t_n h_n(k)) - \max_{k \in \{1,\ldots,K\}} \theta(k)\right).$$

Let $\Theta_{\max} = \{j \in \{1, \ldots, K\} : \theta(j) = \max_{k \in \{1,\ldots,K\}} \theta(k)\}$ and let $j_{\max}$ be an element of

$\Theta_{\max}$. Then, $\max_{k\in\{1,\ldots,K\}}\theta(k)=\theta(j_{\max})$ and thus

$$\frac{\phi(\theta+t_nh_n)-\phi(\theta)}{t_n}=\max\left\{\frac{\theta(1)-\theta(j_{\max})}{t_n}+h_n(1),\ldots,\frac{\theta(K)-\theta(j_{\max})}{t_n}+h_n(K)\right\}.$$

For each $j\in\Theta_{\max}$, $(\theta(j)-\theta(j_{\max}))/t_n+h_n(j)=h_n(j)\to h(j)$. For each $j\notin\Theta_{\max}$, $(\theta(j)-\theta(j_{\max}))/t_n\to-\infty$ since $\theta(j)-\theta(j_{\max})<0$ and $t_n\searrow 0$. Therefore, by continuity of the maximum operator in its arguments, $t_n^{-1}(\phi(\theta+t_nh_n)-\phi(\theta))\to\max_{j\in\Theta_{\max}}h(j)$. □

*Proof of Theorem C.1.* We begin by showing the consistency of $\widehat{\overline{P}}$ for $\overline{P}$.

**Part 1: Consistency**

The estimator $\frac{1}{n}\sum_{i=1}^n\widehat{a}(X_i)\widehat{w}_0(X_i)$ is consistent for $\mathbb{E}[a(X)w_0(X)]$ since its components are assumed consistent by Assumption C.1, and by the continuous mapping theorem. The consistency of $\frac{1}{n}\sum_{i=1}^n\widehat{w}_0(X_i)$ for $\mathbb{E}[w_0(X)]$ is similarly established.

We now consider the maximum term in the denominator. We can write

$$\max_{i:\widehat{w}_0(X_i)>c_n}\widehat{a}(X_i)=\max_{x:\frac{1}{n}\sum_{i=1}^n\mathbb{1}(X_i=x)>0,\widehat{w}_0(x)>c_n}\widehat{a}(x).$$

Let $\mathcal{X}^+:=\mathrm{supp}(X\mid W_0=1)=\{x\in\mathrm{supp}(X):w_0(x)>0\}$ and let $\widehat{\mathcal{X}}^+=\{x:\frac{1}{n}\sum_{i=1}^n\mathbb{1}(X_i=x)>0,\widehat{w}_0(x)>c_n\}$. We first show that $\mathbb{P}(\widehat{\mathcal{X}}^+=\mathcal{X}^+)\to 1$ as $n\to\infty$. To see this, first consider $x_j\in\mathcal{X}^+$. Then,

$$\begin{aligned}\mathbb{P}(x_j\in\widehat{\mathcal{X}}^+)&=\mathbb{P}\left(\left\{\frac{1}{n}\sum_{i=1}^n\mathbb{1}(X_i=x_j)>0\right\}\cap\{\widehat{w}_0(x_j)>c_n\}\right)\\&\geq\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n\mathbb{1}(X_i=x_j)>0\right)+\mathbb{P}\left(\widehat{w}_0(x_j)>c_n\right)-1,\end{aligned}$$

following an application of Bonferroni's inequality.

We have that $\mathbb{P}(\frac{1}{n}\sum_{i=1}^n\mathbb{1}(X_i=x_j)>0)\to 1$ since $\frac{1}{n}\sum_{i=1}^n\mathbb{1}(X_i=x_j)\xrightarrow{p}p_j>0$. We also have that $\mathbb{P}(\widehat{w}_0(x_j)>c_n)=\mathbb{P}(\widehat{w}_0(x_j)-c_n>0)\to 1$ because $\widehat{w}_0(x_j)-c_n\xrightarrow{p}w_0(x_j)>0$ by $c_n=o(1)$ and $w_0(x_j)>0$, which follows from $x_j\in\mathcal{X}^+$. Therefore, $\mathbb{P}(x_j\notin\widehat{\mathcal{X}}^+)=1-\mathbb{P}(x_j\in\widehat{\mathcal{X}}^+)\leq 1-(\mathbb{P}(\frac{1}{n}\sum_{i=1}^n\mathbb{1}(X_i=x_j)>0)+\mathbb{P}(\widehat{w}_0(x_j)>c_n)-1)\to 0$ as $n\to\infty$.

