Partial Identification of Policy-Relevant Treatment Effects with General Treatments via Optimal Transport

Jiyuan Tan Jose Blanchet Vasilis Syrgkanis
{jiyuantan, jose.blanchet, vsyrgk}@stanford.edu
Management Science and Engineering, Stanford University

October 29, 2026

Abstract

This note extends the Constrained Conditional Optimal Transport (CCOT) approach to partial identification of Policy-Relevant Treatment Effects developed in the companion paper [Tan et al., 2026] from a binary treatment to a general treatment taking values in a compact set $\mathcal{W} \subseteq \mathbb{R}$. Replacing threshold crossing by a conditional quantile representation of the selection equation, we show that the resulting CCOT problem again decomposes into independent stratum-level sub-problems, that the observational-equivalence (sharpness) guarantee of the binary case carries over, and that two structural settings deliver closed-form sharp bounds: strictly monotonic treatment selection, where the recovered rank is an observable control variable and the bound is a rank regression on an identified support, and the finite ordered-choice model, where the latent preimages form a π-system and the overlapping observational constraints can be disentangled into disjoint isolated regions, each contributing a one-dimensional optimal transport problem with product cost. The general case is a mixture of the two.

1 Overview

The CCOT formulation of Tan et al. [2026] does not rely on the treatment being binary. This note records the generalization to a compact treatment domain $\mathcal{W} \subseteq \mathbb{R}$ and the two structural

settings in which the resulting problem still admits a closed-form solution. Section 2 restates, for reference, the objects of the companion paper that the development uses; Section 3 gives the formal development; and Section 4 collects the proofs.

Threshold crossing is replaced by the conditional quantile representation

$$W = F^{-1}_{W|Z,X}(U \mid Z, X), \qquad U \mid X \sim \mathrm{Unif}(0,1),$$

together with consistency $Y = Y(W)$ and conditional instrumental exogeneity $Z \perp\!\!\!\perp (U, \{Y(w)\}_{w\in\mathcal{W}}) \mid X$. Here U is a latent rank ordering treatment intensity within each (Z, X) cell. If W is binary, this reduces to Assumption 2.1 after the harmless reparameterization $U' = 1 - U$; if $\mathcal{W}$ is finite, it is the discrete ordered choice model [Heckman and Vytlacil, 2007]; and in the continuous treatment regime it provides the structural foundation for the control function approach in nonseparable models [Imbens and Newey, 2009, Blundell and Powell, 2003, Chesher, 2003].

Let λ be a σ-finite measure dominating $\mathcal{L}(W \mid Z = z, X = x)$ for $\mathbb{P}_{Z,X}$-almost every (z, x), and let $\pi_{w,x}$ be the structural conditional law of $(U, Y(w))$ given $X = x$. The target parameter generalizes to

$$\theta_\omega = \int_{\mathcal{W}} \mathbb{E}_X \left[\int_0^1 \mathbb{E}_{\pi_{w,X}}[Y(w) \mid U = u]\, \omega(w, X, u)\, \mathrm{d}u \right] \lambda(\mathrm{d}w),$$

where λ is Lebesgue measure for an absolutely continuous treatment, counting measure for a discrete treatment, and an appropriate common dominating measure for a mixed treatment. Taking $\mathcal{W} = \{0, 1\}$ recovers (2) once the weights are composed with $U' = 1 - U$, and the counterfactual policy weight of (3) generalizes to a Radon–Nikodym derivative of the signed measure that the policy induces on $[0, 1] \times \mathcal{W}$.

Because both the objective functional and the observational constraints are additively separable across the support of X and W, the global optimization again decomposes into a family of independent CCOT sub-problems, one for λ-a.e. w and $\mathbb{P}_X$-a.e. x, and the sharp bounds for θ_ω are obtained by aggregating the stratum-level optimal values against $\mathbb{E}_X$ and λ. The observational equivalence result of Proposition 2.5 extends to this setting (Proposition 3.3), so these bounds remain sharp.

Two structural settings deliver closed forms. When the selection function is strictly increasing in U [Imbens and Newey, 2009], the recovered rank $R := F_{W|Z,X}(W \mid Z, X)$ equals U almost surely and therefore serves as an observable control variable; $\mathbb{E}_{\pi_{w,x}}[Y(w) \mid U = u]$ is then point-identified by a rank regression on the identified support, under a rank overlap condition requiring the target weight to be dominated there (Theorem 3.6), and outside that

support the bound assigns the worst feasible outcome value pointwise. When $\mathcal{W}$ is finite, the latent preimage of each (z, w) cell is an interval, and these intervals may overlap. If they form a π-system (Assumption 3.7), the overlapping constraints can be disentangled into disjoint isolated regions carrying isolated measures, and each region contributes a one-dimensional OT problem with product cost exactly as in the binary closed-form bounds of Tan et al. [2026] (Theorem 3.8). The general case is a mixture of the two: at continuity points of $F_{W|Z,X}(\cdot \mid z, x)$ the latent preimage is a singleton and the rank regression applies, whereas at mass points it is a nondegenerate interval and the OT sub-problem applies.

2 Background from the Companion Paper

This section restates, without proof, the objects of Tan et al. [2026] that the development below refers to. Numbering is local to this note.

The observable variables are the treatment W, the instrument $Z \in \mathcal{Z}$, covariates $X \in \mathcal{X}$, and the outcome $Y \in \mathcal{Y}$, where $\mathcal{X} \subset \mathbb{R}^{d_X}$, $\mathcal{Z}$, and $\mathcal{Y} \subset \mathbb{R}$ are compact and standard Borel. The unobservables are the potential outcomes and the latent resistance U. In the binary case $\mathcal{W} = \{0, 1\}$ the companion paper imposes the following.

Assumption 2.1 (Structural Assumptions)**.** Throughout, $Y(w) \in \mathcal{Y}$, $Z \in \mathcal{Z}$, and $X \in \mathcal{X}$ almost surely, and

1. (Consistency) $Y = WY(1) + (1 - W)Y(0)$.
2. (Conditional Instrumental Exogeneity) $Z \perp\!\!\!\perp (U, Y(0), Y(1)) \mid X$.
3. (Threshold Crossing) The treatment is selected by

$$W = \mathbb{1}(U \leqslant p(Z, X)), \tag{1}$$

   where $p(Z, X) = \mathbb{P}(W = 1 \mid Z, X)$. Moreover, $U \mid X \sim \text{Unif}(0, 1)$.
4. (Common Instrument Support) $\text{Supp}(Z \mid X = x) = \mathcal{Z}$ for all $x \in \mathcal{X}$.

Assumption 2.2 (Boundedness of the Potential Outcomes)**.** The outcome domain satisfies $y_{\min}, y_{\max} \in \mathcal{Y} \subseteq [y_{\min}, y_{\max}]$ for known constants $y_{\min} < y_{\max}$. Thus, the specified lower and upper endpoints are feasible outcome values.

In Tan et al. [2026] this is not a separate assumption: it holds with $y_{\min} = \min \mathcal{Y}$ and $y_{\max} = \max \mathcal{Y}$ by compactness of $\mathcal{Y}$, with the outcome domain treated as known. We state it separately because the results below invoke it explicitly.

With $\mathrm{MTE}(x, u) = \mathbb{E}[Y(1) - Y(0) \mid X = x, U = u]$, the target parameter is

$$\theta_\omega = \mathbb{E}_X \int_0^1 \mathrm{MTE}(X, u)\omega(X, u)\,\mathrm{d}u, \tag{2}$$

where ω is the target-specific weight, subject to the following standing condition.

Assumption 2.3 (Target-Weight Regularity)**.** The function $\omega : \mathcal{X} \times [0, 1] \to \mathbb{R}$ is jointly measurable and identifiable from the observed law, and

$$\mathbb{E}_X \left[\int_0^1 |\omega(X, u)|\,\mathrm{d}u \right] < \infty.$$

Different PRTEs are obtained by taking

$$\omega(x, u) \propto \mathbb{E}[\mathbb{1}(u \leqslant q_1(Z, X)) - \mathbb{1}(u \leqslant q_0(Z, X)) \mid X = x], \tag{3}$$

for given propensity functions q_0 and q_1, where $q_0 = p$ and $q_1 = q$ correspond to the baseline and alternative propensity score functions in the PRTE case; the coefficient hidden inside the proportionality symbol is identifiable.

The closed-form reduction rests on the following classical one-dimensional transport result.

