Manuscript submitted to *The Econometrics Journal*, pp. 1–21.

Distributional Policy Evaluation via Optimal Transport

Stanisław M. S. Halkiewicz[†]

[†] *AGH University of Cracow, al. Adama Mickiewicza 30, 30-054 Kraków, Lesser Poland, Poland*
E-mail: smsh@student.agh.edu.pl

Summary We propose an optimal-transport framework for analyzing policy effects on outcome distributions. Let $\mu_d(\cdot \mid x)$ denote the conditional distribution of outcomes under policy state $d \in 0, 1$ given covariates $X = x$. Under standard causal assumptions, we show that the counterfactual distributional shift induced by a policy can be represented by an optimal transport map T_x satisfying $\mu_1(\cdot \mid x) = T_x \# \mu_0(\cdot \mid x)$. This representation provides a unified and interpretable description of how policies reallocate probability mass across the outcome distribution. We establish identification of the transport map and develop feasible estimation procedures based on empirical optimal transport. Focusing on policy-relevant functionals of the distributional shift—such as tail effects, inequality measures, and welfare criteria—we derive asymptotic theory and construct confidence intervals. Monte Carlo simulations demonstrate good finite-sample performance, including in designs where policy effects are localized in the tails of the outcome distribution.

Keywords: *optimal transport, Wasserstein distance, distributional treatment effects, policy evaluation, causal inference, nonparametric inference*

1. INTRODUCTION

Let Y denote an outcome of interest and let D indicate exposure to a policy intervention. Empirical policy evaluation traditionally focuses on scalar summaries of treatment effects, such as average treatment effects or quantile treatment effects. While these objects are informative in many settings, they can be inadequate when policies primarily affect the *shape* of the outcome distribution rather than its location. Examples include interventions that alter tail risks, compress or expand dispersion, or reallocate probability mass across different regions of the outcome distribution.

In such environments, mean effects may be small or even zero despite substantial distributional changes, and quantile effects—while informative at specific points—do not provide a coherent representation of the overall transformation induced by the policy. More generally, standard treatment-effect parameters do not directly encode how the entire counterfactual outcome distribution under treatment relates to the distribution absent the policy. This limitation motivates the development of econometric methods that treat policy effects as genuinely *distributional objects* rather than collections of unrelated scalar summaries.

1.1. Contribution and preview

This paper proposes a transport-based framework for causal policy analysis that models the effect of a policy as a structured transformation between counterfactual outcome distributions. Our approach builds on optimal transport theory to provide a unified representation of distributional policy effects that is both interpretable and amenable to statistical inference.

We make three main contributions.

1 **Identification.** Under standard identifying assumptions—consistency, overlap, and unconfoundedness—we show that the policy-induced distributional shift can be identified as the optimal transport map T_x pushing the counterfactual distribution $\mu_0(\cdot \mid x)$ to $\mu_1(\cdot \mid x)$ for each covariate profile x. This map provides a canonical, economically meaningful representation of how the policy reallocates outcomes across the distribution.
2 **Estimation.** We propose feasible estimators of the transport map based on empirical conditional distributions and regularized optimal transport. Rather than requiring estimation of the full map in high resolution, we focus on policy-relevant functionals of T_x that summarize distributional changes in terms of welfare, tail probabilities, or inequality measures.
3 **Inference.** We derive asymptotic theory for plug-in estimators of transport-based policy functionals $\Delta_\phi(x)$. We establish consistency and asymptotic normality under mild conditions and develop bootstrap procedures for constructing confidence intervals and confidence bands. Monte Carlo evidence demonstrates good finite-sample performance, including in designs where policy effects are localized in the tails of the outcome distribution.

1.2. Related literature

This paper contributes to the literature on *distributional policy effects* that studies how policies reshape entire outcome distributions, not only means. A large applied and methodological tradition constructs counterfactual distributions by reweighting or by modeling conditional distributions, including the reweighting approach of DiNardo et al. (1996) and the *changes-in-changes* framework of Athey and Imbens (2006). Related approaches include unconditional quantile methods Firpo et al. (2009) and econometric inference for counterfactual distributions Chernozhukov et al. (2013). Our framework is complementary: we represent the counterfactual shift as a structured mapping between distributions and then conduct inference on policy-relevant functionals of that mapping.

We also build on the growing use of optimal transport and Wasserstein distances in economics, econometrics, and statistics. For economic applications and methodology, a central reference is Galichon (2016). On the computational side, Peyré and Cuturi (2019) provides a modern treatment of scalable algorithms that underpin empirical OT. From a statistical perspective, the Wasserstein space viewpoint and its implications for inference are surveyed in Panaretos and Zemel (2020), while standard OT theory is covered in, e.g., Villani (2009) and Santambrogio (2015).

The remainder of the paper is organized as follows. Section 2 introduces the causal framework and transport-based estimands. Section 3 establishes identification of the transport map and policy functionals. Section 4 presents estimation and inference procedures. Section 6 reports simulation results illustrating finite-sample performance.

2. SETUP AND CAUSAL ESTIMANDS

This section introduces the causal framework and defines the policy effects studied in the paper. We formalize policy impacts as transformations of outcome distributions and identify economically interpretable scalar functionals that will be the focus of statistical inference.

2.1. Potential outcomes and counterfactual distributions

Let $(Y(0), Y(1), X, D)$ be defined on a common probability space, where $Y(d)$ denotes the potential outcome under policy state $d \in \{0, 1\}$, D is the observed policy indicator, and X is a vector of observed covariates. The realized outcome satisfies $Y = Y(D)$.

For each $d \in \{0, 1\}$ and covariate value x, define the conditional counterfactual outcome distribution

$$\mu_d(\cdot \mid x) := \mathcal{L}\big(Y(d) \mid X = x\big). \tag{2.1}$$

Our objective is to characterize and estimate how the entire outcome distribution changes as the policy moves from $d = 0$ to $d = 1$, rather than focusing solely on low-dimensional summaries such as means or quantiles.

2.2. Optimal transport representation of policy effects

Fix a covariate value x. To describe the distributional effect of the policy, we consider the quadratic-cost Wasserstein distance between $\mu_0(\cdot \mid x)$ and $\mu_1(\cdot \mid x)$,

$$W_2^2\big(\mu_0(\cdot \mid x), \mu_1(\cdot \mid x)\big) = \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int \|y_0 - y_1\|^2 \, \pi(\,\mathrm{d}y_0, \,\mathrm{d}y_1), \tag{2.2}$$

where $\Pi(\mu_0, \mu_1)$ denotes the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\mu_0(\cdot \mid x)$ and $\mu_1(\cdot \mid x)$.

Under standard regularity conditions—most importantly, absolute continuity of $\mu_0(\cdot \mid x)$—the optimal coupling is induced by a measurable transport map $T_x : \mathbb{R}^d \to \mathbb{R}^d$ satisfying

$$\mu_1(\cdot \mid x) = T_x \# \mu_0(\cdot \mid x), \qquad T_x \in \operatorname{argmin}_{T:\, T\#\mu_0 = \mu_1} \int \|y - T(y)\|^2 \, \mu_0(\,\mathrm{d}y \mid x). \tag{2.3}$$

The map T_x provides a structural representation of the policy effect: for an individual with baseline outcome y and covariates x, the vector $T_x(y) - y$ describes the policy-induced outcome shift.

2.3. Policy-relevant functionals

Inference on the full function T_x is neither necessary nor desirable for most policy questions. Instead, we focus on scalar functionals that summarize economically meaningful aspects of the distributional change.

Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. We define the *transport-based distributional policy effect*

$$\Delta_\phi(x) := \int \phi\big(T_x(y) - y\big) \, \mu_0(\,\mathrm{d}y \mid x). \tag{2.4}$$

This formulation nests a wide range of policy-relevant parameters. In particular:

$$\text{(Average treatment effect)} \quad \phi(u) = u \;\Rightarrow\; \Delta_\phi(x) = \mathbb{E}[Y(1) - Y(0) \mid X = x], \tag{2.5}$$

$$\text{(Tail probability effect)} \quad \Delta_{\text{tail}}(x; c) := \int \Big(\mathbf{1}\{T_x(y) \le c\} - \mathbf{1}\{y \le c\} \Big)\, \mu_0(\,\mathrm{d}y \mid x), \tag{2.6}$$

$$\text{(Welfare or loss effect)} \quad \Delta_\ell(x) := \int \big(\ell(T_x(y)) - \ell(y) \big)\, \mu_0(\,\mathrm{d}y \mid x), \quad \text{for a loss or utility function } \ell. \tag{2.7}$$

Thus, classical average and quantile-based policy effects arise as special cases or projections of the transport-based representation. As a complementary summary measure, we also consider the magnitude of the overall distributional change as measured by the Wasserstein distance itself,

$$\Delta_W(x) := W_2\big(\mu_0(\cdot \mid x), \mu_1(\cdot \mid x)\big). \tag{2.8}$$

Unlike (2.4), $\Delta_W(x)$ is not a causal effect in the sense of a mean or quantile difference, but it provides a natural scalar index of how strongly the policy reshapes the outcome distribution.

3. IDENTIFICATION

This section establishes identification of policy-induced distributional effects. We proceed in two steps. First, we show that counterfactual outcome distributions are identified from observables under standard causal assumptions. Second, we demonstrate that the policy effect can be represented as an optimal transport map between these distributions and that policy-relevant functionals of this map are identified.

3.1. Baseline identifying assumptions

We work within the standard potential outcomes framework. Let $Y(d)$ denote the potential outcome under policy state $d \in \{0, 1\}$, and let $Y = Y(D)$ be the observed outcome.

Assumption 3.1. (Consistency) *The observed outcome satisfies $Y = Y(D)$ almost surely.*

Assumption 3.2. (Overlap) *For all x in the support of X, $0 < \Pr(D = 1 \mid X = x) < 1$.*

Assumption 3.3. (Unconfoundedness) $(Y(0), Y(1)) \perp D \mid X$.

Assumptions 3.1–3.3 are standard in the treatment effects literature and ensure that, conditional on observed covariates, treatment assignment is as good as random. They are imposed solely for identification and do not restrict the functional form of outcome distributions.

3.2. Identification of counterfactual distributions

Let

$$\mu_d(\cdot \mid x) := \mathcal{L}(Y(d) \mid X = x), \qquad d \in \{0, 1\},$$

denote the counterfactual outcome distribution under policy state d for individuals with covariates $X = x$.

Proposition 3.1. (Identification of counterfactual distributions) *Under Assumptions 3.1–3.3, for each $d \in \{0, 1\}$ and each x in the support of X,*

$$\mu_d(\cdot \mid x) = \mathcal{L}(Y \mid D = d, X = x).$$

Proposition 3.1 follows directly from standard arguments. Unconfoundedness implies that conditioning on $X = x$ suffices to recover the distribution of potential outcomes from the observed data, while overlap ensures that these distributions are identified for all relevant covariate values.

3.3. Identification of the transport map and policy functionals

Having identified the counterfactual distributions, we now formalize the policy effect as a transformation between them. Throughout, we work with the quadratic-cost Wasserstein distance.

Assumption 3.4. (Regularity for optimal transport) *For each x, the distribution $\mu_0(\cdot \mid x)$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$ and has finite second moment. The distribution $\mu_1(\cdot \mid x)$ has finite second moment.*

Assumption 3.4 is standard in optimal transport theory and ensures existence and uniqueness of an optimal transport map between $\mu_0(\cdot \mid x)$ and $\mu_1(\cdot \mid x)$. Importantly, it places no parametric restrictions on outcome distributions and is compatible with skewness, multimodality, and heavy tails.

Under Assumption 3.4, there exists a unique (up to μ_0-null sets) measurable map $T_x : \mathbb{R}^d \to \mathbb{R}^d$ solving

$$\mu_1(\cdot \mid x) = T_x \# \mu_0(\cdot \mid x),$$

where $T_x \# \mu_0$ denotes the pushforward of μ_0 under T_x.

Theorem 3.1. (Identification of distributional policy effects) *Under Assumptions 3.1–3.4, for each x the optimal transport map T_x is uniquely defined $\mu_0(\cdot \mid x)$-almost everywhere and is identified from the observable data. Consequently, for any measurable function $\phi : \mathbb{R}^d \to \mathbb{R}$ satisfying*

$$\int |\phi(T_x(y) - y)| \, \mu_0(\,\mathrm{d}y \mid x) < \infty,$$

the policy functional

$$\Delta_\phi(x) = \int \phi(T_x(y) - y) \, \mu_0(\,\mathrm{d}y \mid x)$$

is identified.

The proof is delegated to B.1. Theorem 3.1 establishes that, under standard causal assumptions, policy effects can be identified not only through scalar summaries but through the entire distributional transformation induced by the policy. The object $T_x(y) - y$ can be interpreted as the policy-induced outcome shift at baseline outcome y for individuals with covariates x.

3.4. Relation to average and quantile treatment effects

Classical treatment effect parameters arise as special cases or projections of the transport-based representation. In particular, the average treatment effect satisfies

$$\mathbb{E}[Y(1) - Y(0) \mid X = x] = \int \big(T_x(y) - y\big)\, \mu_0(\,\mathrm{d}y \mid x),$$

corresponding to the choice $\phi(u) = u$.

In the univariate case ($d = 1$), the optimal transport map admits a closed-form expression. Let $F_{d|x}$ denote the conditional cumulative distribution function of $\mu_d(\cdot \mid x)$. Then

$$T_x(y) = F_{1|x}^{-1}\big(F_{0|x}(y)\big), \tag{3.9}$$

which coincides with the monotone rearrangement underlying quantile treatment effects. Thus, quantile treatment effects correspond to evaluating the transport map at specific quantile indices, while the present framework captures the entire distributional reallocation induced by the policy.

This representation is crucial to understand the relationship between traditional scalar treatment effects and the distributional policy effect studied in this paper. Namely, that mean and quantile effects summarize particular aspects of T_x, whereas optimal transport provides a unified description of how the policy reshapes the outcome distribution.

4. ESTIMATION

4.1. Sampling framework

Let $\{(Y_i, D_i, X_i)\}_{i=1}^n$ be an independent and identically distributed sample drawn from the joint distribution of (Y, D, X). We treat the distribution of X as unrestricted and allow the outcome Y to be multivariate. Throughout this section, estimation is described conditionally on a fixed covariate value x. Unconditional estimators are obtained as special cases by omitting the conditioning on X.

Our objective is to construct feasible estimators of the counterfactual distributions

$$\mu_d(\cdot \mid x) = \mathcal{L}(Y(d) \mid X = x), \qquad d \in \{0, 1\},$$

the associated optimal transport map T_x, and policy-relevant functionals $\Delta_\phi(x)$ defined in Section 2.

