Estimating Functionals of the Joint Distribution of Potential Outcomes with Optimal Transport

Daniel Ober-Reynolds*

This draft: 12 November, 2023

[Click here for the most recent version]

Abstract

Many causal parameters depend on a moment of the joint distribution of potential outcomes. Such parameters are especially relevant in policy evaluation settings, where noncompliance is common and accommodated through the model of Imbens & Angrist (1994). This paper shows that the sharp identified set for these parameters is an interval with endpoints characterized by the value of optimal transport problems. Sample analogue estimators are proposed based on the dual problem of optimal transport. These estimators are $\sqrt{n}$-consistent and converge in distribution under mild assumptions. Inference procedures based on the bootstrap are straightforward and computationally convenient. The ideas and estimators are demonstrated in an application revisiting the National Supported Work Demonstration job training program. I find suggestive evidence that workers who would see below average earnings without treatment tend to see above average benefits from treatment.

Keywords: potential outcomes, treatment effects, partial identification, bounds, optimal transport

JEL Codes: C21, C31

*Email: doberreynolds [at] gmail.com, website: https://danieloberreynolds.com. I want to thank my advisor Andres Santos for his continuous guidance and support. I also want to thank Denis Chetverikov and Jinyong Hahn for their helpful comments and suggestions.

1 Introduction

Researchers studying the causal effects of a binary treatment see an observation's treated or untreated outcome, but never both. As a result, the data identify the marginal distributions of each potential outcome, but not their joint distribution. This "fundamental problem of causal inference" (Holland, 1986) leaves parameters depending on the joint distribution partially identified.

In this paper I study a wide class of parameters that depend on a moment of the joint distribution of potential outcomes. My setting is the canonical potential outcomes framework with binary treatment, a binary instrument satisfying a monotonicity restriction, and finitely supported covariates (Imbens & Angrist, 1994; Abadie, 2003). In this setting, I show the sharp identified set for such parameters is an interval with endpoints characterized by the value of optimal transport problems. I propose sample analogue estimators based on the dual problem of optimal transport, which facilitates both computation and asymptotic analysis. Through the functional delta method, I show these estimators converge in distribution allowing for straightforward inference procedures based on the bootstrap.

The proposed estimators are especially attractive due to their wide applicability and computational simplicity. The class of parameters under study is broad, including the correlation between potential outcomes, the probability of benefitting from treatment, and many more examples discussed in section 2. As argued in Heckman et al. (1997), such parameters are of particular interest to policymakers and economists carrying out econometric policy evaluation. Noncompliance with the assigned treatment status is common in these settings. Most studies accomodate noncompliance with the same framework adopted in this paper, and could make use of these estimators with no additional identifying assumptions. Computing the estimator and constructing confidence sets entails nothing more challenging than solving linear programming problems, for which there are fast and efficient algorithms readily available.

This paper contributes to a large econometrics literature studying parameters of the joint distribution of potential outcomes. Many papers in this literature focus on a subset of the parameters considered here, especially the cumulative distribution function (cdf) or quantiles of treatment effects (Manski, 1997; Heckman et al., 1997; Firpo, 2007; Fan & Park, 2010, 2012; Firpo & Ridder, 2019; Callaway, 2021; Frandsen & Lefgren, 2021). This limited focus allows greater use of known analytical expressions when deriving sharp bounds, especially the famed Makarov bounds on the cdf and Fréchet-Hoeffding bounds on the joint distribution. Several recent works develop methods applicable to broad parameters classes by employing procedures that do not require analytical expressions for the identified set. Russell (2021) studies continuous functionals of the joint distribution of discrete potential outcomes, through a computationally intensive (sometimes infeasible) search over all permissible distributions of model primitives. Fan et al. (2023) study parameters identified through moment conditions in several incomplete data settings – including potential outcomes – by searching over an infinite dimensional space of smooth copulas. This paper occupies a middle ground: by focusing on parameters that depend on a scalar moment of the joint distribution and working with optimal transport, I obtain expressions for the bounds with tractable sample analogues. This approach allows consideration of a wide variety of parameters while maintaining computational tractability.

This paper also contributes to a growing literature on applications of optimal transport to econometrics; see Galichon (2017) for a recent survey. Several recent working papers utilize optimal transport for issues related to casual inference, including inverse propensity weighting (Dunipace, 2021), matching on covariates (Gunsilius & Xu, 2021), and obtaining counterfactual distributions (Torous et al., 2021). In concurrent and highly complementary work, Ji et al. (2023) consider a very similar class of parameters to the present paper and also propose inference based on the dual problem of optimal transport. Their focus, accomodating non-discrete covariates without resorting to parametric models, leads to theory based on cross fitting and high-level assumptions on first stage estimators. The goal of the present paper is to provide simple, low-level conditions and computationally convenient estimators in the common case where covariates are discrete. This leads to theory based on Hadamard directional differentiability and the functional delta method quite distinct from that of Ji et al. (2023).

The remainder of this paper is organized as follows. Section 2 formalizes the setting and introduces the class of parameters under study. Optimal transport is introduced in section 3, and used in identification in section 4. Section 5 proposes the estimators and contains the asymptotic results. Section 6 contains the application, showing suggestive evidence that the the National Supported Work Demonstration job training program was especially beneficial for workers who would otherwise see below average incomes. Section 7 discusses straightforward extensions, and section 8 concludes.

2 Setting and parameter class

2.1 Setting

Consider a potential outcomes framework with binary treatment, a binary instrument, and finitely supported covariates (Imbens & Angrist (1994), Abadie (2003)). Let Y denote the scalar, real-valued outcome of interest and $D \in \{0,1\}$ indicate treatment status. Further let Y_1 denote the potential outcome when treated and Y_0 the potential outcome when untreated. The observed outcome Y is given by

$$Y = DY_1 + (1-D)Y_0. \tag{1}$$

The difference in potential outcomes, $Y_1 - Y_0$, is called the treatment effect.

The binary instrument is denoted $Z \in \{0,1\}$. Let D_1 denote the treatment status when $Z = 1$, and D_0 the treatment status when $Z = 0$. The observed treatment status D is given by

$$D = ZD_1 + (1-Z)D_0. \tag{2}$$

It is assumed that the instrument itself does not affect the outcome.[1] Units with $1 = D_1 > D_0 = 0$ are known as *compliers*.

Assumption 1 formalizes the setting.

Assumption 1 (Setting). *$\{Y_i, D_i, Z_i, X_i\}_{i=1}^n$ is an i.i.d. sample with $(Y, D, Z, X) \sim P$,*

$$Y \in \mathcal{Y} \subseteq \mathbb{R}, \qquad D \in \{0,1\}, \qquad Z \in \{0,1\}, \qquad X \in \mathcal{X} = \{x_1, \ldots, x_M\} \subseteq \mathbb{R}^{d_x} \tag{3}$$

where Y, D, and Z are related to (Y_1, Y_0, D_1, D_0) through equations (1) and (2), and the random vector $(Y_1, Y_0, D_1, D_0, Z, X)$ satisfies

(i) Instrument independence: $(Y_1, Y_0, D_1, D_0) \perp Z \mid X$,
(ii) Monotonicity: $P(D_1 \geq D_0) = 1$,
(iii) Existence of compliers: $P(D_1 > D_0, X = x) > 0$ for each x, and
(iv) $P(X = x, Z = z) > 0$ for each (x, z).

Assumption 1 is essentially equivalent to assumption 2.1 in Abadie (2003), with the addition that covariates are finitely supported. Instrument independence is sometimes referred to as *ignorability*, and satisfied in most randomized controlled trials, where Z indicates being assigned to treatment. Monotonicity is typically a weak assumption in such settings.

[1] One could hypothesize potential outcomes varying with the value of the instrument, i.e. Y_{dz} for each (d, z). The exposition here implicitly assumes *instrument exclusion*, also known as the *Stable Unit Treatment Value Assumption*: that $P(Y_{d1} = Y_{d0}) = 1$ for each d.

It is worth emphasizing that this setting nests the case where treatment is exogenous. Specifically, when $D_1 = 1$ and $D_0 = 0$ (degenerately), every unit is a complier. In this case equation (2) shows treatment status equals the instrument: $D = Z$. Instrument independence simplifies to $(Y_1, Y_0) \perp D \mid X$, and monotonicity is trivially satisfied.

2.1.1 Distributions of compliers

Interest focuses on the distribution of compliers. Such focus is especially policy relevant when "the policy is the instrument" i.e., the proposed change in policy is to assign $Z = 1$ to all units. Abadie (2003) shows that assumption 1 suffices to identify the marginal distributions of Y_1 and Y_0 for the subpopulation of compliers.

Lemma 2.1 (Abadie (2003))**.** *Suppose assumption 1 holds. Then the marginal distributions of Y_d conditional on $D_1 > D_0$ and $X = x$, denoted $P_{d|x}$, are identified by*

$$\begin{aligned} E_{P_{d|x}}[f(Y_d)] &\equiv E[f(Y_d) \mid D_1 > D_0, X = x] \\ &= \frac{E[f(Y)\mathbb{1}\{D = d\} \mid Z = d, X = x] - E[f(Y)\mathbb{1}\{D = d\} \mid Z = 1 - d, X = x]}{P(D = d \mid Z = d, X = x) - P(D = d \mid Z = 1 - d, X = x)} \end{aligned} \tag{4}$$

for any integrable function f. Furthermore, the distribution of X conditional on $D_1 > D_0$ is identified by

$$\begin{aligned} s_x &\equiv P(X = x \mid D_1 > D_0) \\ &= \frac{[P(D = 1 \mid Z = 1, X = x) - P(D = 1 \mid Z = 0, X = x)]\, P(X = x)}{\sum_{x'} [P(D = 1 \mid Z = 1, X = x') - P(D = 1 \mid Z = 0, X = x')]\, P(X = x')} \end{aligned} \tag{5}$$

The joint distribution of potential outcomes is not identified. This is a result of the fundamental problem of causal inference: there is no unit where both Y_1 and Y_0 are observed, and as a result the joint distribution of (Y_1, Y_0) is not identified for any subpopulation. Let $P_{1,0}$ denote the joint distribution of (Y_1, Y_0) conditional on compliance, and $P_{1,0|x}$ denote the joint distribution conditional on compliance and $X = x$. These are related through the law of iterated expectations; for any function $c(y_1, y_0)$ with values in $\mathbb{R}$,

$$E_{P_{1,0}}[c(Y_1, Y_0)] = E[E[c(Y_1, Y_0) \mid D_1 > D_0, X] \mid D_1 > D_0] = \sum_x s_x E_{P_{1,0|x}}[c(Y_1, Y_0)].$$

This relation can also be expressed as $P_{1,0} = \sum_x s_x P_{1,0|x}$.

A joint distribution with marginals $P_{1|x}$ and $P_{0|x}$ is called a *coupling* of $P_{1|x}$ and $P_{0|x}$. $P_{1,0|x}$ is such a coupling, and is otherwise unrestricted by assumption 1. Thus the identified set for $P_{1,0|x}$ is the set of distributions $\pi_{1,0|x}$ for (Y_1, Y_0) with marginals $\pi_{1|x} = P_{1|x}$ and $\pi_{0|x} = P_{0|x}$, denoted

$$\Pi(P_{1|x}, P_{0|x}) = \left\{\pi_{1,0|x} \ : \ \pi_{1|x} = P_{1|x}, \ \pi_{0|x} = P_{0|x}\right\}. \tag{6}$$

Moreover, the identified set for $P_{1,0}$ is $\left\{\pi_{1,0} = \sum_x s_x \pi_{1,0|x} \ : \ \pi_{1,0|x} \in \Pi(P_{1|x}, P_{0|x})\right\}$.

2.2 Parameter class

The idea at the core of this paper is to bound a moment of the joint distribution of potential outcomes by optimization. Accordingly, the focus is on scalar parameters of the form

$$\gamma = g(\theta, \eta) \tag{7}$$

where g is a known function and $\theta = E_{P_{1,0}}[c(Y_1, Y_0)] \in \mathbb{R}$ is a scalar moment of the joint distribution of (Y_1, Y_0) conditional on compliance. The function c is known, and referred to as a "cost function" in connection with the optimal transport literature. This class of parameters is broad, as illustrated by the examples given below. In each of these examples η is a finite collection of moments of the marginal distributions conditional on compliers: $\eta = (E_{P_1}[\eta_1(Y_1)], E_{P_0}[\eta_0(Y_0)]) \in \mathbb{R}^{K_1+K_0}$. The formal results focus on this case, but could be generalized to allow η to be other point identified nuisance parameters.

The following conditions are stronger than necessary for identification of the sharp identified set of γ, but will be used when constructing and studying estimators. Assumption 2 places restrictions on the cost function to ensure optimal transport can be used characterize and estimate the sharp identified set for θ.

Assumption 2 (Cost function)**.** *Either*

(i) $c(y_1, y_0)$ *is Lipschitz continuous and* $\mathcal{Y}$ *is compact, or*

(ii) $c(y_1, y_0) = \mathbb{1}\{y_1 - y_0 \leq \delta\}$ *for a known* $\delta \in \mathbb{R}$ *and the cumulative distribution functions* $F_{d|x}(y) = P(Y_d \leq y \mid D_1 > D_0, X = x)$ *are continuous.*

Assumption 2 covers every example listed below. Continuous cost functions c are given a unified analysis, but for reasons discussed in section 3 discontinuous cost functions must be handled on a case-by-case basis. I focus on the leading case of interest in applications, $c(y_1, y_0) = \mathbb{1}\{y_1 - y_0 \leq \delta\}$, corresponding to the cumulative distribution of treatment effects. The approach developed in this paper could likely be generalized to cover other discontinuous cost functions; for example, results in the appendix allow estimation of the sharp lower bound of $P((Y_1, Y_0) \in C)$ for any open, convex set $C \subseteq \mathbb{R}^2$.

Assumption 2 (ii) requires the cdfs $F_{d|x}$ be continuous. As discussed in section 4, this ensures the set being estimated is the sharp identified set for the parameter of interest. However, the estimation and inference results of section 5 hold *regardless* of whether the cdfs are continuous or not; when the cdfs are not continuous, the estimand is a valid outer identified set.

Under assumptions 1 and 2, the sharp identified set for θ is an interval $[\theta^L, \theta^H]$. Assumption 3 contains conditions on g and η.

Assumption 3 (Function of moments). *The parameter is $\gamma = g(\theta, \eta) \in \mathbb{R}$, where*

$$\theta = E[c(Y_1, Y_0) \mid D_1 > D_0] \in \mathbb{R}, \qquad \eta = E\left[\eta_1(Y_1), \eta_0(Y_0) \mid D_1 > D_0\right] \in \mathbb{R}^{K_1 + K_0}$$

for known functions g, c, η_1 and η_0 such that

(i) $E[\|\eta_d(Y)\|^2] < \infty$ *for* $d = 1, 0$,

(ii) $g(\cdot, \eta)$ *is continuous, and*

(iii) the functions

$$g^L(t^L, t^H, e) = \min_{t \in [t^L, t^H]} g(t, e), \qquad g^H(t^L, t^H, e) = \max_{t \in [t^L, t^H]} g(t, e)$$

are continuously differentiable at $(t^L, t^H, e) = (\theta^L, \theta^H, \eta)$.

Note that when θ itself is of interest, assumption 3 is satisfied with $g(\theta, \eta) = \theta$. Assumption 3 (ii) ensures the identified set for γ is the interval $[\gamma^L, \gamma^H]$, and assumption 3 (iii) is used to apply the delta method. It is straightforward to show assumption 3 (iii) holds when g is continuously differentiable in both arguments and $g(\cdot, \eta)$ is strictly increasing, as the latter condition implies $g^L(\theta^L, \theta^H, \eta) = g(\theta^L, \eta)$ and $g^H(\theta^L, \theta^H, \eta) = g(\theta^H, \eta)$ and the former condition implies they are continuously differentiable. This argument applies to every parameter listed below. When g is differentiable but $g(\cdot, \eta)$ is not monotonic, it is often possible to use the implicit function theorem applied to first order conditions to derive sufficient conditions for the corresponding $\arg\min$ and $\arg\max$ to be differentiable, and thus for assumption 3 (iii) to hold.

2.2.1 Examples

The following examples are intended both to fix ideas and illustrate the broad scope of the parameter class described above.

Example 2.1 (Summary statistics). *Many summary statistics can be rewritten in the form $\gamma = g(\theta, \eta)$. For example, suppose interest is in the variance of treatment effects for compliers: $\gamma = Var(Y_1 - Y_0 \mid D_1 > D_0)$. This parameter can be rewritten as*

$$\gamma = Var(Y_1 - Y_0 \mid D_1 > D_0) = E_{P_{1,0}}[(Y_1 - Y_0)^2] - (E_{P_1}[Y_1] - E_{P_0}[Y_0])^2,$$

This parameter fits the form $\gamma = g(\theta, \eta)$ required of display (7), with $\theta = E_{P_{1,0}}[(Y_1 - Y_0)^2]$, $\eta = (\eta^{(1)}, \eta^{(2)}) = (E_{P_1}[Y_1], E_{P_0}[Y_0])$, and $g(\theta, \eta) = \theta - (\eta^{(1)} - \eta^{(2)})^2$. The cost function $c(y_1, y_0) = (y_1 - y_0)^2$ satisfies assumption 2 (i) when $\mathcal{Y}$, the support of the outcome Y, is bounded.

Similarly, suppose the researcher is interested in the correlation between Y_1 and Y_0 for compliers. Set $\gamma = Corr(Y_1, Y_0 \mid D_1 > D_0)$, which can be rewritten as

$$\gamma = Corr(Y_1, Y_0 \mid D_1 > D_0) = \frac{E_{P_{1,0}}[Y_1 Y_0] - E_{P_1}[Y_1] E_{P_0}[Y_0]}{\sqrt{E_{P_1}[Y_1^2] - (E_{P_1}[Y_1])^2}\sqrt{E_{P_0}[Y_0^2] - (E_{P_0}[Y_0])^2}}$$

This parameter also fits the form $\gamma = g(\theta, \eta)$ in display (7), with $\theta = E_{P_{1,0}}[Y_1 Y_0]$, $\eta = (\eta^{(1)}, \eta^{(2)}, \eta^{(3)}, \eta^{(4)}) = (E_{P_1}[Y_1], E_{P_1}[Y_1^2], E_{P_0}[Y_0], E_{P_0}[Y_0^2])$, and $g(\theta, \eta) = \frac{\theta - \eta^{(1)} \times \eta^{(3)}}{\sqrt{\eta^{(2)} - (\eta^{(1)})^2}\sqrt{\eta^{(4)} - (\eta^{(3)})^2}}$. The cost function $c(y_1, y_0) = y_1 y_0$ satisfies assumption 2 (i) when $\mathcal{Y}$ is bounded.

Example 2.2 (Expected percent change). *The expected percent change in the outcome can be written as $100 \times E\left[\frac{Y_1 - Y_0}{Y_0} \mid D_1 > D_0\right]\%$. This is a unit-invariant causal parameter that is a natural summary measure when Y_0 exhibits considerably variation. For example, a treatment effect of $Y_1 - Y_0 = 5$ is typically of greater economic significance when the untreated outcome is small, say $Y_0 = 10$, than when $Y_0 = 100$.*

The expected percent change is proportional to

$$\gamma = E\left[\frac{Y_1 - Y_0}{Y_0} \mid D_1 > D_0\right] = E_{P_{1,0}}\left[\frac{Y_1 - Y_0}{Y_0}\right],$$

which fits the form of display (7), with $\gamma = \theta = E_{P_{1,0}}\left[\frac{Y_1 - Y_0}{Y_0}\right]$. The cost function $c(y_1, y_0) = \frac{y_1 - y_0}{y_0}$ satisfies assumption 2 (i) when $\mathcal{Y}$ is bounded and bounded away from zero.

Example 2.3 (Equitable policies). *Policy makers are often interested in whether a policy is equitable – that is, whether the benefits are concentrated among those who would have undesirable outcomes without treatment.*

One parameter that speaks to these concerns is the covariance between treatment effects and untreated outcomes among compliers: $\gamma = Cov(Y_1 - Y_0, Y_0 \mid D_1 > D_0)$. Notice that $\gamma < 0$ implies those with below average Y_0 tend to see above average treatment effects. This parameter can be rewritten as

$$\gamma = Cov(Y_1 - Y_0, Y_0 \mid D_1 > D_0) = E_{P_{1,0}}[(Y_1 - Y_0)Y_0] - (E_{P_1}[Y_1] - E_{P_0}[Y_0])E_{P_0}[Y_0]$$

and fits the form $g(\theta, \eta)$ with $\theta = E_{P_{1,0}}[(Y_1 - Y_0)Y_0]$, $\eta = (E_{P_1}[Y_1], E_{P_0}[Y_0])$, and $g(\theta, \eta) = \theta - (\eta^{(1)} - \eta^{(2)})\eta^{(2)}$. The cost function $c(y_1, y_0) = (y_1 - y_0)y_0$ satisfies assumpion 2 (i) when $\mathcal{Y}$ is bounded.

Many related parameters share a sign with $Cov(Y_1 - Y_0, Y_0 \mid D_1 > D_0)$ and are also suitable for such an analysis. One such example is the OLS slope when regressing $Y_1 - Y_0$ on Y_0 and a constant: $\gamma = \frac{Cov(Y_1-Y_0,Y_0|D_1>D_0)}{Var(Y_0|D_1>D_0)}$. This parameter can be rewritten as

$$\gamma = \frac{Cov(Y_1 - Y_0, Y_0 \mid D_1 > D_0)}{Var(Y_0 \mid D_1 > D_0)} = \frac{E_{P_{1,0}}[(Y_1 - Y_0)Y_0] - (E_{P_1}[Y_1] - E_{P_0}[Y_0])E_{P_0}[Y_0]}{E_{P_0}[Y_0^2] - (E_{P_0}[Y_0])^2}$$

where $\theta = E_{P_{1,0}}[(Y_1 - Y_0)Y_0]$, $\eta = (E_{P_1}[Y_1], E_{P_0}[Y_0], E_{P_0}[Y_0^2])$, and $g(\theta, \eta) = \frac{\theta - (\eta^{(1)} - \eta^{(2)})\eta^{(2)}}{\eta^{(3)} - (\eta^{(2)})^2}$.

Example 2.4 (Proportion that benefit)**.** *The share of compliers benefiting from treatment, written*

$$\gamma = P(Y_1 > Y_0 \mid D_1 > D_0),$$

is naturally of interest in applications where theory gives little indication whether the treatment will have a positive or negative effect. For example, Allcott et al. (2020) study the effect of deactivating facebook on subjective well-being. The authors find significant positive average effects of deactivation, but find substantial heterogeneity in follow-up interviews.

This parameter fits the form of display (7), with $\gamma = \theta = E_{P_{1,0}}[\mathbb{1}\{Y_1 - Y_0 \leq 0\}]$. The cost function $c(y_1, y_0) = \mathbb{1}\{y_1 - y_0 \leq 0\}$ satisfies assumption 2 (ii) if the cdfs $F_{d|x}(y)$ are continuous.

The share benefiting from treatment is also of particular interest when the intervention comes at a financial cost and the outcome of interest is a pecuniary return. Common examples include job training programs intended to increase a worker's income (e.g. the National Supported Work Demonstration studied in Couch (1992)) or management practices intended to raise a firm's accounting profit (e.g. the employee referral program studied in Friebel et al. (2023)). To illustrate, suppose the researcher observes $\{R_i, C_i, D_i, Z_i\}_{i=1}^n$, where R is observed revenue and C is the observed cost. These are related to treatment status $D \in \{0, 1\}$, potential revenues (R_1, R_0), and potential costs (C_1, C_0) by

$$R = DR_1 + (1 - D)R_0, \qquad\qquad C = DC_1 + (1 - D)C_0$$

The observed profit, $Y = R - C$, is related to treatment status by

$$Y = D \underbrace{(R_1 - C_1)}_{:=Y_1} + (1 - D) \underbrace{(R_0 - C_0)}_{:=Y_0}$$

The probability the change in revenue exceeds the change in cost is

$$P(R_1 - R_0 > C_1 - C_0 \mid D_1 > D_0) = P(Y_1 > Y_0 \mid D_1 > D_0)$$

Example 2.5 (Quantiles)**.** *Suppose the parameter of interest is any q_τ solving*

$$P(Y_1 - Y_0 \leq q_\tau) = \tau \tag{8}$$

This parameter has a similar interpretation to the τ-th quantile.[2] q_τ cannot be viewed as $\gamma = g(\theta, \eta)$. However, by viewing $\theta(\delta) = P(Y_1 - Y_0 \leq \delta \mid D_1 > D_0) = E_{P_{1,0}}[\mathbb{1}\{Y_1 - Y_0 \leq \delta\}]$ as a function of δ,

[2]The τ-th quantile is usually defined as the unique value $\tilde{q}_\tau = \inf\{y \; ; \; P(Y_1 - Y_0 \leq y) \geq \tau\}$. When the τ level set of the cumulative distribution function $P(Y_1 - Y_0 \leq \cdot)$ is nonempty, the τ-th quantile has the interpretation

the results below can be adapted to construct a confidence set for the identified set of this parameter as described in section 7.2.

3 Optimal Transport

This section defines and discusses optimal transport, which is used to characterize the identified set and construct estimators.

Given any marginal distributions P_1 and P_0 and a "cost function" $c(y_1, y_0)$, the Monge-Kantorovich formulation of **optimal transport** is the problem of choosing a coupling $\pi \in \Pi(P_1, P_0)$ to minimize $E_\pi[c(Y_1, Y_0)]$:

$$OT_c(P_1, P_0) = \inf_{\pi \in \Pi(P_1, P_0)} E_\pi[c(Y_1, Y_0)]. \tag{9}$$

This minimization problem in (9) is referred to as the **primal problem**, and will be used to characterize the identified set of θ.

The dual problem of optimal transport will be used to construct and analyze estimators. Let Φ_c denote the set of functions $\varphi(y_1)$ and $\psi(y_0)$ whose pointwise sum is less than $c(y_1, y_0)$:

$$\Phi_c = \{(\varphi, \psi)\ ;\ \varphi(y_1) + \psi(y_0) \leq c(y_1, y_0)\}. \tag{10}$$

The **dual problem** chooses a pair of functions in Φ_c to maximize the sum of the corresponding expectations:

$$\sup_{(\varphi,\psi) \in \Phi_c} E_{P_1}[\varphi(Y_1)] + E_{P_0}[\psi(Y_0)]. \tag{11}$$

When the cost function is lower semicontinuous and bounded from below, the primal problem is attained and **strong duality** holds:

$$OT_c(P_1, P_0) = \min_{\pi \in \Pi(P_1, P_0)} E_\pi[c(Y_1, Y_0)] = \sup_{(\varphi,\psi) \in \Phi_c} E_{P_1}[\varphi(Y_1)] + E_{P_0}[\psi(Y_0)]. \tag{12}$$

The dual problem will be used to construct and analyze estimators. Indeed, the identification of $P_{d|x}$ in lemma 2.1 suggests straightforward sample analogues estimating $E_{P_{d|x}}[f(Y_d)]$ for a given f, which makes it possible to form a sample analogue of the dual problem.

Although it is clear how to form a sample analogue of the dual problem, it is not immediately clear how to analyze the resulting estimator. Fortunately, the dual problem can often be simplified

that $100 \times \tau\%$ of the population has treatment effect less than or equal to $\tilde{q}_\tau$. Every q_τ solving (8) has the same interpretation.

by restricting the maximization problem to a smaller set of functions. Estimators based on this restricted dual problem can then be studied with empirical process techniques.

The dual feasible set is restricted with the concept of c-concavity. Notice the dual problem's objective is monotonic, in the sense that $\varphi(y_1) \leq \tilde{\varphi}(y_1)$ for all y_1 implies

$$E_{P_1}[\varphi(Y_1)] + E_{P_0}[\psi(Y_0)] \leq E_{P_1}[\tilde{\varphi}(Y_1)] + E_{P_0}[\psi(Y_0)].$$

Increasing ψ pointwise will also increase the dual objective. Speaking loosely, any function pair $(\varphi, \psi) \in \Phi_c$ for which the constraint $\varphi(y_1) + \psi(y_0) \leq c(y_1, y_0)$ is "slack" cannot be a solution to the dual problem and can therefore be ignored. This motivates the definition of the **c-transforms** of a function φ:

$$\varphi^c(y_0) = \inf_{y_1}\{c(y_1, y_0) - \varphi(y_1)\}, \qquad \varphi^{cc}(y_1) = \inf_{y_0}\{c(y_1, y_0) - \varphi^c(y_0)\}.$$

For any pair of functions $(\varphi, \psi) \in \Phi_c$, these definitions imply $\psi(y_0) \leq \varphi^c(y_0)$, $\varphi(y_1) \leq \varphi^{cc}(y_1)$, and $\varphi^{cc}(y_1) + \varphi^c(y_0) \leq c(y_1, y_0)$. Further c-transformations are irrelevant because $(\varphi^{cc})^c = \varphi^c$, so a function φ is called **c-concave** if $\varphi^{cc} = \varphi$. If the c-transforms are integrable, the dual problem can be restricted to c-concave conjugate pairs, $(\varphi^{cc}, \varphi^c)$. Furthermore, c-concave functions often "inherit" properties of the cost function c; for example, if c is Lipschitz continuous then φ^c and φ^{cc} are Lipschitz continuous as well. These properties can be used to define sets of functions $\mathcal{F}_c$ and $\mathcal{F}_c^c$ (depending on the cost function c but not on the distributions P_1, P_0) such that

$$\sup_{(\varphi,\psi)\in\Phi_c} E_{P_1}[\varphi(Y_1)] + E_{P_0}[\psi(Y_0)] = \sup_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} E_{P_1}[\varphi(Y_1)] + E_{P_0}[\psi(Y_0)]. \tag{13}$$

Two cases suffice for the parameters considered in this paper. When the cost function $c(y_1, y_0)$ is Lipschitz continuous and $\mathcal{Y}$ is compact, define

$$\mathcal{F}_c = \left\{\varphi : \mathcal{Y} \to \mathbb{R}\ ;\ -\|c\|_\infty \leq \varphi(y_1) \leq \|c\|_\infty,\ |\varphi(y_1) - \varphi(y_1')| \leq L|y_1 - y_1'|\right\} \tag{14}$$

$$\mathcal{F}_c^c = \left\{\psi : \mathcal{Y} \to \mathbb{R}\ ;\ -2\|c\|_\infty \leq \psi(y_0) \leq 0,\ |\psi(y_0) - \psi(y_0')| \leq L|y_0 - y_0'|\right\} \tag{15}$$

where $\|c\|_\infty = \sup_{(y_1,y_0)}|c(y_1, y_0)|$ and L is the Lipschitz constant of c. When $c(y_1, y_0) = \mathbb{1}\{(y_1, y_0) \in$

$C\}$ for an open, convex set C, let

$$\mathcal{F}_c = \{\varphi : \mathcal{Y} \to \mathbb{R}\ ;\ \varphi(y_1) = \mathbb{1}\{y_1 \in I\} \text{ for some interval } I\} \tag{16}$$

$$\mathcal{F}_c^c = \{\psi : \mathcal{Y} \to \mathbb{R}\ ;\ \psi(y_0) = -\mathbb{1}\{y_0 \in I^c\} \text{ for some interval } I\} \tag{17}$$

Equation (13) shows the optimal transport functional $OT_c(P_1, P_0)$ depends only on the values of $E_{P_1}[\varphi(Y_1)]$ and $E_{P_0}[\psi(Y_0)]$ for $(\varphi, \psi) \in \mathcal{F}_c \times \mathcal{F}_c^c$. For any set A, let $\ell^\infty(A)$ denote the space of real-valued bounded functions defined on A, equipped with the supremum norm: $\ell^\infty(A) = \{f : A \to \mathbb{R}\ ;\ \|f\|_\infty = \sup_{a \in A}|f(a)| < \infty\}$. Optimal transport can be viewed as the map OT_c : $\ell^\infty(\mathcal{F}_c) \times \ell^\infty(\mathcal{F}_c^c) \to \mathbb{R}$ given by

$$OT_c(P_1, P_0) = \sup_{(\varphi,\psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)} E_{P_1}[\varphi(Y_1)] + E_{P_0}[\psi(Y_0)]. \tag{18}$$

This problem will be referred to as the **restricted dual problem**. Estimators formed with this map can be studied with empirical process techniques.

In summary, $OT_c(P_1, P_0)$ will be viewed as the functional in (9) when considering identification, and as the functional given in (18) when considering estimation. By ensuring c is either Lipschitz continuous or the indicator of an open convex set, strong duality and c-concavity ensures these functionals agree on the space of probability distributions.

4 Identification

Recall the parameter of interest is $\gamma = g(\theta, \eta)$, where η is a point identified parameter, $\theta = E_{P_{1,0}}[c(Y_1, Y_0)] \in \mathbb{R}$, and g and c are known functions.

Begin by rewriting $\theta = E_{P_{1,0}}[c(Y_1, Y_0)] = E[c(Y_1, Y_0) \mid D_1 > D_0]$ with the law of iterated expectations:

$$\theta = E[E[c(Y_1, Y_0) \mid D_1 > D_0, X] \mid D_1 > D_0] = E[\theta_X \mid D_1 > D_0] = \sum_x s_x \theta_x$$

where $s_x = P(X = x \mid D_1 > D_0)$ and $\theta_x = E[c(Y_1, Y_0) \mid D_1 > D_0, X = x] = E_{P_{1,0|x}}[c(Y_1, Y_0)]$. As noted in section 2.1.1, the identified set for $P_{1,0|x}$ is the set of couplings of $P_{1|x}$ and $P_{0|x}$, denoted $\Pi(P_{1|x}, P_{0|x})$. Thus the identified set for θ_x is $\Theta_{I,x} = \big\{t \in \mathbb{R}\ :\ t = E_\pi[c(Y_1, Y_0)] \text{ for some } \pi \in \Pi(P_{1|x}, P_{0|x})\big\}$. $\Pi(P_{1|x}, P_{0|x})$ is convex, implying that $\Theta_{I,x}$ is an interval. Let θ_x^L and θ_x^H denote its lower and upper endpoint respectively.

To ensure the restricted dual problem can be used for estimation, θ_x^L and θ_x^H are characterized through an optimal transport problem with a suitable cost function c. When assumption 2 (i) holds ($c(y_1, y_0)$ is Lipschitz continuous and $\mathcal{Y}$ is compact), define

$$\begin{aligned} &c_L(y_1, y_0) = c(y_1, y_0), && c_H(y_1, y_0) = -c(y_1, y_0) \\ &\theta^L(P_{1|x}, P_{0|x}) = OT_{c_L}(P_{1|x}, P_{0|x}), && \theta^H(P_{1|x}, P_{0|x}) = -OT_{c_H}(P_{1|x}, P_{0|x}). \end{aligned} \tag{19}$$

Note that $\theta_x^L = \theta^L(P_{1|x}, P_{0|x})$ and $\theta_x^H = \theta^H(P_{1|x}, P_{0|x})$.

The cumulative distribution function of $Y_1 - Y_0$ corresponds to the cost function $c(y_1, y_0) = \mathbb{1}\{y_1 - y_0 \le \delta\}$, which is not lower semicontinuous. This challenge is circumvented by a small change in the cost function. When assumption 2 (ii) holds (the cost function is $c(y_1, y_0) = \mathbb{1}\{y_1 - y_0 \le \delta\}$) define

$$\begin{aligned} &c_L(y_1, y_0) = \mathbb{1}\{y_1 - y_0 < \delta\}, && c_H = \mathbb{1}\{y_1 - y_0 > \delta\} \\ &\theta^L(P_{1|x}, P_{0|x}) = OT_{c_L}(P_{1|x}, P_{0|x}), && \theta^H(P_{1|x}, P_{0|x}) = 1 - OT_{c_H}(P_{1|x}, P_{0|x}) \end{aligned} \tag{20}$$

It follows from definitions that $\theta_x^H = \theta^H(P_{1|x}, P_{0|x})$. Moreover, $c_L(y_1, y_0) \le c(y_1, y_0)$ implies $\theta^L(P_{1|x}, P_{0|x})$ is a valid lower bound for θ_x. It is sharp if $P_{1|x}$, $P_{0|x}$ have continuous cumulative distribution functions, in which case $\theta_x^L = \theta^L(P_{1|x}, P_{0|x})$. It is worth emphasizing again that the estimation and inference results of section 5 hold *regardless* of whether the cdfs are continuous or not; when the cdfs are not continuous, the estimand is a valid outer identified set.

Under assumptions 1 and 2, the identified set for $\theta = E_{P_{1,0}}[c(Y_1, Y_0)] = E[c(Y_1, Y_0) \mid D_1 > D_0]$ is the compact interval $[\theta^L, \theta^H]$ with endpoints

$$\theta^L = E[\theta_X^L \mid D_1 > D_0] = \sum_x s_x \theta_x^L, \qquad \theta^H = E[\theta_X^H \mid D_1 > D_0] = \sum_x s_x \theta_x^H$$

Under assumptions 1, 2, and 3, the identified set for γ is $[\gamma^L, \gamma^H]$, with endpoints

$$\gamma^L = g^L(\theta^L, \theta^H, \eta) = \inf_{t \in [\theta^L, \theta^H]} g(t, \eta), \qquad \gamma^H = g^H(\theta^L, \theta^H, \eta) = \sup_{t \in [\theta^L, \theta^H]} g(t, \eta) \tag{21}$$

The following theorem summarizes the discussion above. Let $\theta^L(\cdot, \cdot)$ and $\theta^H(\cdot, \cdot)$ be given by (19) or (20) depending on the cost function, and set

$$\theta_x^L = \theta^L(P_{1|x}, P_{0|x}), \qquad \theta_x^H = \theta^H(P_{1|x}, P_{0|x}), \tag{22}$$

$$\theta^L = \sum_x s_x \theta_x^L, \qquad \theta^H = \sum_x s_x \theta_x^H, \tag{23}$$

$$\gamma^L = g^L(\theta^L, \theta^H, \eta), \qquad \gamma^H = g^H(\theta^L, \theta^H, \eta) \tag{24}$$

Theorem 4.1 (Identification of functions of moments)**.** *Suppose assumptions 1, 2, and 3 are satisfied. Then the sharp identified set for* γ *is* $[\gamma^L, \gamma^H]$.

All results are proven in the appendix.

It is worth pausing to consider the role of covariates. When covariates are available, ignoring them leads to wider bounds that are not sharp. Specifically, the marginal distributions P_1 and P_0 could be used to form a lower bound on θ with $\theta^L(P_1, P_0) = \inf_{\pi \in \Pi(P_1, P_0)} E_\pi[c_L(Y_1, Y_0)]$. This bound minimizes over the whole set $\Pi(P_1, P_0) = \{\pi_{1,0} \,;\, \pi_1 = P_1, \pi_0 = P_0\}$, but the identified set for $P_{1,0}$ is the subset of $\Pi(P_1, P_0)$ given by $\left\{\pi_{1,0} = \sum_x s_x \pi_{1,0|x} \,;\, \pi_{1,0|x} \in \Pi(P_{1|x}, P_{0|x})\right\}$. The bound defined through equations (22) and (23) is found while enforcing the additional constraints that $\pi_{1,0|x} \in \Pi(P_{1|x}, P_{0|x})$ for each x. These additional constraints imply $\theta^L(P_1, P_0) \leq \theta^L$, and similarly $\theta^H \leq \theta^H(P_1, P_0)$.

Extreme cases illustrate when covariates are informative. If X is independent of (Y_1, Y_0) conditional on $D_1 > D_0$, then $P_{d|x} = P_d$ for each x, $\Pi(P_{1|x}, P_{0|x}) = \Pi(P_1, P_0)$, and the inequalities above hold as equalities. On the other hand, if $P_{d|x}$ is degenerate for either $d = 1$ or $d = 0$, then there is only one possible coupling of $P_{1|x}$ and $P_{0|x}$. Since $\Pi(P_{1|x}, P_{0|x})$ is a singleton, $\theta_x^L = \theta_x^H$ and $\theta_x = E[c(Y_1, Y_0) \mid D_1 > D_0, X = x]$ is point identified. If this occurs for all $x \in \mathcal{X}$, θ and γ are point identified.

Remark 4.1 (Makarov bounds)*.* The proof of theorem 4.1 given in the appendix uses properties of optimal transport to argue that under assumptions 1 and 2 (ii), $[\theta^L, \theta^H]$ is the sharp identified set for $P(Y_1 - Y_0 \leq \delta \mid D_1 > D_0)$. Nonetheless, it is interesting to note that the proof shows

$$\theta_x^L = OT_{c_L}(P_{1|x}, P_{0|x}) = \sup_y \{F_{1|x}(y) - F_{0|x}(y - \delta)\}$$
$$\theta_x^H = 1 - OT_{c_H}(P_{1|x}, P_{0|x}) = 1 - \sup_y \{F_{0|x}(y - \delta) - F_{1|x}(y)\} = 1 + \inf_y \{F_{1|x}(y) - F_{0|x}(y - \delta)\}$$

which are the Makarov bounds on $P(Y_1 - Y_0 \leq \delta \mid D_1 > D_0, X = x)$ studied in Fan & Park (2010).

Remark 4.2 (Pointwise vs. uniformly sharp CDF bounds). Under assumptions 1 and 2 (ii), $[\theta^L, \theta^H]$ is the sharp identified set for $P(Y_1 - Y_0 \leq \delta \mid D_1 > D_0)$ at the *point* δ. Viewing these bounds as functions of δ, $\theta^L(\delta)$ and $\theta^H(\delta)$ are not *uniformly* sharp bounds for the cumulative distribution function $P(Y_1 - Y_0 \leq \delta \mid D_1 > D_0)$, in the sense that not every CDF $F(\cdot)$ satisfying $\theta^L(\delta) \leq F(\delta) \leq \theta^H(\delta)$ for all δ could be the CDF of $Y_1 - Y_0$. See Firpo & Ridder (2019) for a detailed discussion of this point.

5 Estimators

Sample analogues of the expressions identifying $P_{1|x}$, $P_{0|x}$, and s_x in lemma 2.1 provide convenient plug-in estimators of γ^L and γ^H.

The following notation simplifies expressions for the sample analogues. Let P denote the distribution of an observation (Y, D, Z, X), and f be a real-valued function. Use $P(f)$ to mean $E_P[f(Y, D, Z, X)]$. Similarly, let $P_{d|x}(f) = E_{P_{d|x}}[f(Y_d)] = E[f(Y_d) \mid D_1 > D_0, X = x]$. Let $\mathbb{P}_n$ denote the empirical distribution formed from the sample $\{Y_i, D_i, Z_i, X_i\}_{i=1}^n$, and $\mathbb{P}_n(f) = \frac{1}{n}\sum_{i=1}^n f(Y_i, D_i, Z_i, X_i)$. The following indicator function notation also simplifies expressions:

$$\mathbb{1}_{d,x,z}(D, X, Z) = \mathbb{1}\{D = d, X = x, Z = z\},$$
$$\mathbb{1}_{x,z}(X, Z) = \mathbb{1}\{X = x, Z = z\}, \qquad \mathbb{1}_x(X) = \mathbb{1}\{X = x\}$$

For example, $P(D = d, X = x, Z = z)$ shortens to $P(\mathbb{1}_{d,x,z})$, and $\frac{1}{n}\sum_{i=1}^n \mathbb{1}\{D_i = 1, X_i = x, Z_i = 0\}$ to $\mathbb{P}_n(\mathbb{1}_{1,x,0})$.

The probabilities $p_{d,x,z} = P(\mathbb{1}_{d,x,z})$, $p_{x,z} = P(\mathbb{1}_{x,z})$, and $p_x = P(\mathbb{1}_x)$ are estimated with empirical analogues:

$$\hat{p}_{d,x,z} = \mathbb{P}_n(\mathbb{1}_{d,x,z}), \qquad \hat{p}_{x,z} = \mathbb{P}_n(\mathbb{1}_{x,z}), \qquad \hat{p}_x = \mathbb{P}_n(\mathbb{1}_x)$$

In this notation, $s_x = P(X = x \mid D_1 > D_0)$ and its empirical analogue $\hat{s}_x$ are

$$s_x = \frac{(p_{1,x,1}/p_{x,1} - p_{1,x,0}/p_{x,0})p_x}{\sum_{x'}(p_{1,x',1}/p_{x',1} - p_{1,x',0}/p_{x',0})p'_x}, \qquad \hat{s}_x = \frac{(\hat{p}_{1,x,1}/\hat{p}_{x,1} - \hat{p}_{1,x,0}/\hat{p}_{x,0})\hat{p}_x}{\sum_{x'}(\hat{p}_{1,x',1}/\hat{p}_{x',1} - \hat{p}_{1,x',0}/\hat{p}_{x',0})\hat{p}_{x'}} \tag{25}$$

The maps $P_{d|x}$ and their empirical analogues are

$$\begin{aligned} P_{d|x}(f) &= \frac{P(\mathbb{1}_{d,x,d} \times f)/p_{x,d} - P(\mathbb{1}_{d,x,1-d} \times f)/p_{x,1-d}}{p_{d,x,d}/p_{x,d} - p_{d,x,1-d}/p_{x,1-d}} \\ \hat{P}_{d|x}(f) &= \frac{\mathbb{P}_n(\mathbb{1}_{d,x,d} \times f)/\hat{p}_{x,d} - \mathbb{P}_n(\mathbb{1}_{d,x,1-d} \times f)/\hat{p}_{x,1-d}}{\hat{p}_{d,x,d}/\hat{p}_{x,d} - \hat{p}_{d,x,1-d}/\hat{p}_{x,1-d}} \end{aligned} \tag{26}$$

Under assumption 3, $\eta = (\eta_1, \eta_0) = (E_{P_1}[\eta_1(Y_1)], E_{P_0}[\eta_0(Y_0)])$. Each vector $\eta_d \in \mathbb{R}^{K_d}$ has coordinates $\eta_d^{(k)} = \sum_x s_x P_{d|x}(\eta_d^{(k)})$. Empirical analogues $\hat{\eta} = (\hat{\eta}_1, \hat{\eta}_0)$ are formed by $\hat{\eta}_d^{(k)} = \sum_x \hat{s}_x \hat{P}_{d|x}(\eta_d^{(k)})$.

Computing $\hat{P}_{d|x}(f)$ for a known f is straightforward:

$$\begin{aligned} \hat{P}_{d|x}(f) &= \frac{\frac{1}{\hat{p}_{x,d}} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{d,x,d}(D_i, X_i, Z_i) f(Y_i) - \frac{1}{\hat{p}_{x,1-d}} \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{d,x,1-d}(D_i, X_i, Z_i) f(Y_i)}{\hat{p}_{d,x,d}/\hat{p}_{x,d} - \hat{p}_{d,x,1-d}/\hat{p}_{x,1-d}} \\ &= \sum_{i=1}^n \omega_{d,x,i} \times f_i \end{aligned}$$

where $f_i = f(Y_i)$ and the weights $\omega_{d,x,i}$ can be computed directly from data:

$$\omega_{d,x,i} = \frac{1}{n} \times \frac{\mathbb{1}_{d,x,d}(D_i, X_i, Z_i)/\hat{p}_{x,d} - \mathbb{1}_{d,x,1-d}(D_i, X_i, Z_i)/\hat{p}_{x,1-d}}{\hat{p}_{d,x,d}/\hat{p}_{x,d} - \hat{p}_{d,x,1-d}/\hat{p}_{x,1-d}} \tag{27}$$

Sample analogue estimators of γ^L and γ^H are based on equations (19), (20), (22), (23), and (24). These expressions involve the optimal transport functional $OT_c(P_{1|x}, P_{0|x})$. The sample analogue of the simplified dual problem discussed in section 3 is written

$$OT_c(\hat{P}_{1|x}, \hat{P}_{0|x}) = \sup_{(\varphi,\psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)} \hat{P}_{1|x}(\varphi) + \hat{P}_{0|x}(\psi) \tag{28}$$

Here $\mathcal{F}_c$, $\mathcal{F}_c^c$, and the functions $\theta^L(\cdot)$, $\theta^H(\cdot)$ are defined according to the cost function:

(i) When assumption 2 (i) holds (the cost function $c(y_1, y_0)$ is Lipschitz continuous and $\mathcal{Y}$ is compact), $\mathcal{F}_c$ and $\mathcal{F}_c^c$ are given by:

$$\begin{aligned} \mathcal{F}_c &= \left\{\varphi : \mathcal{Y} \to \mathbb{R}\ ;\ -\|c\|_\infty \le \varphi(y_1) \le \|c\|_\infty,\ |\varphi(y_1) - \varphi(y_1')| \le L|y_1 - y_1'|\right\} \\ \mathcal{F}_c^c &= \left\{\psi : \mathcal{Y} \to \mathbb{R}\ ;\ -2\|c\|_\infty \le \psi(y_0) \le 0,\ |\psi(y_0) - \psi(y_0')| \le L|y_0 - y_0'|\right\} \end{aligned}$$

and $\theta^L(\hat{P}_{1|x}, \hat{P}_{0|x})$, $\theta^H(\hat{P}_{1|x}, \hat{P}_{0|x})$ are analogues of equation (19):

$$c_L(y_1, y_0) = c(y_1, y_0), \qquad\qquad c_H(y_1, y_0) = -c(y_1, y_0)$$
$$\theta^L(\hat{P}_{1|x}, \hat{P}_{0|x}) = OT_{c_L}(\hat{P}_{1|x}, \hat{P}_{0|x}), \qquad\qquad \theta^H(\hat{P}_{1|x}, \hat{P}_{0|x}) = -OT_{c_H}(\hat{P}_{1|x}, \hat{P}_{0|x}).$$

(ii) When assumption 2 (ii) holds (the cost function is $c(y_1, y_0) = \mathbb{1}\{y_1 - y_0 \leq \delta\}$), $\mathcal{F}_c$ and $\mathcal{F}_c^c$ are given by:

$$\mathcal{F}_c = \{\varphi : \mathcal{Y} \to \mathbb{R}\ ;\ \varphi(y_1) = \mathbb{1}\{y_1 \in I\} \text{ for some interval } I\}$$
$$\mathcal{F}_c^c = \{\psi : \mathcal{Y} \to \mathbb{R}\ ;\ \psi(y_0) = -\mathbb{1}\{y_0 \in I^c\} \text{ for some interval } I\}$$

and $\theta^L(\hat{P}_{1|x}, \hat{P}_{0|x})$, $\theta^H(\hat{P}_{1|x}, \hat{P}_{0|x})$ are analogues of equation (20):

$$c_L(y_1, y_0) = \mathbb{1}\{y_1 - y_0 < \delta\}, \qquad\qquad c_H = \mathbb{1}\{y_1 - y_0 > \delta\}$$
$$\theta^L(\hat{P}_{1|x}, \hat{P}_{0|x}) = OT_{c_L}(\hat{P}_{1|x}, \hat{P}_{0|x}), \qquad\qquad \theta^H(\hat{P}_{1|x}, \hat{P}_{0|x}) = 1 - OT_{c_H}(\hat{P}_{1|x}, \hat{P}_{0|x})$$

The sample analogue estimators are given by

$$\hat{\theta}_x^L = \theta^L(\hat{P}_{1|x}, \hat{P}_{0|x}), \qquad\qquad \hat{\theta}_x^H = \theta^H(\hat{P}_{1|x}, \hat{P}_{0|x}), \tag{29}$$
$$\hat{\theta}^L = \sum_x \hat{s}_x \hat{\theta}_x^L, \qquad\qquad \hat{\theta}^H = \sum_x \hat{s}_x \hat{\theta}_x^H, \tag{30}$$
$$\hat{\gamma}^L = g^L(\hat{\theta}^L, \hat{\theta}^H, \hat{\eta}), \qquad\qquad \hat{\gamma}^H = g^H(\hat{\theta}^L, \hat{\theta}^H, \hat{\eta}) \tag{31}$$

The optimization problems in $\theta^L(\hat{P}_{1|x}, \hat{P}_{0|x})$ and $\theta^H(\hat{P}_{1|x}, \hat{P}_{0|x})$ are especially straightforward when treatment is exogenous. Recall the claim of equation (13): the supremum of $P_{1|x}(\varphi) + P_{0|x}(\psi)$ over the larger set Φ_c is the same value when restricted to $\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$. The argument behind this claim uses monotonicity of the maps $P_{d|x}$. When treatment is exogenous, $\hat{P}_{d|x}$ corresponds to a probability distribution and is therefore also monotonic. Thus the claim holds replacing $P_{d|x}$ with

$\hat{P}_{d|x}$, implying the function classes $\mathcal{F}_c$ and $\mathcal{F}_c^c$ can be ignored in computation:

$$
\begin{aligned}
OT_c(\hat{P}_{1|x}, \hat{P}_{0|x}) &= \sup_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} \hat{P}_{1|x}(\varphi) + \hat{P}_{0|x}(\psi) = \sup_{(\varphi,\psi)\in\Phi_c} \hat{P}_{1|x}(\varphi) + \hat{P}_{0|x}(\psi) \\
&= \sup_{\{\varphi_i,\psi_j\}_{i,j}} \sum_{i=1}^{n} \omega_{1,x,i}\varphi_i + \sum_{j=1}^{n} \omega_{0,x,j}\psi_j \qquad (32) \\
&\quad \text{s.t. } \varphi_i + \psi_j \leq c(Y_i, Y_j) \text{ for all } 1 \leq i, j \leq n
\end{aligned}
$$

the final problem in this display is a linear programming problem with $2n$ choice variables and n^2 constraints, and can be further simplified by removing choice variables (and the corresponding constraints) whose weights $\omega_{d,x,i}$ equal zero. Many weights do equal zero, as only observations with $X_i = x$ correspond to nonzero weights.

When there is noncompliance in the sample, $\hat{P}_{d|x}$ does not correspond to a probability distribution. This is easily seen by noting that for observations i where Z_i differs from D_i, the weight $\omega_{d,x,i}$ defined in (27) is negative. Nonetheless, it remains computationally tractable to search over $\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$. For example, when the cost function is continuous $OT_c(\hat{P}_{1|x}, \hat{P}_{0|x})$ remains a linear programming problem, with additional linear constraints enforcing $|\varphi_i + \psi_j| \leq L|Y_i - Y_j|$, $-\|c\|_\infty \leq \varphi_i \leq \|c\|_\infty$, and $-2\|c\|_\infty \leq \psi_j \leq 0$.

5.1 Asymptotic analysis

The estimators proposed above are especially attractive because they are a (Hadamard directionally) differentiable map of the empirical distribution. Specifically, there exists a collection of functions $\mathcal{F}$ and a map $T : \ell^\infty(\mathcal{F}) \to \mathbb{R}^2$ described by equations (25), (26), (29), (30), and (31) such that

$$
(\hat{\gamma}^L, \hat{\gamma}^H) = T(\mathbb{P}_n), \qquad\qquad (\gamma^L, \gamma^H) = T(P)
$$

The set $\mathcal{F}$ consists of the functions in $\mathcal{F}_c$, $\mathcal{F}_c^c$, and the coordinate functions defining η, multiplied by various indicator functions. It is formally defined in appendix C. Under assumption 1, 2, and 3, $\mathcal{F}$ is a Donsker set and $T(\cdot)$ is continuous at P, which implies the esimators are consistent:

$$
(\hat{\gamma}^L, \hat{\gamma}^H) = T(\mathbb{P}_n) \xrightarrow{p} T(P) = (\gamma^L, \gamma^H) \tag{33}
$$

5.1.1 Weak convergence

The map $T(\cdot)$ is not only continuous under assumptions 1, 2, and 3, but Hadamard directionally differentiable. An application of the functional delta method gives the conclusion $\sqrt{n}((\hat{\gamma}^L, \hat{\gamma}^H) - (\gamma^L, \gamma^H))$ converges in distribution, a result stated formally in theorem 5.2 below.