Now let $x_j \notin \mathcal{X}^+$. Then

$$\begin{aligned}
\mathbb{P}(x_j \notin \widehat{\mathcal{X}}^+) &= \mathbb{P}\left(\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}(X_i = x_j) = 0\right\} \cup \{\widehat{w}_0(x_j) \leq c_n\}\right) \\
&\geq \mathbb{P}(\widehat{w}_0(x_j) \leq c_n) = \mathbb{P}(\sqrt{n}\widehat{w}_0(x_j) \leq \sqrt{n}c_n).
\end{aligned}$$

By Assumption C.1, $\sqrt{n}\widehat{w}_0(x_j) = \sqrt{n}(\widehat{w}_0(x_j) - w_0(x_j)) \xrightarrow{d} \mathbb{Z}_{\mathbf{w}_0}(j) = O_p(1)$, since $w_0(x_j) = 0$ for $x_j \notin \mathcal{X}^+$. Also $\sqrt{n}c_n \to +\infty$ by the theorem's assumption. Therefore, $\mathbb{P}(\sqrt{n}\widehat{w}_0(x_j) \leq \sqrt{n}c_n) \to 1$ and $\mathbb{P}(x_j \in \widehat{\mathcal{X}}^+) \to 0$ as $n \to \infty$. Because of this,

$$\begin{aligned}
\mathbb{P}(\widehat{\mathcal{X}}^+ = \mathcal{X}^+) &= \mathbb{P}\left(\bigcap_{x_j \in \mathcal{X}^+}\{x_j \in \widehat{\mathcal{X}}^+\} \cap \bigcap_{x_j \notin \mathcal{X}^+}\{x_j \notin \widehat{\mathcal{X}}^+\}\right) \\
&= 1 - \mathbb{P}\left(\bigcup_{x_j \in \mathcal{X}^+}\{x_j \notin \widehat{\mathcal{X}}^+\} \cup \bigcup_{x_j \notin \mathcal{X}^+}\{x_j \in \widehat{\mathcal{X}}^+\}\right) \\
&\geq 1 - \left(\sum_{j : x_j \in \mathcal{X}^+}\mathbb{P}(x_j \notin \widehat{\mathcal{X}}^+) + \sum_{j : x_j \notin \mathcal{X}^+}\mathbb{P}(x_j \in \widehat{\mathcal{X}}^+)\right) \\
&\to 1 - 0 = 1.
\end{aligned}$$

Thus, we obtain $\mathbb{P}\left(\max_{x \in \widehat{\mathcal{X}}^+}\widehat{a}(x) = \max_{x \in \mathcal{X}^+}\widehat{a}(x)\right) \geq \mathbb{P}(\widehat{\mathcal{X}}^+ \in \mathcal{X}^+) \to 1$, which yields

$$\max_{i : \widehat{w}_0(X_i) > c_n}\widehat{a}(X_i) = \max_{x \in \widehat{\mathcal{X}}^+}\widehat{a}(x) = \max_{x \in \mathcal{X}^+}\widehat{a}(x) + o_p(1).$$

By the consistency of $\widehat{\mathbf{a}}$ for $\mathbf{a}$, the continuity of the maximum operator, and the continuous mapping theorem, $\max_{x \in \mathcal{X}^+}\widehat{a}(x) \xrightarrow{p} \max_{x \in \mathcal{X}^+} a(x)$. Because $\mathcal{X}^+ = \mathrm{supp}(X \mid W_0 = 1)$ is a finite set, we also have that $\max_{x \in \mathcal{X}^+} a(x) = \sup(\mathrm{supp}(a(X) \mid W_0 = 1))$. Another application of the continuous mapping theorem suffices to show that $\widehat{\overline{P}}$ is consistent for $\overline{P}$.

**Part 2: Asymptotic Distribution**

We first establish the joint limiting distribution of terms (i) $\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n}\widehat{a}(X_i)\widehat{w}_0(X_i) - \mathbb{E}[a(X)w_0(X)])$, (ii) $\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n}\widehat{w}_0(X_i) - \mathbb{E}[w_0(X_i)])$, and (iii) $\sqrt{n}(\max_{i:\widehat{w}_0(X_i)>c_n}\widehat{a}(X_i) -$