4.2. Estimation of counterfactual distributions

By Proposition 3.1, the counterfactual distribution $\mu_d(\cdot \mid x)$ coincides with the observed conditional distribution of Y given $(D = d, X = x)$. Accordingly, estimation of $\mu_d(\cdot \mid x)$ reduces to nonparametric estimation of a conditional distribution.

We consider estimators of the form

$$\widehat{\mu}_d(\cdot \mid x) = \sum_{i=1}^{n} \widehat{w}_{i,d}(x)\, \delta_{Y_i}(\cdot), \qquad d \in \{0,1\}, \tag{4.10}$$

where δ_y denotes the Dirac measure at y and $\widehat{w}_{i,d}(x)$ are nonnegative weights satisfying $\sum_{i=1}^{n} \widehat{w}_{i,d}(x) = 1$.

A convenient choice is kernel weighting:

$$\widehat{w}_{i,d}(x) = \frac{\mathbf{1}\{D_i = d\} K_h(X_i - x)}{\sum_{j=1}^{n} \mathbf{1}\{D_j = d\} K_h(X_j - x)}, \tag{4.11}$$

where $K_h(u) = h^{-p} K(u/h)$ for a kernel function K and bandwidth $h > 0$, with $p = \dim(X)$. Alternative implementations include series estimators, local polynomial methods, or inverse-propensity weighted estimators based on a first-stage estimate of $\Pr(D = 1 \mid X)$. The theoretical results developed below apply to a broad class of estimators $\widehat{\mu}_d(\cdot \mid x)$ satisfying standard consistency and smoothness conditions. The representation in (4.10) yields discrete empirical measures with finite support.

4.3. Empirical optimal transport and estimation of the transport map

Given estimators $\widehat{\mu}_0(\cdot \mid x)$ and $\widehat{\mu}_1(\cdot \mid x)$ constructed as in Section 4.2, we estimate the policy-induced distributional shift via an empirical optimal transport problem.

Because $\widehat{\mu}_0(\cdot \mid x)$ and $\widehat{\mu}_1(\cdot \mid x)$ are discrete measures with finite support, the quadratic-cost optimal transport problem admits a finite-dimensional formulation. Let

$$\widehat{\mu}_d(\cdot \mid x) = \sum_{i=1}^{n} \widehat{w}_{i,d}(x)\, \delta_{Y_i}(\cdot), \qquad d \in \{0,1\}.$$

Define the set of admissible couplings

$$\Pi\big(\widehat{\mu}_0(\cdot \mid x), \widehat{\mu}_1(\cdot \mid x)\big) = \Big\{\pi \in \mathbb{R}_+^{n \times n} : \sum_{j=1}^{n} \pi_{ij} = \widehat{w}_{i,0}(x),\ \sum_{i=1}^{n} \pi_{ij} = \widehat{w}_{j,1}(x)\Big\}.$$

The empirical optimal transport plan $\widehat{\pi}_x$ is defined as a solution to

$$\widehat{\pi}_x \in \operatorname{argmin}_{\pi \in \Pi(\widehat{\mu}_0, \widehat{\mu}_1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \|Y_i - Y_j\|^2 \pi_{ij}. \tag{4.12}$$

The optimization problem (4.12) is a linear program and admits at least one solution. When multiple solutions exist, any selection yields identical values for the policy functionals considered below.

To recover a transport map from the empirical coupling, we use the barycentric projection. For each support point Y_i of $\widehat{\mu}_0(\cdot \mid x)$ with $\widehat{w}_{i,0}(x) > 0$, define

$$\widehat{T}_x(Y_i) = \frac{1}{\widehat{w}_{i,0}(x)} \sum_{j=1}^{n} \pi_{ij} Y_j, \qquad \pi = \widehat{\pi}_x. \tag{4.13}$$

This construction yields a measurable map $\widehat{T}_x$ defined on the support of $\widehat{\mu}_0(\cdot \mid x)$ and provides a natural finite-sample analogue of the population transport map T_x.

In practice, computation of (4.12) may be regularized for numerical stability. A common approach is entropic regularization, which replaces (4.12) with

$$\widehat{\pi}_{x,\varepsilon} \in \operatorname{argmin}_{\pi\in\Pi(\widehat{\mu}_0,\widehat{\mu}_1)} \left\{ \sum_{i=1}^{n}\sum_{j=1}^{n} \|Y_i - Y_j\|^2 \pi_{ij} + \varepsilon \sum_{i=1}^{n}\sum_{j=1}^{n} \pi_{ij} \log \pi_{ij} \right\}, \tag{4.14}$$

for a regularization parameter $\varepsilon > 0$. The corresponding transport map estimator is obtained by replacing $\widehat{\pi}_x$ with $\widehat{\pi}_{x,\varepsilon}$ in (4.13). All subsequent analysis applies to both the unregularized and regularized estimators, provided ε is chosen such that regularization bias is asymptotically negligible.

The estimator $\widehat{T}_x$ captures the estimated policy-induced outcome shift conditional on $X = x$.

4.4. Estimation of policy-relevant functionals

The objects of primary interest are scalar functionals of the policy-induced distributional shift defined in Section 2. Given the estimated transport map $\widehat{T}_x$ and the estimated baseline distribution $\widehat{\mu}_0(\cdot \mid x)$, we estimate these functionals by plug-in methods.

Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be a measurable function satisfying the integrability condition in Theorem 3.1. Recall that the population policy functional is

$$\Delta_\phi(x) = \int \phi\big(T_x(y) - y\big)\, \mu_0(\,\mathrm{d}y \mid x).$$

The corresponding estimator is defined as

$$\widehat{\Delta}_\phi(x) := \int \phi\big(\widehat{T}_x(y) - y\big)\, \widehat{\mu}_0(\,\mathrm{d}y \mid x). \tag{4.15}$$

Under the discrete representation (4.10), this estimator admits the explicit form

$$\widehat{\Delta}_\phi(x) = \sum_{i=1}^{n} \widehat{w}_{i,0}(x)\, \phi\big(\widehat{T}_x(Y_i) - Y_i\big), \tag{4.16}$$

where $\widehat{w}_{i,0}(x)$ are the weights defining $\widehat{\mu}_0(\cdot \mid x)$.

This formulation accommodates a wide range of economically relevant parameters. For example:

- choosing $\phi(u) = u$ yields an estimator of the conditional average treatment effect;
- indicator-type functions $\phi(u) = \mathbf{1}\{y + u \le c\} - \mathbf{1}\{y \le c\}$ target changes in tail probabilities;
- loss-based functions $\phi(u) = \ell(y + u) - \ell(y)$ measure welfare or risk effects induced by the policy.

In addition, the magnitude of the overall distributional change can be summarized by the estimated Wasserstein distance

$$\widehat{\Delta}_W(x) := W_2\big(\widehat{\mu}_0(\cdot \mid x), \widehat{\mu}_1(\cdot \mid x)\big), \tag{4.17}$$

which is obtained directly as the optimal value of the empirical transport problem in (4.12).

For unconditional policy effects, the estimators in (4.15)–(4.17) are defined analogously by replacing conditional distributions with their unconditional counterparts. More generally, integration over x with respect to an estimated distribution of X yields average policy effects across subpopulations.

5. INFERENCE

5.1. Regularity conditions

This section establishes large-sample properties of the plug-in estimators defined in Section 4.4. We impose high-level conditions that are standard in nonparametric treatment effects and empirical optimal transport. Throughout, we consider a fixed covariate value x and suppress explicit dependence on x when this causes no confusion.

