São Paulo School of
ECONOMICS

Working Paper

Sep 2011

CENTER FOR APPLIED MICROECONOMICS

20/2011

C-MICRO Series

Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures

Sérgio Firpo
Cristine Pinto

Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures*

Sergio Firpo† Cristine Pinto‡

September 2011

Abstract

This paper presents semiparametric estimators of distributional impacts of interventions (treatment) when selection to the program is based on observable characteristics. Distributional impacts of a treatment are calculated as differences in inequality measures of the marginal distributions of potential outcomes of receiving and not receiving the treatment. These differences are called "Inequality Treatment Effects" (ITE). The estimation procedure involves a first non-parametric step in which the probability of receiving treatment given covariates, the propensity-score, is estimated. In the second step weighted sample versions of inequality measures are computed using weights based on the estimated propensity-score. Root-N consistency, asymptotic normality, semiparametric efficiency and validity of inference based on the bootstrap are shown for the semiparametric estimators proposed. In addition of being easily implementable and computationally simple, results from a Monte Carlo exercise reveal that its relatively good performance in small samples is robust to several data generating processes. Finally, as an illustration of the method, we apply the estimator to a real data set collected for the evaluation of a job training program and use several popular inequality measures to capture distributional impacts of the program.

JEL: C1, C3. Keywords: Inequality Measures, Treatment Effects, Semiparametric Efficiency, Reweighting Estimator.

---

*An earlier version of this paper circulated under the names "Treatment Effects on Some Inequality Measures" and "Inequality Treatment Effects". We have benefited from helpful comments from anonymous referees and the editor on an earlier draft, and from comments from David Green, Jim Heckman, Guido Imbens, Thomas Lemieux, Jeff Smith and participants at the 2004 Meeting of the Sociedade Brasileira de Econometria, 2005 World Congress of the Econometric Society, seminars at UBC, USC, and EPGE-FGV. Hugo Jales, Rafael Cayres and Breno Braga helped us with superb research assistance. Financial support from CNPq-Brazil for Firpo is acknowledged. The usual disclaimer applies.

†Corresponding author. Escola de Economia de São Paulo and C-Micro, Fundação Getulio Vargas (Brazil), and IZA (Germany). Mailing address: EESP-FGV, Rua. Itapeva 474/1215, São Paulo - SP, Brazil, 01332-000. E-mail: sergio.firpo@fgv.br. Webpage: http://sites.google.com/site/sergiopfirpo

‡Escola de Economia de São Paulo and C-Micro, Fundação Getulio Vargas (Brazil). Mailing address: EESP-FGV, Rua. Itapeva 474/1200, São Paulo - SP, Brazil, 01332-000. E-mail: cristine.pinto@fgv.br. Webpage: https://sites.google.com/site/cristinepinto/

1 Introduction

For the evaluation of a social program, the policy-maker may want to learn about the distributional effects of the program and not only its mean impact. For example, it is reasonable to assume that the policy-maker is interested in the effect of the treatment on the dispersion of the outcome, which can be captured by commonly used inequality measures such as the Gini coefficient, the interquartile range or other inequality indices, as those belonging to the Generalized Entropy Class.[1]

The distributional impact of the program on the outcome can be measured by what we call in this paper *Inequality Treatment Effects* (ITE), which are defined as differences in inequality measures of the marginal distributions of the potential outcome of joining the program (receiving the treatment) and not joining it (not receiving the treatment).

We follow an increasing part of the literature of program evaluation that is interested in distributional impacts of a treatment. That recent literature could be divided into two branches, depending on how exactly one defines "distributional impacts of a treatment". If that is understood to be the "distribution of individual treatment effects", then key parameters are features of the distribution of the difference of potential outcomes.[2]

The second branch, which this paper contributes to, defines "distributional impacts of a treatment" as treatment impacts on distributions. In that case, one is interested in learning how a program changes the distribution of the outcome under two scenarios: with and without the program. For that goal, one may also look at the entire cumulative distribution functions (c.d.f.) or all quantiles, as Imbens and Rubin (1997), Abadie (2002), Abadie, Angrist and Imbens (2002), Firpo (2007) and Frölich and Melly (2008).