$\max_{x\in\mathcal{X}^+} a(x))$. The terms (i) and (ii) can be expanded as follows:

$$
\begin{aligned}
&\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{a}(X_i)\widehat{w}_0(X_i) - \mathbb{E}[a(X)w_0(X)]\right)\\
&\quad= \sqrt{n}\left(\sum_{j=1}^{K}\widehat{a}(x_j)\widehat{w}_0(x_j)\widehat{p}_j - \sum_{j=1}^{K}a(x_j)w_0(x_j)p_j\right)\\
&\quad= \sum_{j=1}^{K} w_0(x_j)p_j\sqrt{n}(\widehat{a}(x_j) - a(x_j)) + \sum_{j=1}^{K} a(x_j)p_j\sqrt{n}(\widehat{w}_0(x_j) - w_0(x_j))\\
&\quad+ \sum_{j=1}^{K} a(x_j)w_0(x_j)\sqrt{n}(\widehat{p}_j - p_j)) + o_p(1) \qquad \text{(G.1)}
\end{aligned}
$$

and

$$
\begin{aligned}
\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{w}_0(X_i) - \mathbb{E}[w_0(X)]\right) &= \sqrt{n}\left(\sum_{j=1}^{K}\widehat{w}_0(x_j)\widehat{p}_j - \sum_{j=1}^{K}w_0(x_j)p_j\right)\\
&= \sum_{j=1}^{K}\left(p_j\sqrt{n}(\widehat{w}_0(x_j) - w_0(x_j)) + w_0(x_j)\sqrt{n}(\widehat{p}_j - p_j)\right) + o_p(1). \qquad \text{(G.2)}
\end{aligned}
$$

For term (iii), we use the expansion

$$
\begin{aligned}
\sqrt{n}\left(\max_{i:\widehat{w}_0(X_i)>c_n}\widehat{a}(X_i) - \max_{x\in\mathcal{X}^+}a(x)\right) &= \sqrt{n}\left(\max_{i:\widehat{w}_0(X_i)>c_n}\widehat{a}(X_i) - \max_{x\in\mathcal{X}^+}\widehat{a}(x)\right) \qquad \text{(G.3)}\\
&+ \sqrt{n}\left(\max_{x\in\mathcal{X}^+}\widehat{a}(x) - \max_{x\in\mathcal{X}^+}a(x)\right). \qquad \text{(G.4)}
\end{aligned}
$$

The term in (G.3) is of order $o_p(1)$ because

$$
\begin{aligned}
\mathbb{P}\left(\sqrt{n}\left(\max_{i:\widehat{w}_0(X_i)>c_n}\widehat{a}(X_i) - \max_{x\in\mathcal{X}^+}\widehat{a}(x)\right) = 0\right) &= \mathbb{P}\left(\max_{x\in\widehat{\mathcal{X}}^+}\widehat{a}(x) = \max_{x\in\mathcal{X}^+}\widehat{a}(x)\right)\\
&\geq \mathbb{P}(\widehat{\mathcal{X}}^+ \in \mathcal{X}^+) \to 1,
\end{aligned}
$$

as shown earlier.

The term in (G.4) can be analyzed using Theorem 2.1 in Fang and Santos (2019), which generalizes the delta method to the class of Hadamard directionally differentiable functions. Using Lemma G.1, we have that

$$\sqrt{n}\left(\max_{x\in\mathcal{X}^+}\widehat{a}(x)-\max_{x\in\mathcal{X}^+}a(x)\right)=\max_{x_j\in\arg\max_{x\in\mathcal{X}^+}a(x)}\sqrt{n}\left(\widehat{a}(x_j)-a(x_j)\right)+o_p(1)$$
$$=:\max_{j\in\Psi_{\mathcal{X}^+}}\sqrt{n}\left(\widehat{a}(x_j)-a(x_j)\right)+o_p(1). \tag{G.5}$$

Combining the expressions in (G.1), (G.2), and (G.5) with the delta method yields

$$\begin{aligned}\sqrt{n}(\widehat{\overline{P}}-\overline{P}) &= \frac{1}{\mathbb{P}(W_0=1)\max_{x\in\mathcal{X}^+}a(x)}\sum_{j=1}^{K}\big(w_0(x_j)p_j\sqrt{n}(\widehat{a}(x_j)-a(x_j))\\
&\quad+a(x_j)p_j\sqrt{n}(\widehat{w}_0(x_j)-w_0(x_j))+a(x_j)w_0(x_j)\sqrt{n}(\widehat{p}_j-p_j)\big)\\
&\quad-\frac{\mathbb{E}[a(X)\mid W_0=1]}{\mathbb{P}(W_0=1)\max_{x\in\mathcal{X}^+}a(x)}\sum_{j=1}^{K}\big(p_j\sqrt{n}(\widehat{w}_0(x_j)-w_0(x_j))+w_0(x_j)\sqrt{n}(\widehat{p}_j-p_j)\big)\\
&\quad-\frac{\mathbb{E}[a(X)\mid W_0=1]}{\max_{x\in\mathcal{X}^+}a(x)^2}\max_{j\in\Psi_{\mathcal{X}^+}}\sqrt{n}\left(\widehat{a}(x_j)-a(x_j)\right)+o_p(1)\\
&=\psi(\sqrt{n}(\widehat{\mathbf{a}}-\mathbf{a}),\sqrt{n}(\widehat{\mathbf{w}}_0-\mathbf{w}_0),\sqrt{n}(\widehat{\mathbf{p}}-\mathbf{p}))+o_p(1)\\
&\xrightarrow{d}\psi(\mathbb{Z})\end{aligned}$$