ASSUMPTION 5.1. (SAMPLING) *The observations $\{(Y_i, D_i, X_i)\}_{i=1}^n$ are independent and identically distributed.*

ASSUMPTION 5.2. (MOMENTS AND SUPPORT) *For $d \in \{0,1\}$, the conditional distributions $\mu_d(\cdot \mid x)$ have finite moments of order $2+\delta$ for some $\delta > 0$. The support of $\mu_0(\cdot \mid x)$ is compact.*

ASSUMPTION 5.3. (SMOOTHNESS) *For each $d \in \{0,1\}$, $\mu_d(\cdot \mid x)$ admits a density $f_{d|x}$ that is bounded and continuously differentiable on the interior of its support.*

ASSUMPTION 5.4. (FIRST-STAGE ESTIMATION) *The estimators $\widehat{\mu}_d(\cdot \mid x)$ satisfy*

$$\sup_{A\in\mathcal{A}} \left|\widehat{\mu}_d(A \mid x) - \mu_d(A \mid x)\right| = o_p(n^{-1/4}),$$

for a sufficiently rich class of measurable sets $\mathcal{A}$.

ASSUMPTION 5.5. (TRANSPORT REGULARITY) *The optimal transport map T_x is continuously differentiable $\mu_0(\cdot \mid x)$-almost everywhere, and its Jacobian is bounded.*

ASSUMPTION 5.6. (FUNCTIONAL REGULARITY) *The function $\phi : \mathbb{R}^d \to \mathbb{R}$ is Lipschitz continuous and satisfies*

$$\mathbb{E}\left[\|\phi(T_x(Y(0)) - Y(0))\|^2 \mid X = x\right] < \infty.$$

Assumptions 5.2–5.5 are standard in the analysis of empirical optimal transport and ensure that perturbations of the estimated distributions translate smoothly into perturbations of the transport map. Assumption 5.4 allows for a wide class of nonparametric estimators and can be satisfied, for instance, by kernel or series estimators under conventional bandwidth choices.

5.2. Asymptotic distribution of policy effect estimators

We now state the main large-sample result for inference on policy-relevant functionals.

THEOREM 5.1. (ASYMPTOTIC NORMALITY OF TRANSPORT-BASED POLICY EFFECTS) *Suppose Assumptions 3.1–5.6 hold. Then, for any fixed covariate value x,*

$$\sqrt{n}\big(\widehat{\Delta}_\phi(x) - \Delta_\phi(x)\big) \;\Rightarrow\; \mathcal{N}(0,\, V_\phi(x))\,,$$

where $V_\phi(x) \in (0,\infty)$ denotes the asymptotic variance.

The asymptotic variance $V_\phi(x)$ admits an influence function representation. Specifically, there exists a mean-zero function $\psi_\phi(Y,D,X;x)$ such that

$$\sqrt{n}\big(\widehat{\Delta}_\phi(x) - \Delta_\phi(x)\big) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \psi_\phi(Y_i,D_i,X_i;x) + o_p(1), \tag{5.18}$$

and

$$V_\phi(x) = \mathrm{Var}\left(\psi_\phi(Y,D,X;x)\right).$$

The influence function ψ_ϕ reflects two sources of sampling uncertainty: (i) estimation error in the counterfactual distributions $\mu_d(\cdot \mid x)$ and (ii) the induced perturbation of the optimal transport map T_x. While an explicit closed-form expression for ψ_ϕ depends on the specific implementation of $\widehat{\mu}_d(\cdot \mid x)$, its existence and square-integrability are guaranteed under the stated assumptions.

An analogous asymptotic normality result holds for the Wasserstein-based summary measure $\widehat{\Delta}_W(x)$, provided the optimal transport cost is sufficiently smooth in neighborhoods of (μ_0,μ_1).

5.3. Variance estimation and confidence intervals

The asymptotic result in Theorem 5.1 motivates feasible procedures for variance estimation and construction of confidence intervals. We describe two approaches: an influence-function-based estimator and a resampling-based alternative. Both methods are valid under the regularity conditions of Section 5.1.

Influence-function-based variance estimation. Under the representation in (5.18), the asymptotic variance $V_\phi(x)$ can be written as

$$V_\phi(x) = \mathrm{Var}\left(\psi_\phi(Y,D,X;x)\right).$$

If an explicit estimator $\widehat{\psi}_\phi(Y_i,D_i,X_i;x)$ of the influence function is available, a consistent variance estimator is given by

$$\widehat{V}_\phi^{\mathrm{IF}}(x) = \frac{1}{n}\sum_{i=1}^{n} \widehat{\psi}_\phi(Y_i,D_i,X_i;x)^2. \tag{5.19}$$

In practice, the form of $\widehat{\psi}_\phi$ depends on the choice of first-stage estimator for $\widehat{\mu}_d(\cdot \mid x)$ and on the linearization of the empirical transport map. Deriving an explicit expression may be analytically involved, particularly in multivariate settings.

Bootstrap variance estimation. As an alternative, we employ a bootstrap procedure that avoids explicit construction of the influence function. Let $\{(Y_i^*,D_i^*,X_i^*)\}_{i=1}^n$ denote a bootstrap resample drawn with replacement from the original data. For each bootstrap draw, compute

$$\widehat{\Delta}_\phi^*(x)$$

by repeating the full estimation procedure described in Section 4, including estimation of $\widehat{\mu}_d^*(\cdot \mid x)$ and the empirical transport map $\widehat{T}_x^*$.

The bootstrap variance estimator is then

$$\widehat{V}_\phi^{\text{boot}}(x) = \operatorname{Var}^*\left(\widehat{\Delta}_\phi^*(x)\right), \tag{5.20}$$

where $\operatorname{Var}^*(\cdot)$ denotes variance computed over bootstrap repetitions. Under the conditions of Theorem 5.1, the bootstrap consistently estimates $V_\phi(x)$.

Confidence intervals. Given a consistent variance estimator $\widehat{V}_\phi(x)$, either from (5.19) or (5.20), a $(1-\alpha)$ confidence interval for $\Delta_\phi(x)$ is constructed as

$$\left[\widehat{\Delta}_\phi(x) \ \pm \ z_{1-\alpha/2}\sqrt{\frac{\widehat{V}_\phi(x)}{n}}\right], \tag{5.21}$$

where $z_{1-\alpha/2}$ denotes the $(1-\alpha/2)$ quantile of the standard normal distribution.

For functionals involving non-smooth choices of ϕ (e.g. indicator functions targeting tail probabilities), percentile or studentized bootstrap intervals may be preferable. Inference for the Wasserstein-based summary measure $\Delta_W(x)$ proceeds analogously, provided the corresponding asymptotic normality result holds.

Remarks. The bootstrap approach is particularly attractive in the present setting, as it automatically accounts for the interaction between first-stage distributional estimation and the optimal transport step. All empirical results reported below are based on the bootstrap variance estimator (5.20).

6. SIMULATION DESIGN AND EMPIRICAL ILLUSTRATION

This section studies the finite-sample behavior of the proposed transport-based estimators in a stylized policy environment designed to generate distributional effects that are localized in the lower tail of the outcome distribution. The design allows us to assess the ability of the method to detect and quantify policy-induced distributional changes that are not well summarized by mean effects alone. We report Monte Carlo bias, root mean squared error (RMSE), and confidence interval coverage for policy-relevant functionals, and provide a graphical illustration of the implied distributional shift.