Theorem 2.4 (1D Optimal Transport with Product Cost; see, e.g., [Villani, 2021, Galichon, 2016])**.** *Let μ and ν be probability measures on $\mathbb{R}$ with quantile functions Q_μ and Q_ν. Suppose either that μ has bounded support and ν has a finite first moment, or vice versa, and let the cost be the simple product $c(a, b) = a \cdot b$. Then, the minimum and maximum of the expected cost over all valid couplings are completely characterized by the countermonotonic and comonotonic couplings, respectively:*

$$\min_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^2} a \cdot b\,\mathrm{d}\pi(a, b) = \int_0^1 Q_\mu(t)Q_\nu(1 - t)\,\mathrm{d}t,$$
$$\max_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^2} a \cdot b\,\mathrm{d}\pi(a, b) = \int_0^1 Q_\mu(t)Q_\nu(t)\,\mathrm{d}t.$$

The minimum is attained by the countermonotonic coupling $(Q_\mu(\xi), Q_\nu(1 - \xi))$, and the maximum by the comonotonic coupling $(Q_\mu(\xi), Q_\nu(\xi))$, where $\xi \sim \mathrm{Unif}(0, 1)$.

Finally, let $\Gamma(\mathbb{P}_{\text{obs}})$ be the set of pairs (π_0, π_1) of $(Y(w), U, X)$-laws satisfying the CCOT constraints of the binary problem. Sharpness of the binary bounds follows from the following observational-equivalence result, which Proposition 3.3 below extends.

Proposition 2.5 (Sharp Identified Set of the Joint Latent Law)**.** *Suppose the observed distribution $\mathbb{P}_{obs}$ is generated by a true structural model satisfying Assumption 2.1. Let $\tilde{\pi}$ be any probability measure on $\mathcal{Y} \times \mathcal{Y} \times [0,1] \times \mathcal{X}$ whose $(Y(w), U, X)$-marginals $\tilde{\pi}_w$ satisfy $(\tilde{\pi}_0, \tilde{\pi}_1) \in \Gamma(\mathbb{P}_{obs})$. Then there exists a joint distribution of the latent and observable variables $(Y(0), Y(1), U, X, Z, W)$ such that:*

(i) the joint marginal distribution of $(Y(0), Y(1), U, X)$ is exactly $\tilde{\pi}$;

(ii) Assumption 2.1 holds for $(Y(0), Y(1), U, X, Z, W)$;

(iii) the induced distribution of (Z, X, W, Y) exactly matches $\mathbb{P}_{obs}$.

Conversely, every joint marginal $\tilde{\pi}$ arising from a distribution of $(Y(0), Y(1), U, X, Z, W)$ satisfying (ii)–(iii) *has $(Y(w), U, X)$-marginals in $\Gamma(\mathbb{P}_{obs})$ for $w \in \{0, 1\}$.*

3 Extension to General Treatment

This section gives the formal development summarized in Section 1; the proofs are collected in Section 4.

In this section, we generalize our partial identification framework to accommodate general treatments, including continuous and multi-valued treatments. We allow the treatment domain $\mathcal{W}$ to be a compact subset of $\mathbb{R}$. We impose the following generalized structural assumptions.

Assumption 3.1 (Generalized Structural Assumptions)**.** The potential outcomes form a jointly measurable map $(w, \xi) \mapsto Y(w, \xi)$ on $\mathcal{W} \times \Omega$, where Ω is the underlying sample space, and the potential outcomes and the treatment selection mechanism satisfy the following conditions:

1. **Consistency:** $Y = Y(W)$.
2. **Conditional Instrumental Exogeneity:** $Z \perp\!\!\!\perp (U, \{Y(w)\}_{w \in \mathcal{W}}) \mid X$.

3. **Selection Mechanism:** The treatment W is selected via the structural equation

$$W = F^{-1}_{W|Z,X}(U \mid Z, X), \tag{4}$$

where $U \mid X \sim \text{Unif}(0,1)$, and $F^{-1}_{W|Z,X}(\cdot \mid z, x)$ is the conditional quantile function of W given $Z = z$ and $X = x$.

4. **Common Instrument Support:** The conditional support of Z given X does not depend on X: $\text{Supp}(Z \mid X = x) = \mathcal{Z}$ for all $x \in \mathcal{X}$.

Under the standing standard-Borel convention, all conditional laws in this section are regular conditional distributions. We choose jointly measurable versions throughout, and identities involving continuously distributed conditioning variables hold outside the corresponding null sets. Common support does not by itself select pointwise values of such laws, so we formulate observational constraints almost surely; the additional domination condition below determines when those constraints identify the target measure. If W is a binary treatment, Assumption 3.1 reduces to the threshold-crossing model used in Assumption 2.1 after the harmless reparameterization $U' = 1 - U$.[1] If W is multi-valued, this model is equivalent to the discrete ordered choice model [Heckman and Vytlacil, 2007]. In the continuous treatment regime, this assumption provides the structural foundation for the control function approach in nonseparable models [Imbens and Newey, 2009, Blundell and Powell, 2003, Chesher, 2003]. We discuss this connection in detail in Section 3.1.

In this model, U can be interpreted as a latent rank that orders individuals' treatment intensity within each (Z, X) cell. Under the ordered-choice convention in (4), larger values of U correspond to larger realizations of W.

Following the binary treatment case, we consider the following general target parameter. As a standing condition for this section, assume that there exists a fixed σ-finite measure λ that dominates $\mathcal{L}(W \mid Z = z, X = x)$ for $\mathbb{P}_{Z,X}$-almost every (z, x). By iterated conditioning, this condition also implies $\mathcal{L}(W \mid X = x) \ll \lambda$ for $\mathbb{P}_X$-almost every x. Let $\pi_{w,x}$ be the structural conditional law of $(U, Y(w))$ given $X = x$. The generalized weight satisfies the following analogue of Assumption 2.3.

[1] Under the generalized ordered-choice convention, $W = F^{-1}_{W|Z,X}(U \mid Z, X)$ is increasing in U, whereas our earlier binary threshold-crossing convention writes treatment as $W = \mathbb{1}(U \leq p(Z, X))$. Setting $U' = 1 - U$ aligns the two models.

Assumption 3.2 (Generalized Target-Weight Regularity)**.** The function $\omega : \mathcal{W}\times\mathcal{X}\times[0,1] \to \mathbb{R}$ is jointly measurable and identifiable from the observed law, and

$$\int_{\mathcal{W}} \mathbb{E}_X \left[\int_0^1 |\omega(w, X, u)| \, \mathrm{d}u \right] \lambda(\mathrm{d}w) < \infty,$$

We define

$$\theta_\omega = \int_{\mathcal{W}} \mathbb{E}_X \left[\int_0^1 \mathbb{E}_{\pi_{w,X}}[Y(w) \mid U = u] \, \omega(w, X, u) \, \mathrm{d}u \right] \lambda(\mathrm{d}w), \tag{5}$$

Here λ is Lebesgue measure for an absolutely continuous treatment, counting measure for a discrete treatment, and an appropriate common dominating measure for a mixed treatment. This directly generalizes θ_ω from (2). To see this, consider the binary treatment case $\mathcal{W} = \{0, 1\}$, where λ is counting measure and $\int_{\mathcal{W}} \cdot \, \lambda(\mathrm{d}w)$ reduces to the sum over $\{0, 1\}$. The two models order the latent rank oppositely, so the weights must be composed with the reparameterization $U' = 1 - U$ of Assumption 3.1: set

$$\omega(1, x, u) = \omega(x, 1 - u), \qquad \omega(0, x, u) = -\,\omega(x, 1 - u).$$

Then (5) becomes

$$\begin{aligned}
\theta_\omega &= \mathbb{E}_X \int_0^1 \Big(\mathbb{E}_{\pi_{1,X}}[Y(1) \mid U = u] - \mathbb{E}_{\pi_{0,X}}[Y(0) \mid U = u] \Big) \, \omega(X, 1 - u) \, \mathrm{d}u \\
&= \mathbb{E}_X \int_0^1 \Big(\mathbb{E}_{\pi_{1,X}}[Y(1) \mid U = 1 - u'] - \mathbb{E}_{\pi_{0,X}}[Y(0) \mid U = 1 - u'] \Big) \, \omega(X, u') \, \mathrm{d}u' \\
&= \mathbb{E}_X \int_0^1 \mathrm{MTE}(X, u') \, \omega(X, u') \, \mathrm{d}u',
\end{aligned}$$

where the second line substitutes $u = 1 - u'$ and the third uses $U' = 1 - U$, under which $\mathbb{E}_{\pi_{w,X}}[Y(w) \mid U = 1-u'] = \mathbb{E}[Y(w) \mid U' = u', X]$ is the binary-convention marginal treatment response. This recovers (2).