In order to build hypothesis tests or construct confidence intervals based on the asymptotic distribution of $\sqrt{n}((\hat{\gamma}^L, \hat{\gamma}^H) - (\gamma^L, \gamma^H))$, one must be able to estimate the asymptotic distribution. This is possible under assumptions 1, 2, and 3, but involves a more complex procedure described in section 5.2.2. Under an additional assumption, a straightforward bootstrap will do.

For each instance of the restricted dual problem used in defining $T(\cdot)$, the set of maximizers

$$\Psi_c(P_{1|x}, P_{0|x}) = \underset{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)}{\arg\max} P_{1|x}(\varphi) + P_{0|x}(\psi) \tag{34}$$

is nonempty. If the solutions are suitably unique for each instance, the map $T(\cdot)$ is fully Hadamard differentiable at P and a straightforward bootstrap will consistently estimate the asymptotic distribution.

Assumption 4 states this high-level uniqueness condition, while the following lemma 5.1 gives low-level sufficient conditions for it to hold. Let $\mathcal{Y}_{d,x}$ be the support of Y conditional on $D = d$ and $X = x$, and $\mathbb{1}_{\mathcal{Y}_{d,x}}(y) = \mathbb{1}\{y \in \mathcal{Y}_{d,x}\}$ be the indicator function for this set.

Assumption 4. *For each $x \in \mathcal{X}$, each $c \in \{c_L, c_H\}$, and any $(\varphi_1, \psi_1), (\varphi_2, \psi_2) \in \Psi_c(P_{1|x}, P_{0|x})$, there exists $s \in \mathbb{R}$ such that*

$$\mathbb{1}_{\mathcal{Y}_{1,x}} \times \varphi_1 = \mathbb{1}_{\mathcal{Y}_{1,x}} \times (\varphi_2 + s),\ P\text{-a.s.} \qquad \text{and} \qquad \mathbb{1}_{\mathcal{Y}_{0,x}} \times \psi_1 = \mathbb{1}_{\mathcal{Y}_{0,x}} \times (\psi_2 - s),\ P\text{-a.s.}$$

Lemma 5.1. *Suppose that*

(i) assumption 2 (i) holds, with cost function $c(y_1, y_0)$ that is continuously differentiable, and

(ii) for each (d, x), the support of $P_{d|x}$ is $\mathcal{Y}_{d,x}$, which is a bounded interval.

Then assumption 4 holds.

When treatment is exogenous, condition (ii) of lemma 5.1 simplifies to the assumption that the distribution of $Y_d \mid X = x$ has bounded support $[y_{d,x}^{\ell}, y_{d,x}^{u}]$. In general, this condition requires the support of Y_d for the subpopulation of compliers with covariate value x is a bounded interval that contains the support of the relevant subpopulation of non-compliers. Specifically, the support of Y_1

for compliers is a bounded interval containing the support of Y_1 for always-takers, and the support of Y_0 for compliers is a bounded interval containing the support of Y_0 for never-takers.

Assumption 4 can hold even when the conditions of lemma 5.1 do not. For example, when interest is in the cumulative distribution function and assumption 2 (ii) is satisfied, the dual problem is essentially optimizing over the difference of CDFs (see remark 4.1). Although the cost functions are not continuously differentiable, it is still plausible for this optimization problem to have a unique solution in well-behaved cases. For further discussion of uniqueness of the dual solutions of optimal transport, see Staudt et al. (2022).

The following theorem gives the main weak convergence result.

Theorem 5.2. *Suppose assumptions 1, 2, and 3 hold, and let $\mathbb{G}$ be the weak limit of $\sqrt{n}(\mathbb{P}_n - P)$ in $\ell^\infty(\mathcal{F})$. Then T is Hadamard directionally differentiable at P tangentially to the support of $\mathbb{G}$, and*

$$\sqrt{n}((\hat{\gamma}^L, \hat{\gamma}^H) - (\gamma^L, \gamma^H)) = \sqrt{n}(T(\mathbb{P}_n) - T(P)) \xrightarrow{L} T'_P(\mathbb{G})$$

If assumption 4 also holds, then T'_P is linear on the support of $\mathbb{G}$ and $T'_P(\mathbb{G})$ is bivariate normal.

5.2 Inference

To make use of the weak convergence result of theorem 5.2 for inference, this section develops methods of estimating the law of $T'_P(\mathbb{G})$ by utilizing the bootstrap. The "exchangeable bootstrap" procedures discussed in van der Vaart & Wellner (1997) are computationally convenient for reasons discussed below. These procedures define a new map $\mathbb{P}_n^* \in \ell^\infty(\mathcal{F})$ pointwise with

$$\mathbb{P}_n^*(f) = \frac{1}{n}\sum_{i=1}^n W_i f(Y_i, D_i, Z_i, X_i) \tag{35}$$

for nonnegative random variables $\{W_i\}_{i=1}^n$ independent of the data $\{Y_i, D_i, Z_i, X_i\}_{i=1}^n$, and satisfying technical conditions omitted here. I focus on two notable examples, the nonparametric bootstrap of Efron (1979) and the "Bayesian" bootstrap of Rubin (1981). Either bootstrap can be used to estimate the asymptotic distribution. The Bayesian bootstrap may be preferable in small samples for reasons discussed below.

Definition 5.1 (Nonparametric bootstrap)**.** *Let $(W_1, \dots, W_n) \sim$ Multinomial$(n, (1/n, \dots, 1/n))$ be independent of the data $\{Y_i, D_i, Z_i, X_i\}_{i=1}^n$. Define $\mathbb{P}_n^* \in \ell^\infty(\mathcal{F})$ pointwise with* (35)*.*

Definition 5.2 (Bayesian bootstrap)**.** *Let $\{\xi_i\}_{i=1}^n$ be i.i.d. exponentially distributed random variables with mean 1, independent of the data $\{Y_i, D_i, Z_i, X_i\}_{i=1}^n$. Set $W_i = \xi_i/(n^{-1}\sum_{i=1}^n \xi_i)$, and define $\mathbb{P}_n^* \in \ell^\infty(\mathcal{F})$ pointwise with (35).*

The map $\mathbb{P}_n^*$ in (35) can be used to compute $(\hat{\gamma}^{L*}, \hat{\gamma}^{H*}) = T(\mathbb{P}_n^*)$ in much the same way that $T(\mathbb{P}_n)$ is computed. Specifically, bootstrap analogues of $\hat{p}_{d,x,z}$, $\hat{p}_{x,z}$, and $\hat{p}_x$ are given by

$$\hat{p}^*_{d,x,z} = \frac{1}{n}\sum_{i=1}^n W_i \mathbb{1}_{d,x,z}(D_i, X_i, Z_i), \qquad \hat{p}^*_{x,z} = \frac{1}{n}\sum_{i=1}^n W_i \mathbb{1}_{x,z}(X_i, Z_i), \qquad \hat{p}^*_{x} = \frac{1}{n}\sum_{i=1}^n W_i \mathbb{1}_{x}(X_i),$$

and the bootstrap analogue of $\hat{s}_x$ is

$$\hat{s}^*_x = \frac{(\hat{p}^*_{1,x,1}/\hat{p}^*_{x,1} - \hat{p}^*_{1,x,0}/\hat{p}^*_{x,0})\hat{p}^*_x}{\sum_{x'}(\hat{p}^*_{1,x',1}/\hat{p}^*_{x',1} - \hat{p}^*_{1,x',0}/\hat{p}^*_{x',0})\hat{p}^*_{x'}}$$

The maps $\hat{P}_{d|x}$ have bootstrap analogues

$$\hat{P}^*_{d|x}(f) = \frac{\mathbb{P}_n^*(\mathbb{1}_{d,x,d} \times f)/\hat{p}^*_{x,d} - \mathbb{P}_n^*(\mathbb{1}_{d,x,1-d} \times f)/\hat{p}^*_{x,1-d}}{\hat{p}^*_{d,x,d}/\hat{p}^*_{x,d} - \hat{p}^*_{d,x,1-d}/\hat{p}^*_{x,1-d}} = \sum_{i=1}^n \omega^*_{d,x,i} f_i$$

where $f_i = f(Y_i)$ and $\omega^*_{d,x,i}$ are bootstrap versions of the weights in (27):

$$\omega^*_{d,x,i} = \frac{W_i}{n} \times \frac{\mathbb{1}_{d,x,d}(D_i, X_i, Z_i)/\hat{p}^*_{x,d} - \mathbb{1}_{d,x,1-d}(D_i, X_i, Z_i)/\hat{p}^*_{x,1-d}}{\hat{p}^*_{d,x,d}/\hat{p}^*_{x,d} - \hat{p}^*_{d,x,1-d}/\hat{p}^*_{x,1-d}} \tag{36}$$

Finally, $(\hat{\gamma}^{L*}, \hat{\gamma}^{H*})$ can be computed with

$$\hat{\theta}^{L*}_x = \theta^L(\hat{P}^*_{1|x}, \hat{P}^*_{0|x}), \qquad \hat{\theta}^{H*}_x = \theta^H(\hat{P}^*_{1|x}, \hat{P}^*_{0|x}), \tag{37}$$

$$\hat{\theta}^{L*} = \sum_x \hat{s}^*_x \hat{\theta}^{L*}_x, \qquad \hat{\theta}^{H*} = \sum_x \hat{s}^*_x \hat{\theta}^{H*}_x, \tag{38}$$

$$\hat{\gamma}^{L*} = g^L(\hat{\theta}^{L*}, \hat{\theta}^{H*}, \hat{\eta}^*), \qquad \hat{\gamma}^{H*} = g^H(\hat{\theta}^{L*}, \hat{\theta}^{H*}, \hat{\eta}^*) \tag{39}$$

5.2.1 Simple bootstrap with full differentiability

Under assumption 4, estimating the distribution of $T'_P(\mathbb{G})$ is straightforward.

Theorem 5.3. *Suppose assumptions 1, 2, 3, and 4 hold, and let $\mathbb{P}_n^*$ be given by definition 5.1 or 5.2. Then conditional on $\{Y_i, D_i, Z_i, X_i\}_{i=1}^n$,*

$$\sqrt{n}(T(\mathbb{P}_n^*) - T(\mathbb{P}_n)) \xrightarrow{L} T'_P(\mathbb{G})$$

in outer probability.

It is worth emphasizing the computationally convenience of the bootstrap $\mathbb{P}_n^*$ given in (35) when treatment is exogenous. The weights given in display (36) simplify to

$$\omega^*_{d,x,i} = \frac{W_i}{n} \times \frac{\mathbb{1}\{D_i = d, X_i = x\}}{\hat{p}^*_{x,d}} \tag{40}$$

As these weights are nonnegative and sum to one, $\hat{P}^*_{d|x}$ is a probability distribution. Accordingly, $\theta^L(\hat{P}^*_{1|x}, \hat{P}^*_{0|x})$ and $\theta^H(\hat{P}^*_{1|x}, \hat{P}^*_{0|x})$ can be computed ignoring the function classes $\mathcal{F}_c$ and $\mathcal{F}_c^c$ for the same reasons discussed around display (32):

$$\begin{aligned} OT_c(\hat{P}^*_{1|x}, \hat{P}^*_{0|x}) &= \sup_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} \hat{P}^*_{1|x}(\varphi) + \hat{P}^*_{0|x}(\psi) = \sup_{(\varphi,\psi)\in\Phi_c} \hat{P}^*_{1|x}(\varphi) + \hat{P}^*_{0|x}(\psi) \\ &= \sup_{\{\varphi_i,\psi_j\}_{i,j}} \sum_{i=1}^n \omega^*_{1,x,i}\varphi_i + \sum_{j=1}^n \omega^*_{0,x,j}\psi_j \\ &\qquad \text{s.t. } \varphi_i + \psi_j \le c(Y_i, Y_j) \text{ for all } 1 \le i, j \le n \end{aligned}$$

A researcher utilizing the nonparametric bootstrap runs the risk of a boostrap draw including no observations with $\mathbb{1}\{D_i = d, X_i = x\}$. As $\hat{p}^*_{x,d} = \frac{1}{n}\sum_{i=1}^n W_i \mathbb{1}\{D_i = d, X_i = x\}$, this would result in the formula in (40) attempting to divide by zero. This problem cannot arise when using the Bayesian bootstrap suggested in 5.2; in this procedure $W_i > 0$ for each i, and thus $\hat{p}^*_{x,d} = \frac{1}{n}\sum_{i=1}^n W_i \mathbb{1}\{D_i = d, X_i = x\} > 0$ as long as $\hat{p}_{d,x} > 0$.

5.2.2 Alternative for directional differentiability

The solutions to optimal transport may not be unique as assumption 4 requires. As emphasized in the statement of theorem 5.2, assumption 4 is not needed to obtain the asymptotic distribution of the estimators. However, without assumption 4 the procedure suggested by lemma 5.3 may not consistently estimate that limiting distribution. When in doubt, researchers can make use of an alternative procedure based on the results of Fang & Santos (2019) and described below.

Additional notation is needed to describe this alternative. Let $\eta^{(k)}_{d,x} = P_{d|x}(\eta^{(k)}_d)$, and let $T_1(\cdot)$ denote the "first stage" function computing $P_{1|x}$, $P_{0|x}$, $\eta_{1,x}$, $\eta_{0,x}$, and s_x for each x:

$$T_1(P) = \left(\left\{P_{1|x}, P_{0|x}, \eta_{1,x}, \eta_{0,x}, s_x\right\}_{x\in\mathcal{X}}\right)$$

Here $\{a_x\}_{x\in\mathcal{X}} = (a_{x_1}, \ldots, a_{x_M})$. Let $\{\kappa_n\}_{n=1}^{\infty}$ be a sequence in $\mathbb{R}$ satisfying $\kappa_n \uparrow \infty$ and $\kappa_n/\sqrt{n} \to 0$. Define the set of empirical approximate maximizers:

$$\widehat{\Psi}_{c,x} = \left\{ (\varphi, \psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c) \;;\; OT_c(\hat{P}_{1|x}, \hat{P}_{0|x}) \leq \hat{P}_{1|x}(\varphi) + \hat{P}_{0|x}(\psi) + \frac{\kappa_n}{\sqrt{n}} \right\}$$

and the maps

$$\widehat{OT}'_{c,x}(H_1, H_0) = \sup_{(\varphi,\psi)\in\widehat{\Psi}_{c,x}} H_1(\varphi) + H_0(\psi),$$

and

$$\begin{aligned}
&\widehat{T}'_{2,T_1(P)}\left(\{H_{1,x}, H_{0,x}, h_{\eta_1,x}, h_{\eta_0,x}, h_{s,x}\}_{x\in\mathcal{X}}\right) \\
&\quad = \left(\left\{ \widehat{OT}'_{c_L,x}(H_{1,x}, H_{0,x}), -\widehat{OT}'_{c_H,x}(H_{1,x}, H_{0,x}), h_{\eta_1,x}, h_{\eta_0,x}, h_{s,x} \right\}_{x\in\mathcal{X}} \right)
\end{aligned}$$

The alternative procedure uses the conditional law of

$$\hat{D}_4 \hat{D}_3 \widehat{T}'_{2,T_1(P)}\left(\sqrt{n}(T_1(\mathbb{P}_n^*) - T_1(\mathbb{P}_n))\right)$$

given the data, where $\hat{D}_4$ and $\hat{D}_3$ are matrices given by

$$\hat{D}_3 = \underbrace{\begin{bmatrix} \hat{D}_{3,x_1} & \hat{D}_{s,x_2} & \ldots & \hat{D}_{s,x_M} \end{bmatrix}}_{(2+d_\eta)\times M(3+d_\eta)}, \qquad \hat{D}_{3,x} = \underbrace{\begin{bmatrix} \hat{s}_x & 0 & 0 & 0 & \hat{\theta}_x^L \\ 0 & \hat{s}_x & 0 & 0 & \hat{\theta}_x^H \\ 0 & 0 & \hat{s}_x I_{K_1} & 0 & \hat{\eta}_{1,x} \\ 0 & 0 & 0 & \hat{s}_x I_{K_0} & \hat{\eta}_{0,x} \end{bmatrix}}_{(2+d_\eta)\times(3+d_\eta)},$$

$$D_4 = \underbrace{\begin{bmatrix} \nabla g^L(\hat{\theta}^L, \hat{\theta}^H, \hat{\eta})^\intercal \\ \nabla g^H(\hat{\theta}^L, \hat{\theta}^H, \hat{\eta})^\intercal \end{bmatrix}}_{2\times(2+d_\eta)},$$

Theorem 5.4. *Suppose assumptions 1, 2, and 3 hold, let $\mathbb{P}_n^*$ be given by definition 5.1 or 5.2, and $\{\kappa_n\}_{n=1}^{\infty} \subseteq \mathbb{R}$ satisfy $\kappa_n \to \infty$ and $\kappa_n/\sqrt{n} \to 0$. Then conditional on $\{Y_i, D_i, Z_i, X_i\}_{i=1}^n$,*

$$\hat{D}_4 \hat{D}_3 \widehat{T}_{2,T_1(P)}(\sqrt{n}(T_1(\mathbb{P}_n^*) - T_1(\mathbb{P}_n))) \xrightarrow{L} T'_P(\mathbb{G})$$

in outer probability.

5.2.3 Confidence sets

Theorems 5.3 and 5.4 make it straightforward to conduct inference. For example, a simple confidence set for the identified set $[\gamma^L, \gamma^H]$ is given by

$$\left[\hat{\gamma}^L - \hat{c}_{1-\alpha}/\sqrt{n}, \hat{\gamma}^H + \hat{c}_{1-\alpha}/\sqrt{n}\right]$$

where $\hat{c}_{1-\alpha}$ is a consistent estimator of the $1-\alpha$ quantile of $\max\{T'_P(\mathbb{G})^{(1)}, -T'_P(\mathbb{G})^{(2)}\}$. When assumptions 1 through 4 hold, let $(\hat{\gamma}^{L*}, \hat{\gamma}^{H*}) = T(\mathbb{P}_n^*)$. When assumptions 1 through 3 hold but assumption 4 is doubtful, let $(\hat{\gamma}^{L*}, \hat{\gamma}^{H*}) = (\hat{\gamma}^L, \hat{\gamma}^H) + \frac{1}{\sqrt{n}}\hat{D}_4\hat{D}_3\widehat{T}_{2,T_1(P)}(\sqrt{n}(T_1(\mathbb{P}_n^*) - T_1(\mathbb{P}_n)))$. In either case, compute

$$\hat{c}_{1-\alpha} = \inf\left\{c\ ;\ P\left(\max\left\{\sqrt{n}(\hat{\gamma}^{L*} - \hat{\gamma}^L), -\sqrt{n}(\hat{\gamma}^{H*} - \hat{\gamma}^H)\right\} \leq c \mid \{Y_i, D_i, Z_i, X_i\}_{i=1}^n\right) \geq 1-\alpha\right\}$$

through simulation:

1. Compute $(\hat{\gamma}^L, \hat{\gamma}^H) = T(\mathbb{P}_n)$ and, if necessary, $\hat{D}_4$, and $\hat{D}_3$.
2. Generate N boostrap samples, $\{W_{i,b}\}_{i=1}^n$ for each $b = 1, \ldots, N$ according to definition 5.1 or 5.2. For each bootstrap sample b, compute $(\hat{\gamma}_b^{L*}, \hat{\gamma}_b^{H*})$ as described above.
3. Let $\hat{c}_{1-\alpha}$ be the $1-\alpha$ quantile of $\{\max\{\sqrt{n}(\hat{\gamma}_b^{L*} - \hat{\gamma}^L), -\sqrt{n}(\hat{\gamma}_b^{H*} - \hat{\gamma}^H)\}_{b=1}^N$.

Under the further assumption that the cumulative distribution function of $\max\{T'_P(\mathbb{G})^{(1)}, -T'_P(\mathbb{G})^{(2)}\}$ is continuous and strictly increasing at its $1-\alpha$ quantile,

$$\lim_{n\to\infty} P\left([\gamma^L, \gamma^H] \subseteq \left[\hat{\gamma}^L - \hat{c}_{1-\alpha}/\sqrt{n}, \hat{\gamma}^H + \hat{c}_{1-\alpha}/\sqrt{n}\right]\right) = 1-\alpha$$

Confidence sets for the parameter could be constructed following Imbens & Manski (2004).

6 Application: job training experiment

In this section I demonstrate the estimators in revisiting the famous National Supported Work Demonstration program (LaLonde (1986)). This program was implemented in the 1970s with the aim of helping socially and economically disadvantaged workers obtain job skills. Those randomly selected into the program were guaranteed a job lasting six to eighteen months, and frequently met with a counselor to discuss performance.

I make use of the "LaLonde" sample studied in Diamond & Sekhon (2013). This sample consists of male participants and includes 297 treated and 425 control observations. The outcome of interest is real earnings in 1978. Observed covariates include age, years of education, real earnings in months 13 to 24 prior to randomization, and indicators for whether a participant is a high school dropout, black, hispanic, or married. Averages and standard deviations of these covariates by treatment status are reported in table 1:

Table 1: Balance table

	base inc.	age	yrs. educ	HS dropout	black	hispanic	married	N
control	3672.49	24.45	10.19	0.81	0.80	0.11	0.16	425
	(6521.53)	(6.59)	(1.62)	(0.39)	(0.40)	(0.32)	(0.36)	
treated	3571.00	24.63	10.38	0.73	0.80	0.09	0.17	297
	(5773.13)	(6.69)	(1.82)	(0.44)	(0.40)	(0.29)	(0.37)	

Note: Standard deviations in parentheses.

There is no reported noncompliance, so I interpret the setting as one of exogenous treatment. The parameter of interest is the OLS slope coefficient of regressing treatment effects on a constant and Y_0:

$$\gamma = \frac{\text{Cov}(Y_1 - Y_0, Y_0)}{\text{Var}(Y_0)} = \frac{E_{P_{1,0}}[(Y_1 - Y_0)Y_0] - (E_{P_1}[Y_1] - E_{P_0}[Y_0])E_{P_0}[Y_0]}{E_{P_0}[Y_0^2] - (E_{P_0}[Y_0])^2}$$

as described in example 2.3, the sign of this parameter describes who receives larger benefits from treatment: $\gamma < 0$ implies those with below average untreated outcomes tend to see above average treatment effects.

Discretized versions of baseline income and age are found to be informative covariates. Baseline income is binned as: $[0, 0]$ or $(0, \infty)$, while age is binned as $(16, 20]$, $(20, 26]$, or $(26, \infty)$. X is the cartesian product of bins. The resulting (d, x) bins have a minimum of 31 observations per bin, and an average of 60.2 observations per bin.

The point estimates are $(\hat{\gamma}^L, \hat{\gamma}^H) = (-1.73, -0.004)$. The negative upper bound point estimates suggests that the treatment was especially beneficial for participants who would otherwise have incomes below average (for the eligible population). Covariates are found to be informative, especially for the upper bound. Ignoring covariates, the lower bound point estimate is -1.78 and the upper bound point estimate is 0.189. The 95% confidence set for the identified based on 500 bootstrap draws is $[-1.94, 0.20]$, suggesting γ may still be zero or slightly positive once accounted for sample uncertainty.

7 Extensions

This section briefly describes simple extensions.

7.1 Conditioning on $X \in A$

In many applications parameters conditional on a covariate taking a particular value are of interest. For example, the share of compliers of a particular demographic benefiting from treatment is $P(Y_1 > Y_0 \mid D_1 > D_0, \text{demographic})$.

Such parameters can be written in the form

$$\gamma_A = g(\theta_A, \eta_A)$$

where for a known set $A \subseteq \mathcal{X}$,

$$\theta_A \equiv E[c(Y_1, Y_0) \mid D_1 > D_0, X \in A], \qquad \eta_A \equiv E[\eta_1(Y_1), \eta_0(Y_0) \mid D_1 > D_0, X \in A]$$

The identified set for γ_A is straightforward to characterize and estimate. First note that

$$\theta_A = E[\theta_X \mid D_1 > D_0, X \in A] = \frac{1}{s_A} \sum_{x \in A} s_x \theta_x$$

where $s_A = \sum_{x \in A} s_x$. The proof of theorem 4.1 shows that the sharp identified set for $(\theta_{x_1}, \ldots, \theta_{x_M})$ is in fact $[\theta_{x_1}^L, \theta_{x_1}^H] \times \ldots \times [\theta_{x_M}^L, \theta_{x_M}^H]$. It follows that the sharp identified set for θ_A is $[\theta_A^L, \theta_A^H]$, where

$$\theta_A^L = \frac{1}{s_A} \sum_{x \in A} s_x \theta_x^L, \qquad \theta_A^H = \frac{1}{s_A} \sum_{x \in A} s_x \theta_x^H$$

and the sharp identified set for γ_A is $[\gamma_A^L, \gamma_A^H]$ where

$$\gamma_A^L = \min_{t \in [\theta_A^L, \theta_A^H]} g(t, \eta_A), \qquad \gamma_A^H = \max_{t \in [\theta_A^L, \theta_A^H]} g(t, \eta_A),$$

Let $\hat{s}_x$, $\hat{\theta}_x^L$, and $\hat{\theta}_x^H$ be as defined in section 5. Let $\hat{s}_A = \sum_{x \in A} \hat{s}_x$ and

$$\hat{\theta}_A^L = \frac{1}{\hat{s}_A} \sum_{x \in A} \hat{s}_x \hat{\theta}_x^L, \qquad \hat{\theta}^H(A) = \frac{1}{\hat{s}_A} \sum_{x \in A} \hat{s}_x \hat{\theta}_x^H$$

$$\hat{\gamma}_A^L = \min_{t \in [\hat{\theta}_A^L, \hat{\theta}_A^H]} g(t, \hat{\eta}_A), \qquad \hat{\gamma}_A^H = \max_{t \in [\hat{\theta}_A^L, \hat{\theta}_A^H]} g(t, \hat{\eta}_A),$$

Under assumptions 1, 2, and 3, $\sqrt{n}((\hat{\gamma}_A^L, \hat{\gamma}_A^H) - (\gamma_A^L, \gamma_A^H)$ will converge weakly. With assumption 4 the straightforward bootstrap will consistently estimate its asymptotic distribution.

7.2 Quantiles

Example 2.5 considers the parameter q_τ solving

$$P(Y_1 - Y_0 \le q_\tau \mid D_1 > D_0) = \tau$$

As noted in that example, the sharp identification results for $P(Y_1 - Y_0 \le \delta \mid D_1 > D_0)$ can be adapted to characterize the sharp identified set for q_τ. First view the bounds on the cumulative distribution function as functions of δ:

$$c_{L,\delta}(y_1, y_0) = \mathbb{1}\{y_1 - y_0 < \delta\}, \qquad c_{H,\delta}(y_1, y_0) = \mathbb{1}\{y_1 - y_0 > \delta\},$$

$$\theta_x^L(\delta) = OT_{c_{L,\delta}}(P_{1|x}, P_{0|x}), \qquad \theta_x^H(\delta) = 1 - OT_{c_{H,\delta}}(P_{1|x}, P_{0|x})$$

$$\theta^L(\delta) = \sum_x s_x \theta_x^L(\delta) \qquad \theta^H(\delta) = \sum_x s_x \theta_x^H(\delta)$$

Let $Q_{I,\tau}$ denote the sharp identified set for q_τ.

Lemma 7.1 (Identification of q_τ)**.** *Suppose assumptions 1 and 2 (ii) hold. Then $q \in Q_{I,\tau}$ if and only if $\theta^L(q) \le \tau \le \theta^H(q)$.*

Lemma 7.1 implies that inverting a test of $H_0 : \theta^L(q) \le \tau \le \theta^H(q)$ against the alternative $H_1 : \tau < \theta^L(q)$ or $\theta^H(q) < \tau$ will lead to valid confidence sets for q_τ.

Remark 7.1. Consider instead defining q_τ to be the closed subset of $\mathbb{R}$ given by

$$q_\tau = [\inf\{y \,;\, P(Y_1 - Y_0 \le y) \ge \tau\}, \inf\{y \,;\, P(Y_1 - Y_0 \le y) > \tau\}]$$

Note that this q_τ is the singleton $\inf\{y \,;\, P(Y_1 - Y_0 \le y) \ge \tau\}$, unless $P(Y_1 - Y_0 \le \cdot)$ is flat when equal to τ, in which case it equals the τ-level set $\{y \,;\, P(Y_1 - Y_0 \le y) = \tau\}$. (Compare Ehm et al. (2016), who define the τ-th quantile equivalently as $q_\tau = [\sup\{y \,;\, P(Y_1 - Y_0 \le y) <$

$\tau\}, \sup\{y\ ;\ P(Y_1 - Y_0 \leq y) \leq \tau\}].$) Let $Q_{I,\tau}$ denote the identified set of q_τ as defined in this remark. Lemma A.2 in appendix A shows that under assumptions 1 and 2 (ii), $q \in Q_{I,\tau}$ if and only if $\theta^L(q) \leq \tau \leq \theta^H(q)$.

7.3 Multiple treatment arms with exogenous treatment

The identification results and estimators proposed above are easily extended to a setting with multiple treatment arms and exogenous treatment. Let the mutually exclusive treatment arms indexed by $d \in \{0, 1, \ldots, J\}$, with $d = 0$ indicating control. Let Y_d be the potential outcome with treatment d, D_d equal one if the unit has treatment d and zero otherwise. The observed outcome is

$$Y = \sum_{d=0}^{J} D_d Y_d$$

Let $D = (D_0, D_1, \ldots, D_J)$ and assume

$$(Y_0, Y_1, \ldots, Y_J) \perp D \mid X$$

Note that the marginal distributions of $Y_d \mid X = x$, denoted $P_{d|x}$, are identified with the relation

$$E_{P_{d|x}}[f(Y_d)] = E[f(Y_d) \mid X = x] = \frac{E[f(Y)D_d \mid X = x]}{P(D_d = 1 \mid X = x)}$$

Let $\gamma_d = g(\theta_d, \eta_d)$ where $\theta_d = E[c(Y_d, Y_0)]$. Consider estimating the sharp identified set for $(\gamma_1, \ldots \gamma_J)$. For example, an RCT with two treatment arms may have similar average treatment effects. The treatment arms may be further distinguished by comparing $P(Y_1 - Y_0 > 0)$ with $P(Y_2 - Y_0 > 0)$, or $\text{Cov}(Y_1 - Y_0, Y_0)$ with $\text{Cov}(Y_2 - Y_0, Y_0)$.

Let $\theta_{d,x} = E[c(Y_1, Y_0) \mid X = x]$. The sharp identified set for $(\theta_{1,x}, \ldots, \theta_{J,x})$ is given by

$$[\theta^L_{1,x}, \theta^H_{1,x}] \times \ldots \times [\theta^L_{J,x}, \theta^H_{J,x}]$$

where $\theta^L_{d,x} = \theta^L(P_{d|x}, P_{0|x})$ and $\theta^H_{d,x} = \theta^H(P_{d|x}, P_{0|x})$ as in section 4.[3] The sharp identified set for θ_d

[3] This follows from existing results and the *gluing lemma*, found in Villani (2009) (pp. 11-12).

is $[\theta_d^L, \theta_d^H]$ where $\theta_d^L = \sum_x s_x \theta_{d,x}^L$ and $\theta_d^H = \sum_x s_x \theta_{d,x}^H$, and the sharp identified set for $(\gamma_1, \ldots \gamma_J)$ is

$$[\gamma_1^L, \gamma_1^H] \times \ldots \times [\gamma_J^L, \gamma_J^H]$$

Sample analogues $(\hat{\gamma}_1^L, \hat{\gamma}_1^H, \ldots, \hat{\gamma}_J^L, \hat{\gamma}_J^H)$ can be formed just as in section 5. Under natural adjustments to assumptions 2, 3, and 4, the same arguments work to show

$$\sqrt{n}((\hat{\gamma}_1^L, \hat{\gamma}_1^H, \ldots, \hat{\gamma}_J^L, \hat{\gamma}_J^H) - (\gamma_1^L, \gamma_1^H, \ldots, \gamma_J^L, \gamma_J^H))$$

is asymptotically Gaussian and the bootstrap consistently estimates its asymptotic distribution.

8 Conclusion

This paper studies a large class of causal parameters that depend on a moment of the joint distribution of potential outcomes. The sharp identified set of such parameters is characterized with optimal transport. Estimators based on this identification are $\sqrt{n}$-consistent and converge in distribution under mild assumptions, and inference procedures based on the bootstrap are straightforward and computationally convenient.

References

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of econometrics*, *113*(2), 231–263.

Aliprantis, C. D., & Border, K. C. (2006). *Infinite dimensional analysis: A Hitchhiker's Guide*. Springer.

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, *110*(3), 629–676.

Callaway, B. (2021). Bounds on distributional treatment effect parameters using panel data with an application on job displacement. *Journal of Econometrics*, *222*(2), 861–881.

Couch, K. (1992). Long-term effects of the national supported work experiment, and parametric and nonparametric tests of model specification and the estimation of treatment effects. *Unpublished*

Ph. D. dissertation, University of Wisconsin-Madison. 1992b." New Evidence on the Long-Term Effects of Employment Training Programs." Journal of Labor Economics, *10*(4), 380–88.

Crauel, H. (2002). *Random probability measures on Polish spaces*, vol. 11. CRC press.

Diamond, A., & Sekhon, J. S. (2013). Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, *95*(3), 932–945.

Dunipace, E. (2021). Optimal transport weights for causal inference. *arXiv preprint arXiv:2109.01991*.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife," the annals of statistics, 7, 1–26. freedman, da (1981). *Bootstrapping Regression Models," The Annals of Statistics*, *9*, 1218–1228.

Ehm, W., Gneiting, T., Jordan, A., & Krüger, F. (2016). Of quantiles and expectiles: consistent scoring functions, choquet representations and forecast rankings. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *78*(3), 505–562.

Fan, Y., & Park, S. S. (2010). Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, *26*(3), 931–951.

Fan, Y., & Park, S. S. (2012). Confidence intervals for the quantile of treatment effects in randomized experiments. *Journal of Econometrics*, *167*(2), 330–344.

Fan, Y., Shi, X., & Tao, J. (2023). Partial identification and inference in moment models with incomplete data. *Journal of Econometrics*, *235*(2), 418–443.

Fang, Z., & Santos, A. (2019). Inference on directionally differentiable functions. *The Review of Economic Studies*, *86*(1), 377–412.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica*, *75*(1), 259–276.

Firpo, S., & Ridder, G. (2019). Partial identification of the treatment effect distribution and its functionals. *Journal of Econometrics*, *213*(1), 210–234.

Frandsen, B. R., & Lefgren, L. J. (2021). Partial identification of the distribution of treatment effects with an application to the knowledge is power program (kipp). *Quantitative Economics*, *12*(1), 143–171.

Friebel, G., Heinz, M., Hoffman, M., & Zubanov, N. (2023). What do employee referral programs do? measuring the direct and overall effects of a management practice. *Journal of Political Economy*, *131*(3), 633–686.

Galichon, A. (2017). A survey of some recent applications of optimal transport methods to econometrics. *The Econometrics Journal*, *20*(2), C1–C11.

Gunsilius, F., & Xu, Y. (2021). Matching for causal effects via multimarginal unbalanced optimal transport. *arXiv preprint arXiv:2112.04398*.

Heckman, J. J., Smith, J., & Clements, N. (1997). Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies*, *64*(4), 487–535.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, *81*(396), 945–960.

Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*(2), 467–475.
URL http://www.jstor.org/stable/2951620

Imbens, G. W., & Manski, C. F. (2004). Confidence intervals for partially identified parameters. *Econometrica*, *72*(6), 1845–1857.

Ji, W., Lei, L., & Spector, A. (2023). Model-agnostic covariate-assisted inference on partially identified causal effects. *arXiv preprint arXiv:2310.08115*.

Kitagawa, T. (2015). A test for instrument validity. *Econometrica*, *83*(5), 2043–2063.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American economic review*, (pp. 604–620).

Manski, C. F. (1997). Monotone treatment response. *Econometrica: Journal of the Econometric Society*, (pp. 1311–1334).

Pollard, D. (2002). *A user's guide to measure theoretic probability*. 8. Cambridge University Press.

Rubin, D. B. (1981). The bayesian bootstrap. *The annals of statistics*, (pp. 130–134).

Russell, T. M. (2021). Sharp bounds on functionals of the joint distribution in the analysis of treatment effects. *Journal of Business & Economic Statistics*, *39*(2), 532–546.

Santambrogio, F. (2015). Optimal transport for applied mathematicians. *Birkäuser, NY*, *55*(58-63), 94.

Staudt, T., Hundrieser, S., & Munk, A. (2022). On the uniqueness of kantorovich potentials. *arXiv preprint arXiv:2201.08316*.

Torous, W., Gunsilius, F., & Rigollet, P. (2021). An optimal transport approach to causal inference. *arXiv preprint arXiv:2108.05858*.

van der Vaart, A., & Wellner, J. A. (1997). *Weak convergence and empirical processes with applications to statistics*. London: Royal Statistical Society, 1988-.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, vol. 3. Cambridge university press.

Villani, C. (2003). *Topics in optimal transportation*, vol. 58. American Mathematical Soc.

Villani, C. (2009). *Optimal transport: old and new*, vol. 338. Springer.

A Appendix: identification

Following Kitagawa (2015), let T denote the "type" of a unit:

$$T = \begin{cases} a, \text{ always-taker}, & \text{if } (D_1, D_0) = (1,1) \\ c, \text{ complier}, & \text{if } (D_1, D_0) = (1,0) \\ n, \text{ never-taker}, & \text{if } (D_1, D_0) = (0,0) \\ df, \text{ defier}, & \text{if } (D_1, D_0) = (0,1) \end{cases} \tag{41}$$

Note that the primitives $(Y_1, Y_0, D_1, D_0, Z, X)$ are equivalent to (Y_1, Y_0, T, Z, X).

Lemma A.1 (Identification of moments)**.** *Suppose assumptions 1 and 2 hold. Then the sharp identified set for θ is $[\theta^L, \theta^H]$.*

Proof. Let T be as defined in (41), and note that the primitives of the model $(Y_1, Y_0, D_1, D_0, Z, X)$ are equivalent to (Y_1, Y_0, T, Z, X). Moreover, the event $D_1 > D_0$ is the event $T = c$; thus $P_{d|x}$ is the distribution of $Y_d \mid T = c, X = x$.

In steps:

1. The identified set for $(P_{1,0|x_1}, \ldots, P_{1,0|x_M})$, the conditional distributions of $(Y_1, Y_0) \mid T = c, X = x$ for each $x \in \mathcal{X} = \{x_1, \ldots, x_M\}$, is $\Pi(P_{1|x_1}, P_{0|x_1}) \times \ldots \times \Pi(P_{1|x_M}, P_{0|x_M})$.

   That $(P_{1,0|x_1}, \ldots, P_{1,0|x_M}) \in \Pi(P_{1|x_1}, P_{0|x_1}) \times \ldots \times \Pi(P_{1|x_M}, P_{0|x_M})$ is immediate. To see that any element of $\Pi(P_{1|x_1}, P_{0|x_1}) \times \ldots \times \Pi(P_{1|x_M}, P_{0|x_M})$ is possible given the assumptions and distribution of the observables (Y, D, Z, X), fix a distribution of the observables generated by a distribution of the primitives consistent with the assumptions. Note that the distribution of observables is summarized by $P(D = d, Z = z, X = x)$ for each (d, z, x) and the conditional distributions

   $$Y \mid D = d, Z = z, X = x$$

   Use this observation and the claims of lemma A.4 to see that any two distributions of the primitives (Y_1, Y_0, T, Z, X) (consistent with the assumptions), sharing the same distribution of (T, Z, X), and the same marginal, conditional distributions for

   $$\begin{aligned} &Y_1 \mid T = a, X = x && Y_0 \mid T = n, X = x \\ &Y_1 \mid T = c, X = x, && Y_0 \mid T = c, X = x \end{aligned}$$

   will produce this distribution of observables. Thus, replacing $(P_{1,0|x_1}, \ldots, P_{1,0|x_M})$ from the distribution of primitives with any

   $$(\pi_{x_1}, \ldots, \pi_{x_M}) \in \Pi(P_{1|x_1}, P_{0|x_1}) \times \ldots \times \Pi(P_{1|x_M}, P_{0|x_M})$$

   will generate the same observed distribution of (Y, D, Z, X), without violating assumption 1 or 2. The claim follows.

2. The identified set for $(\theta_{x_1}, \ldots, \theta_{x_M}) \in \mathbb{R}^M$ is $[\theta^L_{x_1}, \theta^H_{x_1}] \times \ldots \times [\theta^L_{x_M}, \theta^H_{x_M}]$.

   Recall that $\theta_x = E[c(Y_1, Y_0) \mid X = x]$, and let $\Theta_{I,x}$ denote its identified set. Note that the previous step implies

$$\Theta_{I,x} = \left\{ t \in \mathbb{R} \; ; \; t = E_{\pi_x}[c(Y_1, Y_0)] \text{ for some } \pi_x \in \Pi(P_{1|x}, P_{0|x}) \right\}$$

   $\Pi(P_{1|x}, P_{0|x})$ is convex. Notice that for any $\lambda \in (0,1)$ and $\pi^1_x, \pi^0_x \in \Pi(P_{1|x}, P_{0|x})$, $E_{\lambda \pi^1_x + (1-\lambda)\pi^0_x}[c(Y_1, Y_0)] = \lambda E_{\pi^1_x}[c(Y_1, Y_0)] + (1-\lambda) E_{\pi^0_x}[c(Y_1, Y_0)]$. Together these imply $\Theta_{I,x}$ is convex.

   It suffices to show that for any x, $\Theta_{I,x} = [\theta^L_x, \theta^H_x]$ There are two cases:

   (i) If assumption 2 (i) holds, then for each x,

$$\begin{aligned}
\theta^L_x &= OT_c(P_{1|x}, P_{0|x}) = \inf_{\pi_x \in \Pi(P_{1|x}, P_{0|x})} E_{\pi_x}[c(Y_1, Y_0)] \\
\theta^H_x &= -OT_{-c}(P_{1|x}, P_{0|x}) = \sup_{\pi_x \in \Pi(P_{1|x}, P_{0|x})} E_{\pi_x}[c(Y_1, Y_0)]
\end{aligned}$$

   Since c is continuous, lemma E.1 implies the optimal transport problems are attained, say by π^L_x and π^H_x respectively. It follows that $\theta^L_x, \theta^H_x \in \Theta_{I,x}$, and it is clear from their definitions that they bound $\Theta_{I,x}$. Since $\Theta_{I,x}$ is convex, it follows that $\Theta_{I,x} = [\theta^L_x, \theta^H_x]$.

   (ii) If Assumption 2 (ii) holds, then

$$\begin{aligned}
c_L(y_1, y_0) &= \mathbb{1}\{y_1 - y_0 < \delta\}, & c_H(y_1, y_0) &= \mathbb{1}\{y_1 - y_0 > \delta\}, \\
\theta^L_x &= OT_{c_L}(P_{1|x}, P_{0|x}), & \theta^H_x &= 1 - OT_{c_H}(P_{1|x}, P_{0|x})
\end{aligned}$$

   Let $\pi^L_x, \pi^H_x \in \Pi(P_{1|x}, P_{0|x})$ be such that $\theta^L_x = E_{\pi^L_x}[\mathbb{1}\{Y_1 - Y_0 < \delta\}] = P_{\pi^L_x}(Y_1 - Y_0 < \delta)$ and $\theta^H_x = 1 - E_{\pi^H_x}[\mathbb{1}\{Y_1 - Y_0 > \delta\}] = P_{\pi^H_x}(Y_1 - Y_0 \leq \delta)$. Notice that $\theta^H_x \in \Theta_{I,x}$. Furthermore, $\mathbb{1}\{y_1 - y_0 < \delta\} \leq \mathbb{1}\{y_1 - y_0 \leq \delta\}$ implies

$$\theta^L_x = \inf_{\pi_x \in \Pi(P_{1|x}, P_{0|x})} E_{\pi_x}[\mathbb{1}\{Y_1 - Y_0 < \delta\}] \leq \inf_{\pi_x \in \Pi(P_{1|x}, P_{0|x})} E_{\pi_x}[\mathbb{1}\{Y_1 - Y_0 \leq \delta\}]$$

   and thus θ^L_x is a lower bound for $\Theta_{I,x}$. Since $\Theta_{I,x}$ is convex, it suffices to show that $\theta^L_x \in \Theta_{I,x}$.

   Corollary E.15 implies that $\theta^L_x = P_{\pi^L_x}(Y_1 - Y_0 < \delta) = \sup_y \left\{ F_{1|x}(y) - F_{0|x}(y - \delta) \right\}$. Moreover, Villani (2009) theorem 5.10 part (iii) implies the dual problem $\sup_y \left\{ F_{1|x}(y) - F_{0|x}(y - \delta) \right\}$ is attained as well, say by y^*. Thus

$$\int \mathbb{1}\{y_1 - y_0 \leq \delta\} d\pi^L_x(y_1, y_0) = \int \mathbb{1}\{y_1 \leq y^*\} dP_{1|x}(y_1) - \int \mathbb{1}\{y_0 \leq y^* - \delta\} dP_{0|x}(y_0) \quad (42)$$

   Next, notice that

$$\mathbb{1}\{y_1 \leq y^*\} - \mathbb{1}\{y_0 \leq y^* - \delta\} \leq \mathbb{1}\{y_1 - y_0 < \delta\} \quad (43)$$

   which holds for all (y_1, y_0), must hold with equality π^L_x-almost surely. Indeed, let N be the set where the inequality in (43) is strict and suppose N is π^L_x-non-negligible. Since

$\pi_x^L \in \Pi(P_{1|x}, P_{0|x})$,

$$\begin{aligned}
&\int \mathbb{1}\{y_1 \leq y^*\} dP_{1|x}(y_1) - \int \mathbb{1}\{y_0 \leq y^* - \delta\} dP_{0|x}(y_0) = \int \mathbb{1}\{y_1 \leq y^*\} - \mathbb{1}\{y_0 \leq y^* - \delta\} d\pi_x^L(y_1, y_0) \\
&= \int_N \mathbb{1}\{y_1 \leq y^*\} - \mathbb{1}\{y_0 \leq y^* - \delta\} d\pi_x^L(y_1, y_0) + \int_{N^c} \mathbb{1}\{y_1 \leq y^*\} - \mathbb{1}\{y_0 \leq y^* - \delta\} d\pi_x^L(y_1, y_0) \\
&< \int_N \mathbb{1}\{y_1 - y_0 < \delta\} d\pi_x^L(y_1, y_0) + \int_{N^c} \mathbb{1}\{y_1 - y_0 < \delta\} d\pi_x^L(y_1, y_0) \\
&= \int \mathbb{1}\{y_1 - y_0 \leq \delta\} d\pi_x^L(y_1, y_0)
\end{aligned}$$

contradicts (42). This implies that π_x^L concentrates on

$$\underbrace{\{(y_1, y_0) \; ; \; y_1 \leq y^*, y_0 > y^* - \delta, y_1 - y_0 < \delta\}}_{\text{both sides of (43) equal 1}} \cup \underbrace{\{(y_1, y_0) \; ; \; y_1 > y^*, y_0 > y^* - \delta, y_1 - y_0 \geq \delta\}}_{\text{both sides of (43) equal 0}}$$
$$\cup \underbrace{\{(y_1, y_0) \; ; \; y_1 \leq y^*, y_0 \leq y^* - \delta, y_1 - y_0 \geq \delta\}}_{\text{both sides of (43) equal 0}}$$

Notice the only point in the set $\{(y_1, y_0) \; ; \; y_1 - y_0 = \delta\}$ where π_x^L could put positive mass is the point $(y_1, y_0) = (y^*, y^* - \delta)$. But since $P_{1|x}$ has a continuous CDF,

$$0 \leq \pi_x^L(\{(y^*, y^* - \delta)\}) \leq \pi_x^L(\{y^*\} \times \mathcal{Y}_0) = P_{1|x}(\{y^*\}) = 0$$

Thus $P_{\pi_x^L}(Y_1 - Y_0 = \delta) = 0$, and so $P_{\pi_x^L}(Y_1 - Y_0 \leq \delta) = P_{\pi_x^L}(Y_1 - Y_0 < \delta) = \theta^L(x)$. Thus $\theta_x^L \in \Theta_{I,x}$, and hence $\Theta_{I,x} = [\theta^L(x), \theta^H(x)]$.

Therefore the identified set for θ_x is $[\theta_x^L, \theta_x^H]$. It follows from this and step one above that the identified set $(\theta_{x_1}, \ldots, \theta_{x_M})$ is $[\theta_{x_1}^L, \theta_{x_1}^H] \times \ldots \times [\theta_{x_M}^L, \theta_{x_M}^H]$.

3. Recall that $\theta = E[c(Y_1, Y_0)] = E[E[c(Y_1, Y_0) \mid X]] = \sum_x s_x \theta_x$. Since $s_x = P(X = x \mid T = c)$ is point identified for each x, it follows from step two above that the identified set for θ is $[\theta^L, \theta^H]$ where

$$\theta^L = \sum_x s_x \theta_x^L, \qquad \theta^H = \sum_x s_x \theta_x^H$$

This concludes the proof. $\square$

Theorem 4.1 (Identification of functions of moments). *Suppose assumptions 1, 2, and 3 are satisfied. Then the sharp identified set for γ is $[\gamma^L, \gamma^H]$.*

Proof. Lemma A.1 shows that under assumptions 1 and 2, the sharp identified set for θ is $[\theta^L, \theta^H]$. Let Γ_I be the identified set for γ, and note that

$$\Gamma_I = \{\gamma \in \mathbb{R} \; ; \; \gamma = g(t, \eta) \text{ for some } t \in [\theta^L, \theta^H]\}$$

Assumption 2 implies c is bounded; under assumption 2 (i) the continuous $c : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ takes a maximum and minimum on the compact set $\mathcal{Y} \times \mathcal{Y}$, while under assumption 2 (ii) the cost function only takes values 0 or 1. It follows that θ^L and θ^H are finite and thus $[\theta^L, \theta^H]$ is compact.

Assumption 3 (ii) is that $g(\cdot, \eta)$ is continuous, and thus the extreme value theorem implies $\gamma^L = \inf_{t \in [\theta^L, \theta^H]} g(t, \eta)$ and $\gamma^H = \sup_{t \in [\theta^L, \theta^H]} g(t, \eta)$ are both elements of Γ_I. The intermediate value theorem then implies $\Gamma_I = [\gamma^L, \gamma^H]$. $\square$

Lemma 7.1 (Identification of q_τ)**.** *Suppose assumptions 1 and 2 (ii) hold. Then $q \in Q_{I,\tau}$ if and only if $\theta^L(q) \leq \tau \leq \theta^H(q)$.*

Proof. By definition, $q \in \Gamma_{I,\tau}$ if and only if there exists a distribution of the primitives, π, consistent with the observed distribution, such that $P_\pi(Y_1 - Y_0 \leq q) = \tau$. Lemma A.1 shows that $\theta^L(q) \leq \tau \leq \theta^H(q)$ if and only if there exists a distribution of the primitives, π, such that $\theta^L(q) \leq \tau \leq \theta^H(q)$. This concludees the proof. $\square$

Lemma A.2 (Identification: τ-th quantile)**.** *Let q_τ be defined as*

$$q_\tau = [\inf\{y \ ; \ P(Y_1 - Y_0 \leq y) \geq \tau\}, \inf\{y \ ; \ P(Y_1 - Y_0 \leq y) > \tau\}]$$

Suppose assumption 1 and 2 (ii) hold, and let $Q_{I,\tau}$ denote the identified set of q_τ defined above. Then $q \in Q_{I,\tau}$ if and only if $\theta^L(q) \leq \tau \leq \theta^H(q)$.

Proof. Suppose $\theta^L(q) \leq \tau \leq \theta^H(q)$. Lemma A.1 implies there exists a distribution π of the primitives consistent with assumption 2 (ii) such that $P_\pi(Y_1 - Y_0 \leq q) = \tau$. Thus $q \in [\inf\{y \ ; \ P_\pi(Y_1 - Y_0 \leq y) \geq \tau\}, \inf\{y \ ; \ P_\pi(Y_1 - Y_0 \leq y) > \tau\}]$ and hence $q \in Q_{I,\tau}$.

Before showing the other direction, we next show that assumption 2 (ii) implies $\theta^L(\delta)$ is continuous. Specifically, apply corollary E.15 to find $\theta^L_x(\delta) = \sup_y\{F_{1|x}(y) - F_{0|x}(y - \delta)\}$. So for any δ, δ',

$$\begin{aligned}
\theta^L_x(\delta) - \theta^L_x(\delta') &= \sup_y\{F_{1|x}(y) - F_{0|x}(y - \delta)\} - \sup_y\{F_{1|x}(y) - F_{0|x}(y - \delta')\} \\
&\leq \sup_y \left\{F_{0|x}(y - \delta') - F_{0|x}(y - \delta)\right\} \\
&\leq \sup_y \left|F_{0|x}(y - \delta') - F_{0|x}(y - \delta)\right|
\end{aligned}$$

and thus $|\theta^L_x(\delta) - \theta^L_x(\delta')| \leq \sup_y \left|F_{0|x}(y - \delta') - F_{0|x}(y - \delta)\right|$. Recall that any continuous CDF is in fact uniformly continuous, and $F_{0|x}$ is in fact uniformly continuous. Let $\varepsilon > 0$, choose $\eta > 0$ such that for any $y, y' \in \mathbb{R}$ with $|y - y'| < \eta$, one has $|F_{0|x}(y) - F_{0|x}(y')| < \varepsilon/2$, and notice that

$$|\delta - \delta'| < \eta \implies \sup_y \left|F_{0|x}(y - \delta') - F_{0|x}(y - \delta)\right| \leq \varepsilon/2 < \varepsilon$$

This shows $\theta^L_x(\delta)$ is continuous, and so $\theta^L(\delta) = \sum_x s_x \theta^L_x$ is continuous.

Return to showing the other direction, through the contrapositive. Suppose it is not the case that $\theta^L(q) \leq \tau \leq \theta^H(q)$. There are two possibilities:

1. Suppose $\theta^H(q) < \tau$. Then there is no distribution π of the primitives such that $P_\pi(Y_1 - Y_0 \leq q) \geq \tau$, hence there is no distribution where $q \in [\inf\{y \ ; \ P(Y_1 - Y_0 \leq y) \geq \tau\}, \inf\{y \ ; \ P(Y_1 - Y_0 \leq y) > \tau\}]$ and thus $q \notin Q_{I,\tau}$.

2. Suppose $\tau < \theta^L(q)$. If one further supposes that $q \in Q_{I,\tau}$, then $\theta^L(\cdot)$ would have a jump discontinuity at q, contradicting the continuity shown above.

   Specifically, if $\tau < \theta^L(q)$ and $q \in Q_{I,\tau}$, then there exists a distribution π of the primitives such that $P_\pi(Y_1 - Y_0 \leq q) > \tau$ and $q \in [\inf\{y \ ; \ P_\pi(Y_1 - Y_0 \leq y) \geq \tau\}, \inf\{y \ ; \ P_\pi(Y_1 - Y_0 \leq y) > \tau\}]$, implying that $P_\pi(Y_1 - Y_0 \leq \cdot)$ jumps at q from below τ to above $\theta^L(q)$:

$$\lim_{\epsilon \to 0} P_\pi(Y_1 - Y_0 \leq q - \epsilon) < \tau < \theta^L(q) \leq P_\pi(Y_1 - Y_0 \leq q)$$

   This jump discontinuity at q is at least of size $\varepsilon = \theta^L(q) - \tau > 0$. But then $\theta^L(\cdot)$ would have a jump discontinuity of at least size ε at q as well, a contradiction of the continuity of $\theta^L(\cdot)$ shown above.

   Thus if $\tau < \theta^L(q)$, then $q \notin Q_{I,\tau}$.

In either case, $q \notin Q_{I,\tau}$. This completes the proof. □

A.1 Additional identification lemmas

The lemmas below contain results well known in the literature. They are included here with proofs for completeness.