6.1. Stylized policy data-generating process

We consider a univariate outcome ($d = 1$) and a binary covariate $X \in \{0, 1\}$. Conditional on $X = x$, the baseline potential outcome $Y(0) \mid X = x$ follows a skewed, heavy-tailed distribution. The policy intervention is constructed to affect only the lower tail of the baseline outcome distribution. Specifically, we define

$$Y(1) = Y(0) - \tau \cdot \mathbf{1}\{Y(0) \leq q_{0|x}(\kappa)\} + \sigma\varepsilon, \tag{6.22}$$

where $q_{0|x}(\kappa)$ denotes the κ-quantile of $Y(0) \mid X = x$, $\tau > 0$ governs the magnitude of the tail shift, and $\varepsilon \sim \mathcal{N}(0, 1)$ is independent noise. This construction induces a negative shift in outcomes below the κ-quantile, while leaving the center and upper tail largely unaffected.

Treatment assignment is independent of potential outcomes conditional on X: given $X = x$, we draw $D \sim \text{Bernoulli}(p(x))$ independently of $(Y(0), Y(1))$. As a result, the identifying assumptions imposed in Section 3 are satisfied by construction.

6.2. Estimands and performance metrics

We evaluate the performance of estimators for two policy-relevant functionals. First, we consider the mean transport functional $\Delta_\phi(x)$ with $\phi(u) = u$, which corresponds to the average treatment effect conditional on $X = x$. Second, we consider a tail-probability functional

$$\Delta_{\text{tail}}(x; c) = \int \Big(\mathbf{1}\{y + (T_x(y) - y) \leq c\} - \mathbf{1}\{y \leq c\} \Big) \, \mu_0(\mathrm{d}y \mid x),$$

which captures changes in the probability mass below a fixed threshold c. As a global summary of distributional change, we also compute the Wasserstein distance

$$\Delta_W(x) = W_2\big(\mu_0(\cdot \mid x), \mu_1(\cdot \mid x)\big).$$

In each Monte Carlo replication, we estimate these functionals using the plug-in optimal transport estimators developed in Sections 4.3–4.4. Confidence intervals are constructed using the bootstrap procedure described in Section 5.3. Performance is summarized across replications in terms of bias, RMSE, and empirical coverage of nominal 95% confidence intervals. The replication-specific “truth” is computed using the realized potential outcomes $(Y(0), Y(1))$ by applying the same transport-based functional to the empirical counterfactual distributions.

6.3. Distributional shift: graphical illustration

To illustrate the nature of the policy-induced distributional change, Figure 1 plots the estimated quantile shift

$$\delta_x(p) = q_{1|x}(p) - q_{0|x}(p), \qquad p \in (0, 1),$$

for a representative replication and covariate value $X = x$. In one dimension, the optimal transport map corresponds to monotone rearrangement, making the quantile-scale representation particularly transparent.

The figure shows that the estimated shift is concentrated in the lower tail of the outcome distribution, with negligible effects at the median and above. The estimated curve closely tracks the population shift computed from potential outcomes, illustrating that the transport-based estimator accurately recovers the shape and localization of the policy effect.

6.4. Monte Carlo results

Table 1 reports Monte Carlo bias, RMSE, and confidence interval coverage for the mean and tail functionals across sample sizes. For the mean functional, both bias and RMSE decrease with sample size, and bootstrap-normal confidence intervals exhibit coverage close to the nominal level. For the tail functional, which is non-smooth by construction, percentile bootstrap intervals exhibit mild under-coverage in smaller samples, while normal bootstrap intervals remain close to nominal.

Overall, the results indicate that the proposed transport-based estimators deliver reliable finite-sample performance and are sensitive to policy effects that are localized in specific regions of the outcome distribution.

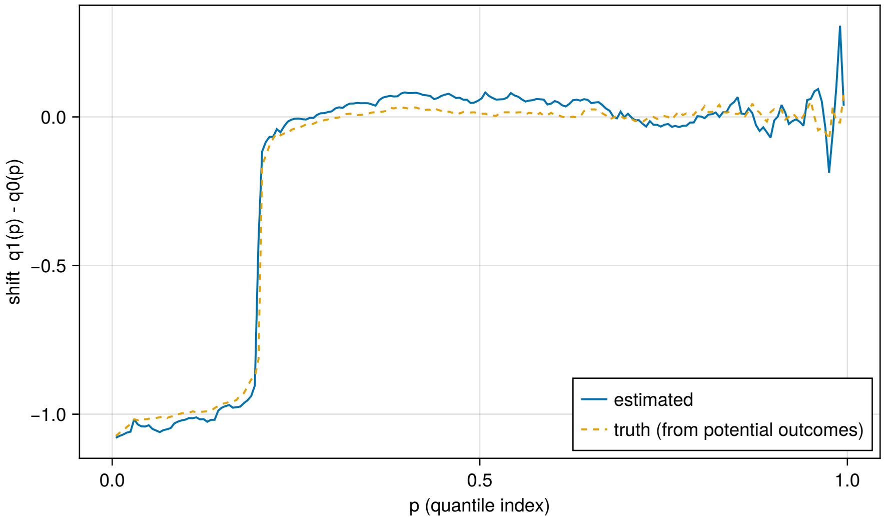

Figure 1. Quantile-scale illustration of the distributional policy effect for $X = 0$. The solid line plots the estimated quantile shift $\widehat{q}_{1|0}(p) - \widehat{q}_{0|0}(p)$, while the dashed line shows the corresponding population shift computed from potential outcomes. The policy effect is concentrated in the lower tail, consistent with the tail-shift design in (6.22).

Table 1. Monte Carlo performance for transport-based policy effects under the tail-shift DGP.

Functional	n	Bias	RMSE	Cov. (normal)	Cov. (percentile)
Δ_ϕ (mean shift)	600	−0.008	0.103	0.965	0.970
Δ_ϕ (mean shift)	1200	−0.003	0.074	0.980	0.975
Δ_ϕ (mean shift)	2400	−0.006	0.063	0.930	0.930
$\Delta_{\text{tail}}(\cdot; c)$	600	−0.005	0.056	0.960	0.965
$\Delta_{\text{tail}}(\cdot; c)$	1200	0.002	0.041	0.930	0.925
$\Delta_{\text{tail}}(\cdot; c)$	2400	0.001	0.028	0.985	0.965

Notes: Results are based on R Monte Carlo replications with B bootstrap draws per replication. The "truth" is computed from realized potential outcomes by applying the same transport-based functional to the empirical counterfactual distributions within each replication.

7. CONCLUSION

This paper proposes an optimal-transport framework for the analysis of policy effects on outcome distributions. By representing the counterfactual distributional change induced by a policy as an optimal transport map, the approach provides a structural and interpretable description of how probability mass is reallocated across the outcome space.

We establish identification of the transport map under standard causal assumptions (consistency, overlap, and unconfoundedness) and show how policy-relevant functionals of the distributional shift can be identified from observed data. Rather than requiring estimation of the full transport map, the analysis focuses on economically meaningful functionals, including mean effects, tail probabilities, and welfare or inequality measures.

We develop feasible estimation procedures based on empirical and regularized optimal transport and derive asymptotic theory for plug-in estimators of transport-based policy functionals. Bootstrap methods provide a practical approach to inference, and Monte Carlo simulations demonstrate good finite-sample performance, including in designs where policy effects are localized in the tails of the outcome distribution.

Relative to traditional policy evaluation methods that focus on mean or quantile effects in isolation, the proposed framework offers a unified representation of distributional policy effects. It is particularly useful in settings where policies primarily affect dispersion, tail risks, or inequality, and where scalar summaries may obscure economically relevant changes. The optimal-transport perspective thus complements existing econometric tools for distributional analysis and provides a flexible foundation for future work on causal inference with rich outcome distributions.