As an example, consider a counterfactual policy that changes the treatment assignment mechanism from $F_{W|Z,X}$ to $\tilde{F}_{W|Z,X}$ while keeping the distribution of (Z, X) fixed. Under this new policy, the counterfactual treatment is $\tilde{W} = \tilde{F}^{-1}_{W|Z,X}(U \mid Z, X)$, and the realized outcome is $\tilde{Y} = Y(\tilde{W})$. Suppose, for $\mathbb{P}_X$-almost every x, that the signed conditional measure $\tilde{F}_{U,\tilde{W}|X}(\cdot \mid x) - F_{U,W|X}(\cdot \mid x)$ is absolutely continuous with respect to $\mathrm{Leb}_{[0,1]} \otimes \lambda$. Then the policy effect $\theta = \mathbb{E}[\tilde{Y} - Y]$ can be expressed as θ_ω, where the weight is a jointly measurable Radon–Nikodym derivative defined by

$$\omega(w, x, u) \, \mathrm{d}u \, \lambda(\mathrm{d}w) = \mathrm{d}\tilde{F}_{U,\tilde{W}|X}(u, w \mid x) - \mathrm{d}F_{U,W|X}(u, w \mid x). \tag{6}$$

For policies whose signed measure is not absolutely continuous, the effect must instead be written as an integral with respect to that signed measure and the density-based separation below does not apply. The policy weight (6) reduces to (3) in the binary case.

The observed data constrain the structural kernel $\pi_{w,x}$. Let

$$r(w \mid z, x) = \frac{\mathrm{d}\mathcal{L}(W \mid Z = z, X = x)}{\mathrm{d}\lambda}(w)$$

be a jointly measurable Radon–Nikodym derivative, and define the effective instrument set, up to $\mathbb{P}_{Z|X=x}$-null sets, by

$$\mathcal{Z}(x, w) \coloneqq \{z \in \mathcal{Z} : r(w \mid z, x) > 0\}.$$

Fix jointly measurable versions of $\mathbb{P}_{\text{obs}}(\mathrm{d}y \mid W = w, Z = z, X = x)$ and of the conditional rank law $F_{U|Z,X,W}(\mathrm{d}u \mid z, x, w)$ induced by (4). For $\lambda(\mathrm{d}w)\mathbb{P}_{Z,X}(\mathrm{d}z, \mathrm{d}x)$-almost every (w, z, x) with $r(w \mid z, x) > 0$, the observational constraint is

$$\int_0^1 \pi_{w,x}(\mathrm{d}y \mid U = u)\mathrm{d}F_{U|Z,X,W}(u \mid z, x, w) = \mathbb{P}_{\text{obs}}(\mathrm{d}y \mid W = w, Z = z, X = x).$$

Crucially, because both the objective functional and the observational constraints are additively separable across the support of X and W, the global optimization over all joint distributions decomposes into a family of independent CCOT sub-problems. Formally, this is equivalent to solving the following CCOT problem for λ-a.e. $w \in \mathcal{W}$ and $\mathbb{P}_X$-a.e. $x \in \mathcal{X}$:

$$\begin{aligned}
\max_{\pi_{w,x}} / \min_{\pi_{w,x}} \quad & \int_0^1 \mathbb{E}_{\pi_{w,x}}[Y(w) \mid U = u]\, \omega(w, x, u)\, \mathrm{d}u \\
\text{s.t.} \quad & \int_0^1 \pi_{w,x}(\mathrm{d}y \mid U = u)\mathrm{d}F_{U|Z,X,W}(u \mid z, x, w) \\
& \qquad = \mathbb{P}_{\text{obs}}(\mathrm{d}y \mid W = w, Z = z, X = x), \quad \mathbb{P}_{Z|X=x}\text{-a.e. } z \in \mathcal{Z}(x, w), \\
& \pi_{w,x}(\mathrm{d}u) = \mathrm{d}u, \\
& \pi_{w,x} \text{ is supported on } \mathcal{Y} \times [0, 1].
\end{aligned} \tag{7}$$

Here $\pi_{w,x}$ is a joint law on $\mathcal{Y} \times [0, 1]$ with a uniform U-marginal; its conditional outcome kernel $\pi_{w,x}(\cdot \mid u)$ enters the observational constraints, including when $F_{U|Z,X,W}$ is singular with respect to Lebesgue measure. The constraints are imposed for λ-a.e. w and $\mathbb{P}_X$-a.e. x under the almost-sure convention above. Let $\overline{\theta}(x, w)$ and $\underline{\theta}(x, w)$ denote the optimal values of the maximization and minimization in (7), respectively. The tight bounds for the global parameter θ_ω are then obtained by aggregating over all (x, w) strata:

$$\underline{\theta}_\omega = \int_{\mathcal{W}} \mathbb{E}_X[\underline{\theta}(X, w)]\, \lambda(\mathrm{d}w), \qquad \overline{\theta}_\omega = \int_{\mathcal{W}} \mathbb{E}_X\left[\overline{\theta}(X, w)\right]\, \lambda(\mathrm{d}w). \tag{8}$$

To formally guarantee that (7) yields sharp bounds for the general treatment setting, we must establish that any family of measures satisfying the generalized observational and marginal constraints corresponds to a valid data-generating process. Let $\Gamma_{\text{gen}}(\mathbb{P}_{\text{obs}})$ denote the set of jointly measurable probability kernels $(w, x) \mapsto \pi_{w,x} \in \mathcal{P}(\mathcal{Y} \times [0,1])$ whose U-marginal is Lebesgue measure and that admit a jointly measurable disintegration $(w, x, u) \mapsto \pi_{w,x}(\mathrm{d}y \mid u)$ satisfying the observational constraints in the CCOT problem above. The following proposition verifies this observational equivalence, extending the sharpness guarantee from the binary framework, Proposition 2.5.

Proposition 3.3 (Generalized Observational Equivalence and Sharpness)**.** *Suppose the observed distribution $\mathbb{P}_{obs}$ is generated by a true structural model satisfying Assumption 3.1. Then, for any candidate family of joint probability kernels $\{\tilde{\pi}_{w,x}\}_{w\in\mathcal{W},x\in\mathcal{X}} \in \Gamma_{gen}(\mathbb{P}_{obs})$, there exists a probability space supporting a jointly measurable process $\{Y(w)\}_{w\in\mathcal{W}}$ together with random variables (U, X, Z, W) such that:*

(i) For λ-a.e. $w \in \mathcal{W}$ and $\mathbb{P}_X$-a.e. $x \in \mathcal{X}$, the conditional distribution of $(Y(w), U) \mid X = x$ is exactly $\tilde{\pi}_{w,x}$. Moreover, conditional on (U, X), the process $\{Y(w)\}_{w\in\mathcal{W}}$ is independent of (Z, W).

(ii) Assumption 3.1 holds for $(\{Y(w)\}_{w\in\mathcal{W}}, U, X, Z, W)$.

(iii) The induced distribution of the observable variables (Z, X, W, Y) exactly matches the observed data distribution $\mathbb{P}_{obs}$.

Conversely, every family of conditional laws $\{\tilde{\pi}_{w,x}\}_{w,x}$ induced by a jointly measurable potential-outcome process and variables satisfying (ii)–(iii) *belongs to $\Gamma_{\text{gen}}(\mathbb{P}_{\text{obs}})$, up to the usual null sets in (w, x).*

This formal CCOT formulation highlights that variation in $F_{U|Z=z,X=x,W=w}$ across different instrument values z imposes multiple marginal constraints on the structural distribution $\pi_{w,x}$. A useful coverage set is the union of these supports,

$$\mathcal{U}_{\text{id}}(x, w) := \bigcup_{z\in\mathcal{Z}(x,w)} \operatorname{supp}(F_{U|Z=z,X=x,W=w}).$$

Each constituent support is either a single point or a closed interval in $[0, 1]$. Throughout this section, we assume that $\mathcal{U}_{\text{id}}(x, w)$ is jointly measurable in (x, w, u). Support coverage alone does not identify a conditional law on null slices; the strictly monotonic case below

therefore adds domination with respect to the target-weighted rank measure. In the finite ordered-choice case, every retained cell has positive probability and hence a positive-length latent interval, so the integral constraints are already version-invariant.

In the following subsections, we demonstrate how to obtain closed-form solutions under specific structural settings. The general case is considerably more involved and can be seen as a mixture of these settings; we leave it for future work.

3.1 Strictly Monotonic Treatment Selection

Imbens and Newey [2009] considers the following assumption for the IV model.

Assumption 3.4 (Strict Monotonicity)**.** The structural treatment function $W = h(Z, X, U)$ is strictly increasing in the latent variable U almost surely.