Lemma A.3. *Let P_1 be any distribution and P_0 be degenerate at $\tilde{y}_0 \in \mathbb{R}$. Then the only possible coupling of P_1 and P_0 is characterized by the cumulative distribution function*

$$P(Y_1 \leq y_1, Y_0 \leq y_0) = \begin{cases} P(Y_1 \leq y_1) & \text{if } y_0 \geq \tilde{y}_0 \\ 0 & \text{if } y_0 < \tilde{y}_0 \end{cases}$$

Proof. First suppose $y_0 < \tilde{y}_0$. Then $0 \leq P(Y_1 \leq y_1, Y_0 \leq y_0) \leq P(Y_0 \leq y_0) = 0$.

Next suppose $y_0 \geq \tilde{y}_0$. Then $1 \geq P(\{Y_1 \leq y_1\} \cup \{Y_0 \leq y_0\}) \geq P(Y_0 \leq y_0) = 1$ implies that

$$\begin{aligned} P(Y_1 \leq y_1, Y_0 \leq y_0) &= P(Y_1 \leq y_1) + \underbrace{P(Y_0 \leq y_0)}_{=1} - \underbrace{P(\{Y_1 \leq y_1\} \cup \{Y_0 \leq y_0\})}_{=1} \\ &= P(Y_1 \leq y_1) \end{aligned}$$

which completes the proof. □

Lemma A.4 below summarizes the empirical content of the model described in assumption 1. In particular, it implies that any two distributions of the primitives consistent with assumption 1 that share the same marginal distribution of (T, Z, X) and marginal, conditional distributions of

$$\begin{aligned} &Y_1 \mid T = a, X = x && Y_0 \mid T = n, X = x \\ &Y_1 \mid T = c, X = x, && Y_0 \mid T = c, X = x \end{aligned}$$

will produce the same distribution of observables.

Lemma A.4. *Suppose assumpion 1 holds. Then*

$$\begin{aligned} P(D = 1 \mid Z = 0, X = x) &= P(T = a \mid X = x) \\ P(D = 0 \mid Z = 1, X = x) &= P(T = n \mid X = x) \\ P(D = 1 \mid Z = 1, X = x) &= P(T \in \{a, c\} \mid X = x) \\ P(D = 0 \mid Z = 0, X = x) &= P(T \in \{c, n\} \mid X = x) \end{aligned}$$

and for any integrable function f,

$$\begin{aligned} E[f(Y) \mid D = 1, Z = 1, X = x] &= E[f(Y_1) \mid T \in \{a, c\}, X = x] \\ E[f(Y) \mid D = 0, Z = 0, X = x] &= E[f(Y_0) \mid T \in \{c, n\}, X = x] \end{aligned}$$

Furthermore,

$$\begin{aligned} &\text{if } P(D = 1 \mid Z = 0, X = x) > 0, \text{ then } E[f(Y) \mid D = 1, Z = 0, X = x] = E[f(Y_1) \mid T = a, X = x] \\ &\text{if } P(D = 0 \mid Z = 1, X = x) > 0, \text{ then } E[f(Y) \mid D = 0, Z = 1, X = x] = E[f(Y_0) \mid T = n, X = x] \end{aligned}$$

Proof. Assumption 1 (ii) implies $\mathbb{1}\{D_1 = 0, D_0 = 1\} = 0$. The definition of T in (41) then implies

$$\begin{aligned} \mathbb{1}\{D_0 = 1\} &= \mathbb{1}\{D_1 = 1, D_0 = 1\} + \cancel{\mathbb{1}\{D_1 = 0, D_0 = 1\}} = \mathbb{1}\{T = a\} \\ \mathbb{1}\{D_1 = 0\} &= \mathbb{1}\{D_1 = 0, D_0 = 0\} + \cancel{\mathbb{1}\{D_1 = 0, D_0 = 1\}} = \mathbb{1}\{T = n\} \\ \mathbb{1}\{D_1 = 1\} &= \mathbb{1}\{D_1 = 1, D_0 = 1\} + \mathbb{1}\{D_1 = 1, D_0 = 0\} = \mathbb{1}\{T \in \{a, c\}\} \\ \mathbb{1}\{D_0 = 0\} &= \mathbb{1}\{D_1 = 1, D_0 = 0\} + \mathbb{1}\{D_1 = 0, D_0 = 0\} = \mathbb{1}\{T \in \{c, n\}\} \end{aligned}$$

These observations, equation (2), and assumption 1 (i) imply

$$\begin{aligned} P(D = 1 \mid Z = 0, X = x) &= P(D_0 = 1 \mid X = x) = P(T = a \mid X = x), \\ P(D = 0 \mid Z = 1, X = x) &= P(D_1 = 0 \mid X = x) = P(T = n \mid X = x), \\ P(D = 1 \mid Z = 1, X = x) &= P(D_1 = 1 \mid X = x) = P(T \in \{a, c\} \mid X = x), \text{ and} \\ P(D = 0 \mid Z = 0, X = x) &= P(D_0 = 0 \mid X = x) = P(T \in \{c, n\} \mid X = x) \end{aligned}$$

Note the first two equalities can be summarized as $P(D = d \mid Z = z, X = x) = P(D_z = d \mid X = x)$.

Next, let $f : \mathbb{R} \to \mathbb{R}$ be integrable. Assumption 1 (i) and equations (1) and (2) imply that for any (d, z, x),

$$\begin{aligned} P(D = d \mid Z = z, X = x)&E[f(Y) \mid D = d, Z = z, X = x] \\ &= P(D_z = d \mid X = x)E[f(Y_d) \mid D_z = d, X = x] \end{aligned}$$

and since $P(D = d \mid Z = z, X = x) = P(D_z = d \mid X = x)$, this implies

$$0 = P(D = d \mid Z = z, X = x)\Big(E[f(Y) \mid D = d, Z = z, X = x] - E[f(Y_d) \mid D_z = d, X = x]\Big) \quad (44)$$

Assumption 1 (iii) implies

$$\begin{aligned}
P(D=1 \mid Z=1, X=x) &= P(T \in \{a,c\} \mid X=x) \geq P(T=c \mid X=x) > 0 \\
P(D=0 \mid Z=0, X=x) &= P(T \in \{c,n\} \mid X=x) \geq P(T=c \mid X=x) > 0
\end{aligned}$$

Use strict positivity of $P(D=1 \mid Z=1, X=x)$ and $P(D=0 \mid Z=0, X=x)$ to see that

$$\begin{aligned}
E[f(Y) \mid D=1, Z=1, X=x] &= E[f(Y_1) \mid D_1=1, X=x] = E[f(Y_1) \mid T \in \{a,c\}, X=x] \\
E[f(Y) \mid D=0, Z=0, X=x] &= E[f(Y_0) \mid D_0=0, X=x] = E[f(Y_0) \mid T \in \{c,n\}, X=x]
\end{aligned}$$

Similarly, (44) implies

$$\begin{aligned}
\text{if } P(D=1 \mid Z=0, X=x) > 0, \text{ then } E[f(Y) \mid D=1, Z=0, X=x] &= E[f(Y_1) \mid T=a, X=x] \\
\text{if } P(D=0 \mid Z=1, X=x) > 0, \text{ then } E[f(Y) \mid D=0, Z=1, X=x] &= E[f(Y_0) \mid T=n, X=x]
\end{aligned}$$

this concludes the proof. □

Lemma 2.1 (Abadie (2003))**.** *Suppose assumption 1 holds. Then the marginal distributions of Y_d conditional on $D_1 > D_0$ and $X=x$, denoted $P_{d|x}$, are identified by*

$$\begin{aligned}
E_{P_{d|x}}[f(Y_d)] &\equiv E[f(Y_d) \mid D_1 > D_0, X=x] \\
&= \frac{E[f(Y)\mathbb{1}\{D=d\} \mid Z=d, X=x] - E[f(Y)\mathbb{1}\{D=d\} \mid Z=1-d, X=x]}{P(D=d \mid Z=d, X=x) - P(D=d \mid Z=1-d, X=x)} \qquad (4)
\end{aligned}$$

for any integrable function f. Furthermore, the distribution of X conditional on $D_1 > D_0$ is identified by

$$\begin{aligned}
s_x &\equiv P(X=x \mid D_1 > D_0) \\
&= \frac{[P(D=1 \mid Z=1, X=x) - P(D=1 \mid Z=0, X=x)]\, P(X=x)}{\sum_{x'} [P(D=1 \mid Z=1, X=x') - P(D=1 \mid Z=0, X=x')]\, P(X=x')} \qquad (5)
\end{aligned}$$

Proof. First notice that using T as defined in (41),

$$E[f(Y_d) \mid D_1 > D_0, X=x] = E[f(Y_d) \mid T=c, X=x] = \frac{E[f(Y_d)\mathbb{1}\{T=c\} \mid X=x]}{P(T=c \mid X=x)} \qquad (45)$$

Now notice that

$$D_1 - D_0 = (1-D_0) - (1-D_1) = \mathbb{1}\{D_d = d\} - \mathbb{1}\{D_{1-d} = d\}$$

for either $d \in \{1, 0\}$. Monotonicity (assumption 1 (ii)) implies that this is an indicator for $T=c$:

$$D_1 - D_0 = \mathbb{1}\{D_1 = 1, D_0 = 0\} = \mathbb{1}\{T=c\}$$

So,

$$
\begin{aligned}
&E[f(Y)\mathbb{1}\{D=d\} \mid Z=d, X=x] - E[f(Y)\mathbb{1}\{D=d\} \mid Z=1-d, X=x] \\
&\qquad = E[f(Y_d)\mathbb{1}\{D_d=d\} \mid X=x] - E[f(Y_d)\mathbb{1}\{D_{1-d}=d\} \mid X=x] \\
&\qquad = E[f(Y_d)(\mathbb{1}\{D_d=d\} - \mathbb{1}\{D_{1-d}=d\}) \mid X=x] \\
&\qquad = E[f(Y_d)\mathbb{1}\{T=c\} \mid X=x] \qquad (46)
\end{aligned}
$$

Lemma A.4 shows that

$$
\begin{aligned}
&P(D=1 \mid Z=1, X=x) - P(D=1 \mid Z=0, X=x) \\
&\qquad = P(T \in \{a,c\} \mid X=x) - P(T=a \mid X=x) = P(T=c \mid X=x)
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
&P(D=0 \mid Z=0, X=x) - P(D=0 \mid Z=1, X=x) \\
&\qquad = P(T \in \{c,n\} \mid X=x) - P(T=n \mid X=x) = P(T=c \mid X=x)
\end{aligned}
$$

Thus for either $d \in \{1,0\}$,

$$
P(D=d \mid Z=d, X=x) - P(D=d \mid Z=1-d, X=x) = P(T=c \mid X=x). \qquad (47)
$$

It follows from (45), (46), and (47) that

$$
\begin{aligned}
E_{P_{d|x}}[f(Y_d)] &= E[f(Y_d) \mid D_1 > D_0, X=x] \\
&= \frac{E[f(Y)\mathbb{1}\{D=d\} \mid X=x, Z=d] - E[f(Y)\mathbb{1}\{D=d\} \mid X=x, Z=1-d]}{P(D=d \mid X=x, Z=d) - P(D=d \mid X=x, Z=1-d)},
\end{aligned}
$$

and from (47) that

$$
\begin{aligned}
s_x &= P(X=x \mid D_1 > D_0) = P(X=x \mid T=c) = \frac{P(T=c \mid X=x)P(X=x)}{\sum_{x'} P(T=c \mid X=x')P(X=x')} \\
&= \frac{[P(D=1 \mid X=x, Z=1) - P(D=1 \mid X=x, Z=0)]\, P(X=x)}{\sum_{x'} [P(D=1 \mid X=x', Z=1) - P(D=1 \mid X=x', Z=0)]\, P(X=x')}.
\end{aligned}
$$

This concludes the proof. □

B Appendix: properties of optimal transport

Suppose that strong duality holds:

$$\inf_{\pi\in\Pi(P_1,P_0)}\int c(y_1,y_0)d\pi(y_1,y_0)=\sup_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)}\int\varphi(y_1)dP_1(y_1)+\int\psi(y_0)dP_0(y_0) \tag{48}$$

for sets of universally bounded functions $\mathcal{F}_c\subseteq L^1(P_1)$ and $\mathcal{F}_c^c\subseteq L^1(P_0)$. See lemmas E.9 and E.13 for examples.[4] Then for suitable sets $\mathcal{F}_1$ and $\mathcal{F}_0$ with $\mathcal{F}_c\subseteq\mathcal{F}_1$ and $\mathcal{F}_c^c\subseteq\mathcal{F}_0$, the map $OT_c(P_1,P_0)=\inf_{\pi\in\Pi(P_1,P_0)}\int c(y_1,y_0)d\pi(y_1,y_0)$ can be viewed as

$$OT_c:\ell^\infty(\mathcal{F}_1)\times\ell^\infty(\mathcal{F}_0)\to\mathbb{R},\qquad OT_c(P_1,P_0)=\sup_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)}P_1(\varphi)+P_0(\psi) \tag{49}$$

where $P_d(f)=\int f(y_d)dP_d(y_d)=E_{P_d}[f(Y_d)]$.

The functional in (49) is defined over the familiar Banach space $\ell^\infty(\mathcal{F}_1)\times\ell^\infty(\mathcal{F}_0)$. This makes it straightforward to show that optimal transport, as a functional from this space to $\mathbb{R}$, has certain desirable properties.

B.1 Continuity

Lemma B.1 (Optimal transport is uniformly continuous)**.** *Suppose that for some universally bounded $\mathcal{F}_c\subseteq L^1(P_1)$ and $\mathcal{F}_c^c\subseteq L^1(P_0)$,* (48) *holds. Then the optimal transport functional, given by* (49)*, is uniformly continuous.*

Proof. Define

$$\begin{aligned}
&\mathcal{S}:\ell^\infty(\mathcal{F}_1)\times\ell^\infty(\mathcal{F}_0)\to\ell^\infty(\mathcal{F}_1\times\mathcal{F}_0), && \mathcal{S}(H_1,H_0)(\varphi,\psi)=H_1(\varphi)+H_0(\psi)\\
&\Xi_c:\ell^\infty(\mathcal{F}_1\times\mathcal{F}_0)\to\mathbb{R}, && \Xi_c[G]=\sup_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)}G(\varphi,\psi)
\end{aligned}$$

[4] $\mathcal{F}_c$ and $\mathcal{F}_c^c$ are typically found with the following steps:

(i) Start with a known strong duality result; for some $\Phi_{cs}\subseteq\Phi_c$,

$$\inf_{\pi\in\Pi(P_1,P_0)}\int c(y_1,y_0)d\pi(y_1,y_0)=\sup_{(\varphi,\psi)\in\Phi_{cs}}\int\varphi(y_1)dP_1(y_1)+\int\psi(y_0)dP_0(y_0)$$

(ii) Compute $\mathcal{F}_c(\Phi_{cs})$ and $\mathcal{F}_c^c(\Phi_{cs})$ defined by (84).

(iii) Notice that $\mathcal{F}_c(\Phi_{cs})\subseteq\mathcal{F}_c$ and $\mathcal{F}_c^c(\Phi_{cs})\subseteq\mathcal{F}_c^c$ for known and easy to study sets $\mathcal{F}_c$, $\mathcal{F}_c^c$

Lemma E.7 and remark E.2 are useful to ensure $\mathcal{F}_c$ and $\mathcal{F}_c^c$ are universally bounded.

and notice that $OT_c(H_1, H_0) = \Xi_c(\mathcal{S}(H_1, H_0))$. Since $s : \mathbb{R}^2 \to \mathbb{R}$ given by $s(h_1, h_2) = h_1 + h_2$ is uniformly continuous, we have that $\mathcal{S}$ is uniformly continuous (see lemma F.1). Lemma F.3 shows that Ξ_c is uniformly continuous. The composition of uniformly continuous functions is uniformly continuous, implying OT_c is uniformly continuous. This completes the proof. $\square$

B.2 Directional Differentiability

The optimal transport functional given by (49) is Hadamard directionally differentiable.[5] The formal result, stated below, requires that $\mathcal{F}_c$ and $\mathcal{F}_c^c$ each be equipped with a semimetric. The semimetrics chosen must be such that $P_1 \in \ell^\infty(\mathcal{F}_c)$ and $P_0 \in \ell^\infty(\mathcal{F}_c^c)$ are continuous and the product space $\mathcal{F}_c \times \mathcal{F}_c^c$ and its subset $\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$ are compact.

The setting suggests a very convenient semimetric. Let P be the distribution of an observation, i.e. $(Y, D, Z, X) \sim P$. Note that under assumption 1, the distributions $P_{d|x}$ are dominated by P with bounded densities $\frac{dP_{d|x}}{dP}$. Specifically, recall that

$$\begin{aligned} E_{P_{d|x}}[f(Y_d)] &= E[f(Y_d) \mid D_1 > D_0, X = x] \\ &= \frac{E[f(Y)\mathbb{1}\{D = d\} \mid Z = d, X = x] - E[f(Y)\mathbb{1}\{D = d\} \mid Z = 1-d, X = x]}{P(D = d \mid Z = d, X = x) - P(D = d \mid Z = 1-d, X = x)} \end{aligned}$$

Let $\mathbb{1}_{d,x,z}(D, X, Z) = \mathbb{1}\{D = d, X = x, Z = z\}$, $p_{d,x,z} = P(D = d, X = x, Z = z)$, and $p_{x,z} = P(X = x, Z = z)$. Observe that

$$\begin{aligned} E[f(Y_d) \mid D_1 > D_0, X = x] &= E\left[f(Y)\frac{\mathbb{1}_{d,x,d}(D, X, Z)/p_{x,d} - \mathbb{1}_{d,x,1-d}(D, X, Z)/p_{x,1-d}}{p_{d,x,d}/p_{x,d} - p_{d,x,1-d}/p_{x,1-d}}\right] \\ &= E\left[f(Y)E\left[\frac{\mathbb{1}_{d,x,d}(D, X, Z)/p_{x,d} - \mathbb{1}_{d,x,1-d}(D, X, Z)/p_{x,1-d}}{p_{d,x,d}/p_{x,d} - p_{d,x,1-d}/p_{x,1-d}} \mid Y\right]\right] \end{aligned}$$

reveals the densities to be $\frac{dP_{d|x}}{dP}(Y) = E\left[\frac{\mathbb{1}_{d,x,d}(D,X,Z)/p_{x,d} - \mathbb{1}_{d,x,1-d}(D,X,Z)/p_{x,1-d}}{p_{d,x,d}/p_{x,d} - p_{d,x,1-d}/p_{x,1-d}} \mid Y\right]$.

We now drop the subscript x for the remainder of this appendix. Because P dominates both

[5] Recall the definition, found in Fang & Santos (2019): let $\mathbb{D}$, $\mathbb{E}$ be Banach spaces (complete, normed, vector spaces), and $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbb{E}$. ϕ is **Hadamard directionally differentiable** at $x_0 \in \mathbb{D}_\phi$ tangentially to $\mathbb{D}_T \subseteq \mathbb{D}$ if there exists a continuous map $\phi'_{x_0} : \mathbb{D}_T \to \mathbb{E}$ such that

$$\lim_{n\to\infty} \left\| \frac{\phi(x_0 + t_n h_n) - \phi(x_0)}{t_n} - \phi'_{x_0}(h) \right\|_{\mathbb{E}} = 0$$

for all sequences $\{h_n\}_{n=1}^\infty \subseteq \mathbb{D}$ and $\{t_n\}_{n=1}^\infty \subseteq \mathbb{R}_+$ such that $h_n \to h \in \mathbb{D}_T$ and $t_n \downarrow 0$ as $n \to \infty$, and $x_0 + t_n h_n \in \mathbb{D}_\phi$ for all n.

P_1 and P_0 with bounded densities, the $L_{2,P}$ semimetric works very well:

$$L_{2,P}(f_1, f_2) = \sqrt{P((f_1 - f_2)^2)} = \sqrt{E_P[(f_1(Y) - f_2(Y))^2]} \tag{50}$$

Equip the product space $\mathcal{F}_1 \times \mathcal{F}_0$ with the product semimetric:

$$L_2((f_1, g_1), (f_2, g_2)) = \sqrt{L_{2,P}(f_1, f_2)^2 + L_{2,P}(g_1, g_2)^2} \tag{51}$$

To apply the $L_{2,P}$ semimetric, each $f \in \mathcal{F}_1$ and $f \in \mathcal{F}_0$ are defined on whole domain $\mathcal{Y}$.

Lemma B.2 (Hadamard directional differentiability of optimal transport)**.** *Let $c : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ be lower semicontinuous, $\mathcal{F}_1, \mathcal{F}_0$ be sets of measurable functions mapping $\mathcal{Y}$ to $\mathbb{R}$, and $\mathcal{F}_c \subseteq \mathcal{F}_1$ and $\mathcal{F}_c^c \subseteq \mathcal{F}_0$ be universally bounded subsets. Suppose that*

1. *Strong duality holds:*

$$\inf_{\pi \in \Pi(P_1, P_0)} \int c(y_1, y_0) d\pi(y_1, y_0) = \sup_{(\varphi, \psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)} \int \varphi(y_1) dP_1(y_1) + \int \psi(y_0) dP_0(y_0),$$

2. *P dominates P_1 and P_0 with bounded densities,*
3. *$\mathcal{F}_d$ is P-Donsker and $\sup_{f \in \mathcal{F}_d} |P(f)| < \infty$ for each $d = 1, 0$, and*
4. *$(\mathcal{F}_1 \times \mathcal{F}_0, L_2)$ and the subset*

$$\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c) = \{(\varphi, \psi) \in \mathcal{F}_c \times \mathcal{F}_c^c \,;\, \varphi(y_1) + \psi(y_0) \leq c(y_1, y_0)\}$$

   are complete.

Then $OT_c : \ell^\infty(\mathcal{F}_1) \times \ell^\infty(\mathcal{F}_0) \to \mathbb{R}$ defined by

$$OT_c(P_1, P_0) = \sup_{(\varphi, \psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)} P_1(\varphi) + P_0(\psi)$$

is Hadamard directionally differentiable at (P_1, P_0) tangentially to

$$\mathbb{D}_{Tan} = \mathcal{C}(\mathcal{F}_1, L_{2,P}) \times \mathcal{C}(\mathcal{F}_0, L_{2,P}). \tag{52}$$

The set of maximizers $\Psi_c(P_1, P_0) = \arg\max_{(\varphi, \psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)} P_1(\varphi) + P_0(\psi)$ is nonempty, and the derivative $OT'_{c,(P_1, P_0)} : \mathbb{D}_{Tan} \to \mathbb{R}$ is given by

$$OT'_{c,(P_1, P_0)}(H_1, H_0) = \sup_{(\varphi, \psi) \in \Psi_c(P_1, P_0)} H_1(\varphi) + H_0(\psi)$$

Proof. For legibility, the proof is broken down into four steps:

1. Define

$$\begin{aligned} &\mathcal{S} : \ell^\infty(\mathcal{F}_1) \times \ell^\infty(\mathcal{F}_0) \to \ell^\infty(\mathcal{F}_1 \times \mathcal{F}_0), && \mathcal{S}(H_1, H_0)(\varphi, \psi) = H_1(\varphi) + H_0(\psi) \\ &\Xi_c : \ell^\infty(\mathcal{F}_1 \times \mathcal{F}_0) \to \mathbb{R}, && \Xi_c[G] = \sup_{(\varphi,\psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)} G(\varphi, \psi) \end{aligned}$$

   and notice that $OT_c(H_1, H_0) = \Xi_c(\mathcal{S}(H_1, H_0))$. This suggests application of the chain rule.

2. $\mathcal{S}$ is linear and continuous at every point of $\ell^\infty(\mathcal{F}_1) \times \ell^\infty(\mathcal{F}_0)$, which implies it is (fully) Hadamard differentiable at any $(H_1, H_0) \in \ell^\infty(\mathcal{F}_1) \times \ell^\infty(\mathcal{F}_0)$ tangentially to $\ell^\infty(\mathcal{F}_1) \times \ell^\infty(\mathcal{F}_0)$, and is its own derivative. Indeed, for any $(H_{1n}, H_{0n}) \to (H_1, H_0) \in \ell^\infty(\mathcal{F}_1) \times \ell^\infty(\mathcal{F}_0)$ and any $t_n \downarrow 0$,

$$\begin{aligned} \lim_{n\to\infty} &\left\| \frac{\mathcal{S}((H_1, H_0) + t_n(H_{1n}, H_{0n})) - \mathcal{S}(H_1, H_0)}{t_n} - \mathcal{S}(H_1, H_0) \right\|_{\mathcal{F}_c \times \mathcal{F}_c^c} \\ &= \lim_{n\to\infty} \|\mathcal{S}(H_{1n}, H_{0n}) - \mathcal{S}(H_1, H_0)\|_{\mathcal{F}_c \times \mathcal{F}_c^c} = 0 \end{aligned}$$

3. Consider Ξ_c. Verify the conditions of lemma F.9:

   (a) $(\mathcal{F}_1 \times \mathcal{F}_0, L_2)$ and the subset $\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$ are compact.

   First recall that a subset of semimetric space is compact if and only if it is totally bounded and complete.[6] Completeness of both sets is assumed, so it suffices to show they are totally bounded. Since $\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$ is a subset of $\mathcal{F}_1 \times \mathcal{F}_0$, it suffices to show the latter set is totally bounded.

   Using the assumption that $\mathcal{F}_d$ is P-Donsker and $\sup_{f \in \mathcal{F}_d} |P(f)| < \infty$, we have that $\sup_{\varphi \in \mathcal{F}_c} |P(\varphi)| < \infty$ and $(\mathcal{F}_d, L_{2,P})$ is totally bounded (see van der Vaart & Wellner (1997) problem 2.1.2.). It follows that the product space $(\mathcal{F}_1 \times \mathcal{F}_0, L_2)$ is totally bounded.[7]

   (b) $\mathcal{S}(P_1, P_0) \in \mathcal{C}(\mathcal{F}_1 \times \mathcal{F}_0, L_2)$.

   Notice that

$$|P_1(f_1) - P_1(f_2)| \leq P_1(|f_1 - f_2|) \leq \sqrt{P_1((f_1 - f_2)^2)} = L_{2,P_1}(f_1, f_2)$$

   where the second inequality is an applications of Jensen's inequality. This implies $P_1 \in \mathcal{C}(\mathcal{F}_1, L_{2,P_1})$. Moreover, since $P_1 \ll P$ and $\frac{dP_1}{dP} \leq K_1 < \infty$ for some $K_1 \in \mathbb{R}$,

$$L_{2,P_1}(f_1, f_2) = \left(\int (f_1 - f_2)^2 \frac{dP_1}{dP} dP \right)^{1/2} \leq K_1^{1/2} \left(\int (f_1 - f_2)^2 dP \right)^{1/2} = K_1^{1/2} L_{2,P}(f_1, f_2)$$

   shows that $\mathcal{C}(\mathcal{F}_1, L_{2,P_1}) \subseteq \mathcal{C}(\mathcal{F}_1, L_{2,P})$ and so $P_1 \in \mathcal{C}(\mathcal{F}_1, L_{2,P})$. A similar argument shows $P_0 \in \mathcal{C}(\mathcal{F}_0, L_{2,P})$.

[6] See van der Vaart & Wellner (1997), footnote on p. 17.

[7] For $\varepsilon > 0$, let $(f_1, \ldots, f_K)$ be the centers of $L_{2,P}$-balls of radius $\varepsilon/\sqrt{2}$ that cover $\mathcal{F}_1$, and $(g_1, \ldots, g_M)$ be the center of $L_{2,P}$-balls of radius $\varepsilon/\sqrt{2}$ that cover $\mathcal{F}_0$. Then for any $(f, g) \in \mathcal{F}_1 \times \mathcal{F}_0$, there exists f_k and g_m such that $L_{2,P}(f, f_k) < \varepsilon/\sqrt{2}$ and $L_{2,P}(g, g_m) < \varepsilon/\sqrt{2}$, and so

$$L_2((f, g), (f_k, g_m)) = \sqrt{L_{2,P}(f, f_k)^2 + L_{2,P}(g, g_m)^2} < \sqrt{(\varepsilon/\sqrt{2})^2 + (\varepsilon/\sqrt{2})^2} = \varepsilon$$

and thus the KM balls in $(\mathcal{F}_1 \times \mathcal{F}_0)$ of radius ε centered at (f_k, g_m) for some k, m cover $\mathcal{F}_1 \times \mathcal{F}_0$.

Use the inequalities above to see that

$$\begin{aligned}
|\mathcal{S}(P_1,P_0)(f_1,g_1)-\mathcal{S}(P_1,P_0)(f_2,g_2)| &= |P_1(f_1)-P_1(f_2)+P_0(g_1)-P_0(g_2)| \\
&\le L_{2,P_1}(f_1,f_2)+L_{2,P_0}(g_1,g_2) \le K_1^{1/2}L_{2,P}(f_1,f_2)+K_0^{1/2}L_{2,P}(\psi_1,\psi_2) \\
&\le 2\max\{K_1^{1/2},K_0^{1/2}\}\max\{L_{2,P}(f_1,f_2),L_{2,P}(g_1,g_2)\} \\
&= 2\max\{K_1^{1/2},K_0^{1/2}\}\sqrt{\max\{L_{2,P}(f_1,f_2)^2,L_{2,P}(g_1,g_2)^2\}} \\
&\le 2\max\{K_1^{1/2},K_0^{1/2}\}\sqrt{L_{2,P}(f_1,f_2)^2+L_{2,P}(g_1,g_2)^2} \\
&= 2\max\{K_1^{1/2},K_0^{1/2}\}L_2((f_1,g_1),(f_2,g_2))
\end{aligned}$$

hence $L_2((f_1,g_1),(f_2,g_2))<\varepsilon/(2\max\{K_1^{1/2},K_0^{1/2}\})$ implies

$$|\mathcal{S}(P_1,P_0)(f_1,g_1)-\mathcal{S}(P_1,P_0)(f_2,g_2)|<\varepsilon$$

and therefore $\mathcal{S}(P_1,P_0)\in\mathcal{C}(\mathcal{F}_1\times\mathcal{F}_0,L_2)$.

Lemma F.9 shows that Ξ_c is Hadamard directionally differentiable at $\mathcal{S}(P_1,P_0)$ tangentially to $\mathcal{C}(\mathcal{F}_1\times\mathcal{F}_0,L_2)$, with derivative

$$\Xi'_{c,\mathcal{S}(P_1,P_0)}:\mathcal{C}(\mathcal{F}_1\times\mathcal{F}_0,L_2)\to\mathbb{R},\qquad \Xi'_{c,\mathcal{S}(P_1,P_0)}(H)=\sup_{(\varphi,\psi)\in\Psi_c(P_1,P_0)}H(\varphi,\psi)$$

where $\Psi_c(P_1,P_0)=\arg\max_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)}P_1(\varphi)+P_0(\psi)$ is nonempty, because $P_1+P_0=\mathcal{S}(P_1,P_0)$ is continuous and $\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)$ is compact.

4. Now consider the tangent spaces to ensure the composition of the derivatives is well defined. Observe that if $(H_1,H_0)\in\mathcal{C}(\mathcal{F}_1,L_{2,P})\times\mathcal{C}(\mathcal{F}_0,L_{2,P})$ then $\mathcal{S}(H_1,H_0)=H_1+H_0\in\mathcal{C}(\mathcal{F}_1\times\mathcal{F}_0,L_2)$.[8] It follows from the chain rule (lemma F.4) that OT_c is Hadamard directionally differentiable at (P_1,P_0) tangentially to $\mathcal{C}(\mathcal{F}_1,L_{2,P})\times\mathcal{C}(\mathcal{F}_0,L_{2,P})$ with derivative $OT_c:\mathcal{C}(\mathcal{F}_1,L_{2,P})\times\mathcal{C}(\mathcal{F}_0,L_{2,P})\to\mathbb{R}$ given by

$$OT'_{c,(P_1,P_0)}(H_1,H_0)=\Xi'_{c,\mathcal{S}(P_1,P_0)}(\mathcal{S}'_{(P_1,P_0)}(H_1,H_0))=\sup_{(\varphi,\psi)\in\Psi_c(P_1,P_0)}H_1(\varphi)+H_0(\psi)$$

□

B.3 Full differentiability

The property distinguishing directional from full differentiability on a subspace is linearity of the derivative (Fang & Santos (2019), proposition 2.1). In the case of optimal transport, the derivative

[8] Fix $(f,g)\in\mathcal{F}_1\times\mathcal{F}_0$ and let $\delta_1>0$ and $\delta_0>0$ be such that $L_{2,P_1}(f,\tilde f)<\delta_1$ implies $H_1(f,\tilde f)<\varepsilon/2$ and $L_{2,P_0}(g,\tilde g)<\delta_0$ implies $H_0(g,\tilde g)<\varepsilon/2$. The inequality

$$\begin{aligned}
L_{2,P}(f,\tilde f)+L_{2,P}(g,\tilde g) &\le 2\max\{L_{2,P}(f,\tilde f),L_{2,P}(g,\tilde g)\} \\
&= 2\sqrt{\max\{L_{2,P}(f,\tilde f)^2,L_{2,P}(g,\tilde g)^2\}}=2L_2((f,g),(\tilde f,\tilde g))
\end{aligned}$$

implies that if $L_2((f,g),(\tilde f,\tilde g))<\min\{\delta_1,\delta_2\}/2$ then $|\mathcal{S}(H_1,H_0)(f,g)-\mathcal{S}(H_1,H_0)(\tilde f,\tilde g)|\le|H_1(f)-H_1(\tilde f)|+|H_0(g)-H_0(\tilde g)|<\varepsilon$.

found in lemma B.2 is linear on a large subspace of the tangent space when the solution to the dual problem is suitably unique. When it holds, this is sufficient for simpler bootstrap procedures to work for inference.

The dual solutions

$$(\varphi, \psi) \in \Psi_c(P_1, P_0) = \underset{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)}{\arg\max} P_1(\varphi) + P_0(\psi)$$

are referred to as **Kantorovich potentials**. Notice that for any $s \in \mathbb{R}$,

$$P_1(\varphi + s) + P_0(\psi - s) = P_1(\varphi) + P_0(\psi)$$

shows the most one can hope for is uniqueness up to a constant; if $(\varphi, \psi) \in \Psi_c(P_1, P_0)$, then $(\varphi + s, \psi - s) \in \Psi_c(P_1, P_0)$ as well.[9] It is well known in the optimal transport literature that when the distributions P_1, P_0 have full support on a convex, compact subset of $\mathbb{R}$ and c is differentiable, the Kantorovich potential is indeed unique in this way on the supports of P_1 and P_0.

Lemma B.3. *Suppose that*

1. *$c(y_1, y_0)$ is continuously differentiable.*
2. *P_d has compact support $\mathcal{Y}_d = [y_d^\ell, y_d^u] \subseteq \mathbb{R}$, and*

Let $\mathcal{F}_c$ and $\mathcal{F}_c^c$ be defined by (14) and (15) respectively, and

$$\Psi_c(P_1, P_0) = \underset{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)}{\arg\max} P_1(\varphi) + P_0(\psi)$$

Then for any $(\varphi_1, \psi_1), (\varphi_2, \psi_2) \in \Psi_c(P_1, P_0)$, there exists $s \in \mathbb{R}$ such that for all $(y_1, y_0) \in \mathcal{Y}_1 \times \mathcal{Y}_0$

$$\varphi_1(y_1) - \varphi_2(y_1) = s, \qquad \psi_1(y_0) - \psi_2(y_0) = -s$$

Proof. The proof is quite similar to that of Santambrogio (2015) proposition 7.18.

Let $(\varphi_1, \psi_1), (\varphi_2, \psi_2) \in \Psi_c(P_1, P_0)$. For $k = 1, 2$, φ_k and ψ_k (being elements of $\mathcal{F}_c$ and $\mathcal{F}_c^c$ respectively) are L-Lipschitz and hence absolutely continuous. This implies all four functions are differentiable Lebesgue-almost everywhere, and that for any $(y_1, y_0) \in \mathcal{Y}_1 \times \mathcal{Y}_0$,

$$\varphi_k(y_1) = \varphi_k(y_1^\ell) + \int_{y_1^\ell}^{y_1} \varphi_k'(y)dy \qquad \psi_k(y_0) = \psi_k(y_0^\ell) + \int_{y_0^\ell}^{y_0} \psi_k'(y)dy$$

Notice that the subset of $\mathcal{Y}_1$ where both φ_1 and φ_2 are differentiable also has full Lebesgue measure. It suffices to show that $\varphi_1'(y_1) = \varphi_2'(y_1)$ on this set (and $\psi_1'(y_0) = \psi_2'(y_0)$ on the subset of $\mathcal{Y}_0$ where

[9] See Staudt et al. (2022) for extended discussion on uniqueness of Kantorovich potentials.

both ψ_1 and ψ_2 are differentiable, which also has full Lebesgue measure), from which it will follow that for any $(y_1, y_0) \in \mathcal{Y}_1 \times \mathcal{Y}_0$,

$$\varphi_1(y_1) - \varphi_2(y_1) = \varphi_1(y_1^\ell) - \varphi_2(y_1^\ell) + \int_{y_1^\ell}^{y_1} (\varphi_1'(y) - \varphi_2'(y))dy = \underbrace{\varphi_1(y_1^\ell) - \varphi_2(y_1^\ell)}_{:=s_\varphi}$$

$$\psi_1(y_0) - \psi_2(y_0) = \psi_1(y^\ell) - \psi_2(y^\ell) + \int_{y_0^\ell}^{y_0} (\psi_1'(y) - \psi_2'(y))dy = \underbrace{\psi_1(y_0^\ell) - \psi_2(y_0^\ell)}_{:=s_\psi}$$

Finally, observe that $P_1(\varphi_2) + P_0(\varphi_2) = P_1(\varphi_1) + P_0(\psi_1) = P_1(\varphi_2 + s_\varphi) + P_0(\psi_2 + s_\psi) = P_1(\varphi_2) + P_0(\psi_2) + s_\varphi + s_\psi$ implies $s_\varphi = -s_\psi$.

The remainder of the proof shows that for any $\bar{y}_1$ in the set where both φ_1 and φ_2 are differentiable, $\varphi_1'(\bar{y}_1) = \varphi_2'(\bar{y}_1)$. The same arguments work to show the corresponding claim regarding ψ_1 and ψ_2.

There exists $\pi \in \Pi(P_1, P_0)$ that solves the primal problem (see lemma E.1). For any such π,

1. $\mathrm{Supp}(P_1) = \{y_1 \in \mathcal{Y}_1 \ ; \ \exists y_0 \in \mathcal{Y}_0 \text{ s.t. } (y_1, y_0) \in \mathrm{Supp}(\pi)\}$

   This follows because $\mathrm{Pr}_1(\mathrm{Supp}(\pi)) := \{y_1 \in \mathcal{Y}_1 \ ; \ \exists y_0 \in \mathcal{Y}_0 \text{ s.t. } (y_1, y_0) \in \mathrm{Supp}(\pi)\}$ is dense in $\mathrm{Supp}(P_1)$, and $\mathrm{Pr}_1(\mathrm{Supp}(\pi))$ is closed because $\mathcal{Y}_0$ is compact.[10]

2. For all $(y_1, y_0) \in \mathrm{Supp}(\pi)$, $\varphi_k(y_1) + \psi_k(y_0) = c(y_1, y_0)$.

   It is easy to see that the equality holds π-almost surely. To see it holds specifically on the support, notice that optimality of π and (φ_k, ψ_k) implies that

$$\int c(y_1, y_0) d\pi(y_1, y_0) = \int \varphi_k(y_1) dP(y_1) + \int \psi_k(y_0) dP_0(y_0)$$

   and recall that $\varphi_k(y_1) + \psi_k(y_0) \leq c(y_1, y_0)$ holds for all $(y_1, y_0) \in \mathcal{Y} \times \mathcal{Y}$. If the inequality were strict for some $(y_1', y_0') \in \mathrm{Supp}(\pi)$, then continuity of φ_k, ψ_k, and c would imply the inequality is sharp on a ball centered at (y_1, y_0) of some positive radius, denoted B, leading

[10] Specifically, for any $A \subseteq \mathcal{Y}_1 \times \mathcal{Y}_0 \subseteq \mathbb{R}^2$, let $\mathrm{Pr}_1(A) = \{y_1 \in \mathcal{Y}_1 \ ; \ \exists y_0 \in \mathcal{Y}_0 \text{ s.t. } (y_1, y_0) \in A\}$ be the cartesian projection of the set A onto the first coordinate. Let $P_1 \in \mathcal{P}(\mathcal{Y}_1)$, $P_0 \in \mathcal{P}(\mathcal{Y}_0)$, and $\pi \in \Pi(P_1, P_0)$. As noted in Staudt et al. (2022) (Remark 1), $\mathrm{Pr}_1(\mathrm{Supp}(\pi)) \subseteq \mathrm{Supp}(P_1)$ with the possibility that inclusion is strict.

However, $\mathrm{Pr}_1(\mathrm{Supp}(\pi))$ is always dense in $\mathrm{Supp}(P_1)$: let $y_1 \in \mathrm{Supp}(P_1)$ and $\delta > 0$ be arbitrary, and suppose for contradiction that $B_\delta(y_1) \cap \mathrm{Pr}_1(\mathrm{Supp}(\pi)) = \varnothing$. Then $\big(B_\delta(y_1) \times \mathcal{Y}_0\big) \cap \mathrm{Supp}(\pi) = \varnothing$ follows from the definition of $\mathrm{Pr}_1(\mathrm{Supp}(\pi))$, and thus

$$\begin{aligned} 0 &= \pi\left(\big(B_\delta(y_1) \times \mathcal{Y}_0\big) \cap \mathrm{Supp}(\pi)\right) = \pi\left(\big(B_\delta(y_1) \times \mathcal{Y}_0\big)\right) + \pi\left(\mathrm{Supp}(\pi)\right) - \pi\left(\big(B_\delta(y_1) \times \mathcal{Y}_0\big) \cup \mathrm{Supp}(\pi)\right) \\ &= \pi\left(\big(B_\delta(y_1) \times \mathcal{Y}_0\big)\right) = P_1(B_\delta(y_1)) > 0 \end{aligned}$$

a contradiction showing $B_\delta(y_1) \cap \mathrm{Pr}_1(\mathrm{Supp}(\pi)) \neq \varnothing$. Thus $\mathrm{Pr}_1(\mathrm{Supp}(\pi))$ is dense in $\mathrm{Supp}(P_1)$.

Moreover, if $\mathcal{Y}_0$ is compact then the map Pr_1 is closed: suppose $A \subseteq \mathcal{Y}_1 \times \mathcal{Y}_0 \subseteq \mathbb{R}^2$ is closed, and $\{y_{1n}\}_{n=1}^\infty \subseteq \mathrm{Pr}_1(A)$ converges to y_1. Then there exists $\{y_{0n}\}_{n=1}^\infty \subseteq \mathcal{Y}_0$ such that $(y_{1n}, y_{0n}) \in A$ for each n. Since $\mathcal{Y}_0$ is compact, there exists a subsequence $\{y_{0n_k}\}_{k=1}^\infty$ and y_0 such that $\lim_{k\to\infty} y_{0n_k} = y_0$. Then notice that $\lim_{k\to\infty}(y_{1n_k}, y_{0n_k}) = (y_1, y_0)$. Since A is closed, $(y_1, y_0) \in A$.

$\mathrm{Supp}(\pi)$ is closed by definition, hence $\mathrm{Pr}_1(\mathrm{Supp}(\pi))$ is closed and dense in $\mathrm{Supp}(P_1)$, from which it follows that $\mathrm{Supp}(\pi) = \mathrm{Supp}(P_1)$.

to the contradiction

$$\begin{aligned}\int c(y_1,y_0)d\pi(y_1,y_0) &= \int_B c(y_1,y_0)d\pi(y_1,y_0) + \int_{B^c} c(y_1,y_0)d\pi(y_1,y_0)\\ &> \int_B \varphi_k(y_1)+\psi_k(y_0)d\pi(y_1,y_0) + \int_{B^c} \varphi_k(y_1)+\psi_k(y_0)d\pi(y_1,y_0)\\ &= \int \varphi_k(y_1)+\psi_k(y_0)d\pi(y_1,y_0) = \int \varphi_k(y_1)dP_1(y_1) + \int \psi_k(y_0)dP_0(y_0)\end{aligned}$$

3. For any $\bar{y}_1 \in \text{Supp}(P_1)$, the above implies there there exists $\bar{y}_0 \in \mathcal{Y}_0$ such that $(\bar{y}_1, \bar{y}_0) \in \text{Supp}(\pi)$, and hence $\varphi_k(\bar{y}_1) + \psi_k(\bar{y}_0) = c(\bar{y}_1, \bar{y}_0)$. For any such $\bar{y}_0$,

$$y_1 \mapsto \varphi_k(y_1) - c(y_1, \bar{y}_0) \text{ is maximized at } \bar{y}_1 \tag{53}$$

Indeed, if there were $y_1' \in \mathcal{Y}_1$ such that $\varphi_k(y_1') - c(y_1', \bar{y}_0) > \varphi_k(\bar{y}_1) - c(\bar{y}_1, \bar{y}_0)$, then by adding $\psi_k(\bar{y}_0)$ to both sides we find

$$\varphi_k(y_1') + \psi_k(\bar{y}_0) - c(y_1', \bar{y}_0) > \varphi_k(\bar{y}_1) + \psi_k(\bar{y}_0) - c(\bar{y}_1, \bar{y}_0) = 0$$

This implies $\varphi_k(y_1') + \psi_k(\bar{y}_0) > c(y_1', \bar{y}_0)$, which contradicts $\varphi_k(y_1') + \psi_k(\bar{y}_0) \leq c(y_1', \bar{y}_0)$ for all $(y_1, y_0) \in \mathcal{Y}_1 \times \mathcal{Y}_0$.

4. Now observe that if $\bar{y}_1 \in (y_1^\ell, y_1^u)$ is a point at which φ_k is differentiable, then (53) implies $\varphi_k'(\bar{y}_1) = \frac{\partial c}{\partial y_1}(\bar{y}_1, \bar{y}_0)$.[11] Thus if $\bar{y}_1 \in (y_1^\ell, y_1^u)$ is a point at which both φ_1 and φ_2 are differentiable, then

$$\varphi_1(\bar{y}_1) = \frac{\partial c}{\partial y_1}(\bar{y}_1, \bar{y}_0) = \varphi_2(\bar{y}_1)$$

This completes the proof. □

To specify the subset of the tangent space on which $OT'_{c,(P_1,P_0)}$ is linear, let $\mathcal{Y}_d \subseteq \mathcal{Y}$ and $\mathbb{1}_{\mathcal{Y}_d}(y) = \mathbb{1}\{y \in \mathcal{Y}_d\}$. Let $\mathcal{G}$ denote a set of real-valued functions $g : \mathcal{Y} \to \mathbb{R}$ with the following property: if $g \in \mathcal{G}$, then $\mathbb{1}_{\mathcal{Y}_d} \times g \in \mathcal{G}$.[12] Let $\ell^\infty_{\mathcal{Y}_d}(\mathcal{G})$ be the set of bounded, linear functions $H : \mathcal{G} \to \mathbb{R}$ that evaluate constant functions to zero and "ignore" the value of functions outside of $\mathcal{Y}_d$. Specifically, define

$$\begin{aligned}\ell^\infty_{\mathcal{Y}_d}(\mathcal{G}) = \Big\{ H \in \ell^\infty(\mathcal{G}) \;;\; &\text{for all } a,b \in \mathbb{R} \text{ and } f,g \in \mathcal{G},\\ &(i)\ H(f) = H(\mathbb{1}_{\mathcal{Y}_d} \times f),\ \ (ii)\ \text{ if } a \in \mathcal{G} \text{ then } H(a) = 0, \text{ and}\\ &(iii)\ \text{ if } af + bg \in \mathcal{G} \text{ then } H(af+bg) = aH(f) + bH(g) \Big\}\end{aligned} \tag{54}$$

Here we slightly abuse notation; $a \in \mathcal{G}$ refers to the function mapping each point in $\mathcal{Y}$ to the

[11] Notice that the "choice" of π or $\bar{y}_0$ doesn't matter, because $\varphi_k'(\bar{y}_1)$ can take only one value.

[12] If we have a set $\tilde{\mathcal{G}}$ that does not satisfy this property, the set $\mathcal{G} = \tilde{\mathcal{G}} \cup \left\{ \mathbb{1}_{\mathcal{Y}_d} \times g \;;\; g \in \tilde{\mathcal{G}} \right\}$ will satisfy it.

constant $a \in \mathbb{R}$. Equip $\ell^\infty_{\mathcal{Y}_d}(\mathcal{G})$ with the supremum norm, $\|H\|_{\mathcal{G}} = \|H\|_\infty = \sup_{g\in\mathcal{G}}|H(g)|$. As shown in appendix C, first stage estimators of (P_1, P_0) based on the empirical distribution have weak limits concentrated on $\ell^\infty_{\mathcal{Y}_1}(\mathcal{F}_c) \times \ell^\infty_{\mathcal{Y}_0}(\mathcal{F}_c^c)$ where $\mathcal{Y}_d$ is the support of P_d.

Lemma B.4. *$\ell^\infty_{\mathcal{Y}_d}(\mathcal{G})$ defined by (54) is closed.*

Proof. Let $\{H_n\}_{n=1}^\infty \subseteq \ell^\infty_{\mathcal{Y}_d}(\mathcal{G})$ be Cauchy, and let H be its limit in the Banach space $\ell^\infty(\mathcal{G})$. It suffices to show $H \in \ell^\infty_{\mathcal{Y}_d}(\mathcal{G})$.

Toward this end, first notice that $\|H_n - H\|_{\mathcal{G}} \to 0$ implies that for any $f \in \mathcal{G}$, $|H_n(f) - H(f)| \to 0$. Next observe that if the constant function $a \in \mathcal{G}$, then $0 = \lim_{n\to\infty}|H_n(a) - H(a)| = \lim_{n\to\infty}|H(a)| = |H(a)|$. For any function $f \in \mathcal{G}$, since $H_n(f) = H_n(\mathbb{1}_{\mathcal{Y}_d} \times f)$,

$$0 \le |H(f) - H(\mathbb{1}_{\mathcal{Y}_d} \times f)| \le |H(f) - H_n(f)| + |H(\mathbb{1}_{\mathcal{Y}_d} \times f) - H_n(\mathbb{1}_{\mathcal{Y}_d} \times f)| \to 0$$

and thus $H(\mathbb{1}_{\mathcal{Y}_d} \times f) = H(f)$. Finally, suppose $a, b \in \mathbb{R}$ and $f, g \in \mathcal{G}$ are such that $af + bg \in \mathcal{G}$. Similar to the argument above, since $H_n(af + bg) = aH_n(f) + bH_n(g)$,

$$\begin{aligned}
0 &\le |H(af + bg) - aH(f) - bH(g)| \\
&\le |H(af + bg) - H_n(af + bg)| + |aH_n(f) + bH_n(f) - aH(f) - bH_n(g)| \\
&\le |H(af + bg) - H_n(af + bg)| + |a||H_n(f) - H(f)| + |b||H_n(g) - H_n(g)| \to 0
\end{aligned}$$

and thus $H(af + bg) = aH(f) + bH(g)$.

This shows $H \in \ell^\infty_{\mathcal{Y}_d}(\mathcal{G})$, and completes the proof. □

Lemma B.5 (Full differentiability of optimal transport)**.** *Let $c : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ be lower semicontinuous, $\mathcal{F}_1, \mathcal{F}_0$ be sets of measurable functions mapping $\mathcal{Y}$ to $\mathbb{R}$, and $\mathcal{F}_c \subseteq \mathcal{F}_1$ and $\mathcal{F}_c^c \subseteq \mathcal{F}_0$ be universally bounded subsets. Suppose that*

1. *Strong duality holds:*

$$\inf_{\pi\in\Pi(P_1,P_0)} \int c(y_1, y_0)d\pi(y_1, y_0) = \sup_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} \int \varphi(y_1)dP_1(y_1) + \int \psi(y_0)dP_0(y_0),$$

2. *P dominates P_1 and P_0 with bounded densities,*
3. *$\mathcal{F}_d$ is P-Donsker and $\sup_{f\in\mathcal{F}_d}|P(f)| < \infty$ for each $d = 1, 0$, and*
4. *$(\mathcal{F}_1 \times \mathcal{F}_0, L_2)$ and the subset*

$$\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c) = \{(\varphi, \psi) \in \mathcal{F}_c \times \mathcal{F}_c^c \; ; \; \varphi(y_1) + \psi(y_0) \le c(y_1, y_0)\}$$

are complete.

Let $\mathcal{Y}_1, \mathcal{Y}_0 \subseteq \mathcal{Y}$ and $\Psi_c(P_1, P_0) = \arg\max_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} P_1(\varphi) + P_0(\psi)$, and further assume

4. *For any $(\varphi_1, \psi_1), (\varphi_2, \psi_2) \in \Psi_c(P_1, P_0)$, there exists $s \in \mathbb{R}$ such that*

$$\mathbb{1}_{\mathcal{Y}_1} \times \varphi_1 = \mathbb{1}_{\mathcal{Y}_1} \times (\varphi_2 + s), \; P\text{-a.s.} \qquad \text{and} \qquad \mathbb{1}_{\mathcal{Y}_0} \times \psi_1 = \mathbb{1}_{\mathcal{Y}_0} \times (\psi_2 - s), \; P\text{-a.s.}$$

Then $OT_c : \ell^\infty(\mathcal{F}_1) \times \ell^\infty(\mathcal{F}_0) \to \mathbb{R}$ defined by

$$OT_c(P_1, P_0) = \sup_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} P_1(\varphi) + P_0(\psi)$$

is fully Hadamard differentiable at (P_1, P_0) tangentially to

$$\mathbb{D}_{Tan,Full} = \Big(\ell^\infty_{\mathcal{Y}_1}(\mathcal{F}_c) \times \ell^\infty_{\mathcal{Y}_0}(\mathcal{F}_c^c)\Big) \cap \Big(\mathcal{C}(\mathcal{F}_1, L_{2,P}) \times \mathcal{C}(\mathcal{F}_0, L_{2,P})\Big) \tag{55}$$

with derivative $OT'_{c,(P_1,P_0)} : \mathbb{D}_{Tan,Full} \to \mathbb{R}$ given by

$$OT'_{c,(P_1,P_0)}(H_1, H_0) = \sup_{(\varphi,\psi)\in\Psi_c(P_1,P_0)} H_1(\varphi) + H_0(\psi)$$

Proof. The first four assumptions allow application of lemma B.2 to find that $OT_c : \ell^\infty(\mathcal{F}_1) \times \ell^\infty(\mathcal{F}_0) \to \mathbb{R}$ given by

$$OT_c(P_1, P_0) = \sup_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} P_1(\varphi) + P_0(\psi)$$

is Hadamard directionally differentiable at (P_1, P_0) tangentially to $\mathbb{D}_{Tan} = \mathcal{C}(\mathcal{F}_1, L_{2,P}) \times \mathcal{C}(\mathcal{F}_0, L_{2,P})$. The set of maximizers $\Psi_c(P_1, P_0) = \arg\max_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} P_1(\varphi) + P_0(\psi)$ is nonempty, and the derivative $OT'_{c,(P_1,P_0)} : \mathbb{D}_{Tan} \to \mathbb{R}$ is given by

$$OT'_{c,(P_1,P_0)}(H_1, H_0) = \sup_{(\varphi,\psi)\in\Psi_c(P_1,P_0)} H_1(\varphi) + H_0(\psi)$$

Next observe that for any $(H_1, H_0) \in \mathbb{D}_{Tan,Full}$, $H_1 + H_0$ is flat on $\Psi_c(P_1, P_0)$. Specifically, for any $(\varphi_1, \psi_1), (\varphi_2, \psi_2) \in \Psi_c(P_1, P_0)$, let s be such that

$$\mathbb{1}_{\mathcal{Y}_1} \times \varphi_1 = \mathbb{1}_{\mathcal{Y}_1} \times (\varphi_2 + s),\ P\text{-a.s.} \qquad \text{and} \qquad \mathbb{1}_{\mathcal{Y}_0} \times \psi_1 = \mathbb{1}_{\mathcal{Y}_0} \times (\psi_2 - s),\ P\text{-a.s.}$$

Then

$$\begin{aligned}
H_1(\varphi_1) + H_0(\psi_1) &= H_1(\mathbb{1}_{\mathcal{Y}_1} \times \varphi_1) + H_0(\mathbb{1}_{\mathcal{Y}_0} \times \psi_1) \\
&= H_1(\mathbb{1}_{\mathcal{Y}_1} \times (\varphi_2 + s)) + H_0(\mathbb{1}_{\mathcal{Y}_0} \times (\psi_2 - s)) \\
&= H_1(\varphi_2 + s) + H_0(\psi_2 - s) \\
&= H_1(\varphi_2) + H_1(s) + H_0(\psi_2) - H_0(s) \\
&= H_1(\varphi_2) + H_0(\psi_2)
\end{aligned}$$

where the first, third, fourth, and fifth equalities hold because $(H_1, H_0) \in \ell^\infty_{\mathcal{Y}_1}(\mathcal{F}_c) \times \ell^\infty_{\mathcal{Y}_0}(\mathcal{F}_c^c)$, and the second because $(H_1, H_0) \in \mathcal{C}(\mathcal{F}_1, L_{2,P}) \times \mathcal{C}(\mathcal{F}_0, L_{2,P})$.