We discuss identification of inequality treatment effects parameters under the assumption termed by Rubin (1977) as *treatment unconfoundedness*, which is also known as the *selection on observables* assumption.[3] The unconfoundedness assumption is a conditional independence assumption: Given observable characteristics, the decision to be treated is independent of the potential outcome of being treated and the one of not being treated. This assumption is crucial as it allows that functionals of the potential outcome distributions be identified from the observed data.

We propose a two step estimation procedure. In the first step, weighting functions are nonparametrically estimated; in the second step inequality measures are calculated using the weighted data. The effect of the program is estimated, therefore, as a simple difference in weighted inequality measures.

Weighted or inverse probability weighted (IPW) estimators are largely used in the missing data and treatment effects literatures and leading examples where IPW estimator is used are Robins and Rotnitzky (1995) and Wooldridge (2007) in the missing data literature and Hirano, Imbens and Ridder (2003) in the treatment effects literature. Recently, Tarozzi (2007), Chen,

[1] For a detailed discussion of several inequality measures see, for example, Cowell (2000).

[2] Some contributions to that branch of the literature are the papers by Heckman (1992), Heckman, Smith and Clements (1997), Heckman and Smith (1998), Carneiro, Hansen and Heckman (2001, 2003), Cunha, Heckman and Navarro (2005), Aakvik, Heckman and Vytlacil (2005), Firpo and Ridder (2008), Fan and Park (2010).

[3] Important examples in which this assumption has been used are, among others, Rosenbaum and Rubin (1983), Heckman, Ichimura, Smith and Todd (1998), Dehejia and Wahba (1999) and Hirano, Imbens and Ridder (2003).

Hong and Tarozzi (2008) and Cattaneo (2009) have shown how to generalize treatment effects identification and estimation under unconfoundedness for a class of parameters that satisfy certain moment conditions. In all these papers, weighted estimators have been presented and used in the context of M-estimation, as parameters of interest solve some moment condition. One main exception is DiNardo, Fortin and Lemieux (1996) who analyzed over time changes in wage densities controlling for covariates through a weighting scheme.

We provide a weighted estimator that can be expressed as a functional of the empirical weighted marginal distribution function. We focus our analysis on the class of Hadamard differentiable functionals. That class encompasses many interesting inequality measures, which are highly non-linear functionals of the distribution but that may admit a linear functional derivative. We show that four popular inequality measures belong to that class: the coefficient of variation, the interquartile range, the Theil index and the Gini coefficient.

Under the unconfoundedness assumption and mild regularity conditions, we show that our weighted estimators are consistent, asymptotically normal and semiparametrically efficient. We provide a consistent estimator for the asymptotic variance. In addition, inference based on the bootstrap is shown to be a valid procedure for estimating the variance.

Under failure of unconfoundedness we may not identify ITE parameters. Nevertheless, the estimation method proposed here can also be used for the goal of comparing inequality measures controlling for the distribution of covariates (observables). Applied researchers are often interested in comparing features of two or more outcome distributions. For example, we might be interested in comparing the Gini coefficient, a widely used inequality measure, for two different wage distributions (e.g. two different countries). Acknowledging for the fact that there are many observed factors whose distributions differ across countries, such as schooling and job experience, leads us to try to control for these factors when comparing Gini coefficients. By doing so, we would be able to quantify how differences in distribution of covariates explain differences in Gini coefficients between the two countries.

In the literature of wage gap decomposition, controlling for observables is achieved by the construction of "counterfactual" wage distributions and Juhn, Murphy and Pierce (1993) and DiNardo, Fortin and Lemieux (1996) provide estimation methods for some features of the counterfactual. DiNardo, Fortin and Lemieux (1996) propose a method for estimation of counterfactual densities, using some of the weights we use in this paper, while Juhn, Murphy and Pierce (1993) construct counterfactual distributions using fitted values and residuals from linear regressions. These methods have been generalized in many ways and recent contributions are the papers by Gosling, Machin and Meghir (2000), Donald, Green and Paarsch (2000), Machado and Mata (2005) and Melly (2006).