by the continuity of $\psi$ and Assumption C.1. □

*Proof of Theorem C.2.* We verify the validity of the bootstrap by appealing to Theorem 3.2 in Fang and Santos (2019). We show that their Assumption 4 holds by showing the mapping $\widehat{\psi}$ satisfies $|\widehat{\psi}(h')-\widehat{\psi}(h)|\leq C_n\|h'-h\|$ for any $h',h\in\mathbb{R}^{3K}$ and for $C_n=O_p(1)$, and by showing that $\widehat{\psi}(h)\xrightarrow{p}\psi(h)$ for all $h\in\mathbb{R}^{3K}$.

Let $h=(h_1,h_2,h_3)$ and $h'=(h_1',h_2',h_3')$.

$$\begin{aligned}|\widehat{\psi}(h')-\widehat{\psi}(h)| &\leq \left|\sum_{j=1}^{K}\frac{\widehat{w}_0(x_j)\widehat{p}_j}{\widehat{\mathbb{P}}(W_0=1)\max_{i:\widehat{w}_0(X_i)>c_n}\widehat{a}(X_i)}(h_1'(j)-h_1(j))\right|\\
&\quad+\left|\frac{\widehat{\mathbb{E}}[a(X)\mid W_0=1]}{\max_{i:\widehat{w}_0(X_i)>c_n}\widehat{a}(X_i)^2}\left(\max_{j\in\widehat{\Psi}_{\mathcal{X}^+}}h_1'(j)-\max_{j\in\widehat{\Psi}_{\mathcal{X}^+}}h_1(j)\right)\right|\\
&\quad+\left|\sum_{j=1}^{K}\frac{(\widehat{a}(x_j)-\widehat{\mathbb{E}}[a(X)\mid W_0=1])\widehat{p}_j}{\widehat{\mathbb{P}}(W_0=1)\max_{i:\widehat{w}_0(X_i)>c_n}\widehat{a}(X_i)}(h_2'(j)-h_2(j))\right|\\
&\quad+\left|\sum_{j=1}^{K}\frac{(\widehat{a}(x_j)-\widehat{\mathbb{E}}[a(X)\mid W_0=1])\widehat{w}_0(x_j)}{\widehat{\mathbb{P}}(W_0=1)\max_{i:\widehat{w}_0(X_i)>c_n}\widehat{a}(X_i)}(h_3'(j)-h_3(j))\right|\end{aligned}$$

$$
\begin{aligned}
&\leq \left(\sum_{j=1}^{K} \frac{\widehat{w}_0(x_j)^2 \widehat{p}_j^2}{\widehat{\mathbb{P}}(W_0=1)^2 \max_{i:\widehat{w}_0(X_i)>c_n} \widehat{a}(X_i)^2}\right)^{1/2} \|h_1' - h_1\| && \text{(G.6)}\\
&+ \frac{|\widehat{\mathbb{E}}[a(X) \mid W_0=1]|}{\max_{i:\widehat{w}_0(X_i)>c_n} \widehat{a}(X_i)^2} \left| \max_{j \in \widehat{\Psi}_{\mathcal{X}^+}} h_1'(j) - \max_{j \in \widehat{\Psi}_{\mathcal{X}^+}} h_1(j) \right| && \text{(G.7)}\\
&+ \left(\sum_{j=1}^{K} \frac{(\widehat{a}(x_j) - \widehat{\mathbb{E}}[a(X) \mid W_0=1])^2 \widehat{p}_j^2}{\widehat{\mathbb{P}}(W_0=1)^2 \max_{i:\widehat{w}_0(X_i)>c_n} \widehat{a}(X_i)^2}\right)^{1/2} \|h_2' - h_2\| && \text{(G.8)}\\
&+ \left(\sum_{j=1}^{K} \frac{(\widehat{a}(x_j) - \widehat{\mathbb{E}}[a(X) \mid W_0=1])^2 \widehat{w}_0(x_j)^2}{\widehat{\mathbb{P}}(W_0=1)^2 \max_{i:\widehat{w}_0(X_i)>c_n} \widehat{a}(X_i)^2}\right)^{1/2} \|h_3' - h_3\|, && \text{(G.9)}
\end{aligned}
$$