Now use this "flatness" to observe the derivative is linear. Let $(H_1, H_0), (G_1, G_0) \in \mathbb{D}_{Tan,Full}$,

$a, b \in \mathbb{R}$, and $(\tilde{\varphi}, \tilde{\psi}) \in \Psi_c(P_1, P_0)$, and notice that

$$
\begin{aligned}
OT'_{c,(P_1,P_0)}(a(H_1,H_0)+b(G_1,G_0)) &= \sup_{(\varphi,\psi)\in\Psi(P_1,P_0)} (aH_1+bG_1)(\varphi)+(aH_0+bG_0)(\psi) \\
&= aH_1(\tilde{\varphi})+bG_1(\tilde{\varphi})+aH_0(\tilde{\psi})+bG_0(\tilde{\psi}) = a(H_1(\tilde{\varphi})+H_0(\tilde{\psi}))+b(G_1(\tilde{\varphi})+G_0(\tilde{\psi}) \\
&= a\times \sup_{(\varphi,\psi)\in\Psi(P_1,P_0)} \{H_1(\varphi)+H_0(\psi)\} + b\times \sup_{(\varphi,\psi)\in\Psi(P_1,P_0)} \{G_1(\varphi)+G_0(\psi)\} \\
&= aOT'_{c,(P_1,P_0)}(H_1,H_0)+bOT'_{c,(P_1,P_0)}(G_1,G_0)
\end{aligned}
$$

Since $OT'_{c,(P_1,P_0)}$ is linear on the subspace $\mathbb{D}_{Tan,Full}$, Fang & Santos (2019) proposition 2.1 implies OT_c is fully Hadamard differentiable at (P_1, P_0) tangentially to $\mathbb{D}_{Tan,Full}$. □

C Appendix: weak convergence

Recall that

$$\theta_x^L = \theta^L(P_{1|x}, P_{0|x}), \qquad \theta_x^H = \theta^H(P_{1|x}, P_{0|x})$$
$$\theta^L = \sum_x s_x \theta_x^L, \qquad \theta^L = \sum_x s_x \theta_x^H$$
$$\gamma^L = \inf_{t\in[\theta^L,\theta^H]} g(t,\eta) \qquad \gamma^H = \sup_{t\in[\theta^L,\theta^H]} g(t,\eta)$$

where $\eta = (\eta_1, \eta_0)$, with $\eta_d \in \mathbb{R}^{K_d}$ having coordinates

$$\eta_d^{(k)} = \sum_x P(X = x \mid D_1 > D_0) E[\eta_d^{(k)}(Y_d) \mid D_1 > D_0, X = x] = \sum_x s_x \eta_{d,x}^{(k)}$$

Here $\eta_{d,x}^{(k)} = P_{d|x}(\eta_d^{(k)})$, which are collected as $\eta_{d,x} = (\eta_{d,x}^{(1)}, \ldots, \eta_{d,x}^{(K_d)})$.

Define the following sets of functions:

$$\tilde{\mathcal{F}}_1 = \Big\{f : \mathcal{Y} \to \mathbb{R}\ ;\ f = \varphi \text{ for some } \varphi \in \mathcal{F}_c, \text{ or } f = \eta_1^{(k)} \text{ for some } k = 1, \ldots, K_1\Big\} \tag{56}$$
$$\tilde{\mathcal{F}}_0 = \Big\{f : \mathcal{Y} \to \mathbb{R}\ ;\ f = \psi \text{ for some } \psi \in \mathcal{F}_c^c, \text{ or } f = \eta_0^{(k)} \text{ for some } k = 1, \ldots, K_0\Big\}$$
$$\mathcal{F}_{d,x} = \Big\{f : \mathcal{Y} \to \mathbb{R}\ ;\ f = g \text{ or } \mathbb{1}_{\mathcal{Y}_{d,x}} \times g \text{ for some } g \in \tilde{\mathcal{F}}_d\Big\}$$

where $\mathcal{Y}_{d,x}$ is the support of $Y \mid D = d, X = x$, and $\mathbb{1}_{\mathcal{Y}_{d,x}}(y) = \mathbb{1}\{y \in \mathcal{Y}_{d,x}\}$. The additional functions of the form $f(y) = \mathbb{1}_{\mathcal{Y}_{d,x}}(y)g(y)$ are used to characterize the support of the weak limit of $\sqrt{n}(\hat{P}_{d|x} - P_{d|x})$ in $\ell^\infty(\mathcal{F}_{d,x})$. The maps $P_{d|x}$ can be written as

$$P_{d|x} : \mathcal{F}_{d,x} \to \mathbb{R}, \qquad P_{d|x}(f) = \frac{P(\mathbb{1}_{d,x,d} \times f)/P(\mathbb{1}_{x,d}) - P(\mathbb{1}_{d,x,1-d} \times f)/P(\mathbb{1}_{x,1-d})}{P(\mathbb{1}_{d,x,d})/P(\mathbb{1}_{x,d}) - P(\mathbb{1}_{d,x,1-d})/P(\mathbb{1}_{x,1-d})} \tag{57}$$

and finally, define the set

$$\mathcal{F} = \bigcup_{d,x,z} \{\mathbb{1}_{d,x,z} \times f\ ;\ f \in \mathcal{F}_{d,x}\} \cup \{\mathbb{1}_{d,x,z}, \mathbb{1}_{x,z}, \mathbb{1}_x\}. \tag{58}$$

This appendix defines and studies the map $T : \mathbb{D}_C \subseteq \ell^\infty(\mathcal{F}) \to \mathbb{R}^2$ given by $(\gamma^L, \gamma^H) = T(P)$. The coming results show that $\mathcal{F}$ is P-Donsker, and the map T is Hadamard directionally differentiable at P. Together these imply, through the functional delta method, the weak convergence of

$\sqrt{n}(T(\mathbb{P}_n) - T(P))$ (Fang & Santos (2019)).

Several operations in the definition of the map T are repeated for each $x \in \mathcal{X} = \{x_1, \ldots, x_M\}$, leading to large expressions. These are shortened with the notation $\{a_x\}_{x\in\mathcal{X}}$, which refers to $(a_{x_1}, \ldots, a_{x_M})$. For example,

$$\left(\left\{P_{1|x}, P_{0|x}, \eta_{1,x}, \eta_{0,x}, s_x\right\}_{x\in\mathcal{X}}\right) = (P_{1|x_1}, P_{0|x_1}, \eta_{1,x_1}, \eta_{0,x_1}, s_{x_1}, \ldots, P_{1|x_M}, P_{0|x_M}, \eta_{1,x_M}, \eta_{0,x_M}, s_{x_M})$$

is an element of $\prod_{m=1}^M \ell^\infty(\mathcal{F}_{1,x_m}) \times \ell^\infty(\mathcal{F}_{0,x_m}) \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}$.

The function T is viewed as the composition of four functions: $T(P) = T_4(T_3(T_2(T_1(P))))$.

1. T_1 is the map to the conditional distributions and $\eta_{d,x}$: $T_1(P) = (\{P_{1|x}, P_{0|x}, \eta_{1,x}, \eta_{0,x}, s_x\}_{x\in\mathcal{X}})$,
2. T_2 involves optimal transport: $T_2(\{(P_{1|x}, P_{0|x}, \eta_{1,x}, \eta_{0,x}, s_x)\}_{x\in\mathcal{X}}) = (\{\theta_x^L, \theta_x^H, \eta_{1,x}, \eta_{0,x}, s_x\}_{x\in\mathcal{X}})$,
3. T_3 takes expectations over covariates: $T_3(\{(\theta_x^L, \theta_x^H, \eta_{1,x}, \eta_{0,x}, s_x)\}_{x\in\mathcal{X}}) \mapsto (\theta^L, \theta^H, \eta)$,
4. T_4 optimizes over $t \in [\theta^L, \theta^H]$: $T^4(\theta^L, \theta^H, \eta) = (\gamma^L, \gamma^H)$.

C.1 Verifying Donsker conditions

Before studying this map, this subsection shows the relevant sets are Donsker. The function classes $\mathcal{F}_c$ and $\mathcal{F}_c^c$ given by (14) and (15), or by (16) and (17), are well known Donsker classes as noted below. The results of van der Vaart & Wellner (1997) chapter 2.10 allow these to be extended to show $\mathcal{F}_{1,x}$ and $\mathcal{F}_{0,x}$ are Donsker. It follows quickly that $\mathcal{F}$ is Donsker.

Lemma C.1. *Suppose that $\mathcal{Y} \subset \mathbb{R}$ is compact and $c : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is L-Lipschitz. Let $\mathcal{F}_c$, $\mathcal{F}_c^c$ be given by* (14) *and* (15) *respectively. Then $\mathcal{F}_c$ and $\mathcal{F}_c^c$ are universally Donsker.*

Proof. Note that any distribution defined on the compact $\mathcal{Y}$ has a finite $2 + \delta$ moment. The result follows from the bracketing number bound given by van der Vaart & Wellner (1997) corollary 2.7.4. □

Lemma C.2. *$\mathcal{F}_c$ and $\mathcal{F}_c^c$ given by* (16) *and* (17) *are universally Donsker.*

Proof. The intervals (convex subsets of $\mathbb{R}$) form a well-known VC class with VC-dimension at most 3. Consider an arbitrary set of three real numbers $\{y_1, y_2, y_3\}$ with $y_1 < y_2 < y_3$, and notice that no interval can pick out the set $\{y_1, y_3\}$; that is, there does not exist an interval I with $\{y_1, y_3\} = \{y_1, y_2, y_3\} \cap I$. Since the intervals cannot shatter finite sets of size 3, the VC-dimension of the intervals is at most 3.

Similarly, the complements of intervals form a VC class of VC-dimension at most 4. Consider $\{y_1, y_2, y_3, y_4\}$ with $y_1 < y_2 < y_3 < y_4$ and notice that no complement of an interval can pick out $\{y_1, y_3\}$. Since the complements of intervals cannot shatter finite sets of size 4, the VC-dimension of the complements of intervals is at most 4.

The claim follows, because any (suitably measurable) VC class is Donsker for any probability measure (van der Vaart & Wellner (1997) section 2.6.1). $\square$

Lemma C.3. *Let $\mathcal{G}$ be P-Donsker and $\mathbb{1}_A$ be the indicator function for the set A. Then the set $\{\mathbb{1}_A \times g \; ; \; g \in \mathcal{G}\}$ is P-Donsker.*

Proof. The proof is an application of van der Vaart & Wellner (1997) theorem 2.10.6. Specifically, let $\phi : \mathcal{G} \times \{\mathbb{1}_A\} \to \mathbb{R}$ be the map $\phi(g, \mathbb{1}_a) = \mathbb{1}_A \times g$. Notice that for any $f, g \in \mathcal{G}_1 \times \{\mathbb{1}_A\}$,

$$\begin{aligned}
|\phi \circ f(w) - \phi \circ g(w)|^2 &= |\mathbb{1}_A(w) \times f_1(w) - \mathbb{1}_A(w) \times g_1(w)|^2 \\
&= \mathbb{1}_A(w) \times |f_1(w) - g_2(w)|^2 \\
&\leq |f_1(w) - g_1(w)|^2 = \sum_{\ell=1}^{k} (f_\ell(w) - g_\ell(w))^2
\end{aligned}$$

and thus van der Vaart & Wellner (1997) condition (2.10.5) holds. Moreover, notice that for any $g \in \mathcal{G}$, $(\mathbb{1}_A \times g)^2 \leq g^2$ and P-square integrability of $g \in \mathcal{G}$ implies $\mathbb{1}_A \times g$ is P-square integrable. Thus van der Vaart & Wellner (1997) theorem 2.10.6 implies $\{\mathbb{1}_A \times g \; ; \; g \in \mathcal{G}\}$ is P-Donsker. $\square$

Lemma C.4 ($\mathcal{F}_{d,x}$ are P-Donsker)**.** *Suppose assumptions 1, 2, and 3 hold. Let $\mathcal{F}_c$ and $\mathcal{F}_c^c$ be given by* (14) *and* (15)*, or by* (16) *and* (17)*. Let $\mathcal{F}_{d,x}$ be as defined in* (56)*. Then $\mathcal{F}_{d,x}$ is P-Donsker and $\sup_{f \in \mathcal{F}_{d,x}} |P(f)| < \infty$.*

Proof. 1. We first show $\tilde{\mathcal{F}}_d$ is P-Donsker and $\sup_{g \in \tilde{F}_d} |P(f)| < \infty$. The argument shows the argument for $\tilde{\mathcal{F}}_1$, as the same argument works when applied to $\tilde{\mathcal{F}}_0$.

Begin by noticing that

$$\begin{aligned}
\tilde{\mathcal{F}}_1 &= \left\{ f : \mathcal{Y} \to \mathbb{R} \; ; \; f = \varphi \text{ for some } \varphi \in \mathcal{F}_c, \text{ or } f = \eta_1^{(k)} \text{ for some } k = 1, \ldots, K_1 \right\} \\
&= \mathcal{F}_c \cup \left\{ \eta_1^{(1)}, \ldots, \eta_1^{(K_1)} \right\}
\end{aligned}$$

Since $\left\{ \eta_1^{(1)}, \ldots, \eta_1^{(K_1)} \right\}$ is a finite number of functions which, by assumption 3 (i), have finite second P-moment: $P((\eta_1^{(k)})^2) < \infty$. Thus $\left\{ \eta_1^{(1)}, \ldots, \eta_1^{(K_1)} \right\}$ is Donsker. $\mathcal{F}_c$ is Donsker by lemma C.1 or C.2, and so $\tilde{\mathcal{F}}_1 = \mathcal{F}_c \cup \left\{ \eta_1^{(1)}, \ldots, \eta_1^{(K_1)} \right\}$ is the union of two P-Donsker sets. Since

$$\|P\|_{\tilde{\mathcal{F}}_1} = \max\{ \sup_{\varphi \in \mathcal{F}_c} |P(\varphi)|, |P(\eta_1^{(1)})|, \ldots, |P(\eta_1^{(K_1)})| \} < \infty$$

van der Vaart & Wellner (1997) example 2.10.7 shows $\tilde{\mathcal{F}}_1$ is P-Donsker. Note we have also shown that $\sup_{g \in \tilde{F}_1} |P(f)| < \infty$.

2. Now notice that

$$\begin{aligned}\mathcal{F}_{d,x} &= \left\{ f : \mathcal{Y} \to \mathbb{R} \; ; \; f = g \text{ or } \mathbb{1}_{\mathcal{Y}_{d,x}} \times g \text{ for some } g \in \tilde{\mathcal{F}}_d \right\} \\ &= \tilde{\mathcal{F}}_d \cup \left\{ \mathbb{1}_{\mathcal{Y}_{d,x}} \times g \; ; \; g \in \tilde{\mathcal{F}}_d \right\}\end{aligned}$$

Lemma C.3 shows $\left\{ \mathbb{1}_{\mathcal{Y}_{d,x}} \times g \; ; \; g \in \tilde{\mathcal{F}}_d \right\}$ is P-Donsker. Moreover, since $\mathcal{F}_c$ is universally bounded,

$$\|P\|_{\left\{\mathbb{1}_{\mathcal{Y}_{d,x}} \times g \; ; \; g \in \tilde{\mathcal{F}}_d\right\}} = \max \left\{ \sup_{\varphi \in \mathcal{F}_c} |P(\mathbb{1}_{\mathcal{Y}_{d,x}} \times \varphi)|, |P(\mathbb{1}_{\mathcal{Y}_{d,x}} \times \eta_1^{(1)})|, \ldots, |P(\mathbb{1}_{\mathcal{Y}_{d,x}} \times \eta_1^{(K_1)})| \right\} < \infty$$

It follows that

$$\|P\|_{\mathcal{F}_{d,x}} = \sup_{f \in \mathcal{F}_{d,x}} |P(f)| = \max \left\{ \sup_{f \in \tilde{\mathcal{F}}_d} |P(f)|, \sup_{f \in \{\mathbb{1}_{\mathcal{Y}_{d,x}} \times g \; ; \; g \in \tilde{\mathcal{F}}_d\}} |P(f)| \right\} < \infty$$

Thus van der Vaart & Wellner (1997) example 2.10.7 implies $\mathcal{F}_1$ is P-Donsker.

□

Lemma C.5 ($\mathcal{F}$ is P-Donsker)**.** *Suppose assumptions 1, 2 and 3 hold. Then $\mathcal{F}$ is P-Donsker, implying*

$$\sqrt{n}(\mathbb{P}_n - P) \xrightarrow{L} \mathbb{G} \qquad \text{in } \ell^\infty(\mathcal{F}),$$

where $\mathbb{G}$ is a tight, mean-zero Gaussian process with $P(\mathbb{G} \in \mathcal{C}(\mathcal{F}, L_{2,P}) = 1$.

Proof. Lemma C.3 shows $\{\mathbb{1}_{d,x,z} \times f \; ; \; f \in \mathcal{F}_{d,x}\}$ is P-Donsker. Moreover, $\mathcal{F}_{d,x}$ is the union of a subset of universally bounded functions (in either $\mathcal{F}_c$ or $\mathcal{F}_c^c$) and a finite subset of square integrable functions. It follows that

$$\|P\|_{\left\{\mathbb{1}_{d,x,z} \times g \; ; \; g \in \mathcal{F}_{d,x}\right\}} = \sup_{f \in \left\{\mathbb{1}_{d,x,z} \times g \; ; \; g \in \mathcal{F}_{d,x}\right\}} |P(f)| < \infty$$

Next notice that

$$\mathcal{F} = \bigcup_{d,x,z} \{\mathbb{1}_{d,x,z} \times f \; ; \; f \in \mathcal{F}_{d,x}\} \cup \{\mathbb{1}_{d,x,z}, \mathbb{1}_{x,z}, \mathbb{1}_x\}$$

is the union of a finite number of P-Donsker sets, with

$$\|P\|_{\mathcal{F}} = \max_{d,x,z} \left\{ \max \left\{ \sup_{f \in \left\{\mathbb{1}_{d,x,z} \times g \; ; \; g \in \mathcal{F}_{d,x}\right\}} |P(f)|, |P(\mathbb{1}_{d,x,z})|, |P(\mathbb{1}_{x,z})|, |P(\mathbb{1}_x)|, \right\} \right\} < \infty$$

It follows from van der Vaart & Wellner (1997) example 2.10.7 that $\mathcal{F}$ is P-Donsker, which implies $\sqrt{n}(\mathbb{P}_n - P) \xrightarrow{L} \mathbb{G}$ in $\ell^\infty(\mathcal{F})$, where $\mathbb{G}$ is a tight, mean-zero Gaussian process. Moreover, van der Vaart & Wellner (1997) section 2.1.2 and problem 2.1.2 imply that $P(\mathbb{G} \in \mathcal{C}(\mathcal{F}, L_{2,P}) = 1$. □

C.2 Conditional Distributions, $T_1(P) = (\{P_{1|x}, P_{0|x}, \eta_{1,x}, \eta_{0,x}, s_x\}_{x\in\mathcal{X}})$

Lemma 2.1 shows that the distributions of $Y_d \mid D_1 > D_0, X = x$, denoted $P_{d|x}$, are identified by

$$
\begin{aligned}
P_{d|x}(f) &= E_{P_{d|x}}[f(Y_d)] = E[f(Y_d) \mid D_1 > D_0, X = x] \\
&= \frac{E[f(Y)\mathbb{1}\{D = d\} \mid Z = d, X = x] - E[f(Y)\mathbb{1}\{D = d\} \mid Z = 1-d, X = x]}{P(D = d \mid Z = d, X = x) - P(D = d \mid Z = 1-d, X = x)}
\end{aligned}
$$

and the distribution of X conditional on $D_1 > D_0$ is identified by

$$
\begin{aligned}
s_x &= P(X = x \mid D_1 > D_0) \\
&= \frac{[P(D = 1 \mid Z = 1, X = x) - P(D = 1 \mid Z = 0, X = x)]\, P(X = x)}{\sum_{x'} [P(D = 1 \mid Z = 1, X = x') - P(D = 1 \mid Z = 0, X = x')]\, P(X = x')}
\end{aligned}
$$

Recall the notation shortening indicators

$$
\mathbb{1}_{d,x,z}(D, X, Z) = \mathbb{1}\{D = d, X = x, Z = z\}, \quad \mathbb{1}_{x,z}(X, Z) = \mathbb{1}\{X = x, Z = z\}, \quad \mathbb{1}_x(X) = \mathbb{1}\{X = x\}
$$

and notice that $P_{d|x} : \ell^\infty(\mathcal{F}_d) \to \mathbb{R}$ and $s_x \in \mathbb{R}$, given by

$$
\begin{aligned}
P_{d|x}(f) &= \frac{P(\mathbb{1}_{d,x,d} \times f)/P(\mathbb{1}_{x,d}) - P(\mathbb{1}_{d,x,1-d} \times f)/P(\mathbb{1}_{x,0})}{P(\mathbb{1}_{d,x,d})/P(\mathbb{1}_{x,d}) - P(\mathbb{1}_{d,x,1-d})/P(\mathbb{1}_{x,1-d})}, \\
s_x &= \frac{[P(\mathbb{1}_{1,x,1})/P(\mathbb{1}_{x,1}) - P(\mathbb{1}_{1,x,0})/P(\mathbb{1}_{x,0})]P(\mathbb{1}_x)}{\sum_{x'}[P(\mathbb{1}_{1,x',1})/P(\mathbb{1}_{x',1}) - P(\mathbb{1}_{1,x',0})/P(\mathbb{1}_{x',0})]P(\mathbb{1}_{x'})},
\end{aligned}
$$

are functions of $P \in \ell^\infty(\mathcal{F})$. Moreover, $\eta_{d,x}^{(k)} = E[\eta_d^{(k)}(Y_d) \mid D_1 > D_0, X = x] = P_{d|x}(\eta_d^{(k)})$ and $\eta_{d,x} = (\eta_{d,x}^{(1)}, \ldots, \eta_{d,x}^{(K_1)})$ is simply an evaluation of $P_{d|x}$ at the points $\eta_d^{(k)} \in \mathcal{F}_{d,x}$.

This map is given by

$$
\begin{aligned}
T_1 &: \mathbb{D}_C \subseteq \ell^\infty(\mathcal{F}) \to \prod_{m=1}^{M} \ell^\infty(\mathcal{F}_{1,x_m}) \times \ell^\infty(\mathcal{F}_{0,x_m}) \times \mathbb{R} \times \mathbb{R}^{(K_1)} \times \mathbb{R}^{(K_0)} \\
T_1(P) &= \left(\left\{P_{1|x}, P_{0|x}, \eta_{1,x}, \eta_{0,x}, s_x\right\}_{x\in\mathcal{X}}\right) \\
&= (P_{1|x_1}, P_{0|x_1}, \eta_{1,x_1}, \eta_{0,x_1}, s_{x_1}, \ldots, P_{1|x_M}, P_{0|x_M}, \eta_{1,x_M}, \eta_{0,x_M}, s_{x_M})
\end{aligned}
$$

where the domain, $\mathbb{D}_C \subseteq \ell^\infty(\mathcal{F})$, ensures the map never divide by zero:

$$\mathbb{D}_C = \big\{G \in \ell^\infty(\mathcal{F}) \ ; \ \text{ for all } (d,x,z),\ G(\mathbb{1}_x) > 0,\ G(\mathbb{1}_{x,z}) > 0, \text{ and} \\ G(\mathbb{1}_{d,x,d})/G(\mathbb{1}_{x,d}) - G(\mathbb{1}_{d,x,1-d})/G(\mathbb{1}_{x,1-d}) > 0\big\} \tag{59}$$

Note that assumption 1 implies $P \in \mathbb{D}_C$, a claim shown in the proof of lemma C.7 below.

Lemma F.5 shows that Hadamard differentiable functions with the same domain can be "stacked". Moreover, the coordinates corresponding to the η terms are evaluations of the $P_{d|x}$ at specific coordinates; since evaluation is linear and continuous, the map defining these terms is fully Hadamard differentiable if the other maps are fully Hadamard differentiable. Thus it suffices to ensure the maps $C_{d,x} : \mathbb{D}_C \to \mathbb{R}$ and $C_{s,x} : \mathbb{D}_C \to \mathbb{R}$ given by $C_{d,x}(P) = P_{d|x}$ and $C_{s,x}(P) = s_x$ are fully Hadamard differentiable at P tangentially to $\ell^\infty(\mathcal{F})$.

Lemma C.6 (Maps to conditional distributions are fully Hadamard differentiable)**.** *Let $\mathcal{F}$ be defined by* (58)*, and $\mathbb{D}_C$ be defined by* (59)*. Define the functions $C_{1,x}$, $C_{0,x}$, and $C_{s,x}$ with*

$$C_{d,x} : \mathbb{D}_C \to \ell^\infty(\mathcal{F}_{d,x}), \qquad C_{d,x}(G)(f) = \frac{G(\mathbb{1}_{d,x,d} \times f)/G(\mathbb{1}_{x,d}) - G(\mathbb{1}_{d,x,1-d} \times f)/G(\mathbb{1}_{x,1-d})}{G(\mathbb{1}_{d,x,d})/G(\mathbb{1}_{x,d}) - G(\mathbb{1}_{d,x,1-d})/G(\mathbb{1}_{x,1-d})},$$

$$C_{s,x} : \mathbb{D}_C \to \mathbb{R}, \qquad C_{s,x}(G) = \frac{[G(\mathbb{1}_{1,x,1})/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0})/G(\mathbb{1}_{x,0})]G(\mathbb{1}_x)}{\sum_{x'}[G(\mathbb{1}_{1,x',1})/G(\mathbb{1}_{x',1}) - G(\mathbb{1}_{1,x',0})/G(\mathbb{1}_{x',0})]G(\mathbb{1}_{x'})}$$

All three functions are fully Hadamard differentiable at any $G \in \mathbb{D}_C$ tangentially to $\ell^\infty(\mathcal{F})$, with derivatives $C'_{d,x,G} : \ell^\infty(\mathcal{F}) \to \ell^\infty(\mathcal{F}_{d,x})$ and $C'_{s,x,G} : \ell^\infty(\mathcal{F}) \to \mathbb{R}$ described in the proof.

Proof. In steps:

1. We first show differentiability of $C_{1,x}$. The argument applies the chain rule. An inner function "rearranges" elements of $\mathbb{D}_C \subseteq \ell^\infty(\mathcal{F})$, which can be viewed as a fully Hadamard differentiable mapping (see lemma F.6). An outer function maps that rearrangment to $\ell^\infty(\mathcal{F}_1)$, and is shown fully Hadamard differentiable at $G \in \mathbb{D}_C$ by applying corollary F.8.

   In detailed steps:

   (a) Define $\mathbb{D}_q = \big\{(n_1, p_{11}, p_1, n_0, p_{10}, p_0) \in \mathbb{R}^6 \ ; \ p_1 > 0,\ p_0 > 0,\ p_{11}/p_1 - p_{10}/p_0 > 0\big\}$ and

$$q : \mathbb{D}_q \to \mathbb{R}, \qquad q(n_1, p_{11}, p_1, n_0, p_{10}, p_0) = \frac{n_1/p_1 - n_0/p_0}{p_{11}/p_1 - p_{10}/p_0}$$

Recall the following notation from corollary F.8:

$$\ell^\infty(\mathcal{F}_1, \mathbb{D}_q) = \left\{ r : \mathcal{F}_1 \to \mathbb{R}^6 \; ; \; r(\varphi) \in \mathbb{D}_q, \; \sup_{\varphi \in \mathcal{F}_1} \|r(f)\| < \infty \right\} \subseteq \ell^\infty(\mathcal{F}_1)^6$$

$$\ell_q^\infty(\mathcal{F}_1, \mathbb{D}_q) = \left\{ r \in \ell^\infty(\mathcal{F}_1, \mathbb{D}_q) \; ; \; \sup_{f \in \mathcal{F}_1} |q(r(f))| < \infty \right\}$$

For elements $r \in \ell^\infty(\mathcal{F}_1, \mathbb{D}_q)$, the composition $q(r(\varphi))$ is well defined for any $\varphi \in \mathcal{F}_1$. For elements $r \in \ell_q^\infty(\mathcal{F}_1, \mathbb{D}_q)$, composition defines a bounded map; that is, $\varphi \mapsto q(r(\varphi))$ defines an element of $\ell^\infty(\mathcal{F}_1)$. Finally, define

$$Q : \ell_q^\infty(\mathcal{F}_1, \mathbb{D}_q) \to \ell^\infty(\mathcal{F}_1), \qquad\qquad Q(r)(\varphi) = q(r(\varphi))$$

(b) For the rearrangement, define $\tilde{\mathcal{F}}_{1,x,1} = \{\mathbb{1}_{1,x,1} \times f \; ; \; f \in \mathcal{F}_1\}$, $\tilde{\mathcal{F}}_{1,x,0} = \{\mathbb{1}_{1,x,0} \times f \; ; \; f \in \mathcal{F}_1\}$, and

$$\begin{aligned}
&\tilde{R}_{1,x} : \mathbb{D}_C \to \ell^\infty(\tilde{\mathcal{F}}_{1,x,1}) \times \ell^\infty(\{\mathbb{1}_{1,x,1}\}) \times \ell^\infty(\{\mathbb{1}_{x,1}\}) \times \ell^\infty(\tilde{\mathcal{F}}_{1,x,0}) \times \ell^\infty(\{\mathbb{1}_{1,x,0}\}) \times \ell^\infty(\{\mathbb{1}_{x,0}\}) \\
&\tilde{R}_{1,x}(G)(\mathbb{1}_{1,x,1} \times f, \mathbb{1}_{1,x,1}, \mathbb{1}_{x,1}, \mathbb{1}_{1,x,0} \times f, \mathbb{1}_{1,x,0}, \mathbb{1}_{x,0}) \\
&\qquad = (G(\mathbb{1}_{1,x,1} \times f), G(\mathbb{1}_{1,x,1}), G(\mathbb{1}_{x,1}), G(\mathbb{1}_{1,x,0} \times f), G(\mathbb{1}_{1,x,0}), G(\mathbb{1}_{x,0}))
\end{aligned}$$

Lemma F.6 shows that $\tilde{R}_{1,x}$ is fully Hadamard differentiable tangentially to $\ell^\infty(\mathcal{F})$ and is its own derivative; i.e. $\tilde{R}'_{1,x,g} = \tilde{R}_{1,x}$. Now view $\tilde{R}_{1,x}$ as a map from $\mathbb{D}_C \subseteq \ell^\infty(\mathcal{F})$ to $\ell_q^\infty(\mathcal{F}_1, \mathbb{D}_q)$, i.e. define $R_{1,x} : \mathbb{D}_C \to \ell_q^\infty(\mathcal{F}_1, \mathbb{D}_q)$ pointwise with

$$\begin{aligned}
R_{1,x}(G)(f) &= \tilde{R}_{1,x}(G)(\mathbb{1}_{1,x,1} \times f, \mathbb{1}_{1,x,1}, \mathbb{1}_{x,1}, \mathbb{1}_{1,x,0} \times g, \mathbb{1}_{1,x,0}, \mathbb{1}_{x,0}) \\
&= (G(\mathbb{1}_{1,x,1} \times f), G(\mathbb{1}_{1,x,1}), G(\mathbb{1}_{x,1}), G(\mathbb{1}_{1,x,0} \times f), G(\mathbb{1}_{1,x,0}), G(\mathbb{1}_{x,0}))
\end{aligned}$$

Note that $G \in \mathbb{D}_C$ implies

$$\sup_{f \in \mathcal{F}_1} |q(R_{1,x}(G)(f))| = \sup_{f \in \mathcal{F}_1} \left| \frac{G(\mathbb{1}_{1,x,1} \times f)/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0} \times f)/G(\mathbb{1}_{x,0})}{G(\mathbb{1}_{1,x,1})/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0})/G(\mathbb{1}_{x,0})} \right| < \infty$$

and thus $R_{1,x}(G) \in \ell_q^\infty(\mathcal{F}_1, \mathbb{D}_q)$.

(c) To apply corollary F.8, observe that $q(n_1, p_{11}, p_1, n_0, p_{10}, p_0) = \frac{n_1/p_1 - n_0/p_0}{p_{11}/p_1 - p_{10}/p_0}$ is continuously differentiable on $\mathbb{D}_q$ with gradient $\nabla q : \mathbb{D}_q \to \mathbb{R}^6$ given by

$$
\begin{aligned}
&\nabla q(n_1, p_{11}, p_1, n_0, p_{10}, p_0) = \left(\tfrac{\partial q}{\partial n_1},\ \ \tfrac{\partial q}{\partial p_{11}},\ \ \tfrac{\partial q}{\partial p_1},\ \ \tfrac{\partial q}{\partial n_0},\ \ \tfrac{\partial q}{\partial p_{10}},\ \ \tfrac{\partial q}{\partial p_0}\right)^\intercal, \\
&\frac{\partial q}{\partial n_1} = \frac{1/p_1}{p_{11}/p_1 - p_{10}/p_0} \\
&\frac{\partial q}{\partial p_{11}} = -\frac{n_1/p_1 - n_0/p_0}{(p_{11}/p_1 - p_{10}/p_0)^2}\frac{1}{p_1} = \left[\frac{1/p_1}{p_{11}/p_1 - p_{10}/p_0}\right](-q) \\
&\frac{\partial q}{\partial p_1} = \frac{(p_{11}/p_1 - p_{10}/p_0)(-n_1/p_1^2) - (n_1/p_1 - n_0/p_0)(-p_{11}/p_1^2)}{(p_{11}/p_1 - p_{10}/p_0)^2} \\
&\qquad = \frac{-n_1/p_1^2}{p_{11}/p_1 - p_{10}/p_0} + \frac{q(p_{11}/p_1^2)}{p_{11}/p_1 - p_{10}/p_0} = \left[\frac{1/p_1}{p_{11}/p_1 - p_{10}/p_0}\right]\frac{qp_{11} - n_1}{p_1} \\
&\frac{\partial q}{\partial n_0} = \frac{-1/p_0}{p_{11}/p_1 - p_{10}/p_0} \\
&\frac{\partial q}{\partial p_{10}} = -\frac{n_1/p_1 - n_0/p_0}{(p_{11}/p_1 - p_{10}/p_0)^2}\left(-\frac{1}{p_0}\right) = \left[\frac{-1/p_0}{p_{11}/p_1 - p_{10}/p_0}\right](-q) \\
&\frac{\partial q}{\partial p_0} = \frac{(p_{11}/p_1 - p_{10}/p_0)(n_0/p_0^2) - (n_1/p_1 - n_0/p_0)(p_{10}/p_0^2)}{(p_{11}/p_1 - p_{10}/p_0)^2} \\
&\qquad = \frac{n_0/p_0^2}{p_{11}/p_1 - p_{10}/p_0} - \frac{q(p_{10}/p_0^2)}{p_{11}/p_1 - p_{10}/p_0} = \left[\frac{-1/p_0}{p_{11}/p_1 - p_{10}/p_0}\right]\frac{qp_{10} - n_0}{p_0}
\end{aligned}
$$

Furthermore, there exists $\delta > 0$ such that

$$
R_{1,x}(G)(\mathcal{F}_1) = \left\{ r \in \mathbb{R}^6\ ;\ \inf_{f \in \mathcal{F}_1} \|r - R_{1,x}(G)(\varphi)\| \le \delta \right\} \subseteq \mathbb{D}_q
$$

and so lemma F.8 implies Q is fully Hadamard differentiable at $R_{1,x}(G)$ tangentially to $\ell^\infty(\mathcal{F}_1)^6$ with derivative $Q'_{R_{1,x}(G)} : \ell^\infty(\mathcal{F}_1)^6 \to \ell^\infty(\mathcal{F}_1)$ given pointwise by

$$
Q'_{R_{1,x}(G)}(J)(f) = [\nabla q(R_{1,x}(G)(\varphi))]^\intercal J(f)
$$

(d) Finally, observe that $C_{1,x}(G) = Q(R_{1,x}(G))$ and apply the chain rule (lemma F.4) to find that $C_{1,x}$ is fully Hadamard differentiable at G tangentially to $\ell^\infty(\mathcal{F})$ with derivative

$$
C'_{1,x,G} : \ell^\infty(\mathcal{F}) \to \ell^\infty(\mathcal{F}_{1,x}), \qquad C'_{1,x,G}(H) = Q'_{R_{1,x}(G)}(R_{1,x}(H))
$$

Writing out an evaluation clarifies the notation of the derivative:

$$
\begin{aligned}
C'_{1,x,G}(H)(f) &= Q'_{R_{1,x}(G)}(R_{1,x}(H))(f) = [\nabla q(R_{1,x}(G)(f))]^\intercal R_{1,x}(H)(f) \qquad (60)\\
&= \left[\frac{1/G(\mathbb{1}_{x,1})}{G(\mathbb{1}_{1,x,1})/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0})/G(\mathbb{1}_{x,0})}\right] H(\mathbb{1}_{1,x,1}\times f)\\
&\quad + \left[\frac{1/G(\mathbb{1}_{x,1})}{G(\mathbb{1}_{1,x,1})/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0})/G(\mathbb{1}_{x,0})}\right] (-C_{1,x}(G)(f))H(\mathbb{1}_{1,x,1})\\
&\quad + \left[\frac{1/G(\mathbb{1}_{x,1})}{G(\mathbb{1}_{1,x,1})/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0})/G(\mathbb{1}_{x,0})}\right] \frac{C_{1,x}(G)(f)\times G(\mathbb{1}_{1,x,1}) - G(\mathbb{1}_{1,x,1}\times f)}{G(\mathbb{1}_{x,1})} H(\mathbb{1}_{x,1})\\
&\quad + \left[\frac{-1/G(\mathbb{1}_{x,0})}{G(\mathbb{1}_{1,x,1})/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0})/G(\mathbb{1}_{x,0})}\right] H(\mathbb{1}_{1,x,0}\times f)\\
&\quad + \left[\frac{-1/G(\mathbb{1}_{x,0})}{G(\mathbb{1}_{1,x,1})/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0})/G(\mathbb{1}_{x,0})}\right] (-C_{1,x}(G)(f))H(\mathbb{1}_{1,x,0})\\
&\quad + \left[\frac{-1/G(\mathbb{1}_{x,0})}{G(\mathbb{1}_{1,x,1})/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0})/G(\mathbb{1}_{x,0})}\right] \frac{C_{1,x}(G)(f)\times G(\mathbb{1}_{1,x,0}) - G(\mathbb{1}_{1,x,0}\times f)}{G(\mathbb{1}_{x,0})} H(\mathbb{1}_{x,0})
\end{aligned}
$$

2. The same arguments imply the claim regarding $C_{0,x}$.

   Specifically, notice that $C_{0,x}$ is the same outer transformation applied to a different rearrangement: let

$$
\begin{aligned}
R_{1,x}(G)(\varphi) &= (G(\mathbb{1}_{1,x,1}\times\varphi), G(\mathbb{1}_{1,x,1}), G(\mathbb{1}_{x,1}), G(\mathbb{1}_{1,x,0}\times\varphi), G(\mathbb{1}_{1,x,0}), G(\mathbb{1}_{x,0}))\\
R_{0,x}(G)(\varphi) &= (G(\mathbb{1}_{0,x,0}\times\psi), G(\mathbb{1}_{0,x,0}), G(\mathbb{1}_{x,0}), G(\mathbb{1}_{0,x,1}\times\psi), G(\mathbb{1}_{0,x,1}), G(\mathbb{1}_{x,1}))
\end{aligned}
$$

   observe that

$$
\begin{aligned}
C_{1,x}(G)(f) &= \frac{G(\mathbb{1}_{1,x,1}\times f)/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0}\times f)/G(\mathbb{1}_{x,0})}{G(\mathbb{1}_{1,x,1})/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0})/G(\mathbb{1}_{x,0})} = q(R_{1,x}(G)(f))\\
C_{0,x}(G)(f) &= \frac{G(\mathbb{1}_{0,x,0}\times f)/G(\mathbb{1}_{x,0}) - G(\mathbb{1}_{0,x,1}\times f)/G(\mathbb{1}_{x,1})}{G(\mathbb{1}_{0,x,0})/G(\mathbb{1}_{x,0}) - G(\mathbb{1}_{0,x,1})/G(\mathbb{1}_{x,1})} = q(R_{0,x}(G)(f))
\end{aligned}
$$

   Thus, the same argument shows $C_{0,x} : \mathbb{D}_C \to \ell^\infty(\mathcal{F}_{0,x})$ is fully Hadamard differentiable at any $G \in \mathbb{D}_C$ tangentially to $\ell^\infty(\mathcal{F})$, and $C'_{0,x,G}(H)(f)$ can be found with the appropriate substitutions in (60) above.

3. Finally consider $C_{s,x}$. Notice that

$$
\begin{aligned}
\mathbb{D}_{q_{s,x}} = \Big\{&\{p_{1,x,1}, p_{x,1}, p_{1,x,0}, p_{x,0}, p_x\}_{x\in\mathcal{X}} \in \mathbb{R}^{5M}\ ;\\
&\qquad p_{x,1} > 0, p_{x,0} > 0, p_{1,x,1}/p_{x,1} - p_{1,x,0}/p_{x,0} > 0, p_x > 0 \text{ for all } x \in \mathcal{X}\Big\}
\end{aligned}
$$

$$
q_{s,x} : \mathbb{D}_{q_{s,x}} \to \mathbb{R},
$$

$$
q_{s,x}(\{p_{1,x_m,1}, p_{x_m,1}, p_{1,x_m,0}, p_{x_m,0}\}_{m=1}^M) = \frac{(p_{1,x,1}/p_{x,1} - p_{1,x,0}/p_{x,0})p_x}{\sum_{m=1}^M (p_{1,x_m,1}/p_{x_m,1} - p_{1,x_m,0}/p_{x_m,0})p_{x_m}}
$$

   is continuously differentiable at any point in $\mathbb{D}_{q_{s,x}}$ with gradient

$$
\nabla q(\{p_{1,x_m,1}, p_{x_m,1}, p_{1,x_m,0}, p_{x_m,0}, p_{x_m}\}_{m=1}^M) \in \mathbb{R}^{5M}
$$

Furthermore, notice that for any $G \in \mathbb{D}_C$, $C_{s,x}(G) = q_{s,x}(R_{s,x}(G))$, where

$$R_{s,x} : \ell^\infty(\mathcal{F}) \to \mathbb{R}^{5M}, \quad R_{s,x}(G) = (\{G(\mathbb{1}_{1,x_m,1}), G(\mathbb{1}_{x_m,1}), G(\mathbb{1}_{1,x_m,0}), G(\mathbb{1}_{x_m,0}), G(\mathbb{1}_{x_m})\}_{m=1}^{M})$$

It follows that $C_{s,x} : \mathbb{D}_C \to \mathbb{R}$ is fully Hadamard differentiable at any $G \in \mathbb{D}_C$ tangentially to $\ell^\infty(\mathcal{F})$. The derivative is

$$\begin{aligned} C'_{s,x,G}(H) = \sum_{m=1}^{M} & \frac{\partial q_{s,x}}{\partial p_{1,x_m,1}}(R_{s,x}(G)) \times H(\mathbb{1}_{1,x_m,1}) + \frac{\partial q_{s,x}}{\partial p_{x_m,1}}(R_{s,x}(G)) \times H(\mathbb{1}_{x_m,1}) \\ & + \frac{\partial q_{s,x}}{\partial p_{1,x_m,0}}(R_{s,x}(G)) \times H(\mathbb{1}_{1,x_m,0}) + \frac{\partial q_{s,x}}{\partial p_{x_m,0}}(R_{s,x}(G)) \times H(\mathbb{1}_{x_m,0}) \\ & + \frac{\partial q_{s,x}}{\partial p_{x_m}}(R_{s,x}(G)) \times H(\mathbb{1}_{x_m}) \end{aligned}$$

This completes the proof. ☐

Lemma C.7 (T_1 is fully Hadamard differentiable)**.** *Let $\mathcal{F}$ be defined by* (58) *and $\mathbb{D}_C$ by* (59)*. Let $C_{d,x}$ and $C_{s,x}$ be as defined in lemma C.6, and*

$$\tilde{\eta}_{d,x} : \mathbb{D}_C \to \mathbb{R}^{K_d}, \qquad \tilde{\eta}_{d,x}(G) = \left(C_{d,x}(G)(\eta_{d,x}^{(1)}), \ldots, C_{d,x}(G)(\eta_{d,x}^{(K_d)})\right)$$

Further define

$$\begin{aligned} T_1 : \mathbb{D}_C &\to \prod_{m=1}^{M} \ell^\infty(\mathcal{F}_{1,x_m}) \times \ell^\infty(\mathcal{F}_{0,x_m}) \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R} \\ T_1(G) &= \left(\{C_{1,x}(G), C_{0,x}(G), \tilde{\eta}_{1,x}(G), \tilde{\eta}_{0,x}(G), C_{s,x}(G)\}_{x \in \mathcal{X}}\right) \end{aligned}$$

T_1 is fully Hadamard differentiable at any $G \in \mathbb{D}_C$ tangentially to $\ell^\infty(\mathcal{F})$.

Proof. Lemma C.6 shows that $C_{d,x}$ and $C_{s,x}$ are fully Hadamard differentiable at any $G \in \mathbb{D}_C$ tangentially to $\ell^\infty(\mathcal{F})$.

Define the evaluation maps

$$ev_{\eta_d^{(k)}} : \ell^\infty(\mathcal{F}_{d,x}) \to \mathbb{R}, \qquad ev_{\eta_d^{(k)}}(H) = H(\eta_d^{(k)})$$

Note that each $ev_{\eta_d^{(k)}}$ is continuous and linear, and is therefore fully Hadamard differentiable at any $H \in \ell^\infty(\mathcal{F}_{d,x})$ tangentially to $\ell^\infty(\mathcal{F}_{d,x})$ (and is its own derivative). Moreover,

$$\tilde{\eta}_{d,x}(G) = (ev_{\eta_d^{(1)}}(C_{d,x}(G)), \ldots, ev_{\eta_d^{(K_1)}}(C_{d,x}(G)))$$

is the composition of an inner function that is fully Hadamard differentiable at any $G \in \mathbb{D}_C$, and an other function that is fully differentiable at any $H \in \ell^\infty(\mathcal{F}_{d,x})$. Therefore $\tilde{\eta}_{d,x}$ is fully Hadamard differentiable at any $G \in \mathbb{D}_C$ tangentially to $\ell^\infty(\mathcal{F})$.

Next apply lemma F.5 to find that

$$T_1 : \mathbb{D}_C \to \prod_{m=1}^{M} \ell^\infty(\mathcal{F}_{1,x_m}) \times \ell^\infty(\mathcal{F}_{0,x_m}) \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}$$
$$T_1(G) = \left(\{C_{1,x}(G), C_{0,x}(G), \tilde{\eta}_{1,x}(G), \tilde{\eta}_{0,x}(G), C_{s,x}(G)\}_{x\in\mathcal{X}}\right)$$

is fully Hadamard differentiable at any $G \in \mathbb{D}_C$ tangentially to $\ell^\infty(\mathcal{F})$. □

C.2.1 Support of the weak limit of $\sqrt{n}(T_1(\mathbb{P}_n) - T_1(P))$

The next few lemmas study the support of the asymptotic distribution of $\sqrt{n}(T_1(\mathbb{P}_n) - T_1(P))$; in particular, it concentrates on the tangent set of the next map studied in appendix C.3.

Lemma C.8 (Continuity of $C'_{d,x,G}(H)(\cdot)$)**.** *Let $C_{d,x}$ be as defined in lemma C.6. If $G, H \in \mathcal{C}(\mathcal{F}, L_{2,P})$, then $C'_{d,x,G}(H) \in \mathcal{C}(\mathcal{F}_{d,x}, L_{2,P})$.*

Proof. Consider $C'_{1,x,G}(H)$ first. Fix $f \in \mathcal{F}_{1,x}$ and let $\varepsilon > 0$. Let

$$\text{Coef}_1(G) = \left[\frac{1/G(\mathbb{1}_{x,1})}{G(\mathbb{1}_{1,x,1})/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0})/G(\mathbb{1}_{x,0})}\right]$$
$$\text{Coef}_2(G) = \left[\frac{-1/G(\mathbb{1}_{x,0})}{G(\mathbb{1}_{1,x,1})/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0})/G(\mathbb{1}_{x,0})}\right]$$

and use display (60) to see that

$$\begin{aligned}
&|C'_{1,x,G}(H)(f) - C'_{1,x,G}(H)(g)| \\
&= \Bigg|\text{Coef}_1(G) \times [H(\mathbb{1}_{1,x,1} \times f) - H(\mathbb{1}_{1,x,1} \times g)] \\
&\quad + \text{Coef}_1(G) \times (-[C_{1,x}(G)(f) - C_{1,x}(G)(g)])\, H(\mathbb{1}_{1,x,1}) \\
&\quad + \text{Coef}_1(G) \times \frac{[C_{1,x}(G)(f) - C_{1,x}(G)(g)] \times G(\mathbb{1}_{1,x,1}) - [G(\mathbb{1}_{1,x,1} \times f) - G(\mathbb{1}_{1,x,1} \times g)]}{G(\mathbb{1}_{x,1})} H(\mathbb{1}_{x,1}) \\
&\quad + \text{Coef}_2(G) \times [H(\mathbb{1}_{1,x,0} \times f) - H(\mathbb{1}_{1,x,0} \times g)] \\
&\quad + \text{Coef}_2(G) \times (-[(C_{1,x}(G)(f)) - C_{1,x}(G)(g)])\, H(\mathbb{1}_{1,x,0}) \\
&\quad + \text{Coef}_2(G) \times \frac{[C_{1,x}(G)(f) - C_{1,x}(G)(g)] \times G(\mathbb{1}_{1,x,0}) - [G(\mathbb{1}_{1,x,0} \times f) - G(\mathbb{1}_{1,x,0} \times g)]}{G(\mathbb{1}_{x,0})} H(\mathbb{1}_{x,0})\Bigg|
\end{aligned}$$

Recall that $C_{1,x}(G)(f) = \frac{G(\mathbb{1}_{1,x,1}\times f)/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0}\times f)/G(\mathbb{1}_{x,0})}{G(\mathbb{1}_{1,x,1})/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0})/G(\mathbb{1}_{x,0})}$, and thus

$$C_{1,x}(G)(f) - C_{1,x}(G)(g) = \frac{[G(\mathbb{1}_{1,x,1} \times f) - G(\mathbb{1}_{1,x,1} \times g)]/G(\mathbb{1}_{x,1}) - [G(\mathbb{1}_{1,x,0} \times f) - G(\mathbb{1}_{1,x,0} \times g)]/G(\mathbb{1}_{x,0})}{G(\mathbb{1}_{1,x,1})/G(\mathbb{1}_{x,1}) - G(\mathbb{1}_{1,x,0})/G(\mathbb{1}_{x,0})}$$

use this to see that

$$\begin{aligned}
&|C'_{1,x,G}(H)(f) - C'_{1,x,G}(H)(g)| \\
&\quad \leq A_1 \times |H(\mathbb{1}_{1,x,1} \times f) - H(\mathbb{1}_{1,x,1} \times g)| + A_2 \times |G(\mathbb{1}_{1,x,1} \times f) - G(\mathbb{1}_{1,x,1} \times g)| \\
&\qquad + A_3 \times |H(\mathbb{1}_{1,x,0} \times f) - H(\mathbb{1}_{1,x,0} \times g)| + A_4 \times |G(\mathbb{1}_{1,x,0} \times f) - G(\mathbb{1}_{1,x,0} \times g)| \qquad (61)
\end{aligned}$$

for finite constants A_1, A_2, A_3, and A_4 that depend on G and H, but not on f or g. Now use $G, H \in \mathcal{C}(\mathcal{F}, L_{2,P})$ to choose $\delta_{z,H} > 0$ and $\delta_{z,G} > 0$ such that

$$\begin{aligned}
L_{2,P}(\mathbb{1}_{1,x,1} \times f, \mathbb{1}_{1,x,1} \times g) < \delta_{1,H} \quad &\Longrightarrow \quad |H(\mathbb{1}_{1,x,1} \times f) - H(\mathbb{1}_{1,x,1} \times g)| < \varepsilon/(4A_1) \\
L_{2,P}(\mathbb{1}_{1,x,1} \times f, \mathbb{1}_{1,x,1} \times g) < \delta_{1,G} \quad &\Longrightarrow \quad |G(\mathbb{1}_{1,x,1} \times f) - G(\mathbb{1}_{1,x,1} \times g)| < \varepsilon/(4A_2) \\
L_{2,P}(\mathbb{1}_{1,x,0} \times f, \mathbb{1}_{1,x,0} \times g) < \delta_{0,H} \quad &\Longrightarrow \quad |H(\mathbb{1}_{1,x,0} \times f) - H(\mathbb{1}_{1,x,0} \times g)| < \varepsilon/(4A_3) \\
L_{2,P}(\mathbb{1}_{1,x,0} \times f, \mathbb{1}_{1,x,0} \times g) < \delta_{0,G} \quad &\Longrightarrow \quad |G(\mathbb{1}_{1,x,0} \times f) - G(\mathbb{1}_{1,x,0} \times g)| < \varepsilon/(4A_4) \qquad (62)
\end{aligned}$$

Finally, notice that

$$\begin{aligned}
L_{2,P}(\mathbb{1}_{1,x,z} \times f, \mathbb{1}_{1,x,z} \times g) &= \sqrt{P((\mathbb{1}_{1,x,z} \times f - \mathbb{1}_{1,x,z} \times g)^2)} = \sqrt{P(\mathbb{1}_{1,x,z} \times (f-g)^2)} \\
&\leq \sqrt{P((f-g)^2)} = L_{2,P}(f,g) \qquad (63)
\end{aligned}$$

It follows from (61), (62), and (63) that

$$L_{2,P}(f,g) < \min\{\delta_{1,H}, \delta_{1,G}, \delta_{0,H}, \delta_{0,G}\} \qquad \Longrightarrow \qquad |C'_{1,x,G}(H)(f) - C'_{1,x,G}(H)(g)| < \varepsilon$$

i.e., $C'_{1,x,G}(H)(\cdot)$ is continuous at f. Since $f \in \mathcal{F}_{1,x}$ and $G, H \in \mathcal{C}(\mathcal{F}, L_{2,P})$ were arbitrary, this shows that $G, H \in \mathcal{C}(\mathcal{F}, L_{2,P})$ implies $C'_{1,x,G}(H) \in \mathcal{C}(\mathcal{F}_{1,x}, L_{2,P})$.