More recently, Chernozhukov, Fernandez-Val and Melly (2009) and Rothe (2010) have extended the analysis based on counterfactuals to situations where one may be interested in learning features of the whole marginal distribution of outcomes using a completely new distribution of covariates, the one that could prevail after a policy intervention that affects solely the distribution of covariates. Their approach is semiparametric: in a first stage they estimate nonparametrically the conditional c.d.f. of the outcome given covariates. In the second step, using the new distribution of covariates, they construct marginal counterfactual distributions and recover features from those distributions.

We view our estimation procedure as a computationally simple and alternative way to recover features (inequality measures) of the counterfactual distribution. It is computationally simple since all it is required is a first step that involves estimation of weights and in a second

step calculation of inequality measures using these weights. We do not need to calculate the conditional c.d.f. at several points in the support nor many conditional quantiles, as other alternative semiparametric methods require.

This paper is divided as follows: In the next section we present more formally the ITE class of parameters. Section 3 presents the main identification result. Section 4 discusses estimation and section 5 derives the large sample properties of the inequality treatment effects estimators. Section 6 discusses finite-sample behavior through a Monte Carlo exercise. We present in section 7 a small empirical exercise that uses data on a Brazilian job training program of the late 90's. Although the training program had been designed to be a randomized experiment, randomization was performed at strata (classes) level with different proportions of treated units across strata. Thus controlling for strata is crucial in obtaining consistent estimates of the program impact. Finally, section 8 concludes. Proofs of results are left to the Appendix.[4]

In both sections 6 and 7 we compare our estimation procedure with three other methods: a naive procedure, which computes simple differences in inequality measures with no attempt to control for selection; a method based on regression, which is the one proposed by Juhn, Murphy and Pierce (1993); and a method based on nonparametric estimation of the conditional distribution of the outcome, which is the one proposed by Chernozhukov, Fernandez-Val and Melly (2009). Although evidence coming solely from Monte Carlo exercises is never definitive, as different data generating processes may lead to very different rankings in finite samples, our simulation results reveal that although we may have a much less cumbersome estimation procedure, the costs in terms of bias, variance and coverage rate of our method, when compared to alternative methods, seem to be negligible even in small sample sizes.

2 Inequality Treatment Effects Parameters

Suppose that a random sample of N individuals (units) is available. For each unit i, let X_i be a random vector of observed covariates with support $\mathcal{X} \subset \mathbb{R}^r$. Define $Y_i(1)$ as the potential outcome for individual i if she enters in the program, and $Y_i(0)$ the potential outcome for the same individual if she does not enter. Let the treatment assignment be defined as T_i, which equals one if individual i is exposed to the program and equals zero otherwise. As we only observe each unit at one treatment status, we say that the unobserved outcome is the counterfactual outcome. Thus, the observed outcome can be expressed as:

$$Y_i = T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0), \qquad \forall i.$$

A legitimate way to introduce inequality measures is to assume that there is a social welfare function, W, that depends on a vector of functionals of the outcome distribution. Suppose in particular that W assumes the following form:

$$W(F) = \Omega(\mu(F), \nu(F))$$

where μ is the outcome mean, ν is the inequality measure and F is a distribution function.[5] We define the inequality measure ν as a functional of the marginal distribution, $\nu : \mathcal{F}_\nu \to \mathbb{R}$.

[4] Details for the proofs of all results established in this paper can be found in the supplemental appendix to this paper at authors' web page (http://sites.google.com/site/sergiopfirpo/supp) or upon request.