Under Assumption 3.4, this selection model is observationally equivalent to the conditional quantile representation in (4). The recovered rank $R \coloneqq F_{W|Z,X}(W \mid Z, X)$ equals U almost surely and therefore serves as an observable control variable. Write $\bar{r}(w \mid x) = \mathrm{d}\mathcal{L}(W \mid X = x)/\mathrm{d}\lambda(w)$. For λ-a.e. w and $\mathbb{P}_X$-a.e. x with $\bar{r}(w \mid x) > 0$, let $\rho_{w,x} = \mathcal{L}(R \mid W = w, X = x)$ and set $\mathcal{U}_{\mathrm{id}}(x, w) = \mathrm{supp}(\rho_{w,x})$; when $\bar{r}(w \mid x) = 0$, set $\mathcal{U}_{\mathrm{id}}(x, w) = \emptyset$. Let $\kappa_{\mathrm{obs}}(\mathrm{d}y \mid w, x, u)$ be a jointly measurable version of $\mathcal{L}(Y \mid W = w, X = x, R = u)$ on the observed strata, and define

$$m_{\mathrm{obs}}(w, x, u) \coloneqq \int_{\mathcal{Y}} y \, \kappa_{\mathrm{obs}}(\mathrm{d}y \mid w, x, u). \tag{9}$$

Assumption 3.5 (Rank Overlap for the Target Measure)**.** The graph $\{(x, w, u) : u \in \mathcal{U}_{\mathrm{id}}(x, w)\}$ is measurable and, for λ-a.e. w and $\mathbb{P}_X$-a.e. x with $\bar{r}(w \mid x) > 0$,

$$\Big(|\omega(w, x, \cdot)| \, \mathrm{Leb}_{[0,1]}\Big)\Big|_{\mathcal{U}_{\mathrm{id}}(x,w)} \ll \rho_{w,x}.$$

The regression in (9) is defined $\rho_{w,x}$-a.e.; Assumption 3.5 makes its restriction unique $|\omega(w, x, \cdot)| \, \mathrm{d}u$-a.e. on the identified region. Consistency, conditional instrumental exogeneity, and $R = U$ then give

$$\mathbb{E}_{\pi_{w,x}}[Y(w) \mid U = u] = m_{\mathrm{obs}}(w, x, u) \quad \text{for } |\omega(w, x, \cdot)| \, \mathrm{d}u\text{-a.e. } u \in \mathcal{U}_{\mathrm{id}}(x, w). \tag{10}$$

This a.e. control-function identity replaces conditioning on individual $(W = w, Z = z, U = u)$ slices.

Outside this identifiable support, there is no constraint, and the optimal solution trivially assigns probability mass to the global outcome bounds to minimize or maximize the objective. Substituting this into our separated CCOT formulation yields the closed-form bounds.

Theorem 3.6 (Closed-Form Bounds under Strictly Monotonic Treatment Selection)**.** *Suppose the generalized structural conditions in Assumption 3.1, the strict monotonicity in Assumption 3.4, the rank overlap in Assumption 3.5, the outcome boundedness in Assumption 2.2, and the generalized weight condition in Assumption 3.2 hold. The sharp lower bound for θ_ω is:*

$$\begin{aligned}\underline{\theta}_\omega = &\int_{\mathcal{W}} \mathbb{E}_X \left[\int_0^1 \mathbb{1}\big(u \in \mathcal{U}_{id}(X,w)\big)\, m_{\text{obs}}(w,X,u)\, \omega(w,X,u)\, \mathrm{d}u \right] \lambda(\mathrm{d}w) \\ &+ \int_{\mathcal{W}} \mathbb{E}_X \left[\int_0^1 \mathbb{1}\big(u \notin \mathcal{U}_{id}(X,w)\big) \right. \\ &\qquad \left. \times \Big(y_{\min} \max\{0, \omega(w,X,u)\} + y_{\max} \min\{0, \omega(w,X,u)\} \Big) \mathrm{d}u \right] \lambda(\mathrm{d}w),\end{aligned}$$

The identified integrand is understood $|\omega(w,X,\cdot)|\,\mathrm{d}u$-a.e. under Assumption 3.5. The sharp upper bound $\overline{\theta}_\omega$ is obtained symmetrically by swapping $y_{\min}$ and $y_{\max}$ in the trivial bound term.

3.2 Ordered Choice Model

We next consider the multi-valued treatment setting, known as the ordered choice model [Heckman and Vytlacil, 2007]. In this case, $\mathcal{W}$ is finite and $\text{supp}(F_{U|Z=z,X=x,W=w})$ is an interval for each z. Denote this interval by $I_{x,w}(z)$. Then the first constraint of (7) simplifies to

$$\frac{1}{|I_{x,w}(z)|} \int_{I_{x,w}(z)} \pi_{w,x}(\mathrm{d}y \mid U = u)\mathrm{d}u = \mathbb{P}_{\text{obs}}(\mathrm{d}y \mid W = w, Z = z, X = x), \quad \forall z \in \mathcal{Z}(x,w). \tag{11}$$

Because $\mathcal{W}$ is finite and $z \in \mathcal{Z}(x,w)$, the cell probability $\mathbb{P}(W = w \mid Z = z, X = x)$ is positive and hence $|I_{x,w}(z)| > 0$. Null singleton preimages are handled only through the regular-conditional-distribution constraint in (7), not by division by Lebesgue length.

The main difficulty of (11) is that the intervals in $\{I_{x,w}(z)\}_{z\in\mathcal{Z}(x,w)}$ may overlap, making it difficult to disentangle the constraints as in the binary treatment setting. A closed-form decomposition follows when the interval family is closed under intersections.

Assumption 3.7 (π-System of Intervals)**.** For $\mathbb{P}_X$-a.e. x and every w, $\mathcal{Z}(x,w)$ is finite, the intervals $\{I_{x,w}(z) : z \in \mathcal{Z}(x,w)\}$ are pairwise distinct, and they form a π-system modulo Lebesgue-null boundaries: for any $z, z' \in \mathcal{Z}(x,w)$, the intersection $I_{x,w}(z) \cap I_{x,w}(z')$ is either Lebesgue-null or agrees up to a null set with $I_{x,w}(z'')$ for some $z'' \in \mathcal{Z}(x,w)$. The interval correspondence $(x,w,z) \mapsto I_{x,w}(z)$, the direct-child relation under strict inclusion, and the resulting isolated-region correspondence $(x,w,z) \mapsto J_{x,w}(z)$ have measurable graphs. Moreover, after removing all direct children from an interval, its isolated region $J_{x,w}(z)$ defined below has strictly positive Lebesgue measure.

In particular, intervals sharing the same start point or the same end point are nested and satisfy this property, as in the binary treatment setting; laminar families are therefore a special case. More generally, Assumption 3.7 gives a directed acyclic graph under strict inclusion—z is an ancestor of z' whenever $I_{x,w}(z') \subsetneq I_{x,w}(z)$. Intersections represented by lower nodes allow overlapping constraints to be disentangled without double counting.

For each $z \in \mathcal{Z}(x,w)$, let children(z) and Dec(z) denote its direct children and all of its descendants in this inclusion DAG. Define the disjoint *isolated sub-region*

$$J_{x,w}(z) = I_{x,w}(z) \setminus \bigcup_{z' \in \text{children}(z)} I_{x,w}(z'),$$

and, recursively from the leaves upward, the *isolated measure*

$$\mu_{w,z,x}(\mathrm{d}y) = \frac{1}{|J_{x,w}(z)|}\Bigg(|I_{x,w}(z)|\mathbb{P}_{\text{obs}}(\mathrm{d}y \mid W=w, Z=z, X=x) - \sum_{z' \in \text{Dec}(z)} |J_{x,w}(z')|\mu_{w,z',x}(\mathrm{d}y)\Bigg). \tag{12}$$

The measure $\mu_{w,z,x}$ is the generalized analogue of $\mu_{1,i}$ from the binary case: it isolates the conditional distribution of $Y(w)$ on the disjoint slice $J_{x,w}(z)$. As in the binary setting, $\mu_{w,z,x}$ is a valid (nonnegative) probability measure under correct model specification, and its nonnegativity serves as a testable implication of the structural model. We also let $J_{x,w}(\emptyset) = [0,1] \setminus \bigcup_{z \in \mathcal{Z}(x,w)} I_{x,w}(z)$ denote the unconstrained domain.

In Section 4, we show that under Assumption 3.7 the original constraint system (11) is equivalent to the disentangled marginal constraints

$$\frac{1}{|J_{x,w}(z)|}\int_{J_{x,w}(z)} \pi_{w,x}(\mathrm{d}y \mid U=u)\mathrm{d}u = \mu_{w,z,x}(\mathrm{d}y), \quad \forall z \in \mathcal{Z}(x,w), \tag{13}$$

each acting on a single disjoint region $J_{x,w}(z)$. Because the $\{J_{x,w}(z)\}_{z\in\mathcal{Z}(x,w)}$ partition the identified support $\mathcal{U}_{\text{id}}(x,w)$ and the objective in (7) is additively separable across these regions, the global problem decomposes into independent 1D optimal transport sub-problems with product cost—one per $J_{x,w}(z)$, pairing $\mu_{w,z,x}$ with $\text{Unif}(J_{x,w}(z))$—plus a pointwise trivial bound on the unconstrained domain $J_{x,w}(\emptyset)$. Applying Theorem 2.4 to each sub-problem yields the following closed-form result.