where we applied the Cauchy–Schwarz inequality several times. Note that the maximum function is Lipschitz with Lipschitz constant one and therefore

$$
\begin{aligned}
\left| \max_{j \in \widehat{\Psi}_{\mathcal{X}^+}} h_1'(j) - \max_{j \in \widehat{\Psi}_{\mathcal{X}^+}} h_1(j) \right| &\leq \sum_{j \in \widehat{\Psi}_{\mathcal{X}^+}} |h_1'(j) - h_1(j)| \\
&\leq \sum_{j=1}^{K} |h_1'(j) - h_1(j)| \\
&\leq \sqrt{K} \|h_1' - h_1\|.
\end{aligned}
$$

Combining equations (G.6)–(G.9) with the consistency of $(\widehat{\mathbf{a}}, \widehat{\mathbf{w}}_0, \widehat{\mathbf{p}})$ established in Theorem C.1 shows that $|\widehat{\psi}(h') - \widehat{\psi}(h)| \leq C_n \|h' - h\|$ for any $h', h \in \mathbb{R}^{3K}$ and for $C_n = O_p(1)$. Therefore, by Remark 3.4 in Fang and Santos (2019), showing $\widehat{\psi}(h) \xrightarrow{p} \psi(h)$ for all $h \in \mathbb{R}^{3K}$ suffices.

Thus we now consider the consistency of the different components of $\widehat{\psi}(h)$. Applying Theorem C.1 and the continuous mapping theorem, we can show that terms

$$
\begin{aligned}
&\left(\sum_{j=1}^{K} \frac{\widehat{w}_0(x_j)\widehat{p}_j}{\widehat{\mathbb{P}}(W_0=1) \max_{i:\widehat{w}_0(X_i)>c_n} \widehat{a}(X_i)} h_1(j), \frac{\widehat{\mathbb{E}}[a(X) \mid W_0=1]}{\max_{i:\widehat{w}_0(X_i)>c_n} \widehat{a}(X_i)^2},\right.\\
&\left.\sum_{j=1}^{K} \frac{(\widehat{a}(x_j) - \widehat{\mathbb{E}}[a(X) \mid W_0=1])\widehat{p}_j}{\widehat{\mathbb{P}}(W_0=1) \max_{i:\widehat{w}_0(X_i)>c_n} \widehat{a}(X_i)} h_2(j), \sum_{j=1}^{K} \frac{(\widehat{a}(x_j) - \widehat{\mathbb{E}}[a(X) \mid W_0=1])\widehat{w}_0(x_j)}{\widehat{\mathbb{P}}(W_0=1) \max_{i:\widehat{w}_0(X_i)>c_n} \widehat{a}(X_i)} h_3(j)\right),
\end{aligned}
$$

are all consistent for their counterparts in $\psi(h)$. It remains to show that $\max_{j \in \widehat{\Psi}_{\mathcal{X}^+}} h_1(j) = \max_{j \in \Psi_{\mathcal{X}^+}} h_1(j) + o_p(1)$. This holds if the set $\widehat{\Psi}_{\mathcal{X}^+}$ is consistent for $\Psi_{\mathcal{X}^+}$, which we establish here. Let $k \in \Psi_{\mathcal{X}^+}$. Then,

$$\begin{aligned}
\mathbb{P}(k \in \widehat{\Psi}_{\mathcal{X}^+}) &= \mathbb{P}\left(\widehat{a}(x_k) \geq \max_{j \in \widehat{\mathcal{X}}^+} \widehat{a}(x_j) - \xi_n\right) \\
&= \mathbb{P}\left(\sqrt{n}(\widehat{a}(x_k) - \max_{j \in \widehat{\mathcal{X}}^+} \widehat{a}(x_j)) \geq -\sqrt{n}\xi_n\right) \\
&= \mathbb{P}\left(\sqrt{n}(\max_{j \in \mathcal{X}^+} \widehat{a}(x_j) - \max_{j \in \widehat{\mathcal{X}}^+} \widehat{a}(x_j)) \geq -\sqrt{n}\xi_n\right).
\end{aligned}$$

The third equality follows from $k \in \Psi_{\mathcal{X}^+}$. By the proof of Theorem C.1, $\sqrt{n}(\max_{j \in \mathcal{X}^+} \widehat{a}(x_j) - \max_{j \in \widehat{\mathcal{X}}^+} \widehat{a}(x_j)) = o_p(1)$. Since $-\sqrt{n}\xi_n \to -\infty$, $\mathbb{P}(k \in \widehat{\Psi}_{\mathcal{X}^+}) \to \mathbb{P}(0 \geq -\infty) = 1$ when $k \in \Psi_{\mathcal{X}^+}$.