[5] This is the reduced-form social welfare function discussed by Champernowne and Cowell (1999) and Cowell (2000).

where $F \in \mathcal{F}_\nu$ if $\nu(F) < +\infty$. A particular example of W and ν is the case where ν is the Gini coefficient and W is decreasing in ν. Under this setting, a natural parameter used to compare two marginal distributions functions F and $G \in \mathcal{F}_\nu$ is the simple difference $\nu(F) - \nu(G)$. We discuss three comparisons of distributions that give rise to three different inequality treatment effect parameters.[6]

The first case arises when we want to compare the situation in which everyone is exposed to the program with the situation in which no one is exposed to it. Under the first scenario, the distribution of the outcome equals $F_{Y(1)}$, the distribution of $Y(1)$; while in the second scenario, the outcome distribution equals $F_{Y(0)}$. The difference in a given inequality measure ν between these two hypothetical cases is the **Overall Inequality Treatment Effect** (ITE), Δ_O^ν, defined as:

$$\Delta_O^\nu = \nu\left(F_{Y(1)}\right) - \nu\left(F_{Y(0)}\right) = \nu_1 - \nu_0.$$

Other parameters could be defined for subpopulations. In particular, consider the **Inequality Treatment Effect on the Treated** (ITT), Δ_{TT}^ν:

$$\Delta_{TT}^\nu = \nu\left(F_{Y(1)|T=1}\right) - \nu\left(F_{Y(0)|T=1}\right) = \nu_{11} - \nu_{01}$$

where $F_{Y(1)|T=1}$ and $F_{Y(0)|T=1}$ are respectively the conditional distributions of the potential outcomes of being in the program and of not being in the program for the subpopulation that was actually exposed to the program.

We finally consider a parameter which is a comparison between the current inequality $\nu(F_Y)$ and the inequality that we would encounter if there were no program $\nu\left(F_{Y(0)}\right)$. We call this parameter the **Current Inequality Treatment Effect** (CIT):[7]

$$\Delta_{CIT}^\nu = \nu(F_Y) - \nu\left(F_{Y(0)}\right) = \nu_Y - \nu_0.$$

3 Identification of Inequality Treatment Effects

This section is divided up into three subsections. In the first one, we introduce notation along with definitions of weighted distributions and respective weighting functions. Subsection 3.2 presents the identification assumptions and the main identification results. Finally, in the last subsection we present some examples of popular inequality measures and show how they fit into the framework just presented.

3.1 The Setup

We now set up assumptions for identification of Δ^ν. Remember that because $Y(1)$ and $Y(0)$ are never fully observable, we need to impose some identifying assumptions in order to be able to express functionals of their marginal distributions as functionals of the joint distribution of observable variables $(Y,\ X,\ T)$. Let the data be defined by the sequence $\{Y_i, X_i, T_i\}_{i=1}^N$

[6] Alternative setups to what follows can be found in Manski (1997) and would lead to the definition of some other possible treatment effects parameters. That includes allowing individuals to choose their treatment status and assigning them to treatment based on observed characteristics.

[7] If ν is not decomposable, we cannot write the CIT as linear combination of the previous parameters. Note that in general, $\nu(F_Y) \neq \nu\left(F_{Y|T=1}\right) \cdot \Pr[T=1] + \nu\left(F_{Y|T=0}\right) \cdot \Pr[T=0]$. Note also that many other parameters could be considered, as for example the difference in inequality measures between treated and control subpopulations that where formed following a rule that is a function of pretreatment covariates X.

where each element (Y_i, X_i, T_i) is a random draw from $\mathcal{P}$, the joint distribution of $(Y, X, T) \in \mathcal{Y} \times \mathcal{X} \times \{0, 1\}$, where $\mathcal{Y} \subset \mathbb{R}$.

Identification of Δ^ν will follow after we establish conditions for identification of functionals of the distributions of $Y(1)$ and $Y(0)$, as the parameters Δ^ν are defined as differences between functionals of their marginal distributions.