Theorem 3.8 (Closed-Form Bounds for Multi-Valued Treatment)**.** *Suppose Assumption 3.1, Assumption 3.7, Assumption 2.2, and Assumption 3.2 hold, $\mathcal{W}$ is finite, and for $\mathbb{P}_X$-a.e. x, every w, and every $z \in \mathcal{Z}(x,w)$ the support $\text{supp}(F_{U|Z=z,X=x,W=w})$ is an interval $I_{x,w}(z) \subseteq [0,1]$. Let $Q_{Y,J_{x,w}(z)}(t)$ be the quantile function of the isolated measure $\mu_{w,z,x}$, and let $Q_{\omega,J_{x,w}(z)}(t)$ be the quantile function of $\omega(w,x,U)$ for $U \sim \text{Unif}(J_{x,w}(z))$.*

For a given $x \in \mathcal{X}$ and $w \in \mathcal{W}$, let $J_{x,w}(\emptyset) = [0,1] \setminus \bigcup_{z\in\mathcal{Z}(x,w)} I_{x,w}(z)$ be the unconstrained domain. The sharp lower bound $\underline{\theta}(x,w)$ is given by:

$$\underline{\theta}(x,w) = \sum_{z\in\mathcal{Z}(x,w)} |J_{x,w}(z)| \int_0^1 Q_{Y,J_{x,w}(z)}(t) Q_{\omega,J_{x,w}(z)}(1-t)\,\mathrm{d}t + \int_{J_{x,w}(\emptyset)} \Big(y_{\min} \max\{0, \omega(w,x,u)\} + y_{\max} \min\{0, \omega(w,x,u)\} \Big) \mathrm{d}u.$$

The global sharp lower bound is $\underline{\theta}_\omega = \sum_{w\in\mathcal{W}} \mathbb{E}_X[\underline{\theta}(X,w)]$. The sharp upper bound $\overline{\theta}_\omega$ is obtained symmetrically by taking the comonotonic coupling integral

$$\int_0^1 Q_{Y,J_{x,w}(z)}(t)\, Q_{\omega,J_{x,w}(z)}(t)\,\mathrm{d}t$$

and swapping $y_{\min}$ and $y_{\max}$ in the trivial bound term.

Remark 3.9 (Mixed Continuous-Discrete Treatments). In general, the solution to the global CCOT problem in (7) is a hybrid of the results in Section 3.1 and Section 3.2. The exact nature of the subproblem for a given stratum (x,w) depends directly on the localized behavior of the treatment distribution.

If the conditional CDF $F_{W|Z,X}(\cdot \mid z,x)$ is continuous at the evaluation point w, then w carries no conditional probability mass and its latent preimage $\{u : F^{-1}_{W|Z,X}(u \mid z,x) = w\}$ is a singleton. In these regions, the recovered-rank regression identifies the potential-outcome expectation $|\omega(w,x,\cdot)|\,\mathrm{d}u$-a.e. wherever the rank-overlap condition holds, as in (10). Conversely, if the data exhibit a discrete probability mass at w, the latent preimage is a nondegenerate interval. Over these mass points, the subproblem reduces to the localized 1D optimal

transport problem described in Section 3.2. Thus, our unified CCOT framework naturally accommodates complex empirical settings with mixed continuous-discrete treatments under the corresponding domination conditions.

4 Proofs in Section 3

Proof of Proposition 3.3. Let $\{\tilde{\pi}_{w,x}\}_{w\in\mathcal{W},x\in\mathcal{X}} \in \Gamma_{\text{gen}}(\mathbb{P}_{\text{obs}})$ be an arbitrary feasible family. By the constraints in (7), each $\tilde{\pi}_{w,x}$ is a probability measure on $\mathcal{Y}\times[0,1]$ satisfying:

(C1) $\int_0^1 \tilde{\pi}_{w,x}(\mathrm{d}y \mid u)\,\mathrm{d}F_{U|Z,X,W}(u \mid z,x,w) = \mathbb{P}_{\text{obs}}(\mathrm{d}y \mid W=w, Z=z, X=x)$, for $\lambda(\mathrm{d}w)\mathbb{P}_{Z,X}(\mathrm{d}z,\mathrm{d}x)$-a.e. (w,z,x) with $r(w \mid z,x) > 0$;

(C2) $\tilde{\pi}_{w,x}(\mathrm{d}u) = \mathrm{d}u$ for λ-a.e. $w \in \mathcal{W}$ and $\mathbb{P}_{\text{obs}}$-a.e. $x \in \mathcal{X}$.

Here $\tilde{\pi}_{w,x}(\mathrm{d}y \mid u)$ denotes the regular conditional distribution obtained by disintegrating $\tilde{\pi}_{w,x}$ with respect to the U-marginal, which exists because $\mathcal{Y}$ and $[0,1]$ are standard Borel spaces. Joint measurability in (w,x,u) is part of the admissibility requirement in (7).

Construction. We construct the joint distribution $\mathbb{P}$ of $(\{Y(w)\}_{w\in\mathcal{W}}, U, X, Z, W)$ in three stages.

Stage 1: Base variables. Draw (X,Z,U) as follows:

$$X \sim \mathbb{P}_{\text{obs}}(\mathrm{d}x), \quad Z \mid X \sim \mathbb{P}_{\text{obs}}(\mathrm{d}z \mid x), \quad U \sim \text{Unif}(0,1) \text{ independently of } (Z,X). \tag{14}$$

Note that U is drawn independently of Z given X; this will ensure conditional instrumental exogeneity. The treatment is then deterministically assigned as $W = F^{-1}_{W|Z,X}(U \mid Z,X)$.

Stage 2: Potential outcome process. Let $F_{w,x,u}(y) = \tilde{\pi}_{w,x}((-\infty,y] \mid u)$ and define its conditional quantile

$$Q_{w,x,u}(v) = \inf\{y \in \mathcal{Y} : F_{w,x,u}(y) \geqslant v\}.$$

Because $(w,x,u) \mapsto \tilde{\pi}_{w,x}(\cdot \mid u)$ is a jointly measurable probability kernel on the standard Borel outcome space $\mathcal{Y} \subseteq \mathbb{R}$, the conditional-quantile map $(w,x,u,v) \mapsto Q_{w,x,u}(v)$ can be chosen jointly Borel measurable. Draw one common $V \sim \text{Unif}(0,1)$ independently of (U,X,Z) and set

$$Y(w) = Q_{w,X,U}(V), \qquad w \in \mathcal{W}. \tag{15}$$

Then $(w,\xi) \mapsto Y(w,\xi)$ is jointly measurable, and each coordinate has the required conditional law. The common auxiliary randomizer supplies an explicit coupling across treatment levels; no cross-w independence is required by either the target or the structural assumptions.

Stage 3: Full joint distribution. The joint law $\mathbb{P}$ of $(\{Y(w)\}_{w\in\mathcal{W}}, U, X, Z, W)$ is obtained by combining Stages 1 and 2. Crucially, the process $\{Y(w)\}_{w\in\mathcal{W}}$ depends on (Z, W) only through (U, X), since Stage 2 conditions solely on (U, X).

Verification of (i): conditional distributions. Fix $w \in \mathcal{W}$ in the full-λ-measure set on which (C2) holds, and fix $x \in \mathcal{X}$ in the corresponding full-$\mathbb{P}_{\text{obs}}$-measure set. By (15), the conditional distribution of $Y(w)$ given $(U, X) = (u, x)$ under $\mathbb{P}$ is exactly $\tilde{\pi}_{w,x}(\cdot \mid u)$. Since U has Lebesgue density on $[0, 1]$ independently of X by (14), the joint distribution of $(Y(w), U)$ given $X = x$ is

$$\mathbb{P}(Y(w) \in A,\, U \in B \mid X = x) = \int_B \tilde{\pi}_{w,x}(A \mid u)\, \mathrm{d}u = \tilde{\pi}_{w,x}(Y(w) \in A,\, U \in B),$$

where the last equality uses (C2). This holds for all measurable $A \subseteq \mathcal{Y}$ and $B \subseteq [0, 1]$, so $(Y(w), U) \mid X = x$ has distribution $\tilde{\pi}_{w,x}$ under $\mathbb{P}$. Therefore part (i) of the proposition holds for λ-a.e. w and $\mathbb{P}_{\text{obs}}$-a.e. x.

Verification of (ii): structural assumptions. We verify each part of Assumption 3.1.

- *Consistency.* Under $\mathbb{P}$, the treatment is $W = F^{-1}_{W|Z,X}(U \mid Z, X)$, and the observed outcome is $Y = Y(W)$ by definition.
- *Conditional instrumental exogeneity.* By construction, conditional on $X = x$, the instrument Z is drawn from $\mathbb{P}_{\text{obs}}(\mathrm{d}z \mid x)$ in Stage 1, while the process $(U, \{Y(w)\}_{w\in\mathcal{W}})$ is constructed in Stages 1–2 depending only on (U, X) and not on Z. Since U is drawn independently of Z given X, and the conditional law of $\{Y(w)\}_{w\in\mathcal{W}}$ given (U, X) does not involve Z, we have $Z \perp\!\!\!\perp (U, \{Y(w)\}_{w\in\mathcal{W}}) \mid X$ under $\mathbb{P}$.
- *Selection mechanism.* The equation $W = F^{-1}_{W|Z,X}(U \mid Z, X)$ holds $\mathbb{P}$-a.s. by Stage 1. Moreover, $U \mid X \sim \text{Unif}(0, 1)$ by (14).
- *Common instrument support.* Stage 1 preserves the observed conditional law of Z given X. Because the proposition assumes that $\mathbb{P}_{\text{obs}}$ is generated by a model satisfying Assumption 3.1, the constructed model inherits its common-support property.