Acknowledgments The author thanks colleagues and seminar participants for helpful discussions and comments.

Funding Statement This research received no external funding.

Competing Interests The author declares no competing interests.

Data Availability Statement Replication code is available publicly at

`https://github.com/profsms/ot-policy`.

REFERENCES

Athey, S. and G. W. Imbens (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica 74*(2), 431–497.

Caffarelli, L. A. (1992). The regularity of mappings with a convex potential. *Journal of the American Mathematical Society 5*(1), 99–104.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica 81*(6), 2205–2268.

Dahal, B., A. Havrilla, M. Chen, T. Zhao, and W. Liao (2022). On deep generative models for approximation and estimation of distributions on manifolds. In A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho (Eds.), *Advances in Neural Information Processing Systems*.

DiNardo, J., N. M. Fortin, and T. Lemieux (1996). Labor market institutions and the distribution of wages, 1973–1992: A semiparametric approach. *Econometrica 64*(5), 1001–1044.

Firpo, S., N. M. Fortin, and T. Lemieux (2009). Unconditional quantile regressions. *Econometrica 77*(3), 953–973.

Friesecke, G. (2024). *Optimal Transport: A Comprehensive Introduction to Modeling, Analysis, Simulation, Applications*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Galichon, A. (2016). *Optimal Transport Methods in Economics*. Princeton University Press.

Panaretos, V. M. and Y. Zemel (2020). *An Invitation to Statistics in Wasserstein Space.* Cham: Springer.

Peyré, G. and M. Cuturi (2019). *Computational Optimal Transport.* Now Publishers.

Santambrogio, F. (2015). *Optimal Transport for Applied Mathematicians*, Volume 87 of *Progress in Nonlinear Differential Equations and Their Applications.* Cham: Birkhäuser.

Santambrogio, F. (2017). {Euclidean, metric, and Wasserstein} gradient flows: an overview. *Bulletin of Mathematical Sciences* *7*(1), 87–154.

Villani, C. (2003). *Topics in Optimal Transportation.* Graduate studies in mathematics. American Mathematical Society.

Villani, C. (2009). *Optimal Transport: Old and New*, Volume 338 of *Grundlehren der mathematischen Wissenschaften.* Berlin, Heidelberg: Springer.

A. TECHNICAL BACKGROUND ON OPTIMAL TRANSPORT

This appendix summarizes the elements of optimal transport theory that are used in the identification, estimation, and asymptotic analysis in the main text. We focus on the quadratic Wasserstein distance, existence and uniqueness of optimal transport maps, and the regularity properties required for plug-in estimation and inference. Standard references include Villani (2009), Santambrogio (2015), Santambrogio (2017), and the recent comprehensive treatment in Friesecke (2024).

A.1. Quadratic Wasserstein distance

Let μ and ν be Borel probability measures on $\mathbb{R}^d$ with finite second moments. The quadratic Wasserstein distance between μ and ν is defined as

$$\mathcal{W}_2^2(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^d\times\mathbb{R}^d} \|x-y\|^2 \,\pi(\mathrm{d}x,\mathrm{d}y), \tag{1.23}$$

where $\Pi(\mu,\nu)$ denotes the set of probability measures on $\mathbb{R}^d\times\mathbb{R}^d$ with marginals μ and ν. This Kantorovich formulation is standard; see, for example, Villani (2009), Santambrogio (2015), Friesecke (2024), and Dahal et al. (2022).

The metric $\mathcal{W}_2$ metrizes weak convergence together with convergence of second moments and provides a natural geometry on the space of probability measures. It plays a central role in modern statistical analysis of distributions and in the formulation of transport-based estimands.

A.2. Existence and uniqueness of optimal transport maps

A key result underlying our identification strategy is Brenier's theorem. Suppose that μ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$ and that both μ and ν have finite second moments. Then there exists a unique (up to μ-almost everywhere equivalence) optimal transport map $T:\mathbb{R}^d\to\mathbb{R}^d$ solving (1.23). Moreover, the map admits the representation

$$T(x) = \nabla\varphi(x),$$

for some convex function $\varphi:\mathbb{R}^d\to\mathbb{R}$.

These results are classical and are presented in detail in f.e. Villani (2009, 2003) and

Santambrogio (2015). In the univariate case ($d = 1$), the optimal transport map reduces to the monotone rearrangement

$$T(y) = F_\nu^{-1}\big(F_\mu(y)\big),$$

where F_μ and F_ν denote the distribution functions of μ and ν, respectively.

A.3. Regularity properties

Beyond existence and uniqueness, the analysis relies on regularity properties of optimal transport maps. When the source and target measures admit densities that are bounded away from zero and infinity on compact, convex supports, the optimal transport map enjoys additional smoothness properties.

In particular, under such conditions the transport map is Hölder continuous and, in many cases, continuously differentiable; see Caffarelli (1992) and subsequent developments summarized in Santambrogio (2017) and Friesecke (2024). These regularity results imply stability of the transport map with respect to perturbations of the underlying distributions and justify the use of plug-in estimators and functional delta-method arguments for transport-based functionals.

A.4. Pushforward measures and transport-based functionals

Given a measurable map $T : \mathbb{R}^d \to \mathbb{R}^d$ and a probability measure μ, the pushforward measure $T\#\mu$ is defined by

$$(T\#\mu)(A) = \mu\big(T^{-1}(A)\big) \quad \text{for all Borel sets } A \subseteq \mathbb{R}^d.$$

Optimal transport maps are characterized precisely by the condition $T\#\mu = \nu$.

B. PROOFS

B.1. Proof of Theorem 3.1

We prove identification of the optimal transport map T_x and of the induced policy functional $\Delta_\phi(x)$.

Step 1: Identification of the conditional counterfactual distributions. Fix x in the support of X and $d \in \{0, 1\}$. By definition, $\mu_d(\cdot \mid x) = \mathcal{L}(Y(d) \mid X = x)$. Under Assumption 3.3, we have $Y(d) \perp D \mid X$, hence

$$\mathcal{L}(Y(d) \mid X = x, D = d) = \mathcal{L}(Y(d) \mid X = x). \tag{2.24}$$

Under Assumption 3.1, on the event $\{D = d\}$ we have $Y = Y(d)$ almost surely, so conditioning on $(X = x, D = d)$ yields

$$\mathcal{L}(Y \mid X = x, D = d) = \mathcal{L}(Y(d) \mid X = x, D = d). \tag{2.25}$$

Combining (2.24) and (2.25) gives

$$\mu_d(\cdot \mid x) = \mathcal{L}(Y(d) \mid X = x) = \mathcal{L}(Y \mid X = x, D = d). \tag{2.26}$$

Assumption 3.2 guarantees $\Pr(D = d \mid X = x) > 0$ for $d \in \{0, 1\}$, so the conditional laws on the right-hand side of (2.26) are well-defined (and hence identified) for all such x.

Step 2: Existence and uniqueness of the optimal transport map. Fix x. Under Assumption 3.4, $\mu_0(\cdot \mid x)$ is absolutely continuous with respect to Lebesgue measure and both $\mu_0(\cdot \mid x)$ and $\mu_1(\cdot \mid x)$ have finite second moments. Consider the quadratic-cost optimal transport problem between $\mu_0(\cdot \mid x)$ and $\mu_1(\cdot \mid x)$:

$$\inf_{\pi \in \Pi(\mu_0(\cdot|x), \mu_1(\cdot|x))} \int \|y_0 - y_1\|^2 \, \pi(\mathrm{d}y_0, \mathrm{d}y_1).$$

By standard results in optimal transport (see, e.g., Villani (2003), Santambrogio (2015) or Friesecke (2024)), there exists a measurable map $T_x : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$\mu_1(\cdot \mid x) = T_x \# \mu_0(\cdot \mid x), \tag{2.27}$$

and the optimal coupling is induced by T_x, i.e., $\pi_x^\star = (\mathrm{Id}, T_x) \# \mu_0(\cdot \mid x)$. Moreover, T_x is unique $\mu_0(\cdot \mid x)$-almost everywhere. (Equivalently, $T_x = \nabla \varphi_x$ for a convex potential φ_x, $\mu_0(\cdot \mid x)$-a.e.)