Now suppose that $k \notin \Psi_{\mathcal{X}^+}$. Then,

$$\mathbb{P}(k \in \widehat{\Psi}_{\mathcal{X}^+}) = \mathbb{P}(\widehat{a}(x_k) \geq \max_{j \in \widehat{\mathcal{X}}^+} \widehat{a}(x_j) - \xi_n) \to \mathbb{P}(a(x_k) \geq \max_{j \in \mathcal{X}^+} a(x_j) - 0) = 0,$$

where the last equality holds from $k \notin \Psi_{\mathcal{X}^+}$. Therefore, $\mathbb{P}(\widehat{\Psi}_{\mathcal{X}^+} = \Psi_{\mathcal{X}^+}) \to 1$ as $n \to \infty$. This implies $\widehat{\psi}(h) \xrightarrow{p} \psi(h)$, which concludes the proof. □

## H Additional Derivations for Difference-in-Differences

Goodman-Bacon (2021) provides the following representation of the two-way fixed effects estimand under the assumption that group-time average treatment effects are constant over time:

$$\beta_{\text{TWFE}} = \sum_{k:\ \text{var}(D|G=k)>0} \left[\sum_{j=1}^{k-1} \sigma_{jk}^k + \sum_{j=k+1}^{K} \sigma_{kj}^k\right] \cdot \mathbb{E}[Y(1) - Y(0) \mid G = k, D = 1],$$

where

$$\sigma_{jk}^k = \frac{\mathbb{P}(G=j) \cdot \mathbb{P}(G=k) \cdot \mathbb{P}(D=1 \mid G=k) \cdot \left[\mathbb{P}(D=1 \mid G=j) - \mathbb{P}(D=1 \mid G=k)\right]}{\text{var}\left(D^{\perp\left(G_{t_1},\ldots,G_{t_{K-1}},P_1,\ldots,P_T\right)}\right)}$$

and

$$\sigma_{kj}^k = \frac{\mathbb{P}(G=j) \cdot \mathbb{P}(G=k) \cdot \left[1 - \mathbb{P}(D=1 \mid G=k)\right]\left[\mathbb{P}(D=1 \mid G=k) - \mathbb{P}(D=1 \mid G=j)\right]}{\text{var}\left(D^{\perp\left(G_{t_1},\ldots,G_{t_{K-1}},P_1,\ldots,P_T\right)}\right)}.$$

Here $A^{\perp B}$ is used to denote the residual in the linear projection of $A$ on $(1, B)$. It is also the case that $\sum_{k:\ \mathrm{var}(D|G=k)>0} \sum_{l>k} \left(\sigma_{kl}^k + \sigma_{kl}^l\right) = 1$.[3] When we compare this representation with Proposition 5.4, that is,

$$
\begin{aligned}
\beta_{\mathrm{TWFE}} &= \frac{\mathbb{E}[a_{\mathrm{TWFE,H}}(G) \cdot \mathbb{P}(D=1 \mid G) \cdot \tau_0(G)]}{\mathbb{E}[a_{\mathrm{TWFE,H}}(G) \cdot \mathbb{P}(D=1 \mid G)]} \\
&= \frac{\sum_{k:\ \mathrm{var}(D|G=k)>0} \mathbb{P}(G=k) \cdot a_{\mathrm{TWFE,H}}(k) \cdot \mathbb{P}(D=1 \mid G=k) \cdot \tau_0(k)}{\sum_{k:\ \mathrm{var}(D|G=k)>0} \mathbb{P}(G=k) \cdot a_{\mathrm{TWFE,H}}(k) \cdot \mathbb{P}(D=1 \mid G=k)},
\end{aligned}
$$

where $\tau_0(k) = \mathbb{E}[Y(1) - Y(0) \mid G=k, D=1]$, it becomes clear that, for each group $k$ other than the always treated and the never treated,

$$
\begin{aligned}
& a_{\mathrm{TWFE,H}}(k) \cdot \mathbb{P}(D=1 \mid G=k) \\
&= \sum_{j=1}^{k-1} \mathbb{P}(G=j) \cdot \mathbb{P}(D=1 \mid G=k) \cdot \big[\mathbb{P}(D=1 \mid G=j) - \mathbb{P}(D=1 \mid G=k)\big] \\
&+ \sum_{j=k+1}^{K} \mathbb{P}(G=j) \cdot \big[1 - \mathbb{P}(D=1 \mid G=k)\big]\big[\mathbb{P}(D=1 \mid G=k) - \mathbb{P}(D=1 \mid G=j)\big],
\end{aligned}
$$