Verification of (iii): observational equivalence. We must show that the distribution of the observables (Y, Z, X, W) induced by $\mathbb{P}$ coincides with $\mathbb{P}_{\text{obs}}$. The marginal of (Z, X) under $\mathbb{P}$ equals $\mathbb{P}_{\text{obs}}(Z, X)$ by (14). We verify the remaining two conditional laws: (1) the conditional distribution of W given (Z, X) and (2) the conditional distribution of Y given (W, Z, X).

Conditional distribution of W given (Z, X). Under $\mathbb{P}$, we have $W = F^{-1}_{W|Z,X}(U \mid Z, X)$ with $U \sim \mathrm{Unif}(0,1)$ independently of (Z, X). By the quantile transformation, $W \mid (Z = z, X = x)$ has cumulative distribution function $F_{W|Z,X}(\cdot \mid z, x)$, which is exactly the observed conditional distribution of W under $\mathbb{P}_{\mathrm{obs}}$.

Conditional distribution of Y given (W, Z, X). Fix (z, x, w) outside the usual null set for a regular conditional distribution. Under the quantile selection rule in Stage 1, let $F_{U|Z,X,W}(\cdot \mid z, x, w)$ denote the resulting regular conditional law of U given $(Z, X, W) = (z, x, w)$. This definition applies equally to atoms, continuous treatment values for which $\{W = w\}$ is null, and mixed laws; it involves no normalization by the Lebesgue length of a latent preimage.

By consistency, $Y = Y(W)$. Since $\{Y(w')\}_{w' \in \mathcal{W}}$ is conditionally independent of Z given (U, X) by Stage 2, the regular conditional distribution of Y given $(W = w, Z = z, X = x)$ under $\mathbb{P}$ is:

$$\mathbb{P}(Y \in A \mid W = w, Z = z, X = x) = \int_0^1 \tilde{\pi}_{w,x}(A \mid u)\, \mathrm{d}F_{U|Z,X,W}(u \mid z, x, w).$$

By constraint (C1), the right-hand side equals $\mathbb{P}_{\mathrm{obs}}(Y \in A \mid W = w, Z = z, X = x)$ for $\mathbb{P}_{W,Z,X}$-almost every (w, z, x).

Since the conditional distributions of $W \mid (Z, X)$ and $Y \mid (W, Z, X)$ agree with those under $\mathbb{P}_{\mathrm{obs}}$ up to the usual null sets in the conditioning variables, the induced distribution of (Y, Z, X, W) under $\mathbb{P}$ is identical to $\mathbb{P}_{\mathrm{obs}}$.

Converse. Now take any jointly measurable potential-outcome process and variables satisfying parts (ii)–(iii), and define $\tilde{\pi}_{w,x}$ as the conditional law of $(Y(w), U)$ given $X = x$. The selection assumption gives $U \mid X = x \sim \mathrm{Unif}(0,1)$, so (C2) holds. Conditional instrumental exogeneity and the fact that W is a measurable function of (U, Z, X) imply, for $\lambda(\mathrm{d}w)\mathbb{P}_{Z,X}(\mathrm{d}z, \mathrm{d}x)$-almost every (w, z, x) with $r(w \mid z, x) > 0$,

$$\mathcal{L}\big(Y(w) \mid U = u, Z = z, X = x, W = w\big) = \tilde{\pi}_{w,x}(\cdot \mid u).$$

By consistency and iterated conditioning,

$$\mathbb{P}(Y \in \mathrm{d}y \mid W = w, Z = z, X = x) = \int_0^1 \tilde{\pi}_{w,x}(\mathrm{d}y \mid u)\, \mathrm{d}F_{U|Z,X,W}(u \mid z, x, w).$$

Part (iii) identifies the left-hand side with the corresponding conditional law under $\mathbb{P}_{\mathrm{obs}}$, which is (C1). Thus $\{\tilde{\pi}_{w,x}\}_{w,x} \in \Gamma_{\mathrm{gen}}(\mathbb{P}_{\mathrm{obs}})$ up to the usual null sets. □

Proof of Theorem 3.6. We prove the sharp lower bound; the upper bound follows by swapping $y_{\min}$ and $y_{\max}$ in the trivial bound term.

By Proposition 3.3, the infimum of θ_ω over all structural models satisfying Assumption 3.1 that are observationally equivalent to $\mathbb{P}_{\text{obs}}$ equals the infimum over the feasible set $\Gamma_{\text{gen}}(\mathbb{P}_{\text{obs}})$. The global minimization problem is:

$$\underline{\theta} = \min_{\{\pi_{w,x}\}\in\Gamma_{\text{gen}}(\mathbb{P}_{\text{obs}})} \int_{\mathcal{W}} \mathbb{E}_X\left[\int_0^1 \mathbb{E}_{\pi_{w,X}}[Y(w) \mid U=u]\,\omega(w,X,u)\,\mathrm{d}u\right]\lambda(\mathrm{d}w),$$

where the feasible set $\Gamma_{\text{gen}}(\mathbb{P}_{\text{obs}})$ consists of jointly measurable probability kernels $(w,x)\mapsto \pi_{w,x} \in \mathcal{P}(\mathcal{Y}\times[0,1])$ whose U-marginal is Lebesgue measure and whose selected jointly measurable disintegrations $\pi_{w,x}(\mathrm{d}y \mid u)$ satisfy, for $\lambda(\mathrm{d}w)\mathbb{P}_X(\mathrm{d}x)$-a.e. (w,x),

(C1) $\int_0^1 \pi_{w,x}(\mathrm{d}y \mid u)\,\mathrm{d}F_{U|Z,X,W}(u \mid z,x,w) = \mathbb{P}_{\text{obs}}(\mathrm{d}y \mid W=w, Z=z, X=x)$, for $\mathbb{P}_{Z|X=x}$-a.e. $z\in\mathcal{Z}(x,w)$;

(C2) $\pi_{w,x}(\mathrm{d}u) = \mathrm{d}u$.

For a given identifiable weight $\omega(w,x,u)$, the objective for each (x,w)-stratum is:

$$\int_0^1 \mathbb{E}_{\pi_{w,x}}[Y(w) \mid U=u]\,\omega(w,x,u)\,\mathrm{d}u.$$

Separation over (x,w) strata. Both the objective and the constraints (C1)–(C2) are separable across (x,w): the kernel $\pi_{w,x}$ at one stratum is constrained independently of all other strata, and the objective decomposes additively. Writing $B_Y = \max\{|y_{\min}|, |y_{\max}|\}$, every feasible stratum objective is dominated in absolute value by $B_Y\int_0^1 |\omega(w,x,u)|\,\mathrm{d}u$. This envelope is integrable with respect to $\lambda(\mathrm{d}w)\mathbb{P}_X(\mathrm{d}x)$ by Assumption 3.2. Therefore, a measurable selection argument [Bertsekas and Shreve, 1996, Proposition 7.50] permits the minimization and the outer integration to be exchanged:

$$\underline{\theta} = \int_{\mathcal{W}} \mathbb{E}_X[\underline{\theta}(X,w)]\,\lambda(\mathrm{d}w),$$

where $\underline{\theta}(x,w)$ denotes the value of the (x,w)-sub-problem in (7). It therefore suffices to solve each sub-problem individually.

Decomposition of the objective for fixed (x,w). Fix (x,w) outside the null sets in Assumption 3.5 and drop them from the notation where unambiguous. Strict monotonicity gives $R = F_{W|Z,X}(W \mid Z,X) = U$ almost surely. The observed conditional rank law is $\rho_{w,x} = \mathcal{L}(R \mid W=w, X=x)$, and its support is $\mathcal{U}_{\text{id}}(x,w)$.