The same argument shows that $G, H \in \mathcal{C}(\mathcal{F}, L_{2,P})$ implies $C'_{0,x,G}(H) \in \mathcal{C}(\mathcal{F}_{0,x}, L_{2,P})$. This completes the proof. □

Lemma C.9 (Support of $T'_{1,P}(\mathbb{G})$)**.** *Let $\mathcal{F}$ be defined by (58) and T_1 be as defined in lemma C.7.*

1. *If assumption 1 holds, $P \in \mathbb{D}_C$ and hence T_1 is fully Hadamard differentiable at P tangentially to $\ell^\infty(\mathcal{F})$.*
2. *If assumptions 1, 2, and 3 hold,*

$$\sqrt{n}(T_1(\mathbb{P}_n) - T_1(P)) \xrightarrow{L} T'_{1,P}(\mathbb{G})$$

where $\mathbb{G}$ is the Gaussian limit of $\sqrt{n}(\mathbb{P}_n - P)$ in $\ell^\infty(\mathcal{F})$ discussed in lemma C.5.

3. *If assumptions 1, 2, and 3, then $P(T'_{1,P}(\mathbb{G}) \in \mathbb{D}_{Tan,Full}) = 1$ where*

$$\mathbb{D}_{Tan,Full} = \prod_{m=1}^{M} \Big(\ell^\infty_{\mathcal{Y}_{1,x_m}}(\mathcal{F}_{1,x_m}) \times \ell^\infty_{\mathcal{Y}_{0,x_m}}(\mathcal{F}_{0,x_m})\Big) \cap \Big(\mathcal{C}(\mathcal{F}_{1,x_m}, L_{2,P}) \times \mathcal{C}(\mathcal{F}_{0,x_m}, L_{2,P})\Big) \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R} \qquad (64)$$

Proof. In steps:

1. $P \in \mathbb{D}_C$ and differentiability of T_1 at P.

   Assumption 1 implies $P \in \mathbb{D}_C$, given by (59). To see this, recall that assumption 1 (iv) is that $P(\mathbb{1}_{x,z}) = P(X = x, Z = z) > 0$ (implying $P(\mathbb{1}_x) = P(X = x) = P(X = x, Z = 1) + P(X = x, Z = 0) > 0$). Furthermore,

$$\begin{aligned} &P(\mathbb{1}_{d,x,d})/P(\mathbb{1}_{x,d}) - P(\mathbb{1}_{d,x,1-d})/P(\mathbb{1}_{x,1-d}) \\ &\qquad = P(D = d \mid X = x, Z = d) - P(D = d \mid X = x, Z = 1 - d) \\ &\qquad = P(D_1 > D_0 \mid X = x) > 0 \end{aligned}$$

   The second equality is shown in the proof of lemma 2.1, and the inequality is assumption 1 (iii). Lemma C.7 thus shows that T_1 is fully Hadamard differentiable at P tangentially to $\ell^\infty(\mathcal{F})$.

2. Functional delta method.

   Under assumptions 1, 2, and 4, lemma C.5 shows that $\sqrt{n}(\mathbb{P}_n - P) \xrightarrow{L} \mathbb{G}$ in $\ell^\infty(\mathcal{F})$. The functional delta method (Van der Vaart (2000) theorem 20.8) then implies

$$\sqrt{n}(T_1(\mathbb{P}_n) - T_1(P)) \xrightarrow{L} T'_{1,P}(\mathbb{G}), \qquad \text{in } \prod_{m=1}^{M} \ell^\infty(\mathcal{F}_{1,x_m}) \times \ell^\infty(\mathcal{F}_{0,x_m}) \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}$$

3. Support of $T'_{1,P}(\mathbb{G})$.

   Notice that $T'_P(\mathbb{G}) = \left(\left\{C'_{1,x,P}(\mathbb{G}), C'_{0,x,P}(\mathbb{G}), \tilde{\eta}'_{1,x,P}(\mathbb{G}), \tilde{\eta}'_{0,x,P}(\mathbb{G}), C'_{s,x,P}(\mathbb{G})\right\}_{x\in\mathcal{X}}\right)$, where $\tilde{\eta}_{d,x}$ are defined in lemma C.7. Let

$$\mathbb{S}_x = \left(\ell^\infty_{\mathcal{Y}_{1,x_m}}(\mathcal{F}_{1,x_m}) \times \ell^\infty_{\mathcal{Y}_{0,x_m}}(\mathcal{F}_{0,x_m})\right) \cap \left(\mathcal{C}(\mathcal{F}_{1,x_m}, L_{2,P}) \times \mathcal{C}(\mathcal{F}_{0,x_m}, L_{2,P})\right) \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}$$

   and note that it suffices to show $P\left(C'_{1,x,P}(\mathbb{G}), C'_{0,x,P}(\mathbb{G}), \tilde{\eta}'_{1,x,P}(\mathbb{G}), \tilde{\eta}'_{0,x,P}(\mathbb{G}), C'_{s,x,P}(\mathbb{G}) \in \mathbb{S}_x\right) = 1$ for each x. Moreover,

$$P\left(\left(\tilde{\eta}'_{1,x,P}(\mathbb{G}), \tilde{\eta}'_{0,x,P}(\mathbb{G}), C'_{s,x,P}(\mathbb{G})\right) \in \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}\right) = 1$$

   is immediate. To complete the proof we must show $P(C'_{d,x,P}(\mathbb{G}) \in \ell^\infty_{\mathcal{Y}_{d,x}}(\mathcal{F}_{d,x})) = P(C'_{d,x,P}(\mathbb{G}) \in \mathcal{C}(\mathcal{F}_{d,x}, L_{2,P})) = 1$.

   (a) To see that $P(C'_{d,x,P}(\mathbb{G}) \in \mathcal{C}(\mathcal{F}_{d,x}, L_{2,P})) = 1$, first note that for any functions $f_1, f_2 \in \mathcal{F}$,

$$|P(f_1) - P(f_2)| \leq P(|f_1 - f_2|) = P(\sqrt{(f_1 - f_2)}) \leq \sqrt{P((f_1 - f_2)^2)} = L_{2,P}(f_1, f_2)$$

   where the second inequality is an application of Jensen's inequality. Thus $P \in \mathcal{C}(\mathcal{F}, L_{2,P})$. Next apply lemma C.8 to see that if $G \in \mathcal{C}(\mathcal{F}, L_{2,P})$ then $C'_{d,x,P}(G) \in \mathcal{C}(\mathcal{F}_{d,x}, L_{2,P})$. It follows that

$$1 = P\left(\mathbb{G} \in \mathcal{C}(\mathcal{F}, L_{2,P})\right) \leq P\left(C'_{d,x,P}(\mathbb{G}) \in \mathcal{C}(\mathcal{F}_{d,x}, L_{2,P})\right)$$

   (b) To see that $P(C'_{d,x,P}(\mathbb{G}) \in \ell^\infty_{\mathcal{Y}_{d,x}}(\mathcal{F}_{d,x})) = 1$, we show that $P(\sqrt{n}(C_{d,x}(\mathbb{P}_n) - C_{d,x}(P)) \in \ell^\infty_{\mathcal{Y}_{d,x}}(\mathcal{F}_{d,x})) = 1$.

First recall the definition given in (54):

$$\ell^\infty_{\mathcal{Y}_{d,x}}(\mathcal{F}_{d,x}) = \Big\{H \in \ell^\infty(\mathcal{F}_{d,x})\ ;\ \text{ for all } a,b \in \mathbb{R} \text{ and } f,g \in \mathcal{F}_{d,x}, \\ H(f) = H(\mathbb{1}_{\mathcal{Y}_{d,x}} \times f),\ \text{if } a \in \mathcal{F}_{d,x} \text{ then } H(a) = 0,\ \text{and} \\ \text{if } af+bg \in \mathcal{F}_{d,x} \text{ then } H(af+bg) = aH(f) + bH(g)\Big\}$$

i. $\sqrt{n}(C_{d,x}(\mathbb{P}_n) - C_{d,x}(P))$ is linear and evaluates constants to zero.

   This follows because $C_{d,x}(\mathbb{P}_n)$ and $C_{d,x}(P)$ are linear and "return constants". To see this, recall that $C_{d,x}(P) \in \ell^\infty(\mathcal{F}_{d,x})$ is given pointwise by

$$C_{d,x}(P)(f) = \frac{P(\mathbb{1}_{d,x,d} \times f)/P(\mathbb{1}_{x,d}) - P(\mathbb{1}_{d,x,1-d} \times f)/P(\mathbb{1}_{x,1-d})}{P(\mathbb{1}_{d,x,d})/P(\mathbb{1}_{x,d}) - P(\mathbb{1}_{d,x,1-d})/P(\mathbb{1}_{x,1-d})}$$

   Use this to see that for any $a,b \in \mathbb{R}$ and $f,g \in \mathcal{F}_{d,x}$. if $af+bg \in \mathcal{F}_{d,x}$, then linearity of P implies $C_{d,x}(P)(af+bg) = aC_{d,x}(P)(f) + bC_{d,x}(P)(g)$ and $C_{d,x}(\mathbb{P}_n)(af+bg) = aC_{d,x}(\mathbb{P}_n)(f) + bC_{d,x}(\mathbb{P}_n)(g)$. Similarly, if $a \in \mathcal{F}_{d,x}$ is the constant function always returning a, then $C_{d,x}(P)(a) = a$. The same observations apply to $C_{d,x}(\mathbb{P}) \in \ell^\infty(\mathcal{F}_{d,x})$.
   Therefore

$$\begin{aligned}
&\sqrt{n}(C_{d,x}(\mathbb{P}_n) - C_{d,x}(P))(af+bg) \\
&\quad = \sqrt{n}(C_{d,x}(\mathbb{P}_n)(af+bg) = -C_{d,x}(P)(af+bg)) \\
&\quad = \sqrt{n}(aC_{d,x}(\mathbb{P}_n)(f) + bC_{d,x}(\mathbb{P}_n)(g) - aC_{d,x}(P)(f) - bC_{d,x}(P)(g)) \\
&\quad = a \times \sqrt{n}(C_{d,x}(\mathbb{P}_n) - C_{d,x}(P))(f) + b \times \sqrt{n}(C_{d,x}(\mathbb{P}_n) - C_{d,x}(P))(g)
\end{aligned}$$

   and furthermore, if $a \in \mathcal{F}_{d,x}$, then

$$\sqrt{n}(C_{d,x}(\mathbb{P}_n) - C_{d,x}(P))(a) = \sqrt{n}(a-a) = 0$$

ii. $C_{d,x}(P)$ "ignores values outside $\mathcal{Y}_{d,x}$"; i.e. $C_{d,x}(P)(f) = C_{d,x}(P)(\mathbb{1}_{\mathcal{Y}_{d,x}} \times f)$.
   To see this,

$$\begin{aligned}
&C_{d,x}(P)(f) \qquad (65)\\
&= \frac{E[f(Y)\mathbb{1}\{D=d\} \mid X=x, Z=d] - E[f(Y)\mathbb{1}\{D=d\} \mid X=x, Z=1-d]}{P(\mathbb{1}_{d,x,d})/P(\mathbb{1}_{x,d}) - P(\mathbb{1}_{d,x,1-d})/P(\mathbb{1}_{x,1-d})} \\
&= \frac{P(D=d \mid X=x, Z=d)E[f(Y) \mid D=d, X=x, Z=d]}{P(\mathbb{1}_{d,x,d})/P(\mathbb{1}_{x,d}) - P(\mathbb{1}_{d,x,1-d})/P(\mathbb{1}_{x,1-d})} \\
&\qquad - \frac{P(D=d \mid X=x, Z=1-d)E[f(Y) \mid D=d, X=x, Z=1-d]}{P(\mathbb{1}_{d,x,d})/P(\mathbb{1}_{x,d}) - P(\mathbb{1}_{d,x,1-d})/P(\mathbb{1}_{x,1-d})}
\end{aligned}$$

Since $\mathcal{Y}_{d,x}$ is the support of $Y \mid D = d, X = x$,

$$
\begin{aligned}
&E[f(Y) \mid D = d, X = x, Z = z] \\
&\qquad = E[f(Y)\mathbb{1}\{Z = z\} \mid D = d, X = x]/P(Z = z \mid D = d, X = x) \\
&\qquad = E[\mathbb{1}\{Y \in \mathcal{Y}_{d,x}\}f(Y)\mathbb{1}\{Z = z\} \mid D = d, X = x]/P(Z = z \mid D = d, X = x) \\
&\qquad = E[\mathbb{1}\{Y \in \mathcal{Y}_{d,x}\}f(Y) \mid D = d, X = x, Z = z]
\end{aligned}
$$

Along with (65), this implies $C_{d,m}(P)(f) = C_{d,m}(P)(\mathbb{1}_{\mathcal{Y}_{d,x}} \times f)$.

iii. Now notice that with probability one the sample is a subset of the support, and when this is so, $C_{d,x}(\mathbb{P}_n)$ ignores values outside of $\mathcal{Y}_{d,x}$.
Specifically, observe that

$$
\begin{aligned}
&C_{d,x}(\mathbb{P}_n)(f) \qquad (66)\\
&= \frac{\left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{D_i = d, X_i = x\}\mathbb{1}\{Z_i = d\}f(Y_i)\right] / \left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i = x, Z_i = d\}\right]}{\mathbb{P}_n(\mathbb{1}_{d,x,d})/\mathbb{P}_n(\mathbb{1}_{x,d}) - \mathbb{P}_n(\mathbb{1}_{d,x,1-d})/\mathbb{P}_n(\mathbb{1}_{x,1-d})} \\
&\qquad - \frac{\left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{D_i = d, X_i = x\}\mathbb{1}\{Z_i = 1-d\}f(Y_i)\right] / \left[\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{X_i = x, Z_i = 1-d\}\right]}{\mathbb{P}_n(\mathbb{1}_{d,x,d})/\mathbb{P}_n(\mathbb{1}_{x,d}) - \mathbb{P}_n(\mathbb{1}_{d,x,1-d})/\mathbb{P}_n(\mathbb{1}_{x,1-d})}
\end{aligned}
$$

Note that because $\mathcal{Y}_{d,x}$ is the support of $Y \mid D = d, X = x$, we have that with probability one, $\{Y_i, D_i, Z_i, X_i\}_{i=1}^{n} \subseteq \mathcal{S} := \bigcup_{d,z,x} \mathcal{Y}_{d,x} \times \{d\} \times \{z\} \times \{x\}$. Indeed, since $\mathcal{Y}_{d,x} \times \{d\} \times \{z\} \times \{x\} \subseteq \mathbb{R}^4$ are disjoint for each distinct (d, z, x),

$$
\begin{aligned}
P((Y_i, D_i, Z_i, X_i) \in \mathcal{S}) &= P\left((Y_i, D_i, Z_i, X_i) \in \bigcup_{d,z,x} \mathcal{Y}_{d,x} \times \{d\} \times \{z\} \times \{x\}\right) \\
&= \sum_{d,z,x} P(Y_i \in \mathcal{Y}_{d,x}, D_i = d, X_i = x, Z_i = z) \\
&= \sum_{d,z,x} P(D_i = d, X_i = x, Z_i = z) \\
&\qquad\qquad \times \underbrace{P(Y_i \in \mathcal{Y}_{d,x}, Z_i = z \mid D_i = d, X_i = x)}_{=P(Z_i = z \mid D_i = d, X_i = x)} / P(Z_i = z \mid D_i = d, X_i = x) \\
&= \sum_{d,z,x} P(D_i = d, X_i = x, Z_i = z) = 1
\end{aligned}
$$

Since $\{Y_i, D_i, Z_i, X_i\}_{i=1}^{n}$ is i.i.d.,

$$
P\left(\{Y_i, D_i, Z_i, X_i\}_{i=1}^{n} \subseteq \mathcal{S}\right) = P\left(\bigcap_{i=1}^{n} \{(Y_i, D_i, Z_i, X_i) \in \mathcal{S}\}\right) = \prod_{i=1}^{n} P\left((Y_i, D_i, Z_i, X_i) \in \mathcal{S}\right) = 1
$$

When $\{Y_i, D_i, Z_i, X_i\}_{i=1}^{n} \subseteq \mathcal{S}$ holds, $\mathbb{1}\{D_i = d, X_i = x\} \leq \mathbb{1}\{Y_i \in \mathcal{Y}_{d,x}\} = \mathbb{1}_{\mathcal{Y}_{d,x}}(Y_i)$ and thus $\mathbb{1}_{\mathcal{Y}_{d,x}}(Y_i) \times \mathbb{1}\{D_i = d, X_i = x\} = \mathbb{1}\{D_i = d, X_i = x\}$. This and (66) implies that when $\{Y_i, D_i, Z_i, X_i\}_{i=1}^{n} \subseteq \mathcal{S}$ holds,

$$
C_{d,x}(\mathbb{P}_n)(f) = C_{d,x}(\mathbb{P}_n)(\mathbb{1}_{\mathcal{Y}_{d,x}} \times f)
$$

iv. Use the facts established above to see that

$$
\begin{aligned}
&P(\sqrt{n}(C_{d,x}(\mathbb{P}_n) - C_{d,x}(P)) \in \ell^\infty_{\mathcal{Y}_{d,x}}(\mathcal{F}_{1,x})) \\
&\quad = P(\sqrt{n}(C_{d,x}(\mathbb{P}_n) - C_{d,x}(P)) \in \ell^\infty_{\mathcal{Y}_{d,x}}(\mathcal{F}_{d,x}) \mid \{Y_i, D_i, Z_i, X_i\}_{i=1}^n \subseteq \mathcal{S}) \\
&\quad = 1
\end{aligned}
$$

Lemma B.4 is that $\ell^\infty_{\mathcal{Y}_{d,x}}(\mathcal{F}_{1,x})$ is closed, so Portmanteau (van der Vaart & Wellner (1997) theorem 1.3.4) implies

$$
1 = \limsup_{n\to\infty} P(\sqrt{n}(C_{d,x}(\mathbb{P}_n) - C_{d,x}(P)) \in \ell^\infty_{\mathcal{Y}_{d,x}}(\mathcal{F}_{1,x})) \leq P(C'_{d,x,P}(\mathbb{G}) \in \ell^\infty_{\mathcal{Y}_{d,x}}(\mathcal{F}_{1,x}))
$$

In summary, we have

$$
\begin{aligned}
1 &= P\left(\left(\tilde{\eta}'_{1,x,P}(\mathbb{G}), \tilde{\eta}'_{0,x,P}(\mathbb{G}), C'_{s,x,P}(\mathbb{G})\right) \in \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}\right) \\
&= P(C'_{d,x,P}(\mathbb{G}) \in \ell^\infty_{\mathcal{Y}_{d,x}}(\mathcal{F}_{d,x})) \\
&= P\left(C'_{d,x,P}(\mathbb{G}) \in \mathcal{C}(\mathcal{F}_{d,x}, L_{2,P})\right)
\end{aligned}
$$

From which it follows that

$$
1 = P\left(C'_{1,x,P}(\mathbb{G}), C'_{0,x,P}(\mathbb{G}), \tilde{\eta}'_{1,x,P}(\mathbb{G}), \tilde{\eta}'_{0,x,P}(\mathbb{G}), C'_{s,x,P}(\mathbb{G}) \in \mathbb{S}_x\right)
$$

for each x, and therefore

$$
\begin{aligned}
&P(T'_{1,P}(\mathbb{G}) \in \mathbb{D}_{Tan,Full}) \\
&\quad = P\left(\bigcap_{x\in\mathcal{X}} \left\{C'_{1,x,P}(\mathbb{G}), C'_{0,x,P}(\mathbb{G}), \tilde{\eta}'_{1,x,P}(\mathbb{G}), \tilde{\eta}'_{0,x,P}(\mathbb{G}), C'_{s,x,P}(\mathbb{G}) \in \mathbb{S}_x\right\}\right) = 1
\end{aligned}
$$

This completes the proof. □

C.3 Optimal transport, $T_2(\{P_{1|x}, P_{0|x}, \eta_{1,x}, \eta_{0,x}, s_x\}_{x\in\mathcal{X}}) = (\{\theta_x^L, \theta_x^H, \eta_{1,x}, \eta_{0,x}, s_x\}_{x\in\mathcal{X}})$

The second map applies the directional differentiability of optimal transport shown in appendix B.2. There are three assumptions in lemma B.2 to verify: strong duality, Donsker conditions, and completeness. Strong duality is shown by lemmas E.9 and E.13, and the Donsker conditions were shown by lemma C.4. It remains to verify the completeness assumptions.

C.3.1 Verifying completeness

Lemma C.10 (Completeness of dual problem feasible set in L_2 for smooth cost functions)**.** *Suppose $\mathcal{Y} \subset \mathbb{R}$ is compact and $c : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is L-Lipschitz. Let $\mathcal{F}_c$, $\mathcal{F}_c^c$ be given by* (14) *and* (15) *respectively:*

$$\mathcal{F}_c = \left\{\varphi : \mathcal{Y} \to \mathbb{R}\ ;\ -\|c\|_\infty \le \varphi(y_1) \le \|c\|_\infty,\ |\varphi(y) - \varphi(y')| \le L|y - y'|\right\},$$
$$\mathcal{F}_c^c = \left\{\psi : \mathcal{Y} \to \mathbb{R}\ ;\ -2\|c\|_\infty \le \psi(y) \le 0,\ |\psi(y) - \psi(y')| \le L|y - y'|\right\},$$

Further let Φ_c be defined by (79), and $\mathcal{F}_d$ defined by (56). Let $L_{2,P}$ be given by (50), and L_2 be given by (51). Then $(\mathcal{F}_{1,x} \times \mathcal{F}_{0,x}, L_2)$ and its subset $\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$ are complete.

Proof. In steps:

1. $(\mathcal{F}_c, L_{2,P})$ and $(\mathcal{F}_c^c, L_{2,P})$ are complete.

   The proof that $(\mathcal{F}_c, L_{2,P})$ is complete is broken into steps:

   (a) Let $\{\varphi_n\}_{n=1}^\infty \subseteq \mathcal{F}_c$ be $L_{2,P}$-Cauchy. The L_p semimetrics are complete for any probability distribution (Pollard (2002) section 2.7 and chapter 2 problem [19]), thus there exists $\tilde{\varphi}$ such that $L_{2,P}(\varphi_n, \tilde{\varphi}) \to 0$. Convergence in $L_{2,P}$ implies convergence almost surely along a subsequence (Pollard (2002) section 2.8). Thus there exists a subsequence $\{\varphi_{n_k}\}_{k=1}^\infty$ such that $\lim_{k\to\infty} \varphi_{n_k}(y) = \tilde{\varphi}(y)$ for P-almost every y. Let $N_1 \subseteq \mathcal{Y}$ be the P-negligible set where this fails.

   (b) Observe that on $N_1^c = \mathcal{Y} \setminus N_1$, $\tilde{\varphi}$ obeys the bounds and Lipschitz continuity of $\mathcal{F}_c$. Specifically,

   $$-\|c\|_\infty \le \lim_{k\to\infty} -\|c\|_\infty \le \underbrace{\lim_{k\to\infty} \varphi_{n_k}(y)}_{\tilde{\varphi}(y)} \le \lim_{k\to\infty} \|c\|_\infty \le \|c\|_\infty$$

   Furthermore, for any $y, y' \in N_1^c$,

   $$\begin{aligned} |\tilde{\varphi}(y) - \tilde{\varphi}(y')| &= |\lim_{k\to\infty} \varphi_{n_k}(y) - \lim_{k\to\infty} \varphi_{n_k}(y')| = \lim_{k\to\infty} |\varphi_{n_k}(y) - \varphi_{n_k}(y')| \\ &\le \lim_{k\to\infty} L|y - y'| = L|y - y'| \end{aligned}$$

   (c) Now define functions $\bar{\varphi}, \varphi : \mathcal{Y} \to \mathbb{R}$ with

   $$\bar{\varphi}(y_1) = \sup_{y_1' \in N_1^c} \{\tilde{\varphi}(y_1') - L|y_1 - y_1'|\}, \qquad \varphi(y_1) = \max\{\bar{\varphi}(y_1), -\|c\|_\infty\}$$

   Then $L_{2,P}(\varphi_n, \varphi) \to 0$ and $\varphi \in \mathcal{F}_c$, which shows $(\mathcal{F}_c, L_{2,P})$ is complete.

      i. $L_{2,P}(\varphi_n, \varphi) \to 0$ follows from $\varphi(y) = \tilde{\varphi}(y)$ for all $y \in N_1^c$. To see this, let $y \in N_1^c$. Since $\tilde{\varphi}$ is L-Lipschitz on N_1^c, it follows that for any $y' \in N_1^c$,

      $$\tilde{\varphi}(y') - L|y - y'| \le \tilde{\varphi}(y)$$

      and thus $\bar{\varphi}(y) = \tilde{\varphi}(y)$. This implies $\bar{\varphi}(y) = \tilde{\varphi}(y) \ge -\|c\|_\infty$, and thus $\varphi(y) = \bar{\varphi}(y) = \tilde{\varphi}(y)$. Thus $\varphi(y) = \tilde{\varphi}(y)$ for P-almost all y, implying $L_{2,P}(\tilde{\varphi}, \varphi) = 0$ and thus $L_{2,P}(\varphi_n, \varphi) \to 0$.

      ii. To see that $\varphi \in \mathcal{F}_c$, first notice that $\bar{\varphi}(y) = \sup_{y' \in N_1^c}\{\tilde{\varphi}(y') - L|y - y'|\} \le \sup_{y' \in N_1^c} \tilde{\varphi}(y) \le \|c\|_\infty$, and hence $\bar{\varphi}$ obeys the upper bound for $\mathcal{F}_c$. It then follows easily that $\varphi(y) = \max\{\bar{\varphi}(y), -\|c\|_\infty\}$ obeys both the upper and lower bound. Next notice that $\bar{\varphi}$ is L-Lipschitz on all of $\mathcal{Y}$:

$$\begin{aligned}
\bar{\varphi}(y) - \bar{\varphi}(y') = &= \sup_{\tilde{y}\in N_1^c}\{\tilde{\varphi}(\tilde{y}) - L|y-\tilde{y}|\} - \sup_{\tilde{y}'\in N_1^c}\{\tilde{\varphi}(\tilde{y}') - L|y'-\tilde{y}'|\} \\
&\leq \sup_{\tilde{y}\in N_1^c}\{\tilde{\varphi}(\tilde{y}) - L|y-\tilde{y}| - \left(\tilde{\varphi}(\tilde{y}) - L|y'-\tilde{y}|\right)\} \\
&= \sup_{\tilde{y}\in N_1^c} L\left(|y'-\tilde{y}| - |y-\tilde{y}|\right) \leq L|y-y'|
\end{aligned}$$

where the last inequality follows from the reverse triangle inequality. It follows that $\varphi(y_1) = \max\{\bar{\varphi}(y_1), -\|c\|_\infty\}$ is also L-Lipschitz, and thus $\varphi \in \mathcal{F}_c$.

2. Very similar steps show that $(\mathcal{F}_c^c, L_{2,P})$ is complete; the only substantial changes are replacing the lower bounds with $-2\|c\|$ and the upper bounds with 0.

3. Note that since $(\mathcal{F}_c \times \mathcal{F}_c^c, L_2)$ is the product space of $(\mathcal{F}_c, L_{2,P})$ and $(\mathcal{F}_c^c, L_{2,P})$, it follows that $(\mathcal{F}_c \times \mathcal{F}_c^c, L_2)$ is complete.

4. $\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$ is complete.

   To see that $\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$ is complete, let $\{(\varphi_n, \psi_n)\}_{n=1}^\infty \subseteq \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$ be L_2-Cauchy, and follow the same steps shown above to define $(\varphi, \psi) \in \mathcal{F}_c \times \mathcal{F}_c^c$ such that $L_2((\varphi_n, \psi_n), (\varphi, \psi)) \to 0$. It remains to show that $\varphi(y_1) + \psi(y_0) \leq c(y_1, y_0)$ for all $(y_1, y_0) \in \mathcal{Y} \times \mathcal{Y} \subseteq \mathbb{R}^2$.

   Since c is L-Lipschitz,

$$c(y_1, y_0) - c(y_1', y_0) \geq -L\|(y_1, y_0) - (y_1', y_0')\| \geq -L|y_1 - y_1'| - L|y_0 - y_0'|$$

   which implies $c(y_1', y_0') - L|y_1 - y_1'| - L|y_0 - y_0'| \leq c(y_1, y_0)$. Thus

$$\begin{aligned}
\bar{\varphi}(y_1) + \bar{\varphi}(y_0) &= \sup_{y_1'\in N_1^c}\{\tilde{\varphi}(y_1') - L|y_1 - y_1'|\} + \sup_{y_0'\in N_0^c}\{\tilde{\psi}(y_0') - L|y_0 - y_0'|\} \\
&= \sup_{(y_1',y_0')\in N_1^c\times N_0^c}\left\{\tilde{\varphi}(y_1') + \tilde{\psi}(y_0') - L|y_1 - y_1'| - L|y_0 - y_0'|\right\} \\
&\leq \sup_{(y_1',y_0')\in N_1^c\times N_0^c}\left\{c(y_1', y_0') - L|y_1 - y_1'| - L|y_0 - y_0'|\right\} \\
&\leq \sup_{(y_1',y_0')\in N_1^c\times N_0^c}\left\{c(y_1, y_0)\right\} = c(y_1, y_0)
\end{aligned}$$

   Finally,

$$\begin{aligned}
\varphi(y_1) + \psi(y_0) &= \max\{\bar{\varphi}(y_1), -\|c\|_\infty\} + \max\{\bar{\psi}(y_0), -2\|c\|\} \\
&= \max\{\bar{\varphi}(y_1) + \bar{\varphi}(y_0), \bar{\varphi}(y_1) - 2\|c\|_\infty, -\|c\|_\infty + \bar{\psi}(y_0), -\|c\|_\infty - 2\|c\|\} \\
&\leq \max\{c(y_1, y_0), -\|c\|_\infty, -\|c\|_\infty, -3\|c\|_\infty\} \\
&\leq c(y_1, y_0)
\end{aligned}$$

   where the first inequality follows from $\bar{\varphi}(y_1) \leq \|c\|_\infty$ and $\bar{\psi}(y_0) \leq 0$.

5. $(\mathcal{F}_{1,x} \times \mathcal{F}_{0,x}, L_2)$ is complete.

   As this is the product space of $(\mathcal{F}_{1,x}, L_{2,P})$ and $(\mathcal{F}_{0,x}, L_{2,P})$, it suffices to show these individual spaces are complete.

Now recall that $\mathcal{F}_{d,x}$ is defined by (56):

$$\begin{aligned}
\tilde{\mathcal{F}}_1 &= \left\{ f : \mathcal{Y} \to \mathbb{R}\ ;\ f = \varphi \text{ for some } \varphi \in \mathcal{F}_c,\ \text{ or } f = \eta_1^{(k)} \text{ for some } k = 1, \ldots, K_1 \right\} \\
\tilde{\mathcal{F}}_0 &= \left\{ f : \mathcal{Y} \to \mathbb{R}\ ;\ f = \psi \text{ for some } \psi \in \mathcal{F}_c^c,\ \text{ or } f = \eta_0^{(k)} \text{ for some } k = 1, \ldots, K_0 \right\} \\
\mathcal{F}_{d,x} &= \left\{ f : \mathcal{Y} \to \mathbb{R}\ ;\ f = g \text{ or } \mathbb{1}_{\mathcal{Y}_{d,x}} \times g \text{ for some } g \in \tilde{\mathcal{F}}_d \right\}
\end{aligned}$$

Recall that the union of a finite number of complete sets is complete. Since $(\mathcal{F}_c, L_{2,P})$ and $\mathcal{F}_c^c, L_{2,P})$ are complete and any finite set is complete, $\tilde{\mathcal{F}}_d$ is complete. Next recognize that $\mathcal{F}_{d,x} = \tilde{\mathcal{F}}_d \cup \left\{ \mathbb{1}_{\mathcal{Y}_{d,x}} \times g\ ;\ g \in \tilde{\mathcal{F}}_d \right\}$ is the union of a finite number of sets, and thus it suffices to show $\left\{ \mathbb{1}_{\mathcal{Y}_{d,x}} \times g\ ;\ g \in \tilde{\mathcal{F}}_d \right\}$ is complete.

Let $\{\mathbb{1}_{\mathcal{Y}_{d,x}} \times g_n\}_{n=1}^\infty \subseteq \left\{ \mathbb{1}_{\mathcal{Y}_{d,x}} \times g\ ;\ g \in \tilde{\mathcal{F}}_d \right\}$ be $L_{2,P}$-Cauchy. Lemma C.4 shows that $\mathcal{F}_{d,x}$ is Donsker and $\sup_{f \in \mathcal{F}_{d,x}} |P(f)| < \infty$, which implies $(\mathcal{F}_{d,x}, L_{2,P})$ is totally bounded (see van der Vaart & Wellner (1997) problem 2.1.2.). Since $\tilde{\mathcal{F}}_d$ is a complete subset of a totally bounded set, it is compact. Thus $\{g_n\}_{n=1}^\infty \subseteq \tilde{\mathcal{F}}_d$ is a sequence in a compact semimetric space, and therefore has a convergent subsequence $\{g_{n_k}\}_{k=1}^\infty$. Let $g \in \tilde{\mathcal{F}}_d$ be its limit, and notice that

$$\begin{aligned}
0 \leq L_{2,P}(\mathbb{1}_{\mathcal{Y}_{d,x}} \times g_{n_k}, \mathbb{1}_{\mathcal{Y}_{d,x}} \times g) &= \sqrt{P((\mathbb{1}_{\mathcal{Y}_{d,x}} \times g_{n_k} - \mathbb{1}_{\mathcal{Y}_{d,x}} \times g)^2)} \\
&\leq \sqrt{P((g_{n_k} - g)^2)} \\
&= L_{2,P}(g_{n_k}, g) \to 0
\end{aligned}$$

and thus $\mathbb{1}_{\mathcal{Y}_{d,x}} \times \varphi_{n_k} \to \mathbb{1}_{\mathcal{Y}_{d,x}} g$. It follows that $\mathbb{1}_{\mathcal{Y}_{d,x}} \times \varphi_n \to \mathbb{1}_{\mathcal{Y}_{d,x}} g$, and thus $\left\{ \mathbb{1}_{\mathcal{Y}_{d,x}} \times g\ ;\ g \in \tilde{\mathcal{F}}_d \right\}$ is complete.

This completes the proof. □

Lemma C.11 (Completeness of dual problem feasible set in L_2 for indicator cost functions)**.** *Let $\mathcal{Y} \subseteq \mathbb{R}$, $C \subseteq \mathcal{Y} \times \mathcal{Y}$ be nonempty, open, and convex, and let $c : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ be given by $c(y_1, y_0) = \mathbb{1}_C(y_1, y_0) = \mathbb{1}\{(y_1, y_0) \in C\}$. Let $\mathcal{F}_c$, $\mathcal{F}_c^c$ be given by* (16) *and* (17) *respectively:*

$$\begin{aligned}
\mathcal{F}_c &= \{\varphi : \mathcal{Y} \to \mathbb{R}\ ;\ \varphi(y_1) = \mathbb{1}_I(y_1) \text{ for some interval } I\}, \\
\mathcal{F}_c^c &= \{\psi : \mathcal{Y} \to \mathbb{R}\ ;\ \psi(y_0) = -\mathbb{1}_{I^c}(y_0) \text{ for some interval } I\},
\end{aligned}$$

Further let Φ_c be defined by (79)*, and $\mathcal{F}_{d,x}$ defined by* (56)*. Let $L_{2,P}$ be given by* (50)*, and L_2 be given by* (51)*. Then $(\mathcal{F}_{1,x} \times \mathcal{F}_{0,x}, L_2)$ and its subset $\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$ are complete.*

Proof. The proof is similar in structure to that of lemma C.10.

1. $(\mathcal{F}_c, L_{2,P})$ is complete.

   Let $\{\varphi_n\}_{n=1}^\infty \subseteq \mathcal{F}_c$ be $L_{2,P}$-Cauchy. Note that $\varphi_n(y) = \mathbb{1}_{I_n}(y)$ for some interval I_n. Just as in the proof of lemma C.10, there exists $\tilde{\varphi}$ such that $L_{2,P}(\varphi_n, \tilde{\varphi}) \to 0$, and a subsequence

$\{\varphi_{n_k}\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty}\varphi_{n_k}(y) = \tilde{\varphi}(y)$ for P-almost every y. Let $N \subset \mathcal{Y}$ be the P-negligible set where this convergence fails.

Let $y \in N^c$, and notice that $\varphi_{n_k}(y) = \mathbb{1}_{I_{n_k}}(y) \in \{0,1\}$ for all k and $\{\varphi_{n_k}(y)\}_{k=1}^{\infty}$ converging in $\mathbb{R}$ implies that $\varphi_{n_k}(y)$ is eventually constant as k grows. This implies $\tilde{\varphi}(y) \in \{0,1\}$, and hence for some set $A \subset \mathcal{Y}$,

$$\tilde{\varphi}(y) = \mathbb{1}_A(y) \qquad\qquad \text{for all } y \in N^c$$

We will show that for some interval I, $A \cap N^c = I \cap N^c$. Let $y_1, y_2, y_3 \in N^c$ satisfy $y_1 < y_2 < y_3$ and $y_1, y_3 \in A$, but be otherwise arbitrary. It suffices to show that $y_2 \in A$; we can then define I to be the interval with endpoints $\inf A$ and $\sup A$ (including the lower endpoint if $\inf A = \min A > -\infty$, and including the upper endpoint if $\sup A = \max A < \infty$), and define the function $\varphi : \mathcal{Y}_1 \to \mathbb{R}$ with $\varphi(y_1) = \mathbb{1}_I(y_1)$.[13]

Notice that $\lim_{k\to\infty}\mathbb{1}_{I_{n_k}}(y_3) = \mathbb{1}_A(y_3) = 1$ and $\lim_{k\to\infty}\mathbb{1}_{I_{n_k}}(y_3) = \mathbb{1}_A(y_3) = 1$ implies that $\mathbb{1}_{I_{n_k}}(y_1)$ and $\mathbb{1}_{I_{n_k}}(y_3)$ are eventually constant and equal to 1, i.e. there exists $K_1, K_3 \in \mathbb{N}$ such that

$$y_1 \in I_{n_k} \text{ for all } k \geq K_1, \text{ and} \qquad\qquad y_3 \in I_{n_k} \text{ for all } k \geq K_3$$

Since I_{n_k} is an interval, this implies

$$y_2 \in I_{n_k} \text{ for all } k \geq \max\{K_1, K_3\}$$

i.e. $\mathbb{1}_{I_{n_k}}(y_2) = 1$ for all such k, and therefore $\mathbb{1}_A(y_2) = \lim_{k\to\infty}\mathbb{1}_{A_n}(y_2) = 1$. Thus $y_2 \in A$.

It follows that $\tilde{\varphi}(y) = \varphi(y) = \mathbb{1}_I(y)$ for all $y \in N^c$. Thus $L_{2,P}(\tilde{\varphi}, \varphi) = 0$, and $L_{2,P}(\varphi_n, \varphi) \to 0$. Since $\varphi \in \mathcal{F}_c$, this completes the proof that $(\mathcal{F}_c, L_{2,P})$ is complete.

2. $(\mathcal{F}_c^c, L_{2,P})$ is complete.

   The argument is similar. Let $\{\psi_n\}_{n=1}^{\infty} \subseteq \mathcal{F}_c$ be $L_{2,P}$-Cauchy. Note that $\psi_n(y) = \mathbb{1}_{I_n^c}(y)$ for some interval I_n. There exists $\tilde{\psi}$ such that $L_{2,P}(\psi_n, \tilde{\psi}) \to 0$, and a subsequence $\{\psi_{n_k}\}_{k=1}^{\infty}$ such that $\lim_{k\to\infty}\psi_{n_k}(y) = \tilde{\psi}(y)$ for P-almost every y. Let $N \subset \mathcal{Y}$ be the P-negligible set where this convergence fails.

   Since $\psi_{n_k}(y) = \mathbb{1}_{I_{n_k}^c}(y) \in \{0,1\}$ for all k and y, and $\lim_{k\to\infty}\psi_{n_k}(y) = \tilde{\psi}(y)$ for all $y \in N^c$, we have $\tilde{\psi}(y) \in \{0,1\}$ for all such y and thus for some set $A \subseteq \mathcal{Y}$,

$$\tilde{\psi}(y) = \mathbb{1}_{A^c}(y) \qquad\qquad \text{for all } y \in N^c$$

   Once again, it suffices to show $A \cap N^c = I \cap N^c$ for some interval I. Consider $y_1, y_2, y_3 \in N^c$, $y_1 < y_2 < y_3$, with $y_1, y_3 \in A$. $\lim_{k\to\infty}\psi_{n_k}(y_1) = \tilde{\psi}(y_1) = 0$ and $\lim_{k\to\infty}\psi_{n_k}(y_3) = \tilde{\psi}(y_3) = 0$ implies that $\psi_{n_k}(y_1) = \mathbb{1}_{I_{n_k}^c}(y_1)$ and $\psi_{n_k}(y_3) = \mathbb{1}_{I_{n_k}^c}(y_3)$ are eventually constant and equal to 0, i.e. for some $K_1, K_3 \in \mathbb{N}$,

$$y_1 \in I_{n_k} \text{ for all } k \geq K_1, \qquad\qquad y_3 \in I_{n_k} \text{ for all } k \geq K_3$$

[13] Explicitly, I is defined as follows: (a) $I = (\ell, u)$ if neither $\ell = \inf A$ nor $u = \sup A$ is attained in $\mathbb{R}$ (b) $I = [\ell, u)$ if $\ell = \inf A = \min A$, but $u = \sup A$ is not attained in $\mathbb{R}$ (c) $I = (\ell, u]$ if $\ell = \inf A$ is not attained in $\mathbb{R}$, but $u = \sup A = \max A$ (d) $I = [\ell, u]$ if both $\ell = \inf A = \min A$ and $u = \sup A = \max A$.

since I_{n_k} is an interval for every k, this implies

$$y_2 \in I_{n_k} \text{ for all } k \geq \max\{K_1, K_3\}$$

thus $\tilde{\psi}(y_2) = \lim_{k\to\infty} \psi_{n_k}(y_2) = 0$. It follows that $A \cap N^c = I \cap N^c$, where I is the interval defined by endpoints $\inf A$ and $\sup A$, which are included if attained and finite. Define $\psi(y) = \mathbb{1}_{I^c}(y)$ and notice $\psi \in \mathcal{F}_c^c$. We have $\psi(y) = \tilde{\psi}(y)$ for all $y \in N^c$ and hence $L_{2,P}(\tilde{\psi}, \psi) = 0$. Thus $L_{2,P}(\psi_n, \psi) \to 0$, showing $(\mathcal{F}_c^c, L_{2,P})$ is complete.

3. Note that $(\mathcal{F}_c \times \mathcal{F}_c^c, L_2)$ is the product space of the complete spaces $(\mathcal{F}_c, L_{2,P})$ and $(\mathcal{F}_c^c, L_{2,P})$, and so is complete.

4. We next show $\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c) = \{(\varphi, \psi) \in \mathcal{F}_c \times \mathcal{F}_c^c \,;\, \varphi(y_1) + \psi(y_0) \leq c(y_1, y_0)\}$ is complete.

   Let $\{(\varphi_n, \psi_n)\}_{n=1}^\infty \subseteq \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$ be L_2-Cauchy, and let $(\tilde{\varphi}, \tilde{\psi})$ be a limit in $\mathcal{F}_c \times \mathcal{F}_c^c$. Since $L_{2,P}(\varphi_n, \tilde{\varphi}) \to 0$ there exists a subsequence $\{(\varphi_{n_k}, \psi_{n_k})\}_{k=1}^\infty$ such that $\lim_{k\to\infty} \varphi_{n_k}(y_1) = \tilde{\varphi}(y_1)$ for P-almost all y_1. Let N_1 be the negligible set where this fails. Furthermore, $L_{2,P}(\psi_{n_k}, \tilde{\psi}) \to 0$ as $k \to \infty$ and so there is a further subsequence $\{(\varphi_{n_{k_j}}, \psi_{n_{k_j}})\}_{j=1}^\infty$ such that $\lim_{j\to\infty} \psi_{n_{k_j}}(y_0) = \tilde{\psi}(y_0)$ for P-almost all y_0. Let N_0 be the negligible set where this fails. It is then clear that if $(y_1, y_0) \in N_1^c \times N_0^c$, then

$$\tilde{\varphi}(y_1) + \tilde{\psi}(y_0) = \lim_{j\to\infty} \{\varphi_{n_{k_j}}(y_1) + \psi_{n_{k_j}}(y_0)\} \leq \lim_{j\to\infty} c(y_1, y_0) = \mathbb{1}_C(y_1, y_0) \tag{67}$$

   Note that $\tilde{\varphi} = \mathbb{1}_{I_{\tilde{\varphi}}}$, and $\tilde{\psi} = -\mathbb{1}_{I_{\tilde{\psi}}^c}$ for some intervals $I_{\tilde{\varphi}}$ and $I_{\tilde{\psi}}$. Let

$$\ell_1 = \inf I_{\tilde{\varphi}} \cap N_1^c, \qquad u_1 = \sup I_{\tilde{\varphi}} \cap N_1^c, \qquad \ell_0 = \inf I_{\tilde{\psi}} \cap N_0^c, \qquad u_0 = \sup I_{\tilde{\psi}} \cap N_0^c$$

   and define $\varphi = \mathbb{1}_{I_\varphi}$ where I_φ is the interval with endpoints ℓ_1, u_1 (included if the inf/sup are finite and attained), and $\psi = -\mathbb{1}_{I_\psi^c}$ where I_ψ^c is the interval with endpoints ℓ_0, u_0 (included if the inf/sup are finite and attained). Notice that $I_\varphi = I_{\tilde{\varphi}}$, P-almost surely and $I_\psi = I_{\tilde{\psi}}$, P-almost surely.

   Notice that for $(y_1, y_0) \in (N_1^c \times N_0^c)^c$ to satisfy $\varphi(y_1) + \psi(y_0) = \mathbb{1}_{I_\varphi}(y_1) - \mathbb{1}_{I_\psi^c}(y_0) > \mathbb{1}_C(y_1, y_0)$, it would have to be the case that $(y_1, y_0) \in (I_{\tilde{\varphi}} \times I_{\tilde{\psi}}) \cap (N_1^c \times N_0^c)^c \setminus C$. Let $(y_1, y_0) \in (I_\varphi \times I_\psi) \cap (N_1^c \times N_0^c)^c$, and note that there exists $y_1^\ell, y_1^u \in I_\varphi \cap N_1^c$ with $y_1^\ell \leq y_1 \leq y_1^u$ and $y_0^\ell, y_0^u \in I_\psi \cap N_0^c$ with $y_0^\ell \leq y_0 \leq y_0^u$. Notice that $[y_1^\ell, y_1^u] \times [y_0^\ell, y_0^u] \subseteq C$, because C is convex and (67) holds for the "corners": $(\ell_1, \ell_0), (\ell_1, u_0), (u_1, \ell_0), (u_1, u_0) \in (I_\varphi \times I_\psi) \cap (N_1^c \times N_0^c)$. Thus $(I_{\tilde{\varphi}} \times I_{\tilde{\psi}}) \cap (N_1^c \times N_0^c)^c \setminus C = \varnothing$, showing that $\varphi(y_1) + \psi(y_0) \leq c(y_1, y_0)$ holds for all $(y_1, y_0) \in \mathcal{Y}_1 \times \mathcal{Y}_0$. This shows $\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$ is complete.

5. The argument thet $(\mathcal{F}_{1,x} \times \mathcal{F}_{0,x}, L_2)$ is complete is identical to the argument given in step 5 of the proof of lemma C.10.

This completes the proof. □

C.3.2 Differentiability of T_2

We first apply lemma B.2 to show show that $\theta^L(\cdot)$ and $\theta^H(\cdot)$, given by either (19) or (20) depending on the function c, are Hadamard differentiable.

Lemma C.12. *Suppose assumptions 1, 2, and 3 hold. Then θ^L and θ^H given by (19) or (20) are Hadamard directionally differentiable at $(P_{1|x}, P_{0|x})$ tangentially to $\mathcal{C}(\mathcal{F}_{1,x}, L_{2,P}) \times \mathcal{C}(\mathcal{F}_{0,x}, L_{2,P})$. The argmax sets*

$$\Psi_{c_L}(P_{1|x}, P_{0|x}) = \underset{(\varphi,\psi)\in\Phi_{c_L}\cap(\mathcal{F}_c\times\mathcal{F}_c^c)}{\arg\max} P_{1|x}(\varphi) + P_{0|x}(\psi)$$
$$\Psi_{c_H}(P_{1|x}, P_{0|x}) = \underset{(\varphi,\psi)\in\Phi_{c_H}\cap(\mathcal{F}_c\times\mathcal{F}_c^c)}{\arg\max} P_{1|x}(\varphi) + P_{0|x}(\psi)$$

are nonempty, and the derivatives $\theta^{L\prime}_{(P_{1|x},P_{0|x})}, \theta^{H\prime}_{(P_{1|x},P_{0|x})} : \mathcal{C}(\mathcal{F}_{1,x}, L_{2,P}) \times \mathcal{C}(\mathcal{F}_{0,x}, L_{2,P}) \to \mathbb{R}$ are given by

$$\theta^{L\prime}_{(P_{1|x},P_{0|x})}(H_1, H_0) = \sup_{(\varphi,\psi)\in\Psi_{c_L}(P_{1|x},P_{0|x})} H_1(\varphi) + H_0(\psi) \tag{68}$$

$$\theta^{H\prime}_{(P_{1|x},P_{0|x})}(H_1, H_0) = -\left[\sup_{(\varphi,\psi)\in\Psi_{c_H}(P_{1|x},P_{0|x})} H_1(\varphi) + H_0(\psi)\right] \tag{69}$$

If assumption 4 also holds, then θ^L and θ^H are fully Hadamard differentiable at $(P_{1|x}, P_{0|x})$ tangentially to

$$\mathbb{D}_{Tan,Full,x} = \left(\ell^\infty_{\mathcal{Y}_{1,x}}(\mathcal{F}_{1,x}) \times \ell^\infty_{\mathcal{Y}_{0,x}}(\mathcal{F}_{0,x})\right) \cap \left(\mathcal{C}(\mathcal{F}_{1,x}, L_{2,P}) \times \mathcal{C}(\mathcal{F}_{0,x}, L_{2,P})\right)$$

with the derivatives $\theta^{L\prime}_{(P_{1|x},P_{0|x})}, \theta^{H\prime}_{(P_{1|x},P_{0|x})} : \mathbb{D}_{Tan,Full,x} \to \mathbb{R}$ also given by (68) and (69).

Proof. We apply lemma B.2. It is clear from inspection that the cost functions c_L and c_H are lower semicontinuous, the sets $\mathcal{F}_{d,x}$ defined by (56) consists of measurable functions mapping $\mathcal{Y}$ to $\mathbb{R}$, and that the subsets $\mathcal{F}_c$ and $\mathcal{F}_c^c$ given by (14) and (15), or by (16) and (17), are universally bounded. Moreover,

1. Strong duality holds.
   (i) If assumption 2 (i) holds, then lemma E.9 shows that strong duality holds.
   (ii) If assumption 2 (ii) holds, then lemma E.13 shows that strong duality holds.

2. Assumption 1 implies P dominates $P_{d|x}$ with bounded densities $\frac{dP_{d|x}}{dP}$. Indeed,

$$\begin{aligned}
E_{P_{d|x}}[f(Y_d)] &= \frac{E_P[f(Y)\mathbb{1}\{D=d\} \mid X=x, Z=d] - E_P[f(Y)\mathbb{1}\{D=d\} \mid X=x, Z=1-d]}{P(D=d \mid X=x, Z=d) - P(D=d \mid X=x, Z=1-d)} \\
&= E_P\left[f(Y)\frac{\mathbb{1}_{d,x,d}(D,X,Z)/p_{x,d} - \mathbb{1}_{d,x,1-d}(D,X,Z)/p_{x,1-d}}{p_{d,x,d}/p_{x,d} - p_{d,x,1-d}/p_{x,1-d}}\right] \\
&= E_P\left[f(Y)E\left[\frac{\mathbb{1}_{d,x,d}(D,X,Z)/p_{x,d} - \mathbb{1}_{d,x,1-d}(D,X,Z)/p_{x,1-d}}{p_{d,x,d}/p_{x,d} - p_{d,x,1-d}/p_{x,1-d}} \mid Y\right]\right]
\end{aligned}$$

   Notice that $\frac{dP_{d|x}}{dP}(Y) = E_P\left[\frac{\mathbb{1}_{d,x,d}(D,X,Z)/p_{x,d} - \mathbb{1}_{d,x,1-d}(D,X,Z)/p_{x,1-d}}{p_{d,x,d}/p_{x,d} - p_{d,x,1-d}/p_{x,1-d}} \mid Y\right]$ must be nonnegative P-almost surely; if the set $A = \left\{y \; ; \; \frac{dP_{d|x}}{dP}(y) < 0\right\}$ was P-non-negligible, the displays above would imply the contradiction $P(Y_d \in A \mid D_1 > D_0, X = x) < 0$. Moreover, it is bounded by $K_{d,x} = \frac{1/p_{x,d}}{p_{d,x,d}/p_{x,d} - p_{d,x,1-d}/p_{x,1-d}}$

3. Lemma C.4 shows that under assumptions 1, 2, and 3, $\mathcal{F}_{d,x}$ is P-Donsker and $\sup_{f\in\mathcal{F}_{d,x}}|P(f)| < \infty$ for $d = 1, 0$, and

4. The set $(\mathcal{F}_1 \times \mathcal{F}_0, L_2)$ and its subset $\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$ are complete.

   (i) If assumption 2 (i) holds, then lemma C.10 shows these sets are complete.

   (ii) If assumption 2 (ii) holds, then lemma C.11 shows these sets are complete.

It follows from the chain rule that θ^L and θ^H are Hadamard directionally differentiable with the claimed directional derivatives.

Now suppose assumptions 1, 2, 3, and 4 hold. Lemma B.5 implies θ^L and θ^H are fully Hadamard differentiable at $(P_{1|x}, P_{0|x})$ tangentially to

$$\mathbb{D}_{T,Full,x} = \left(\ell^\infty_{\mathcal{Y}_{1,x}}(\mathcal{F}_{1,x}) \times \ell^\infty_{\mathcal{Y}_{0,x}}(\mathcal{F}_{0,x}\right) \cap \left(\mathcal{C}(\mathcal{F}_{1,x}, L_{2,P}) \times \mathcal{C}(\mathcal{F}_{0,x}, L_{2,P})\right)$$

with derivatives given by the same expressions. □

We can now show the differentiability properties of T_2.