and this, in turn, implies that

$$
\begin{aligned}
& a_{\mathrm{TWFE,H}}(k) \\
&= \sum_{j=1}^{k-1} \mathbb{P}(G=j) \cdot \big[\mathbb{P}(D=1 \mid G=j) - \mathbb{P}(D=1 \mid G=k)\big] \\
&+ \sum_{j=k+1}^{K} \mathbb{P}(G=j) \cdot \big[\mathbb{P}(D=1 \mid G=k) - \mathbb{P}(D=1 \mid G=j)\big] \cdot \frac{1 - \mathbb{P}(D=1 \mid G=k)}{\mathbb{P}(D=1 \mid G=k)}.
\end{aligned}
\tag{H.1}
$$

[3] The result in Goodman-Bacon (2021) technically also includes a weight $\sigma_{kU}$ attached to the contrast between group $k$ and the never-treated group. We subsume this weight under $\sigma_{kj}^k$, and likewise subsume the weight on the contrast with the always-treated group under $\sigma_{jk}^k$.

**Equivalence of Weight Functions**

We now show that the weights obtained in equation (H.1) are equivalent to those in Proposition 5.4. First, we rewrite the weights in (H.1) as follows:

$$
\begin{aligned}
& a_{\text{TWFE,H}}(k) \\
& = \sum_{j=1}^{k-1} \mathbb{P}(G=j) \cdot \big[\mathbb{P}(D=1 \mid G=j) - \mathbb{P}(D=1 \mid G=k)\big] \\
& + \sum_{j=k+1}^{K} \mathbb{P}(G=j) \cdot \big[\mathbb{P}(D=1 \mid G=k) - \mathbb{P}(D=1 \mid G=j)\big] \cdot \frac{1-\mathbb{P}(D=1 \mid G=k)}{\mathbb{P}(D=1 \mid G=k)} \\
& = \mathbb{P}(D=1, G<k) - \mathbb{P}(G<k)\mathbb{E}[D \mid G=k] + (1-\mathbb{E}[D \mid G=k])\mathbb{P}(G>k) \\
& \quad - \frac{1-\mathbb{E}[D \mid G=k]}{\mathbb{E}[D \mid G=k]}\mathbb{P}(D=1, G>k) \\
& = \mathbb{P}(D=1, G<k) - \mathbb{P}(G<k)\mathbb{E}[D \mid G=k] + \mathbb{P}(G>k) - \mathbb{E}[D \mid G=k]\mathbb{P}(G>k) \\
& \quad - \frac{1}{\mathbb{E}[D \mid G=k]}\mathbb{P}(D=1, G>k) + \mathbb{P}(D=1, G>k) \\
& = \mathbb{P}(D=1, G \neq k) - \mathbb{E}[D \mid G=k]\mathbb{P}(G \neq k) + \mathbb{P}(G>k)\left(1 - \frac{\mathbb{E}[D \mid G>k]}{\mathbb{E}[D \mid G=k]}\right) \\
& = (\mathbb{E}[D] - \mathbb{E}[D \mid G=k]\mathbb{P}(G=k)) - \mathbb{E}[D \mid G=k]\mathbb{P}(G \neq k) \\
& \quad + \mathbb{P}(G>k)\left(1 - \frac{\mathbb{E}[D \mid G>k]}{\mathbb{E}[D \mid G=k]}\right) \\
& = \mathbb{E}[D] - \mathbb{E}[D \mid G=k] + \mathbb{P}(G>k)\left(1 - \frac{\mathbb{E}[D \mid G>k]}{\mathbb{E}[D \mid G=k]}\right).
\end{aligned}
$$

For $k \in \{2, \dots, T\}$, the weights in Proposition 5.4 are equal to

$$
\mathbb{E}[1 - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D] \mid G=k] = 1 - \mathbb{E}[D \mid G=k] - \mathbb{E}[D \mid P \geq k] + \mathbb{E}[D], \tag{H.2}
$$

because they are the average of the weights in Proposition 5.3 conditional on $G=k$. The proof of Proposition 5.4 explicitly shows that

$$
\begin{aligned}
& \mathbb{E}[1 - \mathbb{E}[D \mid G] - \mathbb{E}[D \mid P] + \mathbb{E}[D] \mid G=k] \\
& = \mathbb{P}(D=0 \mid G=k) \cdot (\mathbb{P}(D=0 \mid P \geq k) + \mathbb{P}(D=1 \mid P<g)).
\end{aligned}
$$