We start by writing the weighted marginal distribution of Y, which is a key tool in our identification strategy. The weighted marginal distribution of Y, is

$$F_Y^\omega(y) = E\left[\omega\left(T, p\left(X\right)\right) \cdot \mathbb{1}\{Y \leq y\}\right] \tag{1}$$

where $\omega\left(T, p\left(X\right)\right)$ is a given weighting function that 'converts' marginal (or conditional on T) distribution functions of potential outcomes into weighted marginal distributions of Y. The indicator function is denoted by $\mathbb{1}\{\cdot\}$. Note that the definition of weighted c.d.f of Y subsumes the case of the simple (unweighted) marginal c.d.f. of Y by making $\omega = 1$.

The second argument in the weighting function is $p : \mathcal{X} \to \mathcal{E} \subset [0, 1]$, the propensity score or the conditional probability of being treated, defined as $p\left(x\right) \equiv \Pr\left[T = 1 | X = x\right]$, where $X \in \mathcal{X}$. The unconditional probability of being treated, $\Pr\left[T = 1\right]$, is p, which is assumed to be positive.

Next, we define four weighting functions ω, such that $\omega : \{0, 1\} \times \mathcal{E} \to \mathbb{R}$:

$$\omega_1\left(T, p\left(X\right)\right) = T/p\left(X\right), \qquad \omega_0\left(T, p\left(X\right)\right) = \left(1 - T\right)/\left(1 - p\left(X\right)\right),$$
$$\omega_{11}\left(T, p\left(X\right)\right) = T/p, \quad \text{and} \quad \omega_{01}\left(T, p\left(X\right)\right) = \left(\left(1 - T\right)/\left(1 - p\left(X\right)\right)\right) \cdot \left(p\left(X\right)/p\right).$$

These weighting functions will be used to identify the marginal (conditional on $T = 1$) c.d.f.'s of distributions of $Y(1)$ and $Y(0)$.

3.2 Identification

In this section, we discuss the set of identifying restrictions that will permit that we write the distribution of the unobserved potential outcomes in terms of observable data. Those distributions will actually fall into the category of the weighted distributions just defined. Using these identifying assumptions, we show that we can identify the inequality measures defined in section 2.

Identification is based on *unconfoundedness*, a conditional independence assumption.

Assumption 1 [***Unconfoundedness***] *Let $(Y(1), Y(0), X, T)$ have a joint distribution. For all x in $\mathcal{X}$: $(Y(1), Y(0))$ is jointly independent from T given $X = x$, that is, $(Y(1), Y(0)) \perp\!\!\!\perp T | X = x$.*

Assumption 1 is sometimes a strong assumption and its plausibility has to be analyzed in a case by case basis. This assumption has been used, however, in several studies of the effect of treatments or programs. Prominent examples are Rosenbaum and Rubin (1983), Heckman and Robb (1986), LaLonde (1986), Card and Sullivan (1988), Heckman, Ichimura, and Todd (1997), Heckman, Ichimura, Smith, and Todd (1998), Hahn (1998), Lechner (1999), Dehejia and Wahba (1999) and Becker and Ichino (2002). We present in the empirical section an example where, by design, Assumption 1 is valid.

We also make an assumption on the image set of the propensity score, $\mathcal{E}$.

Assumption 2 [***Common Support***] *For all x in $\mathcal{X}$, there are positive real numbers c_* and c^*, such that $0 < c_* \leq p(x) \leq c^* < 1$.*

Assumption 2 states that with probability one there will be no particular value x in $\mathcal{X}$ that belongs to either the treated group or the control group. Such assumption is important as it allows that groups ($T = 1$ and $T = 0$) become fully comparable in terms of X. Assumptions 1 and 2 are termed together as *strong unconfoundedness*.

Finally, the main identification result will follow from Lemma 1 in Firpo (2007). In the following lemma, we show that we can write the ITE parameters as functions of the observable variables (Y,X,T).