Step 3: Identification of T_x from observables. By Step 1, both $\mu_0(\cdot \mid x)$ and $\mu_1(\cdot \mid x)$ are identified from the observed distribution of (Y, D, X) through (2.26). By Step 2, for each fixed x there is a unique $\mu_0(\cdot \mid x)$-a.e. optimal transport map T_x satisfying (2.27) and minimizing the quadratic transport cost. Therefore T_x is a uniquely determined functional of the pair $(\mu_0(\cdot \mid x), \mu_1(\cdot \mid x))$. Since this pair is identified, the map T_x is identified as well (uniquely up to $\mu_0(\cdot \mid x)$-null sets).

Step 4: Identification of $\Delta_\phi(x)$. Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be measurable and satisfy the integrability condition in Theorem 3.1:

$$\int |\phi(T_x(y) - y)| \, \mu_0(\mathrm{d}y \mid x) < \infty.$$

Then the quantity

$$\Delta_\phi(x) = \int \phi(T_x(y) - y) \, \mu_0(\mathrm{d}y \mid x)$$

is well-defined. Because both T_x (Step 3) and $\mu_0(\cdot \mid x)$ (Step 1) are identified, $\Delta_\phi(x)$ is identified as a functional of the observable distribution of (Y, D, X).

Conclusion. Steps 1–4 establish that, under Assumptions 3.1–3.4, the optimal transport map T_x is uniquely defined $\mu_0(\cdot \mid x)$-almost everywhere and identified from observables, and that any integrable functional $\Delta_\phi(x)$ of the policy-induced shift $T_x(y) - y$ is identified. □

B.2. Proof of Theorem 5.1

The proof follows a standard plug-in / functional delta method argument. We treat the counterfactual distributions $\mu_0(\cdot \mid x)$ and $\mu_1(\cdot \mid x)$ as the primitive objects and view the estimator

$$\widehat{\Delta}_\phi(x) = \int \phi(\widehat{T}_x(y) - y) \, \widehat{\mu}_0(\mathrm{d}y \mid x)$$

as a functional of the estimated pair $(\widehat{\mu}_0(\cdot \mid x), \widehat{\mu}_1(\cdot \mid x))$.

Notation. Fix x and suppress x in the notation. Write $\mu_d := \mu_d(\cdot \mid x)$ and $\widehat{\mu}_d := \widehat{\mu}_d(\cdot \mid x)$ for $d \in \{0,1\}$. Let $T := T_x$ denote the population optimal transport map from μ_0 to μ_1 and $\widehat{T} := \widehat{T}_x$ its empirical analogue (barycentric projection of an optimal empirical coupling, or its entropic-regularized version). Define the population functional

$$\Delta_\phi := \int \phi(T(y) - y)\, \mu_0(\,\mathrm{d}y) \qquad \text{and} \qquad \widehat{\Delta}_\phi := \int \phi(\widehat{T}(y) - y)\, \widehat{\mu}_0(\,\mathrm{d}y).$$

Step 1: Decomposition into a measure-perturbation term and a map-perturbation term. Add and subtract $\int \phi(T(y) - y)\, \widehat{\mu}_0(\,\mathrm{d}y)$ to obtain

$$\widehat{\Delta}_\phi - \Delta_\phi = \underbrace{\int \Big[\phi(\widehat{T}(y) - y) - \phi(T(y) - y)\Big]\, \widehat{\mu}_0(\,\mathrm{d}y)}_{=:A_n} + \underbrace{\int \phi(T(y) - y)\, (\widehat{\mu}_0 - \mu_0)(\,\mathrm{d}y)}_{=:B_n}. \tag{2.28}$$

Term B_n is a standard plug-in term involving only the first-stage estimator $\widehat{\mu}_0$. Term A_n captures the additional uncertainty due to estimating the transport map.

Step 2: Linearization of B_n. Under Assumptions 5.1 and 5.4, the first-stage estimator admits an asymptotic linear representation (possibly after standard smoothing bias correction): there exists a measurable function $\psi_0(Y, D, X; x)$ with $\mathbb{E}[\psi_0(Y, D, X; x)] = 0$ and finite variance such that

$$\int g(y)\, (\widehat{\mu}_0 - \mu_0)(\,\mathrm{d}y) = \frac{1}{n} \sum_{i=1}^{n} \psi_{0,g}(Y_i, D_i, X_i; x) + o_p(n^{-1/2}) \tag{2.29}$$

for each bounded measurable g in a class rich enough to include $g(y) = \phi(T(y) - y)$. Setting $g(y) = \phi(T(y) - y)$ in (2.29) yields

$$B_n = \frac{1}{n} \sum_{i=1}^{n} \psi_{0,\phi\circ(T-\mathrm{Id})}(Y_i, D_i, X_i; x) + o_p(n^{-1/2}). \tag{2.30}$$

Step 3: Linearization of A_n via functional differentiability of the OT map. We now control the difference $\phi(\widehat{T}(y)-y) - \phi(T(y)-y)$. By Assumption 5.6, ϕ is Lipschitz, hence

$$\big|\phi(\widehat{T}(y) - y) - \phi(T(y) - y)\big| \le L_\phi \|\widehat{T}(y) - T(y)\| \tag{2.31}$$

for some constant L_ϕ. Under Assumptions 5.2–5.5, the mapping

$$(\mu_0, \mu_1) \mapsto T(\cdot; \mu_0, \mu_1)$$

is (locally) Hadamard differentiable at (μ_0, μ_1) as a map into an $L^2(\mu_0)$-type space, and the empirical transport map $\widehat{T}$ admits the first-order expansion

$$\widehat{T} - T = \mathcal{D}_0(\widehat{\mu}_0 - \mu_0) + \mathcal{D}_1(\widehat{\mu}_1 - \mu_1) + r_n, \qquad \|r_n\|_{L^2(\mu_0)} = o_p(n^{-1/2}), \tag{2.32}$$

for bounded linear operators $\mathcal{D}_0, \mathcal{D}_1$ (the Fréchet/Hadamard derivatives of the OT map with respect to its arguments). Intuitively, (2.32) states that small perturbations of the marginal distributions induce (to first order) a linear perturbation of the transport map. Inserting (2.32) into (2.31) and integrating with respect to $\widehat{\mu}_0$ yields

$$A_n = \int \nabla\phi(T(y) - y)^\top \big(\widehat{T}(y) - T(y)\big)\, \mu_0(\,\mathrm{d}y) + o_p(n^{-1/2}), \tag{2.33}$$

where $\nabla\phi$ is understood in the a.e. sense (or as any subgradient if ϕ is only directionally differentiable). Combining (2.32) and (2.33) gives

$$A_n = \int \nabla\phi(T(y)-y)^\top \Big(\mathcal{D}_0(\widehat{\mu}_0-\mu_0)(y)+\mathcal{D}_1(\widehat{\mu}_1-\mu_1)(y)\Big)\,\mu_0(\,\mathrm{d}y)+o_p(n^{-1/2}). \tag{2.34}$$

Step 4: Asymptotic linear representation of $\widehat{\Delta}_\phi$. Substituting (2.30) and (2.34) into the decomposition (2.28), we obtain

$$\widehat{\Delta}_\phi-\Delta_\phi=\frac{1}{n}\sum_{i=1}^{n}\psi_\phi(Y_i,D_i,X_i;x)+o_p(n^{-1/2}), \tag{2.35}$$

where the influence function ψ_ϕ is the sum of: (i) the influence component from estimating μ_0 in B_n and in $\mathcal{D}_0(\widehat{\mu}_0-\mu_0)$, and (ii) the influence component from estimating μ_1 in $\mathcal{D}_1(\widehat{\mu}_1-\mu_1)$. By Assumptions 5.2 and 5.6, ψ_ϕ has mean zero and finite variance

$$V_\phi(x)=\mathrm{Var}(\psi_\phi(Y,D,X;x))\in(0,\infty).$$

Multiplying (2.35) by $\sqrt{n}$ yields the representation stated in (5.18).