Let us look at the difference between the weights in (H.1) and (H.2). Fix $k \in \{2, \ldots, T\}$ and write:

$$
\begin{aligned}
&(1 - \mathbb{E}[D \mid G = k] - \mathbb{E}[D \mid P \geq k] + \mathbb{E}[D]) \\
&\quad - \left( \mathbb{E}[D] - \mathbb{E}[D \mid G = k] + \mathbb{P}(G > k) \left( 1 - \frac{\mathbb{E}[D \mid G > k]}{\mathbb{E}[D \mid G = k]} \right) \right) \\
&= 1 - \mathbb{E}[D \mid P \geq k] - \mathbb{P}(G > k) + \frac{\mathbb{E}[D\mathbb{1}(G > k)]}{\mathbb{E}[D \mid G = k]} \\
&= \mathbb{E}[\mathbb{1}(G \leq k)] - \frac{\mathbb{E}[D\mathbb{1}(P \geq k)]}{\mathbb{E}[\mathbb{1}(P \geq k)]} + \frac{\mathbb{E}[D\mathbb{1}(G > k)]}{\mathbb{E}[\mathbb{1}(k \leq P)]} \\
&= \frac{1}{\mathbb{E}[\mathbb{1}(k \leq P)]} \left( F_G(k)\mathbb{E}[\mathbb{1}(k \leq P)] + \mathbb{E}[D\mathbb{1}(G > k)] - \mathbb{E}[D\mathbb{1}(P \geq k)] \right) \\
&= \frac{1}{\mathbb{E}[\mathbb{1}(k \leq P)]} \left( F_G(k)\mathbb{E}[\mathbb{1}(k \leq P)] + \mathbb{E}[\mathbb{1}(k < G \leq P)] - \mathbb{E}[\mathbb{E}[D \mid P]\mathbb{1}(P \geq k)] \right) \\
&= \frac{1}{\mathbb{E}[\mathbb{1}(k \leq P)]} \left( F_G(k)\mathbb{E}[\mathbb{1}(k \leq P)] + \mathbb{E}[\mathbb{E}[\mathbb{1}(k < G \leq P) \mid P]] - \mathbb{E}[F_G(P)\mathbb{1}(P \geq k)] \right) \\
&= \frac{1}{\mathbb{E}[\mathbb{1}(k \leq P)]} \left( F_G(k)\mathbb{E}[\mathbb{1}(k \leq P)] + \mathbb{E}[(F_G(P) - F_G(k))\mathbb{1}(P \geq k)] - \mathbb{E}[F_G(P)\mathbb{1}(P \geq k)] \right) \\
&= \frac{1}{\mathbb{E}[\mathbb{1}(k \leq P)]} \left( F_G(k)\mathbb{E}[\mathbb{1}(k \leq P)] + \mathbb{E}[F_G(P)\mathbb{1}(P \geq k)] - F_G(k)\mathbb{E}[\mathbb{1}(P \geq k)] \right. \\
&\quad \left. - \mathbb{E}[F_G(P)\mathbb{1}(P \geq k)] \right) \\
&= 0.
\end{aligned}
$$

Therefore, the weights in Proposition 5.4 and equations (H.1) and (H.2) are all equal to one another.

## References


Andrews, I., J. Roth, and A. Pakes (2023): "Inference for Linear Conditional Moment Inequalities," *Review of Economic Studies*, 90, 2763–2791.

Bei, X. (2024): "Inference on Union Bounds with Applications to DiD, RDD, Bunching, and Structural Counterfactuals," working paper.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2022): "When Is TSLS Actually LATE?," NBER Working Paper No. 29709.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81, 667–737.

Cho, J., and T. M. Russell (2024): "Simple Inference on Functionals of Set-

Identified Parameters Defined by Linear Moments," *Journal of Business & Economic Statistics*, 42, 563–578.

Cox, G., and X. Shi (2023): "Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models," *Review of Economic Studies*, 90, 201–228.

Fang, Z., and A. Santos (2019): "Inference on Directionally Differentiable Functions," *Review of Economic Studies*, 86, 377–412.

Fang, Z., A. Santos, A. M. Shaikh, and A. Torgovitsky (2023): "Inference for Large-Scale Linear Systems with Known Coefficients," *Econometrica*, 91, 299–327.

Goodman-Bacon, A. (2021): "Difference-in-Differences with Variation in Treatment Timing," *Journal of Econometrics*, 225, 254–277.

Hong, H., and J. Li (2018): "The Numerical Delta Method," *Journal of Econometrics*, 206, 379–394.