Lemma C.13 (T_2 is Hadamard differentiable)**.** *Let $\mathbb{D}_{Tan}$ and $\mathbb{D}_{Tan,Full}$ be given by*

$$\mathbb{D}_{Tan} = \prod_{m=1}^{M} \mathcal{C}(\mathcal{F}_{1,x_m}, L_{2,P}) \times \mathcal{C}(\mathcal{F}_{0,x_m}, L_{2,P}) \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}$$

$$\mathbb{D}_{Tan,Full} = \prod_{m=1}^{M} \left(\ell^\infty_{\mathcal{Y}_{1,x_m}}(\mathcal{F}_{1,x_m}) \times \ell^\infty_{\mathcal{Y}_{0,x_m}}(\mathcal{F}_{0,x_m})\right) \cap \left(\mathcal{C}(\mathcal{F}_{1,x_m}, L_{2,P}) \times \mathcal{C}(\mathcal{F}_{0,x_m}, L_{2,P})\right) \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}$$

and define

$$T_2 : \prod_{m=1}^{M} \ell^\infty(\mathcal{F}_{1,x}) \times \ell^\infty(\mathcal{F}_{0,x}) \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R} \to \prod_{m=1}^{M} \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R},$$
$$T_2(\{P_{1|x}, P_{0|x}, \eta_{1,x}, \eta_{0,x}, s_x\}_{x\in\mathcal{X}}) = \left(\{\theta^L(P_{1|x}, P_{0|x}), \theta^H(P_{1|x}, P_{0|x}), \eta_{1,x}, \eta_{0,x}, s_x\}_{x\in\mathcal{X}}\right)$$

Under assumptions 1, 2, and 3, T_2 is Hadamard directionally differentiable at $T_1(P) = (\{P_{1|x}, P_{0|x}, s_x, \eta_{1,x}, \eta_{0,x}\}_{x\in\mathcal{X}})$ tangentially to $\mathbb{D}_{Tan}$, with derivative

$$T'_{2,T_1(P)} : \mathbb{D}_{Tan} \to \prod_{m=1}^{M} \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}$$
$$T'_{2,T_1(P)}(\{H_{1,x}, H_{0,x}, h_{\eta_1,x}, h_{\eta_0,x}, h_{s,x}\}_{x\in\mathcal{X}})$$
$$= \left(\left\{\theta^{L\prime}_{(P_{1|x},P_{0|x})}(H_{1,x}, H_{0,x}), \theta^{H\prime}_{(P_{1|x},P_{0|x})}(H_{1,x}, H_{0,x}), h_{\eta_1,x}, h_{\eta_0,x}, h_{s,x}\right\}_{x\in\mathcal{X}}\right)$$

If assumption 4 also holds, then T_2 is fully Hadamard differentiable at $T_1(P)$ tangentially to $\mathbb{D}_{Tan,Full}$, with derivative $T_{2,T_1(P)} : \mathbb{D}_{Tan,Full} \to \prod_{m=1}^{M} \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}$ given by the same expression.

Proof. Lemma C.12 shows that under assumptions 1, 2, and 3, $\theta^L(\cdot)$ and $\theta^H(\cdot)$ are Hadamard directionally differentiable at $(P_{1|x}, P_{0|x})$ tangentially to $\mathcal{C}(\mathcal{F}_{1,x}, L_{2,P}) \times \mathcal{C}(\mathcal{F}_{0,x}, L_{2,P})$ for each $x \in \mathcal{X}$. If assumption 4 also holds, lemma C.12 shows these derivatives are linear on the subspace $\mathbb{D}_{Tan,Full}$, and hence $\theta^L(\cdot)$ and $\theta^H(\cdot)$ are fully Hadamard differentiable tangentially to $\mathbb{D}_{Tan,Full}$. The other coordinates are the identity mapping, which is fully Hadamard differentiable. Apply lemma F.5 to obtain the result. □

C.4 Expectations, $T_3(\{\theta_x^L, \theta_x^H, \eta_{1,x}, \eta_{0,x}, s_x\}_{x\in\mathcal{X}}) = (\theta^L, \theta^H, \eta)$

Lemma C.14. *Define*

$$T_3 : \prod_{m=1}^{M} \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0}$$
$$T_3(\{\theta_x^L, \theta_x^H, \eta_{1,x}, \eta_{0,x}, s_x\}_{x\in\mathcal{X}}) = \left(\sum_{x\in\mathcal{X}} s_x\theta_x^L, \sum_{x\in\mathcal{X}} s_x\theta_x^H, \sum_{x\in\mathcal{X}} s_x\eta_{1,x}, \sum_{x\in\mathcal{X}} s_x\eta_{0,x}\right)$$

T_3 is fully (Hadamard) differentiable at any $V = (\{\theta_x^L, \theta_x^H, \eta_{1,x}, \eta_{0,x}, s_x\}_{x\in\mathcal{X}}) \in \prod_{m=1}^{M} \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0}$ tangentially to $\prod_{m=1}^{M} \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}$ with derivative

$$T'_{3,V} : \prod_{m=1}^{M} \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R} \to \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0}$$
$$T'_{3,V}(\{h_x^L, h_x^H, h_{\eta_1,x}, h_{\eta_0,x}, h_{s,x}\}_{x\in\mathcal{X}})$$
$$= \left(\sum_{x\in\mathcal{X}} s_x h_x^L + h_{s,x}\theta^L(x), \sum_{x\in\mathcal{X}} s_x h_x^H + h_{s,x}\theta^H(x), \sum_{x\in\mathcal{X}} s_x h_{\eta_1,x} + h_{s,x}\eta_{1,x}, \sum_{x\in\mathcal{X}} s_x h_{\eta_0,x} + h_{s,x}\eta_{0,x}\right)$$

Proof. The inner product

$$IP : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}, \qquad IP(r_1, r_2) = \langle r_1, r_2\rangle = \sum_{m=1}^{M} r_1^{(m)} r_2^{(m)}$$

is fully Hadamard differentiable at any $(r_1, r_2) \in \mathbb{R}^M \times \mathbb{R}^M$ tangentially to $\mathbb{R}^M \times \mathbb{R}^M$ with derivative

$$IP'_{(r_1,r_2)} : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R},$$
$$IP'_{(r_1,r_2)}(h_1, h_2) = \langle r_1, h_2\rangle + \langle h_1, r_2\rangle = \sum_{m=1}^{M} r_1^{(m)} h_2^{(m)} + h_1^{(m)} r_2^{(m)}$$

Apply lemma F.5 to obtain the result. □

C.5 Optimization over $t \in [\theta^L, \theta^H]$: $T_4(\theta^L, \theta^H, \eta) = (\gamma^L, \gamma^H)$

Lemma C.15. *Let $g^L, g^H : \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \to \mathbb{R}$ be as defined in assumption 3:*

$$g^L(\theta^L, \theta^H, \eta_1, \eta_0) = \inf_{t \in [\theta^L, \theta^H]} g(t, \eta_1, \eta_0), \qquad g^H(\theta^L, \theta^H, \eta_1, \eta_0) = \sup_{t \in [\theta^L, \theta^H]} g(t, \eta_1, \eta_0)$$

Define

$$\begin{aligned} &T_4 : \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \to \mathbb{R} \times \mathbb{R} \\ &T_4(\theta^L, \theta^H, \eta_1, \eta_0) = \big(g^L(\theta^L, \theta^H, \eta_1, \eta_0), g^H(\theta^L, \theta^H, \eta_1, \eta_0)\big) \end{aligned}$$

Under assumption 3, g^L and g^H are continuously differentiable at $(\theta^L, \theta^H, \eta_1, \eta_0) = T_3(T_2(T_1(P)))$ with gradients

$$\nabla g^L = \nabla g^L(\theta^L, \theta^H, \eta_1, \eta_0) \in \mathbb{R}^{2+K_1+K_0}, \qquad \nabla g^H = \nabla g^H(\theta^L, \theta^H, \eta_1, \eta_0) \in \mathbb{R}^{2+K_1+K_0}$$

Therefore T_4 is fully Hadamard differentiable at $(\theta^L, \theta^H, \eta_1, \eta_0)$ tangentially to $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0}$, with derivative

$$\begin{aligned} &T'_{4,T_3(T_2(T_1(P)))} : \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \to \mathbb{R} \times \mathbb{R} \\ &T'_{4,T_3(T_2(T_1(P)))}(h^L, h^H, h_{\eta_1}, h_{\eta_0}) \\ &\quad = \big(\big\langle \nabla g^L, (h^L, h^H, h_{\eta_1}, h_{\eta_0})\big\rangle, \big\langle \nabla g^H, (h^L, h^H, h_{\eta_1}, h_{\eta_0})\big\rangle\big) \end{aligned}$$

Proof. Assumption 3 (iii) is that g^L and g^H are continuously differentiable. The result follows. □

Remark C.1. This remark discusses the derivatives of g^L and g^H. In particular, note that even if $\arg\min_{t \in [\theta^L, \theta^H]} g(t, \eta)$ is within (θ^L, θ^H), the derivative of g^L and g^H are unlikely to be zero because the derivatives with respect to η will not be zero.

Consider $g^H(\theta^L, \theta^H, \eta) = \sup_{t \in [\theta^L, \theta^H]} g(t, \eta)$. The maximization problem has Lagrangian

$$\mathcal{L}(t, \lambda, \theta^L, \theta^H, \eta) = g(t, \eta) + \lambda^L(t - \theta^L) + \lambda^H(\theta^H - t)$$

where $\lambda = (\lambda^L, \lambda^H)$ are Lagrange multipliers. Let $g_\theta(t, \eta) = \frac{\partial g}{\partial \theta}(t, \eta)$. Suppose there is unique solution (θ^*, λ^*). The necessary KKT conditions imply that

$$\begin{aligned} &g_\theta(\theta^*, \eta) + \lambda^{L*} - \lambda^{H*} = 0 \\ &\theta^* - \theta^{L*} \geq 0 \text{ w.e. if } \lambda^{L*} > 0 \\ &\theta^{H^*} - \theta^* \geq 0 \text{ w.e. if } \lambda^{H*} > 0 \\ &\lambda^{L*}, \lambda^{H*} \geq 0 \end{aligned}$$

Notice that at most one of either $\theta^* = \theta^L$ or $\theta^* = \theta^H$ is true. If $\theta^* = \theta^L$, then $\lambda^L > 0$ and $\lambda^H = 0$, and the first KKT implies $-g_\theta(\theta^L, \eta) = \lambda^L$. Similarly, if $\theta^* = \theta^H$ is true then $\lambda^L = 0$ and $g_\theta(\theta^H, \eta) = \lambda^H$.

Now use assumption 3 (iii) to apply the envelope theorem, finding that

$$\begin{aligned}\nabla g^L(\theta^L,\theta^H,\eta)^\intercal &= \begin{pmatrix}\frac{\partial \mathcal{L}}{\partial \theta^L}(\theta^*,\lambda^*,\theta^L,\theta^H,\eta) & \frac{\partial \mathcal{L}}{\partial \theta^H}(\theta^*,\lambda^*,\theta^L,\theta^H,\eta) & \frac{\partial \mathcal{L}}{\partial \eta}(\theta^*,\lambda^*,\theta^L,\theta^H,\eta)\end{pmatrix} \\ &= \begin{pmatrix}-\lambda^{L*} & \lambda^{H*} & g_\eta(\theta^*,\eta)\end{pmatrix}\end{aligned}$$

The linearization of g^L at (θ^L,θ^H,η) is the function $g^{L'}_{(\theta^L,\theta^H,\eta)} : \mathbb{R}\times\mathbb{R}\times\mathbb{R}^{d_1+d_2} \to \mathbb{R}$ given by

$$\begin{aligned} g^{L'}_{(\theta^L,\theta^H,\eta)}(h_L,h_H,h_\eta) = \nabla g^L(\theta^L,\theta^H,\eta)^\intercal h &= \begin{cases} g_\theta(\theta^*,\eta)h_L + g_\eta(\theta^*,\eta)^\intercal h_\eta & \text{if } \theta^* = \theta^L \\ g_\theta(\theta^*,\eta)h_H + g_\eta(\theta^*,\eta)^\intercal h_\eta & \text{if } \theta^* = \theta^H \\ g_\eta(\theta^*,\eta)^\intercal h_\eta & \text{if } \theta^* \in (\theta^L,\theta^H)\end{cases} \\ &= \begin{pmatrix} g_\theta(\theta^*,\eta)\mathbb{1}\{\theta^*=\theta^L\} & g_\theta(\theta^*,\eta)\mathbb{1}\{\theta^*=\theta^H\} & g_\eta(\theta^*,\eta)^\intercal\end{pmatrix}\begin{pmatrix}h_L \\ h_H \\ h_\eta\end{pmatrix}\end{aligned}$$

where $g_\eta(t,\eta) = \frac{\partial g}{\partial \eta}(t,\eta)$. In particular, notice that the first order condition $g_\theta(\theta^*,\eta)=0$, which holds true when $\theta^* \in (0,1)$, does *not* imply this linearization is the zero map, as long as $g_\eta(\theta^*,\eta)$ is not zero.

C.6 The map $T(P) = (\gamma^L,\gamma^H)$, consistency, and weak convergence

Lemma C.16. *Let T_1, T_2, T_3, and T_4 be as defined in lemmas C.7, C.13, C.14, and C.15 respectively. Let*

$$\begin{aligned}\left(\left\{\hat{P}_{1|x},\hat{P}_{0|x},\hat{\eta}_{1,x},\hat{\eta}_{0,x},\hat{s}_x\right\}_{x\in\mathcal{X}}\right) &= T_1(\mathbb{P}_n) \\ \left(\{\hat{\theta}^L_x,\hat{\theta}^H_x,\hat{\eta}_{1,x},\hat{\eta}_{0,x},\hat{s}_x\}_{x\in\mathcal{X}}\right) &= T_2(T_1(\mathbb{P}_n)) \\ (\hat{\theta}^L,\hat{\theta}^H,\hat{\eta}) &= T_3(T_2(T_1(\mathbb{P}_n))), \\ (\hat{\gamma}^L,\hat{\gamma}^H) &= T_4(T_3(T_2(T_1(\mathbb{P}_n))))\end{aligned}$$

be the empirical analogue estimators. If assumptions 1, 2, and 3 hold, then each of these estimators are consistent.

Proof. Lemmas C.7, C.13, C.14, and C.15 show that T_1, T_2, T_3, and T_4 are Hadamard (directionally) differentiable at P, $T_1(P)$, $T_2(T_1(P))$, and $T_3(T_2(T_1(P)))$ respectively, tangentially to sets that include zero. It follows that these functions are continuous at P, $T_1(P)$, $T_2(T_1(P))$, and $T_3(T_2(T_1(P)))$ respectively.[14] Lemma C.5 implies that $\mathbb{P}_n \xrightarrow{p} P$ in $\ell^\infty(\mathcal{F})$, so it follows from the

[14] For normed spaces $\mathbb{D}$, $\mathbb{E}$, $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbb{E}$ is continuous at $\theta \in \mathbb{D}_\phi$ if and only if for every sequence $\{\theta_n\}_{n=1}^\infty \subseteq \mathbb{D}_\phi \setminus \{\theta\}$ with $\theta_n \to \theta$, $\|\phi(\theta_n)-\phi(\theta)\|_{\mathbb{E}} \to 0$. For such a sequence $\{\theta_n\}_{n=1}^\infty$, let $t_n = \|\theta_n - \theta\|_{\mathbb{D}}^{1/2}$ and notice that $t_n \downarrow 0$, $h_n := \frac{\theta_n-\theta}{t_n} \to 0 \in \mathbb{D}_0$, and $\theta + t_n h_n = \theta_n \in \mathbb{D}_\phi$ for all n. The definition of Hadamard directional differentiability then implies $\|\phi(\theta+t_nh_n)-\phi(\theta)-t_n\phi'_\theta(h)\|_{\mathbb{E}} \to 0$, while the reverse traingle inequality implies

$$\begin{aligned}&\|\phi(\theta+t_nh_n)-\phi(\theta)-t_n\phi'_\theta(h)\|_{\mathbb{E}} \geq \left|\|\phi(\theta+t_nh_n)-\phi(\theta)\|_{\mathbb{E}} - t_n\|\phi'_\theta(h)\|_{\mathbb{E}}\right| \geq \|\phi(\theta+t_nh_n)-\phi(\theta)\|_{\mathbb{E}} - t_n\|\phi'_\theta(h)\|_{\mathbb{E}} \\ \Longrightarrow\ & 0 \leq \|\phi(\theta+t_nh_n)-\phi(\theta)\|_{\mathbb{E}} \leq \|\phi(\theta+t_nh_n)-\phi(\theta)-t_n\phi'_\theta(h)\|_{\mathbb{E}} + t_n\|\phi'_\theta(h)\|_{\mathbb{E}} \to 0\end{aligned}$$

showing continuity at θ.

continuous mapping theorem that

$$
\begin{aligned}
T_1(\mathbb{P}_n) &\xrightarrow{p} T_1(P) \\
T_2(T_1(\mathbb{P}_n)) &\xrightarrow{p} T_2(T_1(P)) \\
T_3(T_2(T_1(\mathbb{P}_n))) &\xrightarrow{p} T_3(T_2(T_1(P))) \\
T_4(T_3(T_2(T_1(\mathbb{P}_n)))) &\xrightarrow{p} T_4(T_3(T_2(T_1(P))))
\end{aligned}
$$

In other words, the estimates are all consistent in their respective spaces. □

Lemma C.17 (*T* is Hadamard directionally differentiable)**.** *Let $\mathbb{D}_C$ be defined by* (59)*, and*

$$T : \mathbb{D}_C \to \mathbb{R}^2, \qquad\qquad T(G) = T_4(T_3(T_2(T_1(G))))$$

If assumptions 1, 2, 3 holds, then T is Hadamard directionally differentiable at P tangentially to $\mathcal{C}(\mathcal{F}, L_{2,P})$ with derivative given by

$$T_P' : \mathcal{C}(\mathcal{F}, L_{2,P}) \to \mathbb{R}^2, \qquad T_P'(G) = T'_{4,T_3(T_2(T_1(P)))}(T'_{3,T_2(T_1(P))}(T'_{2,T_1(P)}(T'_{1,P}(G))))$$

If assumption 4 also holds, then T is fully Hadamard differentiable at P tangentially to the support of $\mathbb{G}$ as defined in lemma C.5.

Proof. Lemma C.7 shows that T_1 is fully Hadamard differentiable at any point in $\mathbb{D}_C$ tangentially to $\ell^\infty(\mathcal{F})$. Lemma C.13 shows that under assumptions 1, 2, and 3, T_2 is Hadamard directionally differentiable at $T_1(P)$ tangentially to

$$\mathbb{D}_{Tan} = \prod_{m=1}^{M} \mathcal{C}(\mathcal{F}_{1,x_m}, L_{2,P}) \times \mathcal{C}(\mathcal{F}_{0,x_m}, L_{2,P}) \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}$$

Lemma C.8 implies that if $H \in \mathcal{C}(\mathcal{F}, L_{2,P})$, then $T'_{1,P}(H) \in \mathbb{D}_{Tan}$. It follows from the chain rule (lemma F.4) that $T_2 \circ T_1$ is Hadamard directionally differentiable at P tangentially to $\mathcal{C}(\mathcal{F}, L_{2,P})$. Lemma C.14 shows T_3 is fully differentiable at any point in its domain tangentially to the entire relevant space, and lemma C.15 shows T_4 is fully differentiable at $T_3(T_2(T_1(P)))$ tangentially to the entire relevant space. The chain rule thus implies the first claim: under assumptions 1, 2, and 3, $T = T_4 \circ T_3 \circ T_2 \circ T_1$ is Hadamard directionally differentiable at P tangentially to $\mathcal{C}(\mathcal{F}, L_{2,P})$ with the claimed derivative.

If assumption 4 also holds, lemma C.13 implies that T_2 is fully differentiable at $T_1(P)$ tangentially to $\mathbb{D}_{Tan,Full}$. Lemma C.9 shows the support of $T'_{1,P}(\mathbb{G})$ is contained within $\mathbb{D}_{Tan,Full}$. It follows that $T_P'(\cdot) = T'_{4,T_3(T_2(T_1(P)))}(T'_{3,T_2(T_1(P))}(T'_{2,T_1(P)}(T'_{1,P}(\cdot))))$ is linear on the support of $\mathbb{G}$, and hence Fang & Santos (2019) proposition 2.1 implies T is fully Hadamard differentiable at P tangentially to the support of $\mathbb{G}$. □

Lemma 5.1. *Suppose that*

(i) assumption 2 (i) holds, with cost function $c(y_1, y_0)$ that is continuously differentiable, and

(ii) for each (d, x), the support of $P_{d|x}$ is $\mathcal{Y}_{d,x}$, which is a bounded interval.

Then assumption 4 holds.

Proof. Note that both $c_L(y_1, y_0) = c(y_1, y_0)$ and $c_H(y_1, y_0) = -c(y_1, y_0)$ are continuously differentiable. Moreover, since the support of $P_{d|x}$ is $\mathcal{Y}_{d,x}$ which is a bounded interval, the support can be written as $[y^\ell_{d,x}, y^u_{d,x}]$. So for any $x \in \mathcal{X}$ and either $c \in \{c_L, c_H\}$, lemma B.3 shows that for any $(\varphi_1, \psi_1), (\varphi_2, \psi_2) \in \Psi_c(P_{1|x}, P_{0|x})$, there exists $s \in \mathbb{R}$ such that for all $(y_1, y_0) \in \mathcal{Y}_{1,x} \times \mathcal{Y}_{0,x}$

$$\varphi_1(y_1) - \varphi_2(y_1) = s, \qquad \psi_1(y_0) - \psi_2(y_0) = -s$$

and thus

$$\mathbb{1}_{\mathcal{Y}_{1,x}} \times \varphi_1 = \mathbb{1}_{\mathcal{Y}_{1,x}} \times (\varphi_2 + s),\ P\text{-a.s.} \qquad \text{and} \qquad \mathbb{1}_{\mathcal{Y}_{0,x}} \times \psi_1 = \mathbb{1}_{\mathcal{Y}_{0,x}} \times (\psi_2 - s),\ P\text{-a.s..}$$

Therefore assumption 4 holds. □

Theorem 5.2. *Suppose assumptions 1, 2, and 3 hold, and let $\mathbb{G}$ be the weak limit of $\sqrt{n}(\mathbb{P}_n - P)$ in $\ell^\infty(\mathcal{F})$. Then T is Hadamard directionally differentiable at P tangentially to the support of $\mathbb{G}$, and*

$$\sqrt{n}((\hat{\gamma}^L, \hat{\gamma}^H) - (\gamma^L, \gamma^H)) = \sqrt{n}(T(\mathbb{P}_n) - T(P)) \xrightarrow{L} T'_P(\mathbb{G})$$

If assumption 4 also holds, then T'_P is linear on the support of $\mathbb{G}$ and $T'_P(\mathbb{G})$ is bivariate normal.

Proof. The result is an application of the functional delta method (see Fang & Santos (2019) theorem 2.1) and lemma C.17.

Indeed, $\ell^\infty(\mathcal{F})$ and $\mathbb{R}^2$ are Banach spaces, and under assumptions 1, 2, and 3 lemma C.17 shows T is Hadamard directionally differentiable at P tangentially to $\mathcal{C}(\mathcal{F}, L_{2,P})$. Lemma C.5 shows that $\sqrt{n}(\mathbb{P}_n - P) \xrightarrow{L} \mathbb{G}$ in $\ell^\infty(\mathcal{F})$, where $\mathbb{G}$ is tight and supported in $\mathcal{C}(\mathcal{F}, L_{2,P})$. Fang & Santos (2019) theorem 2.1 gives the result that $\sqrt{n}(T(\mathbb{P}_n) - T(P)) \xrightarrow{L} T'_P(\mathbb{G})$.

If assumption 4 holds as well as assumptions 1, 2, and 3, then lemma C.17 shows that T is fully differentiable on the support of $\mathbb{G}$. Since $\mathbb{G}$ is Gaussian and T'_P is continuous and linear on the support of $\mathbb{G}$, $T'_P(\mathbb{G}) \in \mathbb{R}^2$ is Gaussian. □

D Appendix: inference

D.1 Bootstrap

Lemma D.1. *Suppose assumptions 1, 2, and 3 are satisfied. Let $\mathbb{P}_n^*$ be given by definition 5.1 or 5.2. Then Fang & Santos (2019) assumption 3 is satisfied:*

(i) $\mathbb{P}_n^*$ *is a function of* $\{Y_i, D_i, Z_i, X_i, W_i\}_{i=1}^n$*, with* $\{W_i\}_{i=1}^n$ *independent of* $\{Y_i, D_i, Z_i, X_i\}_{i=1}^n$.

(ii) $\mathbb{P}_n^*$ *satisfies* $\sup_{f \in BL_1} |E[f(\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)) \mid \{Y_i, D_i, Z_i, X_i\}_{i=1}^n] - E[f(\mathbb{G})]| = o_p(1)$.

(iii) $\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)$ *is asymptotically measurable (jointly in* $\{Y_i, D_i, Z_i, X_i, W_i\}_{i=1}^n$*).*

(iv) $f(\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)$ *is a measurable function of* $\{W_i\}_{i=1}^n$ *outer almost surely in* $\{\{Y_i, D_i, Z_i, X_i\}_{i=1}^n$ *for any continuous and bounded real-valued* f.

Proof. Note that assumption 3(i) is satisfied by construction. van der Vaart & Wellner (1997) example 3.6.9, 3.6.10, and theorem 3.6.13 implies assumpion 3(ii) holds:

$$\sup_{f \in \mathrm{BL}_1} \left| E\left[f(\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n)) \mid \{Y_i, D_i, Z_i, X_i\}_{i=1}^n \right] - E[f(\mathbb{G})] \right| \xrightarrow{P^*} 0$$

and further that

$$E\left[f(\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n))^* \right] - E\left[f(\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n))_* \right] = o_p(1)$$

for any $f \in \mathrm{BL}_1$, where $f(\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n))^*$ and $f(\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n))_*$ denote the minimal measurable majorant and maximal measurable minorant of $f(\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n))$, respectively. Note that for any continuous and bounded f, $f(\sqrt{n}(\mathbb{P}_n^* - \mathbb{P}_n))$ is continuous in $\{W_i\}_{i=1}^n$, and is hence measurable satisfying Fang & Santos (2019) assumption 3(iv). Fang & Santos (2019) lemma S.3.9 then implies assumption 3(iii) is satisfied as well. □

Theorem 5.3. *Suppose assumptions 1, 2, 3, and 4 hold, and let $\mathbb{P}_n^*$ be given by definition 5.1 or 5.2. Then conditional on* $\{Y_i, D_i, Z_i, X_i\}_{i=1}^n$,

$$\sqrt{n}(T(\mathbb{P}_n^*) - T(\mathbb{P}_n)) \xrightarrow{L} T_P'(\mathbb{G})$$

in outer probability.

Proof. By application of Fang & Santos (2019) theorem 3.1. There are three numbered assumptions:

1. Fang & Santos (2019) assumption 1 is satisfied; $\ell^\infty(\mathcal{F})$ and $\mathbb{R}^2$ are indeed Banach spaces, and lemma C.17 shows that under this paper's assumptions 1, 2, and 3, the map T is Hadamard directionally differentiable at P tangentially to $\mathcal{C}(\mathcal{F}, L_{2,P})$.

2. Fang & Santos (2019) assumption 2 is satisfied; lemma C.5 shows that $\sqrt{n}(\mathbb{P}_n - P) \xrightarrow{L} \mathbb{G}$ in $\ell^\infty(\mathcal{F})$, where $\mathbb{G}$ is tight and supported in $\mathcal{C}(\mathcal{F}, L_{2,P})$.

3. Lemma D.1 shows that Fang & Santos (2019) assumption 3 is satisfied.

Finally, note that $\mathbb{G}$ is Gaussian and mean zero; it follows that its support is a vector subspace of $\ell^\infty(\mathcal{F})$. Thus Fang & Santos (2019) theorem 3.1 implies T is (fully) Hadamard differentiable tangentially to the support of $\mathbb{G}$ if and only if

$$\sup_{f \in \mathrm{BL}_1} \left| E\left[f\left(\sqrt{n}(T(\mathbb{P}_n^*) - T(\mathbb{P}_n))\right) \mid \{Y_i, D_i, Z_i, X_i\}_{i=1}^n \right] - E\left[f(T_P'(\mathbb{G}))\right] \right| = o_p(1)$$

Since lemma C.17 shows that under assumptions 1, 2, 3, and 4, T is fully Hadamard differentiable tangentially to the support of $\mathbb{G}$, this completes the proof. □

D.2 Alternative procedure

Lemma D.2. *Let assumptions 1, 2, and 3 hold, and $\{\kappa_n\}_{n=1}^\infty \subseteq \mathbb{R}$ satisfy $\kappa_n \to \infty$ and $\kappa_n/\sqrt{n} \to 0$. For $c \in \{c_L, c_H\}$, let*

$$\begin{aligned}
\Psi_c(P_{1|x}, P_{0|x}) &= \operatorname*{arg\,max}_{(\varphi,\psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)} P_{1|x}(\varphi) + P_{0|x}(\psi) \\
\widehat{\Psi}_{c,x} &= \left\{ (\varphi, \psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c) \ ; \ OT_c(\hat{P}_{1|x}, \hat{P}_{0|x}) \leq \hat{P}_{1|x}(\varphi) + \hat{P}_{0|x}(\psi) + \frac{\kappa_n}{\sqrt{n}} \right\}
\end{aligned}$$

and $OT'_{c,(P_{1|x},P_{0|x})}, \widehat{OT}'_{c,x} : \mathcal{C}(\mathcal{F}_{1,x}, L_{2,P}) \times \mathcal{C}(\mathcal{F}_{0,x}, L_{2,P}) \to \mathbb{R}$, be given by

$$\begin{aligned}
OT'_{c,(P_{1,|x},P_{0|x})}(H_1, H_0) &= \sup_{(\varphi,\psi) \in \Psi_c(P_{1|x}, P_{0|x})} H_1(\varphi) + H_0(\psi) \\
\widehat{OT}'_{c,x}(H_1, H_0) &= \sup_{(\varphi,\psi) \in \widehat{\Psi}_{c,x}} H_1(\varphi) + H_0(\psi)
\end{aligned}$$

Then for any $(H_1, H_0) \in \mathcal{C}(\mathcal{F}_{1,x}, L_{2,P}) \times \mathcal{C}(\mathcal{F}_{0,x}, L_{2,P})$,

$$\left| \widehat{OT}'_{c,x}(H_1, H_0) - OT'_{c,(P_{1,|x},P_{0|x})}(H_1, H_0) \right| \xrightarrow{p} 0$$

Proof. The proof is similar that of Fang & Santos (2019) lemma S.4.8. As the subscript x plays no role, we drop it from the notation.

In steps:

1. We first esteablish an inequality used several times below. Note that for any $(\tilde{\varphi}, \tilde{\psi}), (\varphi, \psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$,

$$\begin{aligned}
\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0} &\geq \hat{P}_1(\varphi) - P_1(\varphi) + \hat{P}_0(\psi) - P_0(\psi) \\
\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0} &\geq P_1(\tilde{\varphi}) - \hat{P}_1(\tilde{\varphi}) + P_0(\tilde{\psi}) - \hat{P}_0(\tilde{\psi})
\end{aligned}$$

Add these to obtain

$$\begin{aligned}
&2\left(\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0} \right) \\
&\qquad \geq \hat{P}_1(\varphi) - P_1(\varphi) + \hat{P}_0(\psi) - P_0(\psi) + P_1(\tilde{\varphi}) - \hat{P}_1(\tilde{\varphi}) + P_0(\tilde{\psi}) - \hat{P}_0(\tilde{\psi}), \qquad (70)
\end{aligned}$$

2. We next show

$$\lim_{n\to\infty} P\left(\Psi(P_1,P_0) \subseteq \widehat{\Psi}_c\right) = 1 \tag{71}$$

Let $(\tilde{\varphi},\tilde{\psi}) \in \Psi(P_1,P_0)$, and rearrange (70) to find

$$\begin{aligned}
&2\left(\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0}\right)\\
&\qquad \geq \hat{P}_1(\varphi) + \hat{P}_0(\psi) - \hat{P}_1(\tilde{\varphi}) - \hat{P}(\tilde{\psi}) + \underbrace{P_1(\tilde{\varphi}) + P_0(\tilde{\psi}) - P_1(\varphi) - P_0(\psi)}_{\geq 0}\\
&\qquad \geq \hat{P}_1(\varphi) + \hat{P}_0(\psi) - \hat{P}_1(\tilde{\varphi}) - \hat{P}(\tilde{\psi})
\end{aligned}$$

and therefore

$$\sup_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} \hat{P}_1(\varphi) + \hat{P}_0(\psi) \leq \hat{P}_1(\tilde{\varphi}) + \hat{P}(\tilde{\psi}) + 2\left(\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0}\right)$$

holds for any $(\tilde{\varphi},\tilde{\psi}) \in \Psi_c(P_1,P_0)$. It follows that $2\left(\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0}\right) < \frac{\kappa_n}{\sqrt{n}}$ implies $(\tilde{\varphi},\tilde{\psi}) \in \widehat{\Psi}_c$, and hence

$$P\left(2\frac{\sqrt{n}}{\kappa_n}\left(\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0}\right) < 1\right) \leq P\left(\Psi(P_1,P_0) \subseteq \widehat{\Psi}_c\right)$$

Lemma C.16 implies $\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0} \xrightarrow{p} 0$. Since $\frac{\sqrt{n}}{\kappa_n} \to 0$, this implies that $2\frac{\sqrt{n}}{\kappa_n}\left(\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0}\right) = o_p(1)$ and therefore

$$\lim_{n\to\infty} P\left(\Psi(P_1,P_0) \subseteq \widehat{\Psi}_c\right) \geq \lim_{n\to\infty} P\left(2\frac{\sqrt{n}}{\kappa_n}\left(\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0}\right) < 1\right) = 1$$

as was to be shown.

3. We next show that for any $\delta > 0$,

$$\lim_{n\to\infty} P\left(\widehat{\Psi}_c \subseteq (\Psi(P_1,P_0))^\delta\right) = 1 \tag{72}$$

where $(\Psi(P_1,P_0))^\delta$ is an open δ-enlargement of $\Psi(P_1,P_0)$ under L_2; i.e.

$$(\Psi(P_1,P_0))^\delta = \left\{(f,g)\ ;\ \inf_{(\varphi,\psi)\in\Psi(P_1,P_0)} L_2((\varphi,\psi),(f,g)) < \delta\right\}$$

Toward this end, note that

$$\eta \equiv \left[\sup_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} \{P_1(\varphi) + P_0(\psi)\} - \sup_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)\setminus(\Psi(P_1,P_0))^\delta} \{P_1(\varphi) + P_0(\psi)\}\right] > 0$$

$\eta > 0$ follows from compactness of $\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$ and continuity of $P_1 + P_0$ with respect to L_2 (see the proof of lemma B.2).

Rearrange (70) to find

$$\begin{aligned}
&P_1(\tilde{\varphi}) + P_0(\tilde{\psi}) - P_1(\varphi) - P_0(\psi) \\
&\qquad \leq 2\left(\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0}\right) + \hat{P}_1(\tilde{\varphi}) + \hat{P}_0(\tilde{\psi}) - \hat{P}_1(\varphi) - \hat{P}_0(\psi)
\end{aligned}$$

Take suprema over $(\tilde{\varphi}, \tilde{\psi}) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$ to find

$$\begin{aligned}
&\sup_{(\tilde{\varphi},\tilde{\psi})\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} P_1(\tilde{\varphi}) + P_0(\tilde{\psi}) - P_1(\varphi) - P_0(\psi) \\
&\qquad \leq 2\left(\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0}\right) + \sup_{(\tilde{\varphi},\tilde{\psi})\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} \hat{P}_1(\tilde{\varphi}) + \hat{P}_0(\tilde{\psi}) - \hat{P}_1(\varphi) - \hat{P}_0(\psi)
\end{aligned} \tag{73}$$

Suppose there exists $(\varphi, \psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c) \setminus (\Psi(P_1, P_0))^\delta$ such that $\sup_{(\tilde{\varphi},\tilde{\psi})\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} \hat{P}_1(\tilde{\varphi}) + \hat{P}_0(\tilde{\psi}) \leq \hat{P}_1(\varphi) + \hat{P}_0(\psi) + \frac{\kappa}{\sqrt{n}}$. For any such (φ, ψ), (73) implies

$$\sup_{(\tilde{\varphi},\tilde{\psi})\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} P_1(\tilde{\varphi}) + P_0(\tilde{\psi}) - P_1(\varphi) - P_0(\psi) \leq 2\left(\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0}\right) + \frac{\kappa_n}{\sqrt{n}}$$

from which it follows that

$$\begin{aligned}
&2\left(\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0}\right) + \frac{\kappa_n}{\sqrt{n}} \\
&\qquad \geq \sup_{(\tilde{\varphi},\tilde{\psi})\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} P_1(\tilde{\varphi}) + P_0(\tilde{\psi}) - \sup_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)\setminus(\Psi(P_1,P_0))^\delta} \{P_1(\varphi) + P_0(\psi)\} \\
&\qquad = \eta
\end{aligned}$$

To summarize: if there exists $(\varphi, \psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c) \setminus (\Psi(P_1, P_0))^\delta$ such that $\sup_{(\tilde{\varphi},\tilde{\psi})\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} \hat{P}_1(\tilde{\varphi}) + \hat{P}_0(\tilde{\psi}) \leq \hat{P}_1(\varphi) + \hat{P}_0(\psi) + \frac{\kappa}{\sqrt{n}}$, then $2\left(\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0}\right) + \frac{\kappa_n}{\sqrt{n}} \geq \eta$, from which it follows that

$$\begin{aligned}
&P\left(\widehat{\Psi}_c \not\subseteq (\Psi(P_1, P_0))^\delta\right) \\
&\qquad = P\Bigg(\sup_{(\tilde{\varphi},\tilde{\psi})\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} \hat{P}_1(\tilde{\varphi}) + \hat{P}_0(\tilde{\psi}) \leq \hat{P}_1(\varphi) + \hat{P}_0(\psi) + \frac{\kappa}{\sqrt{n}} \\
&\qquad\qquad\qquad \text{for some } (\varphi, \psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c) \setminus (\Psi(P_1, P_0))^\delta\Bigg) \\
&\qquad \leq P\left(2\left(\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0}\right) + \frac{\kappa_n}{\sqrt{n}} \geq \eta\right) \to 0
\end{aligned}$$

where the final limit claim follows from $\eta > 0$, $\kappa_n/\sqrt{n} \to 0$, and $\|\hat{P}_1 - P_1\|_{\mathcal{F}_1} + \|\hat{P}_0 - P_0\|_{\mathcal{F}_0} = o_p(1)$.

4. (71) and (72) imply that for any $\delta > 0$, $P\left(\Psi_c(P_1, P_0) \subseteq \widehat{\Psi}_c \subseteq \Psi_c(P_1, P_0)^\delta\right) \to 1$. It follows that there exists a sequence $\{\delta_n\}_{n=1}^\infty \subseteq \mathbb{R}_+$ with $\delta_n \downarrow 0$ such that $P\left(\Psi(P_1, P_0) \subseteq \widehat{\Psi}_c \subseteq \Psi(P_1, P_0)^{\delta_n}\right) \to$

1. Notice that when $\Psi(P_1, P_0) \subseteq \widehat{\Psi}_c \subseteq \Psi(P_1, P_0)^{\delta_n}$ holds,

$$
\begin{aligned}
&\left|\widehat{OT}'_{c,x}(H_1, H_0) - OT'_{c,(P_1,P_0)}(H_1, H_0)\right| \\
&\leq \sup_{(\varphi,\psi)\in\Psi_c(P_1,P_0)^{\delta_n}\cap\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c)} \{H_1(\varphi) + H_0(\psi)\} - \sup_{(\varphi,\psi)\in\Psi_c(P_1,P_0)} \{H_1(\varphi) + H_0(\psi)\} \\
&\leq \sup_{(\varphi_1,\psi_1),(\varphi_2,\psi_2)\in\Phi_c\cap(\mathcal{F}_c\times\mathcal{F}_c^c);\, L_2((\varphi_1,\psi_1),(\varphi_2,\psi_2))<\delta_n} \{H_1(\varphi_1) + H_0(\psi_1) - H_1(\varphi_2) - H_0(\psi_0)\} \\
&= o_p(1)
\end{aligned}
$$

where the $o_p(1)$ claim follows from $H_1 + H_0$ being continuous and $\Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)$ being compact, implying $H_1 + H_0$ is in fact uniformly continuous.

This concludes the proof. $\square$

Theorem 5.4. *Suppose assumptions 1, 2, and 3 hold, let $\mathbb{P}_n^*$ be given by definition 5.1 or 5.2, and $\{\kappa_n\}_{n=1}^{\infty} \subseteq \mathbb{R}$ satisfy $\kappa_n \to \infty$ and $\kappa_n/\sqrt{n} \to 0$. Then conditional on $\{Y_i, D_i, Z_i, X_i\}_{i=1}^n$,*

$$\hat{D}_4 \hat{D}_3 \widehat{T}_{2,T_1(P)}(\sqrt{n}(T_1(\mathbb{P}_n^*) - T_1(\mathbb{P}_n))) \xrightarrow{L} T_P'(\mathbb{G})$$

in outer probability.

Proof. The overall strategy is to apply Fang & Santos (2019) theorem 3.2, viewing $T_1(\mathbb{P}_n)$ as the estimator for $T_1(P)$, $T_1(\mathbb{P}_n^*)$ as the bootstrap, and $T_{-1} = T_4 \circ T_3 \circ T_2$ as the directionally differentiable function. There are four assumption to verify.

1. To see that Fang & Santos (2019) assumption 1 holds,
   (i) the map
   $$T_4 \circ T_3 \circ T_2 : \prod_{m=1}^{M} \ell^\infty(\mathcal{F}_{1,x}) \times \ell^\infty(\mathcal{F}_{0,x}) \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R} \to \mathbb{R}^2$$
   is a map between Banach spaces
   (ii) by lemmas C.13, C.14, C.15 and the chain rule (lemma F.4), $T_{-1} = T_4 \circ T_3 \circ T_2$ is Hadamard directionally differentiable at $T_1(P)$ tangentially to
   $$\mathbb{D}_{Tan} = \prod_{m=1}^{M} \mathcal{C}(\mathcal{F}_{1,x_m}, L_{2,P}) \times \mathcal{C}(\mathcal{F}_{0,x_m}, L_{2,P}) \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}$$
2. To see that the estimator $T_1(\mathbb{P}_n)$ satisfies Fang & Santos (2019) assumption 2, note that
   (i) $T_1(P) \in \prod_{m=1}^{M} \ell^\infty(\mathcal{F}_{1,x}) \times \ell^\infty(\mathcal{F}_{0,x}) \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}$ and lemma C.9 shows
   $$T_1(\mathbb{P}_n) : \{Y_i, D_i, Z_i, X_i\}_{i=1}^n \to \prod_{m=1}^{M} \ell^\infty(\mathcal{F}_{1,x}) \times \ell^\infty(\mathcal{F}_{0,x}) \times \mathbb{R}^{K_1} \times \mathbb{R}^{K_0} \times \mathbb{R}$$
   satisfies $\sqrt{n}(T_1(\mathbb{P}_n) - T_1(P)) \xrightarrow{L} T'_{1,P}(\mathbb{G})$.

(ii) $T'_{1,P}(\mathbb{G})$ is tight because $\mathbb{G}$ is tight and $T'_{1,P}$ is continuous. Lemma C.9 also shows the support of $T'_{1,P}(\mathbb{G})$ is included in $\mathbb{D}_{Tan}$.

3. The bootstrap $T_1(\mathbb{P}_n^*)$ satisfies Fang & Santos (2019) assumption 3:

   (i) $T_1(\mathbb{P}_n^*)$ is a function of $\{Y_i, D_i, Z_i, X_i, W_i\}_{i=1}^n$ with $\{W_i\}_{i=1}^n$ independent of $\{Y_i, D_i, Z_i, X_i\}_{i=1}^n$.

   (ii) T_1 is fully Hadamard differentiable at P tangentially to $\ell^\infty(\mathcal{F})$, and hence the functional delta method implies $\sqrt{n}(T_1(\mathbb{P}_n)-T_1(P)) \xrightarrow{L} T'_{1,P}(\mathbb{G})$. Lemma D.1 shows that $\mathbb{P}_n^*$ satisfies Fang & Santos (2019) assumption 3, and thus Fang & Santos (2019) theorem 3.1 implies

$$\sup_{f\in \mathrm{BL}_1} \left| E\left[f(\sqrt{n}(T_1(\mathbb{P}_n^*) - T_1(\mathbb{P}_n))) \mid \{Y_i, D_i, Z_i, X_i\}_{i=1}^n\right] - E[f(T'_{1,P}(\mathbb{G}))]\right| = o_p(1)$$

   (iii) Condition (iv) below holds, and hence Fang & Santos (2019) lemma S.3.9 implies $\sqrt{n}(T_1(\mathbb{P}_n^*) - T_1(\mathbb{P}_n))$ is asymptotically measurable.

   (iv) Note that for any continuous and bounded function f, $f(\sqrt{n}(T_1(\mathbb{P}_n^*) - T_1(\mathbb{P}_n)))$ is continuous in $\{W_i\}_{i=1}^n$ and hence is a measurable function of $\{W_i\}_{i=1}^n$.

4. Fang & Santos (2019) assumption 4 is about the estimator of the derivative.

   Notice that $T'_{-1,T_1(P)} = T'_{4,T_3(T_2(T_1(P)))} \circ T'_{3,T_2(T_1(P))} \circ T'_{2,T_1(P)}$ is given by

$$T'_{-1,T_1(P)} : \mathbb{D}_{Tan} \to \mathbb{R}^2, \qquad\qquad T'_{-1,T_1(P)}(h) = D_4 D_3 T'_{2,T_1(P)}(h)$$

   Estimate this derivative with

$$\widehat{T}'_{-1,T_1(P)} : \mathbb{D}_{Tan} \to \mathbb{R}^2, \qquad\qquad \hat{D}_4 \hat{D}_3 \widehat{T}'_{2,T_1(P)}(h)$$

   The estimator $\widehat{T}'_{-1,T_1(P)}$ satisfies the conditions of Fang & Santos (2019) lemma S.3.6, and therefore Fang & Santos (2019) assumption 4. These conditions are

   (a) Modulus of continuity: $\|\widehat{T}'_{-1,T_1(P)}(h_1) - \widehat{T}'_{-1,T_1(P)}(h_2)\| \leq C_n \|h_1 - h_2\|$ for some $C_n = O_p(1)$.

   (b) Pointwise consistency: for any h, $\|\widehat{T}_{-1,T_1(P)}(h) - T_{-1,T_1(P)}(h)\| = o_p(1)$.

   To see these claims in detail:

   (a) For any matrix A, let $\|A\|_o = \sup_{x;\|x\|_2=1} \|Ax\|_2$ be the operator norm.

$$\begin{aligned}
\|\widehat{T}'_{-1,T_1(P)}(h_1) - \widehat{T}'_{-1,T_1(P)}(h_2)\| &= \|\hat{D}_4 \hat{D}_3 \widehat{T}'_{2,T_1(P)}(h_1) - \hat{D}_4 \hat{D}_3 \widehat{T}'_{2,T_1(P)}(h_2)\| \\
&\leq \|\hat{D}_4 \hat{D}_3\|_o \|\widehat{T}'_{2,T_1(P)}(h_1) - \widehat{T}'_{2,T_1(P)}(h_2)\| \\
&\leq \|\hat{D}_4 \hat{D}_3\| \| \| h_1 - h_2\|
\end{aligned}$$

   where the last claim follows because $\widehat{T}'_{2,T_1(P)}$ is 1-Lipschitz (shown below). Next notice $\hat{D}_4 \xrightarrow{p} D_4$ and $\hat{D}_3 \xrightarrow{p} D_3$ by the CMT, which implies $\|\hat{D}_4 \hat{D}_3\| = O_p(1)$ as required.

To see that $\widehat{T}'_{2,T_1(P)}$ is 1-Lipschitz, recall

$$\begin{aligned}
&\widehat{T}'_{2,T_1(P)}\left(\{H_{1,x}, H_{0,x}, h_{\eta_1,x}, h_{\eta_0,x}, h_{s,x}\}_{x\in\mathcal{X}}\right) \\
&\quad = \left(\left\{\widehat{OT}'_{c_L,x}(H_{1,x}, H_{0,x}), -\widehat{OT}'_{c_H,x}(H_{1,x}, H_{0,x}), h_{\eta_1,x}, h_{\eta_0,x}, h_{s,x}\right\}_{x\in\mathcal{X}}\right)
\end{aligned}$$

The maps $\widehat{OT}_{c_L,x}, -\widehat{OT}_{c_H,x}$ are 1-Lipschitz. Specifically, note that

$$\begin{aligned}
&|\widehat{OT}'_{c_L,x}(H_{1,x}, H_{0,x}) - \widehat{OT}'_{c_L,x}(G_{1,x}, G_{0,x})| \\
&\quad = \left|\sup_{(\varphi,\psi)\in\widehat{\Psi}_{c,x}} \{H_{1,x}(\varphi) + H_{0,x}(\psi)\} - \sup_{(\varphi,\psi)\in\widehat{\Psi}_{c,x}} \{G_{1,x}(\varphi) + G_{0,x}(\psi)\}\right| \\
&\quad \leq \sup_{\varphi\in\mathcal{F}_{1,x}} |H_{1,x}(\varphi) - G_{1,x}(\varphi)| + \sup_{\psi\in\mathcal{F}_{0,x}} |H_{0,x}(\psi) - G_{0,x}(\psi)| \\
&\quad = \|H_{1,x} - G_{1,x}\|_{\mathcal{F}_{1,x}} + \|H_{0,x} - G_{0,x}\|_{\mathcal{F}_{0,x}}
\end{aligned}$$

and similarly, $-\widehat{OT}_{c_H,x}$ is 1-Lipschitz. The other maps in $\widehat{T}_{2,T_1(P)}$ are the identity map, which is also 1-Lipschitz. It follows that $\widehat{T}_{2,T_1(P)}$ is 1-Lipschitz.[15]

(b) To show pointwise consistency, fix $h = (\{H_{1,x}, H_{0,x}, h_{\eta_1,x}, h_{\eta_0,x}, h_{s,x}\}_{x\in\mathcal{X}})$ and note that

$$\begin{aligned}
&\|\widehat{T}'_{-1,T_1(P)}(h) - T_{-1,T_1(P)}\| = \|\hat{D}_4\hat{D}_3\widehat{T}_{2,T_1(P)}(h) - D_4D_3T_{2,T_1(P)}(h)\| \\
&\quad \leq \|(\hat{D}_4\hat{D}_3 - D_4D_3)T'_{2,T_1(P)}(h)\| + \|D_4D_3(\widehat{T}_{2,T_1(P)}(h) - T_{2,T_1(P)}(h))\| \\
&\quad \leq \|\hat{D}_4\hat{D}_3 - D_4D_3\|_o\|T'_{2,T_1(P)}(h)\| + \|D_4D_3\|_o\|\widehat{T}_{2,T_1(P)}(h) - T_{2,T_1(P)}(h)\|
\end{aligned}$$

Since $\hat{D}_4\hat{D}_3 \xrightarrow{p} D_4D_3$ by the CMT, it suffices to show

$$\|\widehat{T}_{2,T_1(P)}(h) - T_{2,T_1(P)}(h)\| = o_p(1)$$

The only nonzero coordinates correspond to $\widehat{OT}^{L\prime}_{c_L,x}(H_{1,x}, H_{0,x})$ and $-\widehat{OT}^{H\prime}_{c_H,x}(H_{1,x}, H_{0,x})$:

$$\begin{aligned}
&\|\widehat{T}_{2,T_1(P)}(h) - T_{2,T_1(P)}(h)\|^2 \\
&\quad = \left(\widehat{OT}'_{c_L,x}(H_{1,x}, H_{0,x}) - OT'_{c_L,(P_{1|x},P_{0|x})}(H_{1,x}, H_{0,x})\right)^2 \\
&\qquad\qquad + \left(\widehat{OT}'_{c_H,x}(H_{1,x}, H_{0,x}) - OT'_{c_H,(P_{1|x},P_{0|x})}(H_{1,x}, H_{0,x})\right)^2 \\
&\quad = o_p(1) + o_p(1)
\end{aligned}$$

where the last $o_p(1)$ claim follows from lemma D.2.

[15] For $k = 1, 2$, let $\mathbb{D}_k$, $\mathbb{E}_k$ be metric spaces. If $f_k : \mathbb{D}_k \to \mathbb{E}_k$ be Lipschitz with constants L_k, then $f : \mathbb{D}_1 \times \mathbb{D}_2 \to \mathbb{E}_1 \times \mathbb{E}_2$ given by $f(x_1, x_2) = (f_1(x_1), f_2(x_2))$ is Lipschitz with constant $\max\{L_1, L_2\}$. To see this, recall $\mathbb{D}_1 \times \mathbb{D}_2$ and $\mathbb{E}_1 \times \mathbb{E}_2$ are metricized with the norms $\|(x_1, x_2)\|_{\mathbb{D}_1\times\mathbb{D}_2} = \|x_1\|_{\mathbb{D}_1} + \|x_2\|_{\mathbb{D}_2}$ and $\|(y_1, y_2)\|_{\mathbb{E}_1\times\mathbb{E}_2} = \|y_1\|_{\mathbb{E}_1} + \|y_2\|_{\mathbb{E}_2}$, and note that

$$\begin{aligned}
\|f(x_1, x_2) - f(x_1', x_2')\|_{\mathbb{E}_1\times\mathbb{E}_2} &= \|(f_1(x_1), f_2(x_2)) - (f_1(x_1'), f_2(x_2'))\|_{\mathbb{E}_1\times\mathbb{E}_2} = \|f_1(x_1) - f_1(x_1')\|_{\mathbb{E}_1} + \|f_2(x_2) - f_2(x_2')\|_{\mathbb{E}_2} \\
&\leq L_1\|x - x_1'\|_{\mathbb{D}_1} + L_2\|x_2 - x_2'\|_{\mathbb{D}_2} \leq \max\{L_1, L_2\}\|x - x_1'\|_{\mathbb{D}_1} + \max\{L_1, L_2\}\|x_2 - x_2'\|_{\mathbb{D}_2} \\
&= \max\{L_1, L_2\} \times \|(x_1, x_2) - (x_1', x_2')\|_{\mathbb{D}_1\times\mathbb{D}_2}
\end{aligned}$$

We conclude through Fang & Santos (2019) lemma S.3.6 that Fang & Santos (2019) assumption 4 is satisfied.

Finally, apply Fang & Santos (2019) theorem 3.2 to find that

$$\sup_{f \in \mathrm{BL}_1} \left| E\left[f(\hat{D}_4 \hat{D}_3 \widehat{T}_{2,T_1(P)}(\sqrt{n}(T_1(\mathbb{P}_n^*) - T_1(\mathbb{P}_n)))) \right] - E\left[f(T_P'(\mathbb{G})) \right] \right| = o_p(1)$$

as was to be shown. □

E Appendix: duality in optimal transport

This appendix contains terminology, notation, and results regarding optimal transport used in this paper. Most of these results can be found in the monographs Villani (2003), Villani (2009), or Santambrogio (2015).

E.1 Primal and dual problems

Let $\mathcal{Y}_1, \mathcal{Y}_0$ be Polish subsets of $\mathbb{R}$, equipped with their Borel sigma algebras. Let $\mathcal{P}(\mathcal{Y}_d)$ be the set of probability distributions defined on $\mathcal{Y}_d$, and $P_d \in \mathcal{P}(\mathcal{Y}_d)$. Let $\mathcal{P}(\mathcal{Y}_1 \times \mathcal{Y}_0)$ be the set of probability distributions on the product space $\mathcal{Y}_1 \times \mathcal{Y}_0$.