Step 5: Central limit theorem. Under Assumption 5.1, $\{\psi_\phi(Y_i,D_i,X_i;x)\}_{i=1}^n$ are i.i.d. with finite second moment, so the classical Lindeberg–Lévy CLT implies

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_\phi(Y_i,D_i,X_i;x)\ \Rightarrow\ \mathcal{N}\big(0,V_\phi(x)\big).$$

Combining this with (2.35) proves

$$\sqrt{n}\big(\widehat{\Delta}_\phi(x)-\Delta_\phi(x)\big)\ \Rightarrow\ \mathcal{N}\big(0,V_\phi(x)\big),$$

which completes the proof. □

C. ALGORITHMIC DESCRIPTION OF THE ESTIMATION PROCEDURE

This appendix presents the estimation and inference procedure in algorithmic form, highlighting the mapping between the population objects in the main text and their empirical counterparts. The description is schematic and abstracts from numerical and software-specific details. A complete and reproducible implementation is provided in the accompanying replication code.

C.1. Empirical distributions, regularized OT, and transport functionals

Algorithm 1 Estimation of conditional empirical distributions $\widehat{\mu}_d(\cdot \mid x)$

Require: Data $\{(Y_i,D_i,X_i)\}_{i=1}^n$, treatment state $d\in\{0,1\}$, stratum value x
Ensure: Empirical conditional measure $\widehat{\mu}_d(\cdot \mid x)$

1: $\mathcal{I}_{d,x}\leftarrow\{i: D_i=d,\ X_i=x\}$; $\quad n_{d,x}\leftarrow|\mathcal{I}_{d,x}|$
2: Define $\widehat{\mu}_d(\cdot \mid x)$ as the empirical measure placing mass $1/n_{d,x}$ on each $\{Y_i\}_{i\in\mathcal{I}_{d,x}}$
3: **return** $\widehat{\mu}_d(\cdot \mid x)$

Algorithm 2 Entropically regularized optimal transport plan $\widehat{\pi}_{x,\varepsilon}$

Require: Supports $\{Y_i^{(0)}\}_{i=1}^{n_{0,x}} \sim \widehat{\mu}_0(\cdot \mid x)$ and $\{Y_j^{(1)}\}_{j=1}^{n_{1,x}} \sim \widehat{\mu}_1(\cdot \mid x)$, regularization $\varepsilon > 0$

Ensure: Regularized coupling $\widehat{\pi}_{x,\varepsilon} \in \Pi(\widehat{\mu}_0, \widehat{\mu}_1)$

1: Form the cost matrix $C \in \mathbb{R}^{n_{0,x} \times n_{1,x}}$ with entries $C_{ij} = \|Y_i^{(0)} - Y_j^{(1)}\|^2$

2: Compute

$$\widehat{\pi}_{x,\varepsilon} \in \operatorname{argmin}_{\pi \in \Pi(\widehat{\mu}_0, \widehat{\mu}_1)} \sum_{i=1}^{n_{0,x}} \sum_{j=1}^{n_{1,x}} C_{ij}\pi_{ij} + \varepsilon \sum_{i=1}^{n_{0,x}} \sum_{j=1}^{n_{1,x}} \pi_{ij} \log \pi_{ij}.$$

3: **return** $\widehat{\pi}_{x,\varepsilon}$

Algorithm 3 Barycentric projection transport map $\widehat{T}_x$

Require: Regularized coupling $\widehat{\pi}_{x,\varepsilon}$, support points $\{Y_i^{(0)}\}_{i=1}^{n_{0,x}}$, $\{Y_j^{(1)}\}_{j=1}^{n_{1,x}}$

Ensure: Empirical transport map $\widehat{T}_x$ on $\operatorname{supp}(\widehat{\mu}_0(\cdot \mid x))$

1: **for** $i = 1, \dots, n_{0,x}$ **do**

2: $\quad w_i \leftarrow \sum_{j=1}^{n_{1,x}} \widehat{\pi}_{ij}$

3: $\quad \widehat{T}_x(Y_i^{(0)}) \leftarrow \frac{1}{w_i} \sum_{j=1}^{n_{1,x}} \widehat{\pi}_{ij}\, Y_j^{(1)}$ ▷ Eq. (4.13)

4: **end for**

5: **return** $\widehat{T}_x$

Algorithm 4 Plug-in estimator of transport functional $\widehat{\Delta}_\phi(x)$

Require: Baseline support $\{Y_i^{(0)}\}_{i=1}^{n_{0,x}}$, transport map $\widehat{T}_x$, functional $\phi : \mathbb{R}^d \to \mathbb{R}$

Ensure: Estimate $\widehat{\Delta}_\phi(x)$

1: Compute

$$\widehat{\Delta}_\phi(x) = \int \phi(\widehat{T}_x(y) - y)\, \widehat{\mu}_0(\mathrm{d}y \mid x) = \frac{1}{n_{0,x}} \sum_{i=1}^{n_{0,x}} \phi\big(\widehat{T}_x(Y_i^{(0)}) - Y_i^{(0)}\big).$$

2: **return** $\widehat{\Delta}_\phi(x)$

C.2. Bootstrap inference

Remark. Algorithms 1–5 describe the conceptual estimation and inference steps. The replication code implements numerically efficient versions of the regularized OT solver, includes tuning choices for ε, and reproduces all tables and figures reported in the paper.

Algorithm 5 Bootstrap confidence interval for $\Delta_\phi(x)$

Require: Data $\{(Y_i, D_i, X_i)\}_{i=1}^n$, stratum x, number of draws B, level α, functional ϕ, regularization ε

Ensure: Bootstrap confidence interval for $\Delta_\phi(x)$

Compute $\widehat{\Delta}_\phi(x)$ from Algorithms 1–4
for $b = 1, \dots, B$ **do**
    Draw a bootstrap resample of $\{(Y_i, D_i, X_i)\}_{i=1}^n$ (with replacement)
    Recompute $\widehat{\Delta}_\phi^{*(b)}(x)$ using Algorithms 1–4
end for
Construct a $(1-\alpha)$ interval using either:
  (i) normal approximation: $\widehat{\Delta}_\phi(x) \pm z_{1-\alpha/2}\, \widehat{\mathrm{se}}^*(x)$, or
  (ii) percentile interval: empirical $(\alpha/2, 1-\alpha/2)$ quantiles of $\{\widehat{\Delta}_\phi^{*(b)}(x)\}_{b=1}^B$
return confidence interval