We decompose the unit interval into the identified region $\mathcal{U}_{\mathrm{id}}(x,w)$ and its complement $[0,1]\setminus\mathcal{U}_{\mathrm{id}}(x,w)$:

$$\int_0^1 \mathbb{E}_\pi[Y(w)\mid U=u]\,\omega(w,x,u)\,\mathrm{d}u = \underbrace{\int_0^1 \mathbb{1}\big(u\in\mathcal{U}_{\mathrm{id}}(x,w)\big)\,\mathbb{E}_\pi[Y(w)\mid U=u]\,\omega(w,x,u)\,\mathrm{d}u}_{\text{identified contribution}} + \underbrace{\int_0^1 \mathbb{1}\big(u\notin\mathcal{U}_{\mathrm{id}}(x,w)\big)\,\mathbb{E}_\pi[Y(w)\mid U=u]\,\omega(w,x,u)\,\mathrm{d}u}_{\text{unidentified contribution}}.$$

By the recovered-rank identity (10), strict monotonicity gives

$$\pi_{w,x}(\mathrm{d}y\mid u)=\kappa_{\mathrm{obs}}(\mathrm{d}y\mid w,x,u)\quad\text{for }\rho_{w,x}\text{-a.e. }u.$$

By Assumption 3.5, the same equality holds for $|\omega(w,x,\cdot)|\,\mathrm{d}u$-a.e. $u\in\mathcal{U}_{\mathrm{id}}(x,w)$. The constraints impose no target-relevant restriction on the complement, so we optimize the two contributions independently.

Identification on $\mathcal{U}_{\mathrm{id}}(x,w)$. The recovered-rank outcome kernel is unique $\rho_{w,x}$-a.e., and rank overlap makes it unique for the target-weighted integral on the identified region. Hence the identified contribution is

$$\begin{aligned}&\int_0^1 \mathbb{1}\big(u\in\mathcal{U}_{\mathrm{id}}(x,w)\big)\,\mathbb{E}_{\pi_{w,x}}[Y(w)\mid U=u]\,\omega(w,x,u)\,\mathrm{d}u\\ &\quad=\int_0^1 \mathbb{1}\big(u\in\mathcal{U}_{\mathrm{id}}(x,w)\big)\,m_{\mathrm{obs}}(w,x,u)\,\omega(w,x,u)\,\mathrm{d}u,\end{aligned}$$

with equality understood $|\omega(w,x,\cdot)|\,\mathrm{d}u$-a.e. on $\mathcal{U}_{\mathrm{id}}(x,w)$.

Trivial bound on $[0,1]\setminus\mathcal{U}_{\mathrm{id}}(x,w)$. For $u\notin\mathcal{U}_{\mathrm{id}}(x,w)$, no instrument value generates treatment level w at latent rank u. The constraints impose no restriction on $\pi_{w,x}(\mathrm{d}y\mid u)$, and the minimization is unconstrained. By Assumption 2.2, the pointwise minimum is achieved by $\pi_{w,x}(\cdot\mid u)=\delta_{y_{\min}}$ when $\omega(w,x,u)\geqslant 0$, and by $\pi_{w,x}(\cdot\mid u)=\delta_{y_{\max}}$ when $\omega(w,x,u)<0$, yielding the contribution:

$$\int_0^1 \mathbb{1}\big(u\notin\mathcal{U}_{\mathrm{id}}(x,w)\big)\Big(y_{\min}\max\{0,\omega(w,x,u)\}+y_{\max}\min\{0,\omega(w,x,u)\}\Big)\mathrm{d}u.$$

Aggregation. Combining the two regions and integrating over all (x, w) strata:

$$\underline{\theta} = \int_{\mathcal{W}} \mathbb{E}_X\Bigg[\int_0^1 \mathbb{1}\big(u \in \mathcal{U}_{\mathrm{id}}(X, w)\big)\, m_{\mathrm{obs}}(w, X, u)\, \omega(w, X, u)\, \mathrm{d}u\Bigg] \lambda(\mathrm{d}w)$$
$$+ \int_{\mathcal{W}} \mathbb{E}_X\Bigg[\int_0^1 \mathbb{1}\big(u \notin \mathcal{U}_{\mathrm{id}}(X, w)\big)$$
$$\times \Big(y_{\min} \max\{0, \omega(w, X, u)\} + y_{\max} \min\{0, \omega(w, X, u)\}\Big)\mathrm{d}u\Bigg] \lambda(\mathrm{d}w).$$

The absolute value of the displayed expression is dominated by $B_Y \int_0^1 |\omega(w, X, u)|\, \mathrm{d}u$, so its λ-integrated expectation is finite by Assumption 3.2.

Sharpness. By Proposition 3.3, the infimum over $\Gamma_{\mathrm{gen}}(\mathbb{P}_{\mathrm{obs}})$ equals the infimum over all observationally equivalent structural models. On $\mathcal{U}_{\mathrm{id}}(x, w)$ the conditional outcome kernel is fixed $|\omega(w, x, \cdot)|\, \mathrm{d}u$-a.e. by κ_{obs} under Assumption 3.5; on the complement the infimum is attained by the Dirac mass $\delta_{y_{\min}}$ where $\omega(w, x, u) \geqslant 0$ and $\delta_{y_{\max}}$ where $\omega(w, x, u) < 0$. The resulting piecewise kernel $\pi^\star_{w,x}(\mathrm{d}y \mid u)$ is jointly measurable in (w, x, u) because κ_{obs}, ω, and the graph of $\mathcal{U}_{\mathrm{id}}$ are jointly measurable by assumption. Therefore $\pi^\star_{w,x}(\mathrm{d}y, \mathrm{d}u) \coloneqq \pi^\star_{w,x}(\mathrm{d}y \mid u)\, \mathrm{d}u$ defines a jointly measurable family of joint laws in $\Gamma_{\mathrm{gen}}(\mathbb{P}_{\mathrm{obs}})$ and attains the global infimum. □

Proof of Theorem 3.8. We prove the sharp lower bound; the upper bound follows by replacing the countermonotonic coupling with the comonotonic one and swapping $y_{\min}$ and $y_{\max}$ in the trivial bound term.

For each (x, w)-stratum, the sub-problem from (7) reads:

$$\underline{\theta}(x, w) = \min_{\pi_{w,x}} \int_0^1 \mathbb{E}_{\pi_{w,x}}[Y(w) \mid U = u]\, \omega(w, x, u)\, \mathrm{d}u$$

subject to the observational constraints: for every $z \in \mathcal{Z}(x, w)$,

$$\frac{1}{|I_{x,w}(z)|} \int_{I_{x,w}(z)} \pi_{w,x}(\mathrm{d}y \mid u)\, \mathrm{d}u = \mathbb{P}_{\mathrm{obs}}(\mathrm{d}y \mid W = w, Z = z, X = x),$$

and the uniform marginal constraint $\pi_{w,x}(\mathrm{d}u) = \mathrm{d}u$, where $I_{x,w}(z) = \{u \in [0, 1] : F^{-1}_{W|Z,X}(u \mid z, x) = w\}$ is the latent support interval and $\omega(w, x, u)$ is the identifiable weight function.

Separation over (x, w) strata. By the same separability argument as in the proof of Theorem 3.6, the minimization and the outer summation can be exchanged, yielding $\underline{\theta}_\omega = \sum_{w \in \mathcal{W}} \mathbb{E}_X[\underline{\theta}(X, w)]$. It therefore suffices to solve each sub-problem individually.

Decomposition of the objective for fixed (x, w). Fix (x, w) and drop them from the notation where unambiguous. By Assumption 3.7, the constraint (11) is a finite system. The identified support is $\mathcal{U}_{\mathrm{id}}(x, w) = \bigcup_{z \in \mathcal{Z}(x,w)} I(z)$, and the unconstrained region is $J(\emptyset) = [0, 1] \setminus \mathcal{U}_{\mathrm{id}}(x, w)$. We decompose the objective accordingly:

$$\int_0^1 \mathbb{E}_\pi[Y(w) \mid U = u]\, \omega(w, x, u)\, \mathrm{d}u = \underbrace{\int_{\mathcal{U}_{\mathrm{id}}(x,w)} \mathbb{E}_\pi[Y(w) \mid U = u]\, \omega(w, x, u)\, \mathrm{d}u}_{\text{constrained contribution}} + \underbrace{\int_{J(\emptyset)} \mathbb{E}_\pi[Y(w) \mid U = u]\, \omega(w, x, u)\, \mathrm{d}u}_{\text{unconstrained contribution}}.$$

On $J(\emptyset)$, the data impose no restriction on $\pi(\mathrm{d}y \mid u)$, so this region is optimized separately (see below). On $\mathcal{U}_{\mathrm{id}}(x, w)$, the intervals $\{I(z)\}_{z \in \mathcal{Z}(x,w)}$ may overlap. We disentangle these constraints using the inclusion DAG from Assumption 3.7.

Disentangling the constraints via the inclusion DAG. Recall that the constraint (11) reads: for every $z \in \mathcal{Z}(x, w)$,

$$\frac{1}{|I(z)|} \int_{I(z)} \pi(\mathrm{d}y \mid u)\, \mathrm{d}u = \mathbb{P}_{\mathrm{obs}}(\mathrm{d}y \mid W = w, Z = z, X = x),$$

where $I(z) = I_{x,w}(z)$ and $\pi = \pi_{w,x}$.