A probability measure $\pi \in \mathcal{P}(\mathcal{Y}_1 \times \mathcal{Y}_0)$ has marginals P_1 and P_0 if

$$\text{For all } A \subset \mathcal{Y}_1 \text{ measurable, } \pi(A \times \mathcal{Y}_0) = P_1(A) = \int \mathbb{1}_A(y_1) dP_1(y_1) \tag{74}$$

$$\text{For all } B \subset \mathcal{Y}_0 \text{ measurable, } \pi(\mathcal{Y}_1 \times B) = P_0(B) = \int \mathbb{1}_B(y_0) dP_0(y_0) \tag{75}$$

The collection of such joint distributions with marginals P_1 and P_0 is denoted

$$\Pi(P_1, P_0) = \{\pi \in \mathcal{P}(\mathcal{Y}_1 \times \mathcal{Y}_0)\ ;\ \pi \text{ satisfies (74) and (75)}\} \tag{76}$$

The **cost function** is a measurable function $c : \mathcal{Y}_1 \times \mathcal{Y}_0 \to \mathbb{R}$. The functional $I : \mathcal{P}(\mathcal{Y}_1 \times \mathcal{Y}_0) \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$I_c[\pi] = \int c(y_1, y_0) d\pi(y_1, y_0) \tag{77}$$

The **optimal cost** $OT_c(P_1, P_0)$ is the infimum of $I_c[\pi]$ over $\Pi(P_1, P_0)$:

$$OT_c(P_1, P_0) = \inf_{\pi \in \Pi(P_1, P_0)} I_c[\pi] = \inf_{\pi \in \Pi(P_1, P_0)} \int c(y_1, y_0) d\pi(y_1, y_0) \tag{78}$$

This minimization problem in (78) is known as **optimal transport**. When attained, a solution to (78) is called an **optimal transference plan** or **optimal coupling**. Attainment is common; Villani (2009) theorem 4.1 implies:

Lemma E.1 (Optimal transport is attained)**.** *Let $c : \mathcal{Y}_1 \times \mathcal{Y}_0 \to \mathbb{R}$ be lower semicontinuous and bounded from below. Then there exists $\pi^* \in \Pi(P_1, P_0)$ such that*

$$E_{\pi^*}[c(Y_1, Y_0)] = \inf_{\pi \in \Pi(P_1, P_0)} \int c(y_1, y_0) d\pi(y_1, y_0)$$

The dual problem will require some additional notation. For any probability measure P let $L^1(P)$ denote the P-integrable functions. Define

$$\Phi_c = \left\{(\varphi,\psi) \in L^1(P_1) \times L^1(P_0)\ ;\ \varphi(y_1) + \psi(y_0) \leq c(y_1, y_0)\right\}, \tag{79}$$

and $J : L^1(P_1) \times L^1(P_0) \to \mathbb{R}$ by

$$J(\varphi,\psi) = \int_{\mathcal{Y}_1} \varphi(y_1) dP_1(y_1) + \int_{\mathcal{Y}_0} \psi(y_0) dP_0(y_0) \tag{80}$$

The **dual problem** of optimal transport is

$$\sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi) = \sup_{(\varphi,\psi)\in\Phi_c} \int \varphi(y_1) dP_1(y_1) + \int \psi(y_0) dP_0(y_0) \tag{81}$$

E.2 Duality

For any topological space $\mathcal{Z}$, let $\mathcal{C}_b(\mathcal{Z})$ denotes the set of functions $f : \mathcal{Z} \to \mathbb{R}$ that are continuous and bounded, and

$$\Phi_c \cap \mathcal{C}_b = \{(\varphi,\psi) \in \mathcal{C}_b(\mathcal{Y}_1) \times \mathcal{C}_b(\mathcal{Y}_0)\ ;\ \varphi(y_1) + \psi(y_0) \leq c(y_1, y_0)\} \tag{82}$$

The following weak duality statement is Villani (2003) proposition 1.5.

Lemma E.2 (Weak duality)**.**

$$\sup_{(\varphi,\psi)\in\Phi_c\cap\mathcal{C}_b} J(\varphi,\psi) \leq \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi) \leq \inf_{\pi\in\Pi(P_1,P_0)} I_c[\pi]$$

The following strong duality statement can be directly inferred from Villani (2009) theorem 5.10, or Santambrogio (2015) theorem 1.42, and so is presented without proof.

Theorem E.3 (Strong duality)**.** *Let $c : \mathcal{Y}_1 \times \mathcal{Y}_0 \to \mathbb{R}$ be lower semi-continuous and bounded from below. Then*

$$\inf_{\pi\in\Pi(P_1,P_0)} I_c[\pi] = \sup_{\varphi,\psi\in\Phi_c} J(\varphi,\psi) = \sup_{(\varphi,\psi)\in\Phi_c\cap\mathcal{C}_b} J(\varphi,\psi) \tag{83}$$

Moreover, the infimum of the left-hand side of (83) *is attained.*

E.3 c-concave functions

For any function $\varphi : \mathcal{Y}_1 \to \mathbb{R}$ and cost function $c(y_1, y_0)$, define the **c-transform** of φ as the function $\varphi^c : \mathcal{Y}_0 \to \mathbb{R}$ given by

$$\varphi^c(y_0) = \inf_{y_1 \in \mathcal{Y}_1} \{c(y_1, y_0) - \varphi(y_1)\}.$$

Similarly, $\psi^c(y_1) = \inf_{y_0 \in \mathcal{Y}_0}\{c(y_1, y_0) - \psi(y_0)\}$ is the c-transform of ψ. φ is called **c-concave** if $\varphi^{cc} = (\varphi^c)^c = \varphi$. If φ is c-concave, then (φ, φ^c) is called a **c-concave conjugate pair**.

The following lemma E.4 is exercise 2.35 found in Villani (2003) and presented without proof.

Lemma E.4 (Villani (2003) exercise 2.35)**.** *Let $\mathcal{Y}_1$ and $\mathcal{Y}_0$ be nonempty sets and $c : \mathcal{Y}_1 \times \mathcal{Y}_0 \to \mathbb{R}$ be an arbitrary function. Let $\varphi : \mathcal{Y}_1 \to \mathbb{R}$. Then*

(i) $\varphi(y_1) + \varphi^c(y_0) \leq c(y_1, y_0)$ *for all* $(y_1, y_0) \in \mathcal{Y}_1 \times \mathcal{Y}_0$

(ii) $\varphi^{cc}(y_1) \geq \varphi(y_1)$ *for all* $y_1 \in \mathcal{Y}_1$*, and*

(iii) $\varphi^{ccc}(y_0) = \varphi^c(y_0)$ *for all* $y_0 \in \mathcal{Y}_0$

It follows that $\varphi^{cc} = \varphi$ if and only if φ is c-concave.

For $H \subseteq \{(f, g)\ ;\ f : \mathcal{Y}_1 \to \mathbb{R},\ \text{and}\ g : \mathcal{Y}_0 \to \mathbb{R}\}$, let

$$\mathcal{F}_c^c(H) = \left\{\varphi^c : \mathcal{Y}_0 \to \mathbb{R}\ ;\ \exists (f, g) \in H \text{ s.t. } \varphi^c(y_0) = \inf_{y_1 \in \mathcal{Y}_1} \{c(y_1, y_0) - f(y_1)\}\right\} \tag{84}$$
$$\mathcal{F}_c(H) = \left\{\varphi : \mathcal{Y}_1 \to \mathbb{R}\ ;\ \exists \varphi^c \in F_c^c(H) \text{ s.t. } \varphi(y_1) = \inf_{y_0 \in \mathcal{Y}_0} \{c(y_1, y_0) - \varphi^c(y_0)\}\right\}$$

$\mathcal{F}_c(H)$ is called the **c-concave functions generated by H**, and $\mathcal{F}_c^c(H)$ the **c-conjugates generated by H.**[16] Notice that not every $(\varphi, \psi) \in \mathcal{F}_c(H) \times \mathcal{F}_c^c(H)$ is a c-concave conjugate pair.

Lemma E.5 (Restricting the dual to c-concave functions)**.** *Let $\Phi_{cs} \subseteq \Phi_c$ be such that*

1. *strong duality holds:* $\inf_{\pi \in \Pi(P_1, P_0)} I_c[\pi] = \sup_{(\varphi,\psi) \in \Phi_{cs}} J(\varphi, \psi)$*, and*

2. *the c-concave functions generated by Φ_{cs} are integrable:* $\mathcal{F}_c(\Phi_{cs}) \times \mathcal{F}_c^c(\Phi_{cs}) \subset L^1(P_1) \times L^1(P_0)$

then

$$\inf_{\pi \in \Pi(P_1, P_0)} I_c[\pi] = \sup_{\varphi \in \mathcal{F}_c(\Phi_{cs})} J(\varphi, \varphi^c) = \sup_{(\varphi,\psi) \in \Phi_c \cap \left(\mathcal{F}_c(\Phi_{cs}) \times \mathcal{F}_c^c(\Phi_{cs})\right)} J(\varphi, \psi).$$

[16] H is a typically a subset of $L^1(P_1) \times L^1(P_0)$. As defined the sets $\mathcal{F}_c(H)$ and $\mathcal{F}_c^c(H)$ only depend on the functions in H that map $\mathcal{Y}_0$ to $\mathbb{R}$. This notational choice is more natural with the reasoning of lemma E.5 below.

Proof. Let $(\varphi, \psi) \in \Phi_{cs}$. $\psi(y_0) \leq c(y_1, y_0) - \varphi(y_1)$ implies $\psi(y_0) \leq \varphi^c(y_0)$, and lemma E.4 shows both that $\varphi(y_1) \leq \varphi^{cc}(y_1)$ and the pair $(\varphi^{cc}, \varphi^c)$ is a c-concave conjugate pair; thus $(\varphi^{cc}, \varphi^c) \in \Phi_c \cap \big(\mathcal{F}_c(\Phi_{cs}) \times \mathcal{F}_c^c(\Phi_{cs})\big)$.

Since φ^{cc} and φ^c are integrable by assumption, $J(\varphi, \psi) \leq J(\varphi^{cc}, \varphi^c)$ and hence

$$\inf_{\pi \in \Pi(P_1,P_0)} I_c[\pi] = \sup_{(\varphi,\psi)\in\Phi_{cs}} J(\varphi,\psi) \leq \sup_{\varphi^{cc}\in\mathcal{F}_c(\Phi_{cs})} J(\varphi^{cc},\varphi^c) \leq \sup_{(\varphi,\psi)\in\Phi_c\cap(\mathcal{F}_c(\Phi_{cs})\times\mathcal{F}_c^c(\Phi_{cs}))} J(\varphi,\psi)$$

Finally, since $\Phi_c \cap (\mathcal{F}_c(\Phi_{cs}) \times \mathcal{F}_c^c(\Phi_{cs})) \subset \Phi_c$, it follows that

$$\sup_{\varphi\in\mathcal{F}_c(\Phi_{cs})} J(\varphi,\varphi^c) \leq \sup_{(\varphi,\psi)\in\Phi_c} J(\varphi,\psi) = \inf_{\pi\in\Pi(P_1,P_0)} I_c[\pi]$$

with the final equality following from strong duality. □

Lemma E.6 (Continuous cost function implies measurability of c-concave functions)**.** *If $c : \mathcal{Y}_1 \times \mathcal{Y}_0 \to \mathbb{R}$ is continuous, then for any $\psi : \mathcal{Y}_0 \to \mathbb{R}$, $\varphi(y_1) = \inf_{y_0\in\mathcal{Y}_0}\{c(y_1, y_0) - \psi(y_0)\}$ and $\varphi^c(y_0) = \inf_{y_1\in\mathcal{Y}_1}\{c(y_1, y_0) - \varphi(y_1)\}$ are upper semicontinuous and hence measurable.*

Proof. The pointwise infimum of a family of upper semicontinuous functions is upper semicontinuous (Aliprantis & Border (2006) Lemma 2.41). Since $c(y_1, y_0)$ is continuous, for any fixed $y_0 \in \mathcal{Y}_0$ the function $y_1 \mapsto c(y_1, y_0) - \psi(y_0)$ is continuous and hence

$$\varphi(y_1) = \inf_{y_0\in\mathcal{Y}_0}\{c(y_1,y_0) - \psi(y_0)\}$$

is upper semicontinuous. Similarly, $\varphi^c(y_0) = \inf_{y_1\in\mathcal{Y}_1}\{c(y_1, y_0) - \varphi(y_1)\}$ is upper semicontinuous. Being upper semicontinuous, φ and φ^c are measurable. □

Remark E.1. Compare lemma E.6 with Villani (2009) Remark 5.5 discussing measurability of c-concave functions. Note that continuity of c is sufficient but not necessary for measurability of c-concave functions; see section E.3.2 for counterexamples.

Lemma E.7 (Universal bound on the the dual problem feasible set)**.** *Suppose $c : \mathcal{Y}_1 \times \mathcal{Y}_0 \to \mathbb{R}$ is bounded, and let $c_L = \inf_{(y_1,y_0)\in\mathcal{Y}_1\times\mathcal{Y}_0} c(y_1, y_0)$, $c_H = \sup_{(y_1,y_0)\in\mathcal{Y}_1\times\mathcal{Y}_0} c(y_1, y_0)$.*

1. *For any bounded functions $\varphi : \mathcal{Y}_1 \to \mathbb{R}$ and $\psi : \mathcal{Y}_0 \to \mathbb{R}$, φ^c and ψ^c are bounded.*
2. *For any bounded, measurable c-conjugate pair (φ, φ^c) there exists $\bar{\varphi}$ such that*
   - *(i) $\bar{\varphi}$ and $\bar{\varphi}^c$ satisfy the bounds:*

     $$c_L \leq \bar{\varphi}(y_1) \leq c_H \qquad\qquad c_L - c_H \leq \bar{\varphi}^c(y_0) \leq 0$$

     for all $(y_1, y_0) \in \mathcal{Y}_1 \times \mathcal{Y}_0$.
   - *(ii) $J(\varphi, \varphi^c) = J(\bar{\varphi}, \bar{\varphi}^c)$.*

Proof. For claim 1, let φ be bounded and note that

$$c_L - \sup\varphi \leq \underbrace{\inf_{y_1\in\mathcal{Y}_1}\{c(y_1,y_0)-\varphi(y_1)\}}_{=\varphi^c(y_0)} \leq c_H - \sup\varphi \tag{85}$$

are finite bounds on φ^c. The upper bound on φ^c follows from the existence of a sequence $\{y_{1j}\}_{j=1}^\infty$ with $\varphi(y_{1j}) \to \sup_{y_1\in\mathcal{Y}_1}\varphi(y_1)$, because $\varphi^c(y_0) = \inf_{y_1\in\mathcal{Y}_1}\{c(y_1,y_0)-\varphi(y_1)\} \leq c(y_{1j},y_0)-\varphi(y_{1j}) \leq c_H - \varphi(y_{1j})$ for all j. The same argument shows ψ^c is bounded, specifically,

$$c_L - \sup\psi \leq \underbrace{\inf_{y_0\in\mathcal{Y}_0}\{c(y_1,y_0)-\psi(y_0)\}}_{=\psi^c(y_1)} \leq c_H - \sup\psi \tag{86}$$

For claim 2, let (φ,φ^c) be a c-conjugate pair, i.e. $\varphi(y_1) = \inf_{y_0\in\mathcal{Y}_0}\{c(y_1,y_0)-\varphi^c(y_0)\}$. Notice that for any $s\in\mathbb{R}$,

$$\begin{aligned}
(\varphi+s)^c(y_0) &= \inf_{y_1\in\mathcal{Y}_1}\{c(y_1,y_0)-\varphi(y_1)-s\} = \varphi^c(y_0)-s\\
(\varphi+s)^{cc}(y_0) &= \inf_{y_0\in\mathcal{Y}_0}\{c(y_1,y_0)-\varphi^c(y_1)+s\} = \varphi(y_1)+s
\end{aligned}$$

Define $\bar\varphi(y_1) = \varphi(y_1) - \sup\varphi + c_H$, and notice that $\sup\bar\varphi = c_H$. Thus (85) implies $c_L - c_H \leq \bar\varphi^c(y_0) \leq 0$ for all $y_0\in\mathcal{Y}_0$, and so (86) implies $c_L \leq \bar\varphi^{cc}(y_1) = \bar\varphi(y_1) \leq c_H$. Finally,

$$\begin{aligned}
J(\varphi,\varphi^c) &= \int \varphi(y_1)dP_1(y_1) + \int \varphi^c(y_0)dP_0(y_0)\\
&= \int \varphi(y_1) - \sup\varphi + c_H dP_1(y_1) + \int \varphi^c(y_0) + \sup\varphi - c_H dP_0(y_0)\\
&= J(\bar\varphi,\bar\varphi^c)
\end{aligned}$$

which completes the proof. □

Remark E.2. Lemma E.7 shows that it is often without loss of generality to restrict the dual to classes of functions sharing universal bounds. For an example, see lemma E.9 below.

Note that when $c_L = 0$, the bounds simplify to

$$0 \leq \bar\varphi(y_1) \leq \|c\|_\infty, \qquad\qquad -\|c\|_\infty \leq \bar\varphi^c(y_0) \leq 0$$

as in Villani (2003) Remark 1.13. Also note that, when any universal bound suffices, one can take

$$-\|c\|_\infty \leq \bar\varphi(y_1) \leq \|c\|_\infty, \qquad\qquad -2\|c\|_\infty \leq \bar\varphi^c(y_0) \leq 0$$

which depend only on $\|c\|_\infty = \sup_{(y_1,y_0)\in\mathcal{Y}_1\times\mathcal{Y}_0}|c(y_1,y_0)|$.

E.3.1 c-concave functions of smooth cost functions

For $\alpha \in (0,1]$ and $L > 0$, $c : \mathcal{Y}_1 \times \mathcal{Y}_0 \to \mathbb{R}$ is called (α, L)-**Hölder continuous** if

$$|c(y_1, y_0) - c(y_1', y_0')| \leq L\|(y_1, y_0) - (y_1', y_0')\|^\alpha$$

for all $(y_1, y_0), (y_1', y_0') \in \mathcal{Y}_1 \times \mathcal{Y}_0$.

Lemma E.8 (Hölder cost implies Hölder c-concave functions)**.** *Let $c : \mathcal{Y}_1 \times \mathcal{Y}_0 \to \mathbb{R}$ be (α, L)-Hölder continuous. For any $g : \mathcal{Y}_0 \to \mathbb{R}$,*

$$\varphi(y_1) = \inf_{y_0 \in \mathcal{Y}_0} \{c(y_1, y_0) - g(y_0)\}, \qquad \varphi^c(y_0) = \inf_{y_1 \in \mathcal{Y}_1} \{c(y_1, y_0) - \varphi(y_1)\}$$

are (α, L)-Hölder continuous.

Proof. Hölder continuity implies $c(y_1, y_0) \leq c(y_1', y_0) + L|y_1 - y_1'|^\alpha$ holds for any $y_0 \in \mathcal{Y}_0$ and any $y_1, y_1' \in \mathcal{Y}_1$. It follows that

$$\varphi(y_1) = \inf_{y_0' \in \mathcal{Y}_0} \{c(y_1, y_0') - g(y_0')\} \leq c(y_1, y_0) - g(y_0) \leq c(y_1', y_0) - g(y_0) + L|y_1 - y_1'|^\alpha$$

implying $\varphi(y_1) - (c(y_1', y_0) - g(y_0)) \leq L|y_1 - y_1'|^\alpha$. Therefore

$$\varphi(y_1) - \varphi(y_1') = \varphi(y_1) - \inf_{y_0 \in \mathcal{Y}_0} \{c(y_1', y_0) - g(y_0)\} \leq L|y_1 - y_1'|^\alpha$$

holds for any $y_1, y_1' \in \mathcal{Y}_1$. This implies $\varphi(y_1') - \varphi(y_1) \leq L|y_1' - y_1|^\alpha$, hence φ is (α, L)-Hölder. The same argument implies φ^c is (α, L)-Hölder. □

Lemmas E.9, C.1, and C.10, are relevant for compact $\mathcal{Y}_1, \mathcal{Y}_0 \subset \mathbb{R}$, and L-Lipscthiz $c : \mathcal{Y}_1 \times \mathcal{Y}_0 \to \mathbb{R}$. Under these assumptions, define

$$\mathcal{F}_c = \left\{\varphi : \mathcal{Y}_1 \to \mathbb{R}\ ;\ -\|c\|_\infty \leq \varphi(y_1) \leq \|c\|_\infty,\ |\varphi(y_1) - \varphi(y_1')| \leq L|y_1 - y_1'|\right\} \tag{87}$$

$$\mathcal{F}_c^c = \left\{\psi : \mathcal{Y}_0 \to \mathbb{R}\ ;\ -2\|c\|_\infty \leq \psi(y_0) \leq 0,\ |\psi(y_0) - \psi(y_0')| \leq L|y_0 - y_0'|\right\} \tag{88}$$

Lemma E.9 (Strong duality for smooth cost functions)**.** *Let $\mathcal{Y}_1, \mathcal{Y}_0 \subset \mathbb{R}$ be compact, $c : \mathcal{Y}_1 \times \mathcal{Y}_0 \to \mathbb{R}$ be L-Lipschitz, and $\mathcal{F}_c$, $\mathcal{F}_c^c$ be given by* (87) *and* (88) *respectively. Then strong duality holds:*

$$\inf_{\pi \in \Pi(P_1, P_0)} I_c[\pi] = \sup_{(\varphi, \psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)} J(\varphi, \psi)$$

Proof. First notice lemma E.8 implies $\mathcal{F}_c(\Phi_c \cap \mathcal{C}_b)$ and $\mathcal{F}_c^c(\Phi_c \cap \mathcal{C}_b)$ consist of L-Lipschitz functions.[17]

[17] Note that $\mathcal{F}_c(\Phi_c \cap \mathcal{C}_b)$ and $\mathcal{F}_c^c(\Phi_c \cap \mathcal{C}_b)$ are not necessarily $\mathcal{F}_c$ and $\mathcal{F}_c^c$ defined in the statement of the lemma.

Since c is continuous and $\mathcal{Y}_1 \times \mathcal{Y}_0$ is compact, $\|c\|_\infty = \sup_{y_1, y_0 \in \mathcal{Y}_1 \times \mathcal{Y}_0} |c(y_1, y_0)| < \infty$. Continuity implies these c-concave functions are measurable, and lemma E.7 shows they are bounded. Thus $\mathcal{F}_c(\Phi_c \cap \mathcal{C}_b) \times \mathcal{F}_c^c(\Phi_c \cap \mathcal{C}_b) \subseteq L^1(P_1) \times L^1(P_0)$, and so lemma E.5 implies

$$\inf_{\pi \in \Pi(P_1, P_0)} I_c[\pi] = \sup_{\varphi \in \mathcal{F}_c(\Phi_c \cap \mathcal{C}_b)} J(\varphi, \varphi^c)$$

Lemma E.7 and remark E.2 further shows that for every $\varphi \in \mathcal{F}_c(\Phi_c \cap \mathcal{C}_b)$, a shifted function $\bar{\varphi}$ is such that $\sup_{y_1 \in \mathcal{Y}_1} |\bar{\varphi}(y_1)| \leq \|c\|_\infty$, $-2\|c\| \leq \bar{\varphi}^c(y_0) \leq 0$, $\bar{\varphi}$ and $\bar{\varphi}^c$ are L-lipschitz, and $J(\varphi, \varphi^c) = J(\bar{\varphi}, \bar{\varphi}^c)$. Thus

$$\sup_{\varphi \in \mathcal{F}_c(\Phi_c \cap \mathcal{C}_b)} J(\varphi, \varphi^c) = \sup_{\varphi \in \mathcal{F}_c} J(\varphi, \varphi^c)$$

Furthermore,

$$\sup_{\varphi \in \mathcal{F}_c} J(\varphi, \varphi^c) \leq \sup_{(\varphi, \psi) \in \Phi_c \cap (\mathcal{F}_c \times \mathcal{F}_c^c)} J(\varphi, \psi) \leq \sup_{(\varphi, \psi) \in \Phi_c} J(\varphi, \psi) = \inf_{\pi \in \Pi(P_1, P_0)} I_c[\pi]$$

completes the proof. □

Remark E.3. Suppose $\mathcal{Y}_1$ and $\mathcal{Y}_0$ are compact and $c(y_1, y_0)$ is continuously differentiable on an open set containing $\mathcal{Y}_1 \times \mathcal{Y}_0$. Then c restricted to $\mathcal{Y}_1 \times \mathcal{Y}_0$ is bounded and Lipschitz.

That $c : \mathcal{Y}_1 \times \mathcal{Y}_0 \to \mathbb{R}$ is bounded follows from c being continuous, $\mathcal{Y}_1 \times \mathcal{Y}_0$ being compact, and the extreme value theorem. To see that c restricted to $\mathcal{Y}_1 \times \mathcal{Y}_0$ is L-Lipschitz, let $(y_1, y_0), (y_1', y_0') \in \mathcal{Y}_1 \times \mathcal{Y}_0$ be arbitrary and note that the mean value theorem applied to $g(t) = c(t(y_1, y_0) + (1-t)(y_1', y_0'))$ implies there exists $s \in (0, 1)$ such that

$$\begin{aligned}
(c(y_1, y_0) - c(y_1', y_0')) &= g(1) - g(0) = g'(s) \\
&= \left\langle \nabla c(s(y_1, y_0) + (1-s)(y_1', y_0')), (y_1, y_0) - (y_1', y_0') \right\rangle
\end{aligned}$$

Notice that Cauchy-Schwarz then implies

$$\begin{aligned}
|c(y_1, y_0) - c(y_1', y_0')| &\leq \|\nabla c(s(y_1, y_0) + (1-s)(y_1', y_0'))\| \|(y_1, y_0) - (y_1', y_0')\| \\
&\leq \sup_{(y_1'', y_0'') \in \mathcal{Y}_1 \times \mathcal{Y}_0} \|\nabla c(y_1'', y_0'')\| \|(y_1, y_0) - (y_1', y_0')\|
\end{aligned}$$

Finally, notice $L = \sup_{(y_1'', y_0'') \in \mathcal{Y}_1 \times \mathcal{Y}_0} \|\nabla c(y_1'', y_0'')\|$ is finite because $\mathcal{Y}_1 \times \mathcal{Y}_0$ is compact and $(y_1, y_0) \mapsto \|\nabla c(y_1, y_0)\|$ is continuous.

E.3.2 c-concave functions when $c(y_1, y_0) = \mathbb{1}\{(y_1, y_0) \in C\}$

Theorem E.10 (Strong duality with indicator costs)**.** *Let C be a nonempty, open subset of $\mathcal{Y}_1 \times \mathcal{Y}_0$, and $c : \mathcal{Y}_1 \times \mathcal{Y}_0 \to \mathbb{R}$ given by $c(y_1, y_0) = \mathbb{1}_C(y_1, y_0) = \mathbb{1}\{(y_1, y_0) \in C\}$. Then*

$$\inf_{\pi \in \Pi(P_1, P_0)} \int \mathbb{1}_C(y_1, y_0) d\pi(y_1, y_0) = \sup_{(A, B) \in \Phi_c^I} \int \mathbb{1}_A(y_1) dP_1(y_1) - \int \mathbb{1}_B(y_0) d\nu(y_0)$$

where

$\Phi_c^I = \{(A, B)\ ;\ A \subset \mathcal{Y}_1$ *is closed and nonempty,* $B \subset \mathcal{Y}_0$ *is measurable, and* $\mathbb{1}_A(y_1) - \mathbb{1}_B(y_0) \leq \mathbb{1}_C(y_1, y_0)\}$

Proof. Villani (2003) Theorem 1.27 implies

$$\inf_{\pi\in\Pi(P_1,P_0)}\int \mathbb{1}_C(y_1,y_0)d\pi(y_1,y_0) = \sup_{A \text{ closed}}\int \mathbb{1}_A(y_1)dP_1(y_1) - \int \mathbb{1}_{A^C}(y_0)dP_0(y_0)$$

where $A^C = \{y \in \mathcal{Y}_0 \,;\, \exists y_1 \in A,\ (y_1,y_0) \notin C\}$ is the **projection** of $(A\times\mathcal{Y}_0)\setminus C$ onto $\mathcal{Y}_0$. Measurability of A^C is guaranteed by the measurable projection theorem; see Crauel (2002) theorem 2.12. It is clear that

$$\sup_{A \text{ closed}}\int \mathbb{1}_A(y_1)dP_1(y_1) - \int \mathbb{1}_{A^C}(y_0)dP_0(y_0) \leq \sup_{A\subseteq\mathcal{Y}_1, B\subseteq\mathcal{Y}_0}\int \mathbb{1}_A(y_1)dP_1(y_1) - \int \mathbb{1}_B(y_0)d\nu(y_0)$$

with A, B measurable. Notice it is without loss to exclude $A=\varnothing$, because $J(\mathbb{1}_\varnothing, -\mathbb{1}_B) \leq 0 = J(\mathbb{1}_{\mathcal{Y}_1}, \mathbb{1}_{\mathcal{Y}_0})$ and $\mathbb{1}_{\mathcal{Y}_1}(y_1) - \mathbb{1}_{\mathcal{Y}_0}(y_0) = 0 \leq \mathbb{1}_C(y_1,y_0)$ for all $(y_1,y_0)\in\mathcal{Y}_1\times\mathcal{Y}_0$. Thus

$$\sup_{A\subseteq\mathcal{Y}_1, B\subseteq\mathcal{Y}_0}\int \mathbb{1}_A(y_1)dP_1(y_1) - \int \mathbb{1}_B(y_0)d\nu(y_0) = \sup_{(A,B)\in\Phi_c^I}\int \mathbb{1}_A(y_1)dP_1(y_1) - \int \mathbb{1}_B(y_0)d\nu(y_0)$$

Weak duality (lemma E.2) implies

$$\sup_{(A,B)\in\Phi_c^I}\int \mathbb{1}_A(y_1)dP_1(y_1) - \int \mathbb{1}_B(y_0)dP_0(y_0) \leq \inf_{\pi\in\Pi(P_1,P_0)}\int \mathbb{1}_C(y_1,y_0)d\pi(y_1,y_0)$$

and the result follows. □

The strong duality result of theorem E.10 is especially useful when combined with a careful characterization of the corresponding c-concave functions. To describe these, let $A\subseteq\mathcal{Y}_1$ be nonempty, and define

$$A^C = \{y_0\in\mathcal{Y}_0 \,;\, \exists y_1\in A,\ (y_1,y_0)\notin C\}, \qquad A^{CC} = \left\{y_1\in\mathcal{Y}_1 \,;\, \forall y_0\in\mathcal{Y}_0\setminus A^C,\ (y_1,y_0)\in C\right\}, \tag{89}$$

$$C_{0m} = \{y_0\in\mathcal{Y}_0 \,;\, \forall y_1\in\mathcal{Y}_1,\ (y_1,y_0)\in C\}, \quad C_{1m} = \{y_1\in\mathcal{Y}_1 \,;\, \forall y_0\in\mathcal{Y}_0,\ (y_1,y_0)\in C\} \tag{90}$$

$$C_{0m}^C = \begin{cases} C_{1m} & \text{if } C_{0m}=\varnothing \\ \varnothing & \text{if } C_{0m}\neq\varnothing \end{cases}, \qquad C_{1m}^C = \begin{cases} C_{0m} & \text{if } C_{1m}=\varnothing \\ \varnothing & \text{if } C_{1m}\neq\varnothing \end{cases} \tag{91}$$

Note that A^C is well defined whenever $A\neq\varnothing$, and to ensure A^{CC} is well defined we require $A^C\neq\mathcal{Y}_0$. C_{0m} is denoted as such because $\mathbb{1}_{C_{0m}}(y_0) = \inf_{y_1\in\mathcal{Y}_1}\mathbb{1}_C(y_1,y_0)$ is the subset of $\mathcal{Y}_{\underline{0}}$ found by <u>m</u>inimizing $\mathbb{1}_C(y_1,y_0)$ over $y_1\in\mathcal{Y}_1$.

Lemma E.11 (c-concave functions for indicator costs)**.** *Let C be a nonempty, open subset of $\mathcal{Y}_1\times\mathcal{Y}_0$, $c:\mathcal{Y}_1\times\mathcal{Y}_0\to\mathbb{R}$ given by $c(y_1,y_0)=\mathbb{1}_C(y_1,y_0)$, $A\subseteq\mathcal{Y}_1$ be closed and nonempty, and*

$\varphi(y_1) = \mathbb{1}_A(y_1) = \mathbb{1}\{y_1 \in A\}$. *Then*

1. $\varphi^c(y_0) = -\mathbb{1}_{A^C}(y_0)$,
2. *if* $A^C \neq \mathcal{Y}_0$, *then* $\varphi^{cc}(y_1) = \mathbb{1}_{A^{CC}}(y_1)$, *and*
3. *If* $A^C = \mathcal{Y}_0$, *then* $J(\varphi^{cc}, \varphi^c) = J(\mathbb{1}_{C_{1m}}, 0)$

Proof. 1. Notice $\mathbb{1}_C(y_1, y_0) - \mathbb{1}_A(y_1) \in \{-1, 0, 1\}$, and

$$\varphi^c(y_0) = \inf_{y_1 \in \mathcal{Y}_1} \{\mathbb{1}_C(y_1, y_0) - \mathbb{1}_A(y_1)\}$$

will never take value 1 because any $y_1 \in A$ implies the objective is at most 0. Furthermore, if there exists $y_1 \in A$ such that $(y_1, y_0) \notin C$, then the infimum attains -1. If there does not exist such y_1, then $\varphi^c(y_0) = 0$. Thus $\varphi^c(y_0) = -\mathbb{1}_{A^C}(y_0)$.

2. Suppose $A^C \neq \mathcal{Y}_0$. Notice that $\mathbb{1}_C(y_1, y_0) + \mathbb{1}_{A^C}(y_0)$ takes values in $\{0, 1, 2\}$, and

$$\varphi^{cc}(y_1) = \inf_{y_0 \in \mathcal{Y}_0} \{\mathbb{1}_C(y_1, y_0) + \mathbb{1}_{A^C}(y_0)\}$$

will never equal 2 because $\mathcal{Y}_0 \setminus A^C \neq \varnothing$. Moreover, the infimum will equal 1 if and only if $(y_1, y_0) \in C$ for all $y_0 \in \mathcal{Y}_0 \setminus A^C$; thus $\varphi^{cc}(y_1) = \mathbb{1}_{A^{CC}}(y_1)$.

3. If $A^C = \mathcal{Y}_0$, then $\varphi^{cc}(y_1) = \inf_{y_0 \in \mathcal{Y}_0}\{\mathbb{1}_C(y_1, y_0) + 1\} = \mathbb{1}_{C_{1m}}(y_1) + 1$ and

$$\varphi^{ccc}(y_0) = \inf_{y_1 \in \mathcal{Y}_1} \{\mathbb{1}_C(y_1, y_0) - \mathbb{1}_{C_{1m}}(y_1) - 1\} = \mathbb{1}_{C_{1m}^C}(y_0) - 1$$

To see that $(\mathbb{1}_{C_{1m}})^c = 0$ if $C_{1m} \neq \varnothing$, notice the objective $\mathbb{1}_C(y_1, y_0) - \mathbb{1}_{C_{1m}}(y_0)$ takes values in $\{-1, 0, 1\}$, and because $C_{1m} \neq \varnothing$ will never take value 1. For the objective to take value -1 at a given y_1, it must be the case that $\mathbb{1}_{C_{1m}}(y_1) = 1$ and there exists y_0 such that $\mathbb{1}_C(y_1, y_0) = 0$, but this contradicts the definition $C_{1m} = \{y_1 \in \mathcal{Y}_1 \ ; \ \forall y_0 \in \mathcal{Y}_0, \ (y_1, y_0) \in C\}$.

However, recall that $\varphi^{ccc}(y_0) = \varphi^c(y_0)$ as shown in lemma E.4. Since $\varphi^c(y_0) = -\mathbb{1}_{A^C}(y_0) = -\mathbb{1}_{\mathcal{Y}_0}(y_0) = -1$, this implies $(\mathbb{1}_{C_{0m}^C})(y_0) = 0$. Then notice that

$$J(\varphi^{cc}, \varphi^c) = J(\mathbb{1}_{C_{1m}} + 1, -1) = J(\mathbb{1}_{C_{1m}}, 0)$$

□

Remark E.4. Compare theorem E.10 and lemma E.11 with Villani (2003) theorem 1.27.

Lemma E.12 (Convex C implies c-concave functions defined with convex sets)**.** *Let C be a nonempty, open, convex subset of $\mathcal{Y}_1 \times \mathcal{Y}_0$, and $c : \mathcal{Y}_1 \times \mathcal{Y}_0 \to \mathbb{R}$ given by $c(y_1, y_0) = \mathbb{1}_C(y_1, y_0)$. Let $A \subseteq \mathcal{Y}_1$ be nonempty.*

1. A^C *equals* $\mathcal{Y}_0 \setminus B$ *for some convex set* B.
2. *If* $A^C \neq \mathcal{Y}_0$, *then* A^{CC} *is convex.*
3. C_{1m} *is convex.*

Proof. For claim 1, notice that

$$A^C = \big\{y_0 \in \mathcal{Y}_0 \; ; \; \exists y_1 \in A, \; (y_1, y_0) \in \big(\mathcal{Y}_1 \times \mathcal{Y}_0\big) \setminus C\big\} = \bigcup_{y_1 \in A} \big\{y_0 \in \mathcal{Y}_0 \; ; \; (y_1, y_0) \in \big(\mathcal{Y}_1 \times \mathcal{Y}_0\big) \setminus C\big\}$$

$$= \bigcup_{y_1 \in A} \mathcal{Y}_0 \setminus \{y_0 \in \mathcal{Y}_0 \; ; \; (y_1, y_0) \in C\} = \mathcal{Y}_0 \setminus \bigcap_{y_1 \in A} \{y_0 \in \mathcal{Y}_0 \; ; \; (y_1, y_0) \in C\}$$

Since C is convex, $\{y \in \mathcal{Y}_0 \; ; \; (y_1, y_0) \in C\}$ is also convex for any y_1. The intersection of an arbitrary collection of convex sets is convex, so $A^C = \mathcal{Y}_0 \setminus B$ for some convex B.

Consider claim 2 next. Notice that

$$A^{CC} = \big\{y_1 \in \mathcal{Y}_1 \; ; \; \forall y_0 \in \mathcal{Y}_0 \setminus A^C, \; (y_1, y_0) \in C\big\} = \bigcap_{y_0 \in \mathcal{Y}_0 \setminus A^C} \{y_1 \in \mathcal{Y}_1 \; ; \; (y_1, y_0) \in C\}$$

Since C is convex, $\{y_1 \in \mathcal{Y}_1 \; ; \; (y_1, y_0) \in C\}$ is convex as well, and thus A^{CC} is convex.

Finally, we show claim 3. Similar to A^{CC}, notice that

$$C_{1m} = \{y_1 \in \mathcal{Y}_1 \; ; \; \forall y_0 \in \mathcal{Y}_0, \; (y_1, y_0) \in C\} = \bigcap_{y_0 \in \mathcal{Y}_0} \{y_1 \in \mathcal{Y}_1 \; ; \; (y_1, y_0) \in C\}$$

is the intersection of convex sets and therefore convex. $\square$

Refer to the convex subsets of $\mathbb{R}$ as **intervals**; specifically, $I \subset \mathbb{R}$ is called an interval if I takes the form

$$(\ell, u) \qquad [\ell, u) \qquad (\ell, u] \qquad [\ell, u]$$

where $\ell = -\infty$ is allowed for (ℓ, u) and $(\ell, u]$ and $u = \infty$ is allowed for (ℓ, u) and $[\ell, u)$. I^c is the complement of the interval I.

Lemmas E.13, C.2, and C.11 are relevant when the cost function is $c(y_1, y_0) = \mathbb{1}\{(y_1, y_0) \in C\}$ for some nonempty, open, convex $C \subseteq \mathcal{Y}_1 \times \mathcal{Y}_0$. When this is so, define

$$\mathcal{F}_c = \{\varphi : \mathcal{Y}_1 \to \mathbb{R} \; ; \; \varphi(y_1) = \mathbb{1}_I(y_1) \text{ for some interval } I\} \tag{92}$$

$$\mathcal{F}_c^c = \{\psi : \mathcal{Y}_0 \to \mathbb{R} \; ; \; \psi(y_0) = -\mathbb{1}_{I^c}(y_0) \text{ for some interval } I\} \tag{93}$$

Lemma E.13 (Strong duality for indicator cost functions of a convex set)**.** *Let $\mathcal{Y}_1, \mathcal{Y}_0 \subseteq \mathbb{R}$, $C \subseteq \mathcal{Y}_1 \times \mathcal{Y}_0$ be nonempty, open, and convex, and let $c : \mathcal{Y}_1 \times \mathcal{Y}_0 \to \mathbb{R}$ be given by $c(y_1, y_0) = \mathbb{1}_C(y_1, y_0)$.*

Let $\mathcal{F}_c$ and $\mathcal{F}_c^c$ be given by (92) and (93) respectively. Then strong duality holds:

$$\inf_{\pi\in\Pi(P_1,P_0)}\int \mathbb{1}_C(y_1,y_0)d\pi(y_1,y_0) = \sup_{(\varphi,\psi)\in\Phi_c\cap\left(\mathcal{F}_c\times\mathcal{F}_c^c\right)}\int \varphi(y_1)dP_1(y_1)+\int\psi(y_0)dP_0(y_0) \tag{94}$$

Proof. Recall that theorem E.10 shows

$$\inf_{\pi\in\Pi(P_1,P_0)}\int \mathbb{1}_C(y_1,y_0)d\pi(y_1,y_0) = \sup_{(A,B)\in\Phi_c^I}\int \mathbb{1}_A(y_1)dP_1(y_1)-\int \mathbb{1}_B(y_0)d\nu(y_0)$$

where

$\Phi_c^I = \{(A,B)\ ;\ A\subset\mathcal{Y}_1$ is closed and nonempty, $B\subset\mathcal{Y}_0$ is measurable, and $\mathbb{1}_A(y_1)-\mathbb{1}_B(y_0)\leq\mathbb{1}_C(y_1,y_0)\}$

We will apply lemma E.5. Let $\varphi(y_1)=\mathbb{1}_A(y_1)$ for some closed and nonempty $A\subset\mathcal{Y}_1$. There are two possibilities:

1. $A^C=\mathcal{Y}_0$, in which case $J(\varphi^{cc},\varphi^c)=J(\mathbb{1}_{C_{1m}},0)$, or
2. $A^C\neq\mathcal{Y}_0$, in which case $J(\varphi^{cc},\varphi^c)=J(\mathbb{1}_{A^{CC}},-\mathbb{1}_{A^C})$.

Since C is convex, C_{1m}, and A^{CC} are convex subsets of $\mathbb{R}$ (i.e., intervals), as shown in lemma E.12. A^C is the complement of an interval, and $0=\mathbb{1}_\varnothing(y_0)$ is the indicator of the complement of $\mathbb{R}$, which is the interval $(-\infty,\infty)$. Since all functions involved are bounded, they are all integrable, and lemma E.5 implies

$$\inf_{\pi\in\Pi(P_1,P_0)}\int \mathbb{1}_C(y_1,y_0)d\pi(y_1,y_0) = \sup_{(\varphi,\psi)\in\Phi_c\cap\left(\mathcal{F}_c(\Phi_c^I)\times\mathcal{F}_c^c(\Phi_c^I)\right)}\int \varphi(y_1)dP_1(y_1)+\int\psi(y_0)dP_0(y_0)$$

Finally, note that $\mathcal{F}_c(\Phi_c^I)\subseteq\mathcal{F}_c$ and $\mathcal{F}_c^c(\Phi_c^I)\subseteq\mathcal{F}_c^c$, which implies the strong duality claim in display (94) holds. □

E.4 Special cases: $c_L(y_1,y_0,\delta)=\mathbb{1}\{y_1-y_0<\delta\}$ and $c_H(y_1,y_0,\delta)=\mathbb{1}\{y_1-y_0>\delta\}$

Lemma E.14. *Let $F_1(y)=P_1(Y_1\leq y)=\int\mathbb{1}\{y_1\leq y\}dP_1(y_1)$ denote the cumulative distribution function (CDF) of P_1, and let F_0 the CDF of P_0. Let $c_L(y_1,y_0,\delta)=\mathbb{1}\{y_1-y_0<\delta\}$. Then*

$$\begin{aligned} OT_{c_L}(P_1,P_0) &= \inf_{\pi\in\Pi(P_1,P_0)}\int \mathbb{1}\{y_1-y_0<\delta\}d\pi(y_1,y_0)\\ &= \max\left\{\sup_y\{F_1(y)-F_0(y-\delta)\}, P_1(Y_1<\min\{\mathcal{Y}_0\}+\delta)\right\} \end{aligned} \tag{95}$$

Proof. Let $C=\{y_1-y_0<\delta\}$. Apply theorem E.10 and lemma E.11 to find that

$$OT_{c_L}(P_1,P_0)=\max\{\sup_{A\in\mathcal{A}}P_1(Y_1\in A^{CC})-P_0(Y_0\in A^C),P_1(Y_1\in C_{1m})\}$$

where

$$A^C = \{y_0 \in \mathcal{Y}_0 \ ; \ \exists y_1 \in A, \ (y_1, y_0) \notin C\}, \quad A^{CC} = \left\{y_1 \in \mathcal{Y}_1 \ ; \ \forall y_0 \in \mathcal{Y}_0 \setminus A^C, \ (y_1, y_0) \in C\right\},$$
$$C_{1m} = \{y_1 \in \mathcal{Y}_1 \ ; \ \forall y_0 \in \mathcal{Y}_0, \ (y_1, y_0) \in C\}.$$

and $\mathcal{A}$ is the collection of closed, nonempty subsets of $\mathcal{Y}_1$ such that $A^C \neq \mathcal{Y}_0$.

First consider $\sup_{A \in \mathcal{A}} P_1(Y_1 \in A^{CC}) - P_0(Y_0 \in A^C)$. Let $A \in \mathcal{A}$ and $\varphi(y_1) = \mathbb{1}_A(y_1)$. Thus

$$\begin{aligned} A^C &= \{y \in \mathcal{Y}_0 \ ; \ \exists y_1 \in A, \ (y_1, y_0) \notin C\} = \{y_0 \in \mathcal{Y}_0 \ ; \ y_0 \leq \max\{A\} - \delta\}, \\ A^{CC} &= \left\{y_1 \in \mathcal{Y}_1 \ ; \ \forall y_0 \in \mathcal{Y}_0 \setminus A^C, \ y_1 - y_0 < \delta\right\} = \{y_1 \in \mathcal{Y}_1 \ ; \ y_1 \leq \max\{A\}\} \end{aligned}$$

where we've used the fact that $A^C \neq \mathcal{Y}_0$ implies $\sup\{A\} < \infty$ and so $\sup\{A\} = \max\{A\}$ because A is closed. Therefore

$$\begin{aligned} J(\varphi^{cc}, \varphi^c) &= P_1(Y_1 \in A^{CC}) - P_0(Y_0 \in A^c) \\ &= P_1(Y_1 \leq \max\{A\}) - P_0(Y_0 \leq \max\{A\} - \delta) \end{aligned}$$

which takes the form $F_1(y) - F_0(y - \delta)$ for $y = \max\{A\}$.

Now consider $P_1(Y_1 \in C_{1m})$, and notice that

$$\begin{aligned} C_{1m} &= \{y_1 \in \mathcal{Y}_1 \ ; \ \forall y_0 \in \mathcal{Y}_0, \ (y_1, y_0) \in C\} = \{y_1 \in \mathcal{Y}_1 \ ; \ \forall y_0 \in \mathcal{Y}_0, \ y_1 - y_0 < \delta\} \\ &= \{y_1 \in \mathcal{Y}_1 \ ; \ \forall y_0 \in \mathcal{Y}_0, \ y_1 < \min\{\mathcal{Y}_0\} + \delta\} \end{aligned}$$

Thus $P_1(Y_1 \in C_{1m}) = P_1(Y_1 < \min\{\mathcal{Y}_0\} + \delta)$. The result follows. □

Remark E.5. C_{1m} may be closed; e.g., let $\mathcal{Y}_1 = [0,1] \cup [3,10]$, let $\mathcal{Y}_0 = [2,10]$, and $\delta = 0$. Then $C_{1m} = \{y_1 \in \mathcal{Y}_1 \ ; \ y_1 < 2\} = [0,1]$.

Corollary E.15. *Let $c_L(y_1, y_0, \delta) = \mathbb{1}\{y_1 - y_0 < \delta\}$ and P_1, P_0 have continuous cumulative distribution functions $F_1(y) = P_1(Y_1 \leq y)$ and $F_0(y) = P_0(Y_0 \leq y)$ respectively. Then*

$$OT_{c_L}(P_1, P_0) = \inf_{\pi \in \Pi(P_1, P_0)} \int \mathbb{1}\{y_1 - y_0 < \delta\} d\pi(y_1, y_0) = \sup_y \{F_1(y) - F_0(y - \delta)\} \tag{96}$$

Proof. Continuity of the cumulative distribution functions implies $P_1(Y_1 = \delta + \min\{\mathcal{Y}_0\}) = P_0(Y_0 = \min\{\mathcal{Y}_0\}) = 0$, and thus

$$P_1(Y_1 < \delta + \min\{\mathcal{Y}_0\}) = P_1(Y_1 \leq \delta + \min\{\mathcal{Y}_0\}) - P_0(Y_0 \leq \min\{\mathcal{Y}_0\})$$

Which takes the form $F_1(y) - F_0(y - \delta)$ for $y = \delta + \min\{\mathcal{Y}_0\}$. It follows that

$$\max\left\{\sup_y \{F_1(y) - F_0(y - \delta)\}, P_1(Y_1 < \min\{\mathcal{Y}_0\} + \delta)\right\} = \sup_y \{F_1(y) - F_0(y - \delta)\}$$

and lemma E.14 gives the result. □

Lemma E.16. *Let $c_H(y_1, y_0, \delta) = \mathbb{1}\{y_1 - y_0 > \delta\}$. Then*

$$
\begin{aligned}
OT_{c_H}(P_1, P_0) &= \inf_{\pi \in \Pi(P_1, P_0)} \int \mathbb{1}\{y_1 - y_0 > \delta\} d\pi(y_1, y_0) \\
&= \max\left\{ \sup_y \{P_1([y, \infty)) - P_0([y - \delta, \infty))\}, P_1((\max\{\mathcal{Y}_0\} + \delta, \infty)) \right\}
\end{aligned}
\tag{97}
$$

Proof. The proof is similar to that of lemma E.14. Let $C = \{y_1 - y_0 > \delta\}$. Apply theorem E.10 and lemma E.11 to find that

$$
OT_{c_L}(P_1, P_0) = \max\{\sup_{A \in \mathcal{A}} P_1(Y_1 \in A^{CC}) - P_0(Y_0 \in A^C), P_1(Y_1 \in C_{1m})\}
$$

where

$$
\begin{gathered}
A^C = \{y_0 \in \mathcal{Y}_0 \; ; \; \exists y_1 \in A, \; (y_1, y_0) \notin C\}, \quad A^{CC} = \left\{y_1 \in \mathcal{Y}_1 \; ; \; \forall y_0 \in \mathcal{Y}_0 \setminus A^C, \; (y_1, y_0) \in C\right\}, \\
C_{1m} = \{y_1 \in \mathcal{Y}_1 \; ; \; \forall y_0 \in \mathcal{Y}_0, \; (y_1, y_0) \in C\}.
\end{gathered}
$$

and $\mathcal{A}$ is the collection of closed, nonempty subsets of $\mathcal{Y}_1$ such that $A^C \neq \mathcal{Y}_0$.

Consider $\sup_{A \in \mathcal{A}} P_1(Y_1 \in A^{CC}) - P_0(Y_0 \in A^C)$. Let $A \in \mathcal{A}$ and $\varphi(y_1) = \mathbb{1}_A(y_1)$, and notice that

$$
\begin{aligned}
A^C &= \{y \in \mathcal{Y}_0 \; ; \; \exists y_1 \in A, \; (y_1, y_0) \notin C\} = \{y_0 \in \mathcal{Y}_0 \; ; \; y_0 \geq \min\{A\} - \delta\}, \\
A^{CC} &= \left\{y_1 \in \mathcal{Y}_1 \; ; \; \forall y_0 \in \mathcal{Y}_0 \setminus A^C, \; y_1 - y_0 < \delta\right\} = \{y_1 \in \mathcal{Y}_1 \; ; \; y_1 \geq \min\{A\}\}
\end{aligned}
$$

Where as in the proof of lemma E.14, $A^C \neq \mathcal{Y}_0$ implies $\inf\{A\} > -\infty$ and so $\inf\{A\} = \min\{A\}$ because A is closed. Thus

$$
\begin{aligned}
J(\varphi^{cc}, \varphi^c) &= P_1(Y_1 \in A^{CC}) - P_0(Y_0 \in A^c) \\
&= P_1(Y_1 \geq \min\{A\}) - P_0(Y_0 \geq \min\{A\} - \delta)
\end{aligned}
$$

which takes the form $P_1([y, \infty)) - P_0([y - \delta, \infty))$ for $y = \min\{A\}$.

Now consider $P_1(Y_1 \in C_{1m})$, and notice that

$$
\begin{aligned}
C_{1m} &= \{y_1 \in \mathcal{Y}_1 \; ; \; \forall y_0 \in \mathcal{Y}_0, \; (y_1, y_0) \in C\} = \{y_1 \in \mathcal{Y}_1 \; ; \; \forall y_0 \in \mathcal{Y}_0, \; y_1 - y_0 > \delta\} \\
&= \{y_1 \in \mathcal{Y}_1 \; ; \; \forall y_0 \in \mathcal{Y}_0, \; y_1 > \max\{\mathcal{Y}_0\} + \delta\}
\end{aligned}
$$

Thus $P_1(Y_1 \in C_{1m}) = P_1(Y_1 > \max\{\mathcal{Y}_0\} + \delta)$. The result follows. □

Corollary E.17. *Let $c_H(y_1, y_0, \delta) = \mathbb{1}\{y_1 - y_0 > \delta\}$ and P_1, P_0 have continuous cumulative distribution functions $F_1(y) = P_1(Y_1 \leq y)$ and $F_0(y) = P_0(Y_0 \leq y)$ respectively. Then*

$$
OT_{c_L}(P_1, P_0) = \inf_{\pi \in \Pi(P_1, P_0)} \int \mathbb{1}\{y_1 - y_0 > \delta\} d\pi(y_1, y_0) = \sup_y \{F_0(y - \delta) - F_1(y)\} \tag{98}
$$

Proof. Continuity of the cumulative distribution functions implies that for any y,

$$\begin{aligned} P_1([y,\infty)) - P_0([y-\delta,\infty)) &= P_1((y,\infty)) - P_0((y-\delta,\infty)) \\ &= (1 - F_1(y)) - (1 - F_0(y-\delta)) \\ &= F_0(y-\delta) - F_1(y) \end{aligned}$$

and furthermore,

$$\begin{aligned} P_1(Y_1 > \delta + \max\{\mathcal{Y}_0\}) &= 1 - F_1(\delta + \min\{\mathcal{Y}_0\}) - (1 - F_0(\max\{\mathcal{Y}_0\}) \\ &= F_0(\max\{\mathcal{Y}_0\}) - F_1(\delta + \max\{\mathcal{Y}_0\}) \end{aligned}$$

equals $F_0(y-\delta) - F_1(y)$ for $y = \max\{\mathcal{Y}_0\} + \delta$. Finally, lemma E.16 gives

$$\begin{aligned} OT_{c_H}(P_1, P_0) &= \max\left\{\sup_y\{P_1([y,\infty)) - P_0([y-\delta,\infty))\}, P_1((\max\{\mathcal{Y}_0\} + \delta, \infty))\right\} \\ &= \sup_y\{F_0(y-\delta) - F_1(y)\} \end{aligned}$$

$\square$

F Appendix: miscellaneous lemmas

F.1 Continuity

Lemma F.1 (Continuity of maps between bounded function spaces)**.** *Let $f : \mathbb{D}_f \subseteq \mathbb{R}^K \to \mathbb{R}^M$ be uniformly continuous. Define the subset of bounded functions on T taking values in $\mathbb{D}_f$:*

$$\ell^\infty(T, \mathbb{D}_f) = \left\{ g : T \to \mathbb{R}^K \; ; \; g(t) \in \mathbb{D}_f, \; \sup_{t \in T} \|g(t)\| < \infty \right\} \subseteq \ell^\infty(T)^K$$

Let $F : \ell^\infty(T, \mathbb{D}_f) \to \ell^\infty(T)^M$ be defined pointwise as $F(g)(t) = f(g(t))$. Then F is uniformly continuous.

Proof. To see that $F : \ell^\infty(T, \mathbb{D}_f) \to \ell^\infty(T)^M$ is well defined, recall that uniform continuity of f implies f is bounded on bounded sets. Since $\{g(t) \; ; \; t \in T\}$ is bounded for any $g \in \ell^\infty(T, \mathbb{D}_f)$, this implies $\sup_t \|f(g(t))\| < \infty$ and hence $F(g) \in \ell^\infty(T)^M$.

To see uniform continuity of F, let $\varepsilon > 0$ and use uniform continuity of f to choose $\delta > 0$ such that for all $x, \tilde{x} \in \mathbb{D}_f$,

$$\|x - \tilde{x}\| < \delta \implies \|f(x) - f(\tilde{x})\| < \varepsilon/2$$

Now let $g, \tilde{g} \in \ell^\infty(T, \mathbb{D}_f)$ satisfy $\|g - \tilde{g}\|_T = \sup_{t \in T} \|g(t) - \tilde{g}(t)\| < \delta$. Then $\|g(t) - \tilde{g}(t)\| < \delta$ for all $t \in T$, and hence $\|f(g(t)) - f(\tilde{g}(t))\| < \varepsilon/2$ for all $t \in T$, and therefore

$$\|F(g) - F(\tilde{g})\|_T = \sup_{t \in T} \|f(g(t)) - f(\tilde{g}(t))\| \leq \frac{\varepsilon}{2} < \varepsilon$$

which completes the proof. □

Corollary F.2. *Let $f : \mathbb{D}_f \subseteq \mathbb{R}^K \to \mathbb{R}^M$ be continuous and bounded on bounded subsets of $\mathbb{D}_f$. Let $g_0 \in \ell^\infty(T, \mathbb{D}_f)$ where $\ell^\infty(T, \mathbb{D}_f)$ is as defined in lemma F.1. Suppose that for some $\delta > 0$,*

$$g(T)^\delta \equiv \left\{ x \in \mathbb{R}^K \; ; \; \inf_{t \in T} \|g_0(t) - x\| \leq \delta \right\}$$

is a subset of $\mathbb{D}_f$. Then $F : \ell^\infty(T, \mathbb{D}_f) \to \ell^\infty(T)^M$ defined pointwise by $F(g)(t) = f(g(t))$ is continuous at g_0.

Proof. For any $g \in \ell^\infty(T, \mathbb{D}_f)$, we have $F(g) \in \ell^\infty(T)^M$ because $\{x \; ; \; x = g(t) \text{ for some } t \in T\}$ is bounded and f is bounded on bounded subsets.

Let $\{g_n\}_{n=1}^\infty \subseteq \ell^\infty(T, \mathbb{D}_f)$ be such that $g_n \to g_0$ in $\ell^\infty(T)^K$. It suffices to show that $F(g_n) \to F(g_0)$ in $\ell^\infty(T)^M$. Let $\tilde{f} : g(T)^\delta \to \mathbb{R}^M$ be the restriction of f to $g_0(T)^\delta$; i.e., $\tilde{f}(x) = f(x)$. Note that because $g_0(T)^\delta$ is a closed and bounded subset of $\mathbb{R}^K$, it is compact, and hence $\tilde{f}$ is uniformly continuous by the Heine-Cantor theorem. Apply lemma F.1 to find that

$$\tilde{F} : \ell^\infty(T, g(T)^\delta) \to \ell^\infty(T)^M, \qquad \tilde{F}(g)(t) = \tilde{f}(g(t)) = f(g(t))$$

is continuous. Since $g_n \to g_0$ in $\ell^\infty(T)^K$, there exists N such that for all $n \geq N$, $\|g_n - g_0\|_T = \sup_{t\in T}\|g_n(t) - g_0(t)\| < \delta$. Let $\tilde{g}_k = g_{k+N}$. Notice that $\tilde{g}_k(T) = \left\{x \in \mathbb{R}^K \; ; \; x = g_k(t) \text{ for some } t \in T\right\} \subseteq g_0(T)^\delta$, and hence $\tilde{g}_k \in \ell^\infty(T, g_0(T)^\delta)$. Continuity of $\tilde{F}$ and $\tilde{g}_k \to g_0$ implies $\tilde{F}(\tilde{g}_k) \to \tilde{F}(\tilde{g}_0)$. Thus

$$0 = \lim_{k\to\infty}\|\tilde{F}(\tilde{g}_k) - \tilde{F}(g_0)\|_T = \lim_{k\to\infty}\|F(g_{k+N}) - F(g_0)\|_T = \lim_{n\to\infty}\|F(g_n) - F(g_0)\|_T$$

which completes the proof. □

Lemma F.3 (Uniform continuity of restricted sup)**.** *For any set X, subset $A \subseteq X$, and bounded real-valued functions $f, g \in \ell^\infty(X)$,*

$$\left|\sup_{x\in A} f(x) - \sup_{x\in A} g(x)\right| \leq \sup_{x\in A}|f(x) - g(x)| \tag{99}$$

and therefore $\sigma_A : \ell^\infty(X) \to \mathbb{R}$ given by $\sigma_A(f) = \sup_{x\in A} f(x)$ is uniformly continuous.

Proof. Observe that

$$\sup_{x\in A} f(x) - \sup_{x\in A} g(x) \leq \sup_{x\in A}\{f(x) - g(x)\} \leq \sup_{x\in A}|f(x) - g(x)|$$

and

$$-\left[\sup_{x\in A} f(x) - \sup_{x\in A} g(x)\right] = \sup_{x\in A} g(x) - \sup_{x\in A} f(x) \leq \sup_{x\in A}\{g(x) - f(x)\} \leq \sup_{x\in A}|f(x) - g(x)|$$

Together these inequalities imply

$$-\sup_{x\in A}|f(x) - g(x)| \leq \sup_{x\in A} f(x) - \sup_{x\in A} g(x) \leq \sup_{x\in A}|f(x) - g(x)|$$

which is equivalent to (99).

To see uniform continuity, let $\varepsilon > 0$ and choose $\delta = \varepsilon$. Whenever $\|f - g\|_X = \sup_{x\in X}|f(x) - g(x)| < \delta$,

$$|\sigma_A(f) - \sigma_A(g)| = \left|\sup_{x\in A} f(x) - \sup_{x\in A} g(x)\right| \leq \sup_{x\in A}|f(x) - g(x)| \leq \sup_{x\in X}|f(x) - g(x)| < \delta = \varepsilon$$

which completes the proof. □

F.2 Differentiability

This appendix reviews definitions and various facts related to Hadamard directional differentiability. The following definitions can be found in Fang & Santos (2019).

Let $\mathbb{D}$, $\mathbb{E}$ be Banach spaces (complete, normed, vector spaces), and $\phi : \mathbb{D}_\phi \subseteq \mathbb{D} \to \mathbb{E}$.

(i) ϕ is (fully) **Hadamard differentiable** at $x_0 \in \mathbb{D}_\phi$ tangentially to $\mathbb{D}_0 \subseteq \mathbb{D}$ if there exists a continuous linear map $\phi'_{x_0} : \mathbb{D}_0 \to \mathbb{E}$ such that

$$\lim_{n\to\infty} \left\| \frac{\phi(x_0 + t_n h_n) - \phi(x_0)}{t_n} - \phi'_{x_0}(h) \right\|_{\mathbb{E}} = 0$$

for all sequences $\{h_n\}_{n=1}^{\infty} \subseteq \mathbb{D}$ and $\{t_n\}_{n=1}^{\infty} \subseteq \mathbb{R}$ such that $h_n \to h \in \mathbb{D}_0$ and $t_n \to 0$ as $n \to \infty$, and $x_0 + t_n h_n \in \mathbb{D}_\phi$ for all n.

(ii) ϕ is **Hadamard directionally differentiable** at $x_0 \in \mathbb{D}_\phi$ tangentially to $\mathbb{D}_0 \subseteq \mathbb{D}$ if there exists a continuous map $\phi'_{x_0} : \mathbb{D}_0 \to \mathbb{E}$ such that

$$\lim_{n\to\infty} \left\| \frac{\phi(x_0 + t_n h_n) - \phi(x_0)}{t_n} - \phi'_{x_0}(h) \right\|_{\mathbb{E}} = 0$$

for all sequences $\{h_n\}_{n=1}^{\infty} \subseteq \mathbb{D}$ and $\{t_n\}_{n=1}^{\infty} \subseteq \mathbb{R}_+$ such that $h_n \to h \in \mathbb{D}_0$ and $t_n \downarrow 0$ as $n \to \infty$, and $x_0 + t_n h_n \in \mathbb{D}_\phi$ for all n.

Fang & Santos (2019) proposition 2.1 shows that linearity is the key property distinguishing directional and full Hadamard differentiability. Specifically, if ϕ is Hadamard directionally differentiable at x_0 tangentially to a subspace $\mathbb{D}_0$, and ϕ'_{x_0} is linear, then ϕ is in fact fully Hadamard differentiable at x_0 tangentially to $\mathbb{D}_0$.

Hadamard directional differentiability obeys the chain rule.

Lemma F.4 (Chain rule)**.** *Let $\mathbb{D}_1$, $\mathbb{D}_2$, and $\mathbb{E}$ be Banach spaces and $\phi_1 : \mathbb{D}_{\phi_1} \subseteq \mathbb{D}_1 \to \mathbb{D}_2$, $\phi_2 : \mathbb{D}_{\phi_2} \subseteq \mathbb{D}_2 \to \mathbb{E}$ be functions. Suppose*

(i) $\phi_1(\mathbb{D}_{\phi_1}) = \{y \in \mathbb{D}_2 \,;\, y = \phi_1(x) \text{ for some } x \in \mathbb{D}_{\phi_1}\} \subseteq \mathbb{D}_{\phi_2}$,

(ii) ϕ_1 is Hadamard directionally differentiable at $x_0 \in \mathbb{D}_{\phi_1}$ tangentially to $\mathbb{D}_1^T \subseteq \mathbb{D}_1$, with derivative $\phi'_{1,x_0}(h)$, and

(iii) ϕ_2 is Hadamard directionally differentiable at $\phi_1(x_0) \in \mathbb{D}_{\phi_2}$ tangentially to $\mathbb{D}_2^T \subseteq \mathbb{D}_2$, with derivative $\phi'_{2,\phi_1(x_0)}(h)$

Let $\mathbb{D}^T = \left\{x \in \mathbb{D}_1^T \,;\, \phi'_{1,x_0}(x) \in \mathbb{D}_2^T\right\}$. The composition function

$$\phi : \mathbb{D}_{\phi_1} \to \mathbb{E}, \qquad \phi(x) = \phi_2(\phi_1(\theta))$$

is Hadamard directionally differentiable at x_0 tangentially to $\mathbb{D}^T$, with

$$\phi'_{x_0} : \mathbb{D}^T \to \mathbb{E}, \qquad \phi'_{x_0}(h) = \phi'_{2,\phi_1(x_0)}(\phi'_{1,x_0}(h))$$

Proof. That ϕ is well defined is clear from assumption (i). To show its Hadamard directional differentiability, let $\{h_n\}_{n=1}^{\infty} \subseteq \mathbb{D}_{\phi_1}$ and $\{t_n\}_{n=1}^{\infty} \subseteq \mathbb{R}_+$ be such that $h_n \to h \in \mathbb{D}^T$, $t_n \downarrow 0$, and $x_0 + t_n h_n \in \mathbb{D}_{\phi_1}$ for all n. Assumption (ii) implies that

$$\lim_{n\to\infty} \left\| \frac{\phi_1(x_0 + t_n h_n) - \phi_1(x_0)}{t_n} - \phi'_{1,x_0}(h) \right\|_{\mathbb{D}_2} = 0 \tag{100}$$

Let $g_n = \frac{1}{t_n}\left[\phi_1(x_0 + t_n h_n) - \phi_1(x_0)\right]$, $g = \phi'_{1,x_0}(h)$, and notice that (100) implies $g_n \to g$ in $\mathbb{D}_2$.

Assumption (i) implies $\phi_1(x_0) + t_n g_n = \phi_1(x_0 + t_n h_n) \in \mathbb{D}_{\phi_2}$ for each n, and the definition of $\mathbb{D}^T$ implies $g \in \mathbb{D}_2^T$. Assumption (iii) implies that

$$\lim_{n\to\infty} \left\| \frac{\phi_2(\phi_1(x_0) + t_n g_n) - \phi_2(\phi_1(x_0))}{t_n} - \phi'_{2,\phi_1(x_0)}(g) \right\|_{\mathbb{E}} = 0 \tag{101}$$

Substitute $\phi_2(\phi_1(x_0) + t_n g_n) = \phi_2(\phi_1(x_0 + t_n h_n))$, and $g = \phi'_{1,x_0}(h)$, into (101) to find

$$\lim_{n\to\infty} \left\| \frac{\phi_2(\phi_1(x_0 + t_n h_n)) - \phi_2(\phi_1(x_0))}{t_n} - \phi'_{2,\phi_1(x_0)}(\phi'_{1,x_0}(h)) \right\|_{\mathbb{E}} = 0$$

which completes the proof. □

Remark F.1. When defining and differentiating composition of functions, the outer function's properties determine restrictions that must be placed on the inner function to ensure the composition function is well defined and differentiable.

A familiar example of this is that the domain of the "inner function" ϕ_1 may need to be restricted to ensure the composition map is well defined. For a simple example, x^3 is well defined and differentiable for any $x \in \mathbb{R}$, but $\log(x^3)$ is only well defined (and differentiable) for $x \in (0, \infty)$.

A less familiar example shows up only when considering Hadamard differentiability tangentially to a set. The tangent spaces of each function jointly determine the tangent space of the derivative of the composition map.

The next lemma shows that Hadamard directionally differentiable functions can be "stacked".

Lemma F.5 (Stacking Hadamard differentiable functions)**.** *Let $\mathbb{D}$, $\mathbb{E}_1$, and $\mathbb{E}_2$ be Banach spaces, and $\mathbb{D}_\phi \subseteq \mathbb{D}$. Suppose $\phi^{(1)} : \mathbb{D}_\phi \to \mathbb{E}_1$ and $\phi^{(2)} : \mathbb{D}_\phi \to \mathbb{E}_2$ are Hadamard directionally differentiable tangentially to $\mathbb{D}_0 \subseteq \mathbb{D}$ at $x_0 \in \mathbb{D}_\phi$ with derivatives $\phi_{x_0}^{(1)\prime} : \mathbb{D}_0 \to \mathbb{E}_1$ and $\phi_{x_0}^{(2)\prime} : \mathbb{D}_0 \to \mathbb{E}_2$. Define*

$$\phi : \mathbb{D}_\phi \to \mathbb{E}_1 \times \mathbb{E}_2, \qquad\qquad \phi(x) = \left(\phi^{(1)}(x), \quad \phi^{(2)}(x)\right)$$

Then ϕ is Hadamard directionally differentiable tangentially to $\mathbb{D}_0$ at x_0, with derivative

$$\phi'_{x_0} : \mathbb{D}_0 \to \mathbb{E}_1 \times \mathbb{E}_2, \qquad\qquad \phi'_{x_0}(h) = \left(\phi_{x_0}^{(1)\prime}(h), \quad \phi_{x_0}^{(2)\prime}(h)\right)$$

Proof. Hadamard directional differentiability of $\phi^{(1)}$ and $\phi^{(2)}$ tangentially to $\mathbb{D}_0$ at x_0 implies that for any sequences $\{h_n\}_{n=1}^\infty \subseteq \mathbb{D}$ and $\{t_n\} \subseteq \mathbb{R}_+$ such that $h_n \to h \in \mathbb{D}_0$, $t_n \downarrow 0$, and $x_0 + t_n h_n \in \mathbb{D}_\phi$ for all n,

$$\lim_{n\to\infty} \left\| \frac{\phi^{(1)}(x_0 + t_n h_n) - \phi^{(1)}(x_0)}{t_n} - \phi_{x_0}^{(1)\prime}(h) \right\|_{\mathbb{E}_1} = 0, \text{ and}$$

$$\lim_{n\to\infty} \left\| \frac{\phi^{(2)}(x_0 + t_n h_n) - \phi^{(2)}(x_0)}{t_n} - \phi_{x_0}^{(2)\prime}(h) \right\|_{\mathbb{E}_2} = 0$$

Since $\|(e_1, e_2) - (\tilde{e}_1, \tilde{e}_2)\|_{\mathbb{E}_1 \times \mathbb{E}_2} = \|e_1 - \tilde{e}_1\|_{\mathbb{E}_1} + \|e_2 - \tilde{e}_2\|_{\mathbb{E}_2}$ metricizes $\mathbb{E}_1 \times \mathbb{E}_2$ (Aliprantis & Border (2006) lemma 3.3), we have

$$
\begin{aligned}
&\left\| \frac{\phi(x_0 + t_n h_n) - \phi(x_0)}{t_n} - \phi'_{x_0}(h) \right\|_{\mathbb{E}_1 \times \mathbb{E}_2} \\
&\quad = \left\| \frac{\left(\phi^{(1)}(x_0 + t_n h_n), \quad \phi^{(2)}(x_0 + t_n h_n)\right) - \left(\phi^{(1)}(x_0), \quad \phi^{(2)}(x_0)\right)}{t_n} - \left(\phi^{(1)}_{x_0}(h), \quad \phi^{(2)}_{x_0}(h)\right) \right\|_{\mathbb{E}_1 \times \mathbb{E}_2} \\
&\quad = \left\| \left(\tfrac{\phi^{(1)}(x_0 + t_n h_n) - \phi^{(1)}(x_0)}{t_n} - \phi^{(1)\prime}_{x_0}(h), \quad \tfrac{\phi^{(2)}(x_0 + t_n h_n) - \phi^{(2)}(x_0 + t_n h_n)}{t_n} - \phi^{(2)\prime}_{x_0} \right) \right\|_{\mathbb{E}_1 \times \mathbb{E}_2} \\
&\quad = \left\| \frac{\phi^{(1)}(x_0 + t_n h_n) - \phi^{(1)}(x_0)}{t_n} - \phi^{(1)\prime}_{x_0}(h) \right\|_{\mathbb{E}_1} + \left\| \frac{\phi^{(2)}(x_0 + t_n h_n) - \phi^{(2)}(x_0)}{t_n} - \phi^{(2)\prime}_{x_0}(h) \right\|_{\mathbb{E}_2}
\end{aligned}
$$

Taking the limit as $n \to \infty$ gives the result. $\square$

F.2.1 Hadamard differentiability in bounded function spaces

It is common to "rearrange" Donsker sets; i.e. view them not as scalar-valued but vector-valued with each coordinate occuring over a particular subset of functions (see Van der Vaart (2000) p. 270). The following lemma shows that one direction of the equivalence can be viewed as an application of the delta method.

Lemma F.6 (Rearranging Donsker sets). *Suppose $\mathcal{F} = \mathcal{F}_1 \cup \ldots \cup \mathcal{F}_K$ is P-Donsker, and $\sqrt{n}(\mathbb{P}_n - P) \xrightarrow{L} \mathbb{G}$ in $\ell^\infty(\mathcal{F})$. The map $\phi : \ell^\infty(\mathcal{F}) \to \ell^\infty(\mathcal{F}_1) \times \ldots \times \ell^\infty(\mathcal{F}_K)$ defined pointwise by*

$$\phi(g)(f_1, \ldots, f_K) = (g(f_1), \ldots, g(f_K))$$

is fully Hadamard differentiable at any $P \in \ell^\infty(\mathcal{F})$ tangentially to $\ell^\infty(\mathcal{F})$, and is its own derivative:

$$\phi'_P : \ell^\infty(\mathcal{F}) \to \ell^\infty(\mathcal{F}_1) \times \ldots \times \ell^\infty(\mathcal{F}_K), \qquad \phi'_P(h) = \phi(h)$$

and hence

$$\sqrt{n}(\phi(\mathbb{P}_n) - \phi(P)) \xrightarrow{L} \phi(\mathbb{G}) \qquad \text{in } \ell^\infty(\mathcal{F}_1) \times \ldots \times \ell^\infty(\mathcal{F}_K)$$

Proof. The map ϕ is linear; let $a, b \in \mathbb{R}$ and $g, h \in \ell^\infty(\mathcal{F})$ and notice that for any $(f_1, \ldots, f_K) \in$

$\mathcal{F}_1 \times \ldots \times \mathcal{F}_K$,

$$\begin{aligned}
\phi(ag+bh)(f_1,\ldots,f_K) &= ((ag+bh)(f_1),\ldots,(ag+bh)(f_K)) \\
&= (ag(f_1)+bh(f_1),\ldots,ag(f_K)+bh(f_K)) \\
&= a(g(f_1),\ldots,g(f_K)) + b(h(f_1)\ldots,h(f_K)) \\
&= a\phi(g)(f_1,\ldots,f_K) + b\phi(h)(f_1,\ldots,f_K) \\
&= (a\phi(g)+b\phi(h))(f_1,\ldots,f_K)
\end{aligned}$$

hence $\phi(ag+bh) = (a\phi(g)+b\phi(h))$, as these functions agree on all of $\mathcal{F}_1 \times \ldots \times \mathcal{F}_K$.

Next observe that ϕ is continuous. Recall that the product topology on $\ell^\infty(\mathcal{F}_1) \times \ldots \times \ell^\infty(\mathcal{F}_K)$ is generated by the norm

$$\|(g_1,\ldots,g_K) - (h_1,\ldots,h_K)\|_{\mathcal{F}_1\times\ldots\times\mathcal{F}_K} = \max\{\|g_1-h_1\|_{\mathcal{F}_1},\ldots,\|g_K-h_K\|_{\mathcal{F}_K}\}$$

see Aliprantis & Border (2006) lemma 3.3. Thus

$$\begin{aligned}
\|\phi(g)-\phi(h)\|_{\mathcal{F}_1\times\ldots\times\mathcal{F}_K} &= \max\left\{\sup_{f_1\in\mathcal{F}_1}|g(f_1)-h(f_1)|,\ldots,\sup_{f_K\in\mathcal{F}_K}|g(f_K)-h(f_K)|\right\} \\
&= \|g-h\|_{\mathcal{F}}
\end{aligned}$$

and hence ϕ is continuous.

Since ϕ is linear and continuous, it is (fully) Hadamard differentiable at any point tangentially to $\ell^\infty(\mathcal{F})$ and is its own Hadamard derivative; indeed, for an: for all sequences $h_n \to h \in \ell^\infty(\mathcal{F})$ and $t_n \downarrow 0 \in \mathbb{R}$, one has $g + t_n h_n \in \ell^\infty(\mathcal{F})$ and

$$\lim_{n\to\infty}\left\|\frac{\phi(g+t_nh_n)-\phi(g)}{t_n} - \phi(h)\right\|_{\mathcal{F}_1\times\ldots\times\mathcal{F}_K} = \lim_{n\to\infty}\|\phi(h_n)-\phi(h)\|_{\mathcal{F}_1\times\ldots\times\mathcal{F}_K} = 0$$

Finally, since $\sqrt{n}(\mathbb{P}_n - P) \xrightarrow{L} \mathbb{G}$ in $\ell^\infty(\mathcal{F})$, the functional delta method (Van der Vaart (2000) theorem 20.8) implies $\sqrt{n}(\phi(\mathbb{P}_n)-\phi(P)) \xrightarrow{L} \phi(\mathbb{G})$ in $\ell^\infty(\mathcal{F}_1) \times \ldots \times \ell^\infty(\mathcal{F}_K)$. □

Although the following lemma and its corollary are stated for functions taking values in $\mathbb{R}$, by combining it with lemma F.5 a similar result can be obtained for functions taking values in $\mathbb{R}^M$, similar to the setting of lemma F.1. Compare van der Vaart & Wellner (1997) lemma 3.9.25.

Lemma F.7 (Hadamard differentiability of maps between bounded function spaces)**.** *Let $f : \mathbb{D}_f \subseteq \mathbb{R}^K \to \mathbb{R}$. Suppose that*

1. *f is continuously differentiable, and*
2. *the gradient of f,*

$$\nabla f : \mathbb{D}_f \to \mathbb{R}^K, \qquad \nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) & \ldots & \frac{\partial f}{\partial x_K}(x)\end{pmatrix}^\intercal,$$

is uniformly continuous.

Define the subset of $\ell^\infty(T)^K$ taking values in $\mathbb{D}_f$,

$$\ell^\infty(T,\mathbb{D}_f) = \left\{ g : T \to \mathbb{R}^K \ ; \ g(t) \in \mathbb{D}_f, \ \sup_{t\in T} \|g(t)\| < \infty \right\} \subseteq \ell^\infty(T)^K$$

and the subset of $\ell^\infty(T,\mathbb{D}_f)$ such that composition with f defines a bounded function:

$$\ell_f^\infty(T,\mathbb{D}_f) = \left\{ g \in \ell^\infty(T,\mathbb{D}_f) \ ; \ \sup_{t\in T} |f(g(t))| < \infty \right\}$$

Then $F : \ell_f^\infty(T,\mathbb{D}_f) \to \ell^\infty(T)$ defined pointwise with $F(g)(t) = f(g(t))$ is (fully) Hadamard differentiable tangentially to $\ell^\infty(T)^K$ at any $g_0 \in \ell_f^\infty(T,\mathbb{D}_f)$, with derivative $F'_{g_0} : \ell^\infty(T)^K \to \ell^\infty(T)$ given pointwise by

$$F'_{g_0}(h)(t) = [\nabla f(g_0(t))]^\intercal \, h(t) = \sum_{k=1}^K \frac{\partial f}{\partial x_k}(g_0(t)) h_k(t)$$

Proof. The domain of $\ell_f^\infty(T,\mathbb{D}_f)$ ensures that $F : \ell_f^\infty(T,\mathbb{D}_f) \to \ell^\infty(T)$ is well defined.

Let $\{h_n\}_{n=1}^\infty \subseteq \ell^\infty(T)^K$ and $\{r_n\}_{n=1}^\infty \subseteq \mathbb{R}$ such that $h_n \to h \in \ell^\infty(T)^K$, $r_n \to 0$, and $g_0 + r_n h_n \in \ell_f^\infty(T,\mathbb{D}_f)$ for each n. For each n and each $t \in T$, apply the mean value theorem to find $\lambda_n(t) \in (0,1)$ such that $g_n(t) \coloneqq \lambda_n(t)(g_0(t) + r_n h_n(t)) + (1-\lambda_n(t)) g_0(t)$ satisfying[18]

$$\begin{aligned} f(x_0(t) + r_n h_n(t)) - f(x_0(t)) &= [\nabla f(g_n(t))]^\intercal \, (x_0(t) + r_n h_n(t) - x_0(t)) \\ &= r_n \, [\nabla f(g_n(t))]^\intercal \, h_n(t) \end{aligned}$$

Use this to see that for all n and all $t \in T$,

$$\begin{aligned} &\left| \frac{f(g_0(t) + r_n h_n(t)) - f(g_0(t))}{r_n} - \nabla f(g_0(t))^\intercal h(t) \right| = |\nabla f(g_n(t))^\intercal h_n(t) - \nabla f(g_0(t))^\intercal h(t)| \\ &\quad \le |\nabla f(g_n(t))^\intercal h_n(t) - \nabla f(g_0(t))^\intercal h_n(t)| + |\nabla f(g_0(t))^\intercal h_n(t) - \nabla f(g_0(t))^\intercal h(t)| \\ &\quad \le \|\nabla f(g_n(t)) - \nabla f(g_0(t))\| \times \|h_n(t)\| + \|\nabla f(g_0(t))\| \times \|h_n(t) - h(t)\| \end{aligned}$$

where the first inequality is by the triangle inequality and the second by Cauchy-Schwarz in $\mathbb{R}^K$. It follows that

$$\sup_{t\in T} \left| \frac{f(g_0(t) + r_n h_n(t)) - f(g_0(t))}{r_n} - \nabla f(g_0(t))^\intercal h(t) \right|$$
$$\le \sup_{t\in T} \|\nabla f(g_n(t)) - \nabla f(g_0(t))\| \times \sup_{t\in T} \|h_n(t)\| \tag{102}$$
$$+ \sup_{t\in T} \|\nabla f(g_0(t))\| \times \sup_{t\in T} \|h_n(t) - h(t)\| \tag{103}$$

[18] The mean value theorem being invoked here is the standard result: for any $x, \tilde{x} \in \mathbb{D}_f$, let $g_{x,\tilde{x}} : [0,1] \to \mathbb{R}$ be given by $g_{x,\tilde{x}}(\lambda) = f(\lambda \tilde{x} + (1-\lambda)x)$. Then $g_{x,\tilde{x}}(0) = f(x)$ and $g_{x,\tilde{x}}(1) = f(\tilde{x})$, and the mean value theorem tells us that there exists $\lambda \in (0,1)$ such that

$$f(\tilde{x}) - f(x) = g_{x,\tilde{x}}(1) - g_{x,\tilde{x}}(0) = g'_{x,\tilde{x}}(\lambda)(1-0) = [\nabla f(\lambda \tilde{x} + (1-\lambda)x)]^\intercal \, (\tilde{x} - x)$$

Consider the term in (102). Recall that for some $\lambda_n(t) \in (0,1)$,

$$\begin{aligned} g_n(t) &= \lambda_n(t)(g_0(t) + r_n h_n(t)) + (1-\lambda_n(t))g_0(t) \\ &= \lambda_n(t) r_n h_n(t) + g_0(t) \end{aligned}$$

and so

$$\|g_n - g_0\|_T = \sup_{t\in T}\|\lambda_n(t) r_n h_n(t)\| \leq |r_n| \times \sup_{t\in T}\|h_n(t)\| \to 0$$

where the limit claim follows from $\sup_{t\in T}\|h_n(t)\| = \|h_n\|_T \to \|h\|_T < \infty$ (implying $\{\sup_{t\in T}\|h_n(t)\|\}_{n=1}^{\infty}$ is bounded) and $r_n \to 0$. Thus $g_n \to g_0$ in $\ell^\infty(T)^K$. Using this and uniform continuity of $\nabla f : \mathbb{D}_f \to \mathbb{R}^K$, lemma F.1 implies $\nabla f(g_n) \to \nabla f(g_0)$ in $\ell^\infty(T)^K$, i.e.

$$\|\nabla f(g_n) - \nabla f(g_0)\|_T = \sup_{t\in T}\|\nabla f(g_n(t)) - \nabla f(g_0(t))\| \to 0$$

Using once again that $\{\sup_{t\in T}\|h_n(t)\|\}_{n=1}^{\infty}$ is bounded, this implies

$$\lim_{n\to\infty} \sup_{t\in T}\|\nabla f(g_n(t)) - \nabla f(g_0(t))\| \times \sup_{t\in T}\|h_n(t)\| = 0 \tag{104}$$

Now consider the term in (103). $\sup_{t\in T}\|\nabla f(g_0(t))\| < \infty$ because $\|\nabla f(\cdot)\|$ is uniformly continuous and $\sup_{t\in T}\|g_0(t)\| < \infty$, just as in the proof of lemma F.1. Furthermore, $\lim_{n\to\infty}\sup_{t\in T}\|h_n(t) - h(t)\| = 0$, so

$$\lim_{n\to\infty} \sup_{t\in T}\|\nabla f(g_0(t))\| \times \sup_{t\in T}\|h_n(t) - h(t)\| = 0 \tag{105}$$

Combining (102) through (105) we obtain

$$\lim_{n\to\infty} \sup_{t\in T}\left|\frac{f(g_0(t) + r_n h_n(t)) - f(g_0(t))}{r_n} - \nabla f(g_0(t))^\intercal h(t)\right| = 0$$

which concludes the proof. $\square$

Remark F.2. Lemma F.7 specifies the domain of F as $\ell^\infty_f(T,\mathbb{D}_f) = \{g \in \ell^\infty(T,\mathbb{D}_f) \; ; \; \sup_{t\in T}|f(g(t))| < \infty\}$. It is often straightforward to clarify the space $\ell^\infty_f(T,\mathbb{D}_f)$ in particular cases; for example, $\ell^\infty_f(T,\mathbb{D}_f) = \ell^\infty(T,\mathbb{D}_f)$ if f satisfies any one of the following: (i) f is bounded, (ii) f is Lipschitz, or (iii) f is bounded on bounded subsets (e.g., $f(x) = x$ is bounded on bounded subsets) See also lemma C.6.

Lemma F.7 requires $\nabla f(\cdot)$ be uniformly continuous, but this often stronger than necessary. When hoping to argue $F : \ell^\infty_f(T,\mathbb{D}_f) \to \ell^\infty(T)$ defined pointwise with $F(g)(t) = f(g(t))$ is (fully) Hadamard differentiable at $g_0 \in \ell^\infty_f(T,\mathbb{D}_f)$, it suffices that f is continuously differentiable on a closed set slightly larger than the (bounded) range of g_0. Compactness of this expanded range and the fact that continuous functions on compact sets are uniformly continuous allow us to apply the preceding lemma. This logic is formalized in the following corollary.

Corollary F.8 (Hadamard differentiability of maps between bounded function spaces, corollary)**.** *Let $f : \mathbb{D}_f \subseteq \mathbb{R}^K \to \mathbb{R}$ be continuously differentiable.*

Define the subset of $\ell^\infty(T)^K$ taking values in $\mathbb{D}_f$,

$$\ell^\infty(T,\mathbb{D}_f) = \left\{g : T \to \mathbb{R}^K \; ; \; g(t) \in \mathbb{D}_f, \; \sup_{t\in T}\|g(t)\| < \infty\right\} \subseteq \ell^\infty(T)^K$$

and the subset of $\ell^\infty(T,\mathbb{D}_f)$ such that composition with f defines a bounded function:

$$\ell^\infty_f(T,\mathbb{D}_f) = \left\{ g \in \ell^\infty(T,\mathbb{D}_f) \ ; \ \sup_{t\in T} |f(g(t))| < \infty \right\}$$

Let $g_0 \in \ell^\infty_f(T,\mathbb{D}_f)$, and suppose that for some $\delta > 0$,

$$g_0(T)^\delta \equiv \left\{ x \in \mathbb{R}^K \ ; \ \inf_{t\in T} \|x - g_0(t)\| \le \delta \right\} \subseteq \mathbb{D}_f.$$

Then $F : \ell^\infty_f(T,\mathbb{D}_f) \to \ell^\infty(T)$ defined pointwise by $F(g)(t) = f(g(t))$ is (fully) Hadamard differentiable at g_0 tangentially to $\ell^\infty(T)^K$, with derivative $F'_{g_0} : \ell^\infty(T)^K \to \ell^\infty(T)$ given pointwise by

$$F'_{g_0}(h)(t) = [\nabla f(g_0(t))]^\intercal h(t) = \sum_{k=1}^K \frac{\partial f}{\partial x_k}(g_0(t)) h_k(t)$$

Proof. Let $\tilde{f} : g_0(T)^\delta \to \mathbb{R}$ be the restriction of f to $g_0(T)^\delta$. Note that $\tilde{f}$ is continuously differentiable on the compact $g_0(T)^\delta \subseteq \mathbb{R}^K$, hence $\nabla\tilde{f}$ is in fact uniformly continuous by the Heine-Cantor theorem. Apply lemma F.7 to find that

$$\tilde{F} : \ell^\infty_f(T, g_0(T)^\delta) \to \ell^\infty(T), \qquad\qquad \tilde{F}(g)(t) = \tilde{f}(g(t)) = f(g(t))$$

is (fully) Hadamard differentiable at g_0, with derivative $\tilde{F}'_{g_0} : \ell^\infty(T)^K \to \ell^\infty(T)$ given pointwise by $\tilde{F}'_{g_0}(h)(t) = [\nabla f(g_0(t))]^\intercal h(t)$. By definition, this means that for any sequences $\{\tilde{h}_n\}_{n=1}^\infty \subseteq \ell^\infty(T)^K$ and $\{\tilde{r}_n\}_{n=1}^\infty \subseteq \mathbb{R}$ such that $\tilde{h}_n \to \tilde{h} \in \ell^\infty(T)^K$, $\tilde{r}_n \to 0$, and $g_0 + \tilde{r}_n\tilde{h}_n \in \ell^\infty(T, g_0(T)^\delta)$ for all n,

$$\lim_{n\to\infty} \left\| \frac{\tilde{F}(g_0 + \tilde{r}_n\tilde{h}_n) - \tilde{F}(g_0)}{\tilde{r}_n} - F'_{g_0}(\tilde{h}) \right\|_T = 0 \tag{106}$$

Let $\{h_n\}_{n=1}^\infty \subseteq \ell^\infty(T)^K$, $\{r_n\}_{n=1}^\infty \subseteq \mathbb{R}$ be such that $h_n \to h \in \ell^\infty(T)^K$, $r_n \to 0$, and $g_0 + r_n h_n \in \ell^\infty(T,\mathbb{D}_f)$ for all n. It suffices to show that

$$\begin{aligned}&\left\| \frac{F(g_0 + r_n h_n) - F(g_0)}{r_n} - F'_{g_0}(h) \right\|_T \\ &\qquad = \sup_{t\in T} \left| \frac{f(g_0(t) + r_n h_n(t)) - f(g_0(t))}{r_n} - [\nabla f(g_0(t))]^\intercal h(t) \right|\end{aligned}$$

has limit zero.

Notice that $g_0 + r_n h_n \to g_0$ in $\ell^\infty(T)^K$, so for some N we have that for all $n \ge N$, $\|g_0 + r_n h_n - g_0\|_T = r_n \sup_{t\in T}\|h_n\| < \delta$. It follows that for $k \in \mathbb{N}$, $g_0 + r_{k+N} h_{k+N} \in \ell^\infty(T, g_0(T)^\delta)$ and hence

$\tilde{r}_k = r_{k+N}$ and $\tilde{h}_k = h_{k+N}$ are sequences for which (106) applies. Therefore,

$$\begin{aligned}
\lim_{n\to\infty}\left\|\frac{F(g_0+r_nh_n)-F(g_0)}{r_n}-F'_{g_0}(h)\right\|_T &= \lim_{k\to\infty}\left\|\frac{F(g_0+r_{k+N}h_{k+N})-F(g_0)}{r_{k+N}}-F'_{g_0}(h)\right\|_T \\
&= \lim_{k\to\infty}\left\|\frac{\tilde{F}(g_0+\tilde{r}_k\tilde{h}_k)-\tilde{F}(g_0)}{\tilde{r}_k}-F'_{g_0}(h)\right\|_T \\
&= 0
\end{aligned}$$

Where the second equality follows from $\tilde{F}(g_0+\tilde{r}_k\tilde{h}_k) = F(g_0+r_{k+N}h_{k+N})$ and $\tilde{F}(g_0) = F(g_0)$. $\square$

The following lemma is lemma S.4.9 from Fang & Santos (2019), but the authors state it for a metric space. The same proof works to show that statement holds in semimetric spaces as well.[19] The statement and proof are included here for completeness.

Lemma F.9 (Hadamard directional differentiability of supremum)**.** *(Fang & Santos (2019) lemma S.4.9)*

Let $(\boldsymbol{A}, d)$ be a compact semimetric space, A a compact subset of $\boldsymbol{A}$, and

$$\psi : \ell^\infty(\boldsymbol{A}) \to \mathbb{R}, \qquad \psi(p) = \sup_{a\in A} p(a)$$

Then ψ is Hadamard directionally differentiable at any $p_0 \in \mathcal{C}(\boldsymbol{A}, d)$ tangentially to $\mathcal{C}(\boldsymbol{A}, d)$. $\Psi_A(p_0) = \arg\max_{a\in A} p_0(a)$ is nonempty, and the directional derivative is given by

$$\psi'_{p_0} : \mathcal{C}(\boldsymbol{A}, d) \to \mathbb{R}, \qquad \psi'_{p_0}(p) = \sup_{a\in\Psi_A(p_0)} p(a)$$

Proof. Let $p_0 \in \mathcal{C}(\mathbf{A})$. Since A is compact, $\Psi_A(p_0) = \arg\max_{a\in A} p_0$ is nonempty (Aliprantis & Border (2006) theorem 2.43). Let $\{p_n\}_{n=1}^\infty \subseteq \ell^\infty(\mathbf{A})$ and $\{t_n\}_{n=1}^\infty \subseteq \mathbb{R}_+$ such that $p_n \to p \in \mathcal{C}(\mathbf{A})$

[19] Some useful facts about semimetrics: (i) A semimetric defines a topology that is first countable (Aliprantis & Border (2006) pp. 70, 72), but this topology is not second countable or Hausdorff. The limits of sequences are not guaranteed to be unique. (ii) In a semimetric space, sequences still characterize the closures of sets, as well as continuity and semicontinuity of functions (Aliprantis & Border (2006), theorems 2.40 and 2.42 on pp. 42-43). (iii) A subset of a semimetric space is compact if and only if it is complete and totally bounded (van der Vaart & Wellner (1997), footnote on p. 17).

and $t_n \downarrow 0$. Notice that

$$\begin{aligned}
&\left|\frac{\psi(p_0+t_np_n)-\psi(p_0)}{t_n}-\psi'_{p_0}(p)\right| \\
&\quad=\left|\frac{\sup_{a\in A}\{p_0(a)+t_np_n(a)\}-\sup_{a\in A}p_0(a)}{t_n}-\sup_{a\in\Psi_A(p_0)}p(a)\right| \\
&\quad=\left|\frac{\sup_{a\in A}\{p_0(a)+t_np_n(a)\}-\sup_{a\in\Psi_A(p_0)}p_0(a)}{t_n}-\sup_{a\in\Psi_A(p_0)}p(a)\right| \\
&\quad\leq\left|\frac{\sup_{a\in\Psi_A(p_0)}\{p_0(a)+t_np(a)\}-\sup_{a\in\Psi_A(p_0)}p_0(a)}{t_n}-\sup_{a\in\Psi_A(p_0)}p(a)\right| &(107)\\
&\qquad+\left|\frac{\sup_{a\in A}\{p_0(a)+t_np_n(a)\}-\sup_{a\in A}\{p_0(a)+t_np(a)\}}{t_n}\right| &(108)\\
&\qquad+\left|\frac{\sup_{a\in A}\{p_0(a)+t_np(a)\}-\sup_{a\in\Psi_A(p_0)}\{p_0(a)+t_np(a)\}}{t_n}\right| &(109)
\end{aligned}$$

First, consider (107). Notice that p_0 is flat on $\Psi_A(p_0)$, so

$$\begin{aligned}
&\left|\frac{\sup_{a\in\Psi_A(p_0)}\{p_0(a)+t_np(a)\}-\sup_{a\in\Psi_A(p_0)}p_0(a)}{t_n}-\sup_{a\in\Psi_F(p_0)}p(a)\right| \\
&\quad=\left|\sup_{a\in\Psi_A(p_0)}p(a)-\sup_{a\in\Psi_A(p_0)}p(a)\right|=0 &(110)
\end{aligned}$$

Next consider (108). Since $p_0+t_np_n$ and p_0+t_np are elements of $\ell^\infty(\mathbf{A})$, lemma F.3 implies

$$\begin{aligned}
&\left|\frac{\sup_{a\in A}\{p_0(a)+t_np_n(a)\}-\sup_{a\in A}\{p_0(a)+t_np(a)\}}{t_n}\right| \\
&\quad\leq\sup_{a\in A}|p_n(a)-p(a)|\leq\|p_n-p\|_{\mathbf{A}}\to 0 &(111)
\end{aligned}$$

Now consider (109). Notice that

$$\varphi:\mathcal{C}(\mathbf{A})\rightrightarrows\mathbf{A}, \qquad \varphi(g)=A$$

is a trivially continuous correspondence with nonempty, compact values. Furthermore,

$$\Gamma_{p_0}:\mathcal{C}(\mathbf{A})\times\mathbf{A}\to\mathbb{R}, \qquad \Gamma_{p_0}(g,a)=p_0(a)+g(a)$$

is continuous on all of $\mathcal{C}(\mathbf{A})\times\mathbf{A}$.[20] Thus $\sup_{a\in A}\{p_0(a)+g(a)\}=\max_{a\in\varphi(g)}\Gamma_{p_0}(g,a)$ satisfies the conditions of the Berge Maximum Theorem (Aliprantis & Border (2006) theorem 17.31), implying the argmax corresondence $\Phi:\mathcal{C}(\mathbf{A})\rightrightarrows\mathbf{A}$ given by $\Phi(g)=\Psi_A(p_0+g)$ is compact valued and upper

[20] To see this, recall that the topology of $\mathcal{C}(\mathbf{A})\times\mathbf{A}$ is generated by the semimetric $\rho((g,a),(\tilde{g},\tilde{a}))=\max\{\|g-\tilde{g}\|_{\mathbf{A}},d(a,\tilde{a})\}$ (Aliprantis & Border (2006) lemma 3.3). Let $\varepsilon>0$ and $g\in\mathcal{C}(\mathbf{A})$. Note that each element of $\mathcal{C}(\mathbf{A})$ is a continuous function defined on a compact set, and is hence uniformly continuous by the Heine-Cantor theorem (lemma F.10). Use uniform continuity of p_0 and g to choose $\delta_{p_0},\delta_g>0$ such that $d(a,\tilde{a})<\delta_{p_0}$ implies $|p_0(a)-p_0(\tilde{a})|<\varepsilon/3$, and $d(a,\tilde{a})<\delta_g$ implies $|g(a)-g(\tilde{a})|<\varepsilon/3$. Let $\delta=\min\{\delta_{p_0},\delta_g,\varepsilon/3\}$, and notice that $\rho((g,a),(\tilde{g},\tilde{a}))<\delta$ implies $|p_0(a)-p_0(\tilde{a})|<\varepsilon/3$, $|g(a)-g(\tilde{a})|<\varepsilon/3$, and $\|g-\tilde{g}\|_{\mathbf{A}}<\varepsilon/3$, and hence $|\Gamma_{p_0}(g,a)-\Gamma_{p_0}(\tilde{g},\tilde{a})|=|p_0(a)+g(a)-p_0(\tilde{a})-\tilde{g}(\tilde{a})|\leq|p_0(a)-p_0(\tilde{a})|+|g(a)-g(\tilde{a})|+|g(\tilde{a})-\tilde{g}(\tilde{a})|<\varepsilon$.

hemicontinuous.

Let $\Psi_A(p_0)^\epsilon = \big\{a \in A\ ;\ \inf_{\tilde{a}\in\Psi_A(p_0)} d(a,\tilde{a}) \le \epsilon\big\}$. Upper hemicontinuity and $\|t_n p\|_{\mathbf{A}} \to 0$ implies that there exists $\delta_n \downarrow 0$ such that $\Psi_A(p_0+t_np) \subseteq \Psi_A(p_0)^{\delta_n}$.[21]

It follows that

$$\begin{aligned}&\left|\frac{\sup_{a\in A}\{p_0(a)+t_np(a)\} - \sup_{a\in\Psi_A(p_0)}\{p_0(a)+t_np(a)\}}{t_n}\right|\\ &\qquad = \frac{1}{t_n}\left(\sup_{a\in\Psi_A(p_0)^{\delta_n}}\{p_0(a)+t_np(a)\} - \sup_{a\in\Psi_A(p_0)}\{p_0(a)+t_np(a)\}\right)\end{aligned}$$

Let $a_{s,n} \in \arg\max_{a\in\Psi_A(p_0)}\{p_0(a)+t_np(a)\}$, which is nonempty because $\Psi_A(p_0)$ is compact and $p_0(a)+t_np(a)$ is continuous. Let $a_{b,n} \in \Psi_A(p_0+t_np) \subseteq \Psi_A(p_0)^{\delta_n}$ satisfy $d(a_{b,n},a_{s,n}) \le \delta_n$, and notice that $\Psi_A(p_0+t_np) \subseteq \Psi_A(p_0)^{\delta_n}$ implies $\sup_{a\in\Psi_A(p_0)^{\delta_n}}\{p_0(a)+t_np(a)\} = p_0(a_{b,n})+t_np(a_{b,n})$. So,

$$\begin{aligned}&\frac{1}{t_n}\left(\sup_{a\in\Psi_A(p_0)^{\delta_n}}\{p_0(a)+t_np(a)\} - \sup_{a\in\Psi_A(p_0)}\{p_0(a)+t_np(a)\}\right)\\ &\qquad = p_0(a_{b,n})+t_np(a_{b,n})-p_0(a_{s,n})-t_np(a_{s,n})\\ &\qquad \le p(a_{b,n})-p(a_{s,n})\end{aligned}$$

where the inequality follows because $a_{s,n}$ maximizes p_0 over A while $a_{b,n}$ may not. Furthermore, $d(a_{b,n},a_{s,n}) \le \delta_n$ implies

$$p(a_{b,n})-p(a_{s,n}) \le \sup_{a,a'\in A,\ d(a,a')\le\delta_n}\big\{p(a)-p(a')\big\}$$

[21] To see this, recall the definition of Φ being upper hemicontinuous (uhc) given in Aliprantis & Border (2006), definition 17.2: Φ is uhc at g if for every neighborhood U of $\Phi(g)$, the upper inverse image

$$\Phi^u(U) = \{h \in \mathcal{C}(\mathbf{A})\ ;\ \Phi(h) \subseteq U\}$$

is a neighborhood of g, i.e. g is in the interior of $\Phi^u(U)$, so there exists $\eta > 0$ such that $\|g-\tilde{g}\|_{\mathbf{A}} < \eta$ implies $\tilde{g} \in \Phi^u(U)$, and hence $\Phi(\tilde{g}) \subseteq U$. Since Ψ_A is uhc and $\Psi_A(p_0)^\epsilon$ is a neighborhood of $\Psi_A(p_0)$, whenever $\|p_0+t_np-p_0\|_{\mathbf{A}} = t_n\|p\|_{\mathbf{A}} < \epsilon$ we have that $\Psi(p_0+t_np) \subseteq \Psi_A(p_0)^\epsilon$.

Let

$$\delta_n = \max_{a\in\Psi_A(p_0+t_np)}\ \min_{\tilde{a}\in\Psi_A(p_0)} d(a,\tilde{a})$$

The inner min is attained because d is continuous and the feasible set is compact. $a \mapsto \max_{\tilde{a}\in\Psi_A(p_0)}\{-d(a,\tilde{a})\}$ is continuous by the Maximum Theorem (Aliprantis & Border (2006) theorem 17.31), which implies $a \mapsto \min_{\tilde{a}\in\Psi_A(p_0)} d(a,\tilde{a})$ is continuous. The outer max is then attained because the feasible set is compact. Notice that δ_n that $\Psi_A(p_0+t_np) \subseteq \Psi_A(p_0)^{\delta_n}$.

Suppose for contradiction that $\delta_n \not\to 0$. Then there exists $\epsilon > 0$ and a subsequence $\{\delta_{n'}\}_{n'=1}^\infty$ such that $\delta_{n'} \ge \epsilon$ for all n', which implies $\Psi_A(p_0+t_{n'}p) \not\subseteq \Psi_A(p_0)^{\epsilon/2}$ for all n'. $\Psi_A(p_0)^{\epsilon/2}$ is a neighborhood of $\Psi_A(p_0) = \Phi(0)$, and Φ is uhc at 0, hence $\Phi^u(\Psi_A(p_0)^{\epsilon/2})$ is a neighborhood of $0 \in \mathcal{C}(\mathbf{A})$. So for some $\eta > 0$, $\|t_{n'}p\| < \eta$ implies $\Phi(t_{n'}p) = \Psi_A(p_0+t_{n'}p) \subseteq \Psi_A(p_0)^{\epsilon/2}$. Since $t_{n'}p \to 0 \in \mathcal{C}(\mathbf{A})$, there exist n' with $\|t_{n'}p\|_{\mathbf{A}} < \eta$, and for such n' we have $\Psi_A(p_0+t_{n'}p) \subseteq \Psi_A(p_0)^{\epsilon/2}$ by upper hemicontinuity. This is the desired contradiction; therefore $\delta_n \to 0$.

If δ_n does not converge monotonically to zero, set $\tilde{\delta}_n = \sup\{\delta_k\ ;\ k \ge n\}$. Note that $\tilde{\delta}_n \downarrow 0$ and $\tilde{\delta}_n \ge \delta_n$, the latter of which implies $\Psi_A(p_0+t_np) \subseteq \Psi_A(p_0)^{\tilde{\delta}_n}$.

and hence

$$
\begin{aligned}
&\left|\frac{\sup_{a\in A}\{p_0(a)+t_np(a)\}-\sup_{a\in\Psi_A(p_0)}\{p_0(a)+t_np(a)\}}{t_n}\right| \\
&\qquad\le \sup_{a,a'\in A,\ d(a,a')\le\delta_n}\{p(a)-p(a')\} \\
&\qquad\to 0
\end{aligned}
\tag{112}
$$

Where the limit claim follows from p being a continuous function defined on a compact set, and so is in fact uniformly continuous by the Heine-Cantor theorem (lemma F.10).

To summarize,

$$
\begin{aligned}
&\left|\frac{\psi(p_0+t_np_n)-\psi(p_0)}{t_n}-\psi'_{p_0}(p)\right| \\
&\quad\le \left|\frac{\sup_{a\in\Psi_A(p_0)}\{p_0(a)+t_np(a)\}-\sup_{a\in\Psi_A(p_0)}p_0(a)}{t_n}-\sup_{a\in\Psi_A(p_0)}p(a)\right| \\
&\qquad+\left|\frac{\sup_{a\in A}\{p_0(a)+t_np_n(a)\}-\sup_{a\in A}\{p_0(a)+t_np(a)\}}{t_n}\right| \\
&\qquad+\left|\frac{\sup_{a\in A}\{p_0(a)+t_np(a)\}-\sup_{a\in\Psi_A(p_0)}\{p_0(a)+t_np(a)\}}{t_n}\right|
\end{aligned}
$$

along with (110), (111), and (112) implies that ψ is Hadamard directionally differentiable at any $p_0\in\mathcal{C}(\mathbf{A})$ tangentially to any $p\in\mathcal{C}(\mathbf{A})$, with $\psi'_{p_0}(p)=\sup_{a\in\Psi_A(p_0)}p(a)$. □

F.3 Other

The Heine-Cantor theorem is usually stated for metric spaces. As it is applied in the proof of lemma F.9 to a setting with semimetric spaces, the statement and standard proof are included here to make clear the result applies to semimetric spaces as well.

Lemma F.10 (Heine-Cantor theorem)**.** *Let (X,d_X) and (Y,d_Y) be semimetric spaces, X compact, and $f:X\to Y$ continuous. Then f is in fact uniformly continuous.*

Proof. Let $\varepsilon>0$. For each $x\in X$, use continuity of f to choose δ_x such that

$$d_X(x,x')<2\delta_x \implies d_Y(f(x),f(x'))<\varepsilon/2$$

Let $B_d(x)\subseteq X$ be the open ball of radius d centered at x. Then $\bigcup_{x\in X}B_{\delta_x}(x)$ is an open cover of X. By compactness of X, there exists $x_1,\ldots,x_n$ such that $\bigcup_{i=1}^n B_{\delta_{x_i}}(x_i)$ covers X. Let $\delta=\min_{i\in\{1,\ldots,n\}}\delta_{x_i}$. As the minimum of a finite number of positive real numbers, we have $\delta>0$.

Suppose $d_X(x,x')<\delta$. Since $\bigcup_{i=1}^n B_{\delta_{x_i}}(x_i)$ covers X, there exists $k\in\{1,\ldots,n\}$ such that $x\in B_{\delta_{x_k}}(x_k)$. Notice that

$$d_X(x',x_k)\le d_X(x',x)+d_X(x,x_k)<\delta+\delta_{x_k}\le 2\delta_{x_k}$$

and thus $d_X(x,x') < \delta$ implies $d_X(x',x_k) < 2\delta_{x_k}$ and $d_X(x,x_k) < \delta_{x_k} < 2\delta_{x_k}$ for whichever k is such that $x \in B_{\delta_{x_k}}(x_k)$. Then the definition of δ_{x_k} implies

$$d_Y(f(x), f(x')) \leq d_Y(f(x), f(x_k)) + d_Y(f(x_k), f(x')) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

$\square$