Under Assumption 3.7, the finite intersection-closed family $\{I(z)\}_{z \in \mathcal{Z}(x,w)}$ admits a directed acyclic ordering by strict inclusion. Recall from (12) the sub-regions $J(z) = I(z) \setminus \bigcup_{z' \in \mathrm{children}(z)} I(z')$ and the recursively defined measures $\mu_{w,z,x}$. These sub-regions are pairwise disjoint and partition $\mathcal{U}_{\mathrm{id}}(x, w)$ up to null boundaries. Indeed, for any u in the identified support, intersecting all family members that contain u gives, by finiteness and the π-system property, the unique smallest interval $I(z_u)$ containing u. No strict child of z_u contains u, so $u \in J(z_u)$. Conversely, if $u \in J(z) \cap J(z')$, intersection closure produces a family member contained in both $I(z)$ and $I(z')$ and containing u; the definition of the isolated regions then forces $z = z'$. Consequently, for every z,

$$I(z) = J(z) \,\dot\cup \bigdot_{z' \in \mathrm{Dec}(z)} J(z') \quad \text{up to null boundaries.}$$

We now verify that the original constraints are equivalent to the simplified marginal constraints (13).

We proceed by induction along the inclusion DAG, beginning with its minimal nodes. For a leaf node z (i.e., $\text{children}(z) = \emptyset$), we have $J(z) = I(z)$, so (11) directly gives

$$\frac{1}{|J(z)|}\int_{J(z)} \pi(\mathrm{d}y \mid u)\,\mathrm{d}u = \mathbb{P}_{\text{obs}}(\mathrm{d}y \mid W=w, Z=z, X=x) = \mu_{w,z,x}(\mathrm{d}y).$$

For an internal node z, suppose (13) holds for all descendants. The disjoint decomposition above and (11) give

$$\begin{aligned}|I(z)|\,\mathbb{P}_{\text{obs}}(\mathrm{d}y \mid w,z,x) &= \int_{I(z)} \pi(\mathrm{d}y \mid u)\,\mathrm{d}u \\ &= \int_{J(z)} \pi(\mathrm{d}y \mid u)\,\mathrm{d}u \;+\; \sum_{z'\in\text{Dec}(z)} \int_{J(z')} \pi(\mathrm{d}y \mid u)\,\mathrm{d}u.\end{aligned}$$

By the inductive hypothesis, the second term equals $\sum_{z'\in\text{Dec}(z)} |J(z')|\,\mu_{w,z',x}(\mathrm{d}y)$. Rearranging and dividing by $|J(z)|$ recovers exactly the definition (12):

$$\frac{1}{|J(z)|}\int_{J(z)} \pi(\mathrm{d}y \mid u)\,\mathrm{d}u = \mu_{w,z,x}(\mathrm{d}y).$$

This proves that the original constraints imply the isolated constraints. Conversely, suppose the isolated constraints hold. Summing them over the disjoint decomposition of $I(z)$ displayed above and using the recursive definition (12) yields

$$\int_{I(z)} \pi(\mathrm{d}y \mid u)\,\mathrm{d}u = |I(z)|\,\mathbb{P}_{\text{obs}}(\mathrm{d}y \mid W=w, Z=z, X=x),$$

which recovers (11). Hence the two constraint systems are equivalent.

Independent OT sub-problems on the constrained regions. Since the regions $\{J(z)\}_{z\in\mathcal{Z}(x,w)}$ are pairwise disjoint and partition the identified support $\mathcal{U}_{\text{id}}(x,w)$, and the simplified constraint (13) couples $\pi(\mathrm{d}y \mid u)$ only within each $J(z)$, the constrained contribution from the decomposition above further decomposes:

$$\int_{\mathcal{U}_{\text{id}}(x,w)} \mathbb{E}_\pi[Y(w) \mid U=u]\,\omega(w,x,u)\,\mathrm{d}u = \sum_{z\in\mathcal{Z}(x,w)} \int_{J(z)} \mathbb{E}_\pi[Y(w) \mid U=u]\,\omega(w,x,u)\,\mathrm{d}u.$$

No cross-region constraints exist, so the optimization on the identified support reduces to independent 1D problems.

For each $z \in \mathcal{Z}(x,w)$, the constraint (13) fixes the marginal distribution of $Y(w)$ on $J(z)$ to be $\mu_{w,z,x}$, while the U-marginal is $\text{Unif}(J(z))$. The sub-problem on $J(z)$ is therefore a standard 1D optimal transport problem:

$$|J(z)| \min_{\gamma_z\in\Pi(\mu_{w,z,x},\,\text{Unif}(J(z)))} \int_{\mathcal{Y}\times J(z)} y\cdot\omega(w,x,u)\,\mathrm{d}\gamma_z(y,u).$$

Closed-form solution via countermonotonic coupling. Let $T_{w,x}(u) = \omega(w,x,u)$ and let $\rho_{z,w,x}$ be the pushforward of $\mathrm{Unif}(J(z))$ under $T_{w,x}$. Pushing a coupling of (Y,U) forward to $(Y, T_{w,x}(U))$, and conversely lifting any coupling of $(Y, T_{w,x}(U))$ through a regular conditional law of U given $T_{w,x}(U)$, shows that this problem is equivalent to product-cost transport between $\mu_{w,z,x}$ and $\rho_{z,w,x}$. Applying Theorem 2.4 gives

$$|J(z)| \min_{\gamma_z \in \Pi(\mu_{w,z,x},\, \mathrm{Unif}(J(z)))} \int y \cdot \omega(w,x,u)\, \mathrm{d}\gamma_z = |J(z)| \int_0^1 Q_{Y,J(z)}(t)\, Q_{\omega,J(z)}(1-t)\, \mathrm{d}t,$$

where $Q_{Y,J(z)}$ is the quantile function of $\mu_{w,z,x}$ and $Q_{\omega,J(z)}$ is the quantile function of $\omega(w,x,U)$ for $U \sim \mathrm{Unif}(J(z))$. The prefactor $|J(z)| = \mathbb{P}(U \in J(z))$ is the Lebesgue mass of the region, since γ_z is a probability measure.

Trivial bound on the unconstrained region. On $J(\emptyset) = [0,1] \setminus \bigcup_{z \in \mathcal{Z}(x,w)} I(z)$, the data impose no restriction on $\pi(\mathrm{d}y \mid u)$. By Assumption 2.2, the pointwise minimum is $\pi(\cdot \mid u) = \delta_{y_{\min}}$ when $\omega(w,x,u) \geqslant 0$ and $\pi(\cdot \mid u) = \delta_{y_{\max}}$ when $\omega(w,x,u) < 0$, contributing:

$$\int_{J(\emptyset)} \Big(y_{\min} \max\{0, \omega(w,x,u)\} + y_{\max} \min\{0, \omega(w,x,u)\} \Big) \mathrm{d}u.$$

Aggregation. Summing over all constrained regions and the unconstrained remainder:

$$\begin{aligned} \underline{\theta}(x,w) = &\sum_{z \in \mathcal{Z}(x,w)} |J_{x,w}(z)| \int_0^1 Q_{Y,J_{x,w}(z)}(t)\, Q_{\omega,J_{x,w}(z)}(1-t)\, \mathrm{d}t \\ &+ \int_{J_{x,w}(\emptyset)} \Big(y_{\min} \max\{0, \omega(w,x,u)\} + y_{\max} \min\{0, \omega(w,x,u)\} \Big) \mathrm{d}u. \end{aligned}$$

The global bound follows by summing over treatment levels:

$$\underline{\theta}_\omega = \sum_{w \in \mathcal{W}} \mathbb{E}_X[\underline{\theta}(X,w)].$$

Writing $B_Y = \max\{|y_{\min}|, |y_{\max}|\}$, its absolute value is dominated by

$$B_Y \sum_{w \in \mathcal{W}} \int_0^1 |\omega(w,X,u)|\, \mathrm{d}u,$$

whose expectation is finite by Assumption 3.2 and counting measure on the finite set $\mathcal{W}$.

Sharpness. Sharpness follows by the same argument as in the proof of Theorem 3.6: on each $J(z)$ the infimum is attained by the countermonotonic coupling, on $J(\emptyset)$ by the appropriate Dirac masses, and a measurable selection argument assembles these into a globally feasible optimizer. □

References

Dimitri P. Bertsekas and Steven E. Shreve. *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific, Belmont, MA, 1996.

Richard Blundell and James L Powell. Endogeneity in nonparametric and semiparametric regression models. *Econometric society monographs*, 36:312–357, 2003.

Andrew Chesher. Identification in nonseparable models. *Econometrica*, 71(5):1405–1441, 2003.

Alfred Galichon. *Optimal transport methods in economics*. Princeton University Press, 2016.

James J Heckman and Edward J Vytlacil. Econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of econometrics*, 6:4875–5143, 2007.

Guido W Imbens and Whitney K Newey. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512, 2009.

Jiyuan Tan, Jose Blanchet, and Vasilis Syrgkanis. Partial identification of policy-relevant treatment effects with instrumental variables via optimal transport. *arXiv preprint arXiv:2604.12263*, 2026.

Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.