LEMMA 1 [***Identification***]:*Under Assumptions 1 and 2 Δ_O^ν, Δ_{TT}^ν, and Δ_{CIT}^ν are identifiable from observable data (Y,X,T).*

The proof of this lemma is based on the fact that under assumptions 1 and 2, we can write the marginal distribution function of $Y(1)$ and $Y(0)$ as a weighted c.d.f of Y. For $j \in \{0,1\}$:

$$F_{Y(j)}(y) \equiv \Pr[Y(j) < y] = E\left[\omega_j(T, p(X)) \cdot \mathbb{I}\{Y \leq y\}\right].$$

Similarly, we can write the conditional distribution functions of $Y(1)$ and $Y(0)$ given $T = 1$ as a weighted c.d.f of Y. For $j \in \{0,1\}$:

$$F_{Y(j)|T}(y|1) \equiv Pr[Y(j) < y|T = 1] = E\left[\omega_{j1}(T, p(X)) \cdot \mathbb{I}\{Y \leq y\}\right].$$

Once we know that the marginal (conditional on $T = 1$) distributions of potential outcomes are identified, functionals of these distributions will also be identified. Thus, our inequality treatment effects are identifiable from data on (Y, X, T).

We can turn our attention to estimation and inference. Before doing so, let us give concrete examples of inequality measures that are considered in this article.

3.3 Some Inequality Measures

We now discuss some concrete examples of inequality measures and express them as functionals of a weighted distribution of Y.

Comparison of inequality measures is often performed on the basis of the attainment of some desirable properties. There is no clear ranking among the measures, but it is common in the welfare literature to check which of the usual properties an inequality measure possesses. Among those properties, the most common and important ones are the *principle of transfers*, *invariance*, *decomposability* and *anonymity*. For a detailed discussion on this topic, see Cowell (2000) and Cowell (2003).[8]

We consider four popular inequality measures: the coefficient of variation, the interquartile range, the Theil index and the Gini coefficient. As discussed in Cowell (2000), the coefficient of variation will satisfy all properties listed before but invariance. The interquartile range will not satisfy any of those properties besides anonymity. The Theil index, being a member of the Generalized Entropy class, will satisfy all four properties, whereas the Gini coefficient, probably the most used inequality measure, is known to be non-decomposable.

We proceed treating those four measures as functionals of a weighted outcome distribution. By doing that, we gain the flexibility necessary to further define the treatment effects as differences in functionals of weighted distributions:[9]

[8] An interesting result in the income distribution literature establishes that any continuous inequality measure that satisfies the principle of transfers, scale invariance, decomposability and the anonymity must be ordinally equivalent to the Generalized Entropy class, which is indexed by a single scalar parameter. See Cowell (2003), Theorem 2.

[9] In what follows we assume that $\mu_Y^\omega \equiv \int y \cdot dF_Y^\omega(y) \neq 0$.

1. **Coefficient of Variation** (CV):

$$\nu^{CV}\left(F_Y^{\omega}\right) = \frac{\left(\int \left(y - \int z \cdot dF_Y^{\omega}(z)\right)^2 \cdot dF_Y^{\omega}(y)\right)^{1/2}}{\int y \cdot dF_Y^{\omega}(y)};$$

2. **Interquartile Range** (IQR):

$$\begin{aligned}\nu^{IQR}\left(F_Y^{\omega}\right) &= \nu^{Q.75}\left(F_Y^{\omega}\right) - \nu^{Q.25}\left(F_Y^{\omega}\right) \\ &= \inf_q \left\{\int_{-\infty}^{q} dF_Y^{\omega}(y) \geq \frac{3}{4}\right\} - \inf_q \left\{\int_{-\infty}^{q} dF_Y^{\omega}(y) \geq \frac{1}{4}\right\};\end{aligned}$$

3. **Theil Index** (TI):[10]

$$\nu^{TI}\left(F_Y^{\omega}\right) = \frac{\int y \cdot \left(\log(y) - \log\left(\int z \cdot dF_Y^{\omega}(z)\right)\right) \cdot dF_Y^{\omega}(y)}{\int y \cdot dF_Y^{\omega}(y)};$$

4. **Gini Coefficient** (GC):

$$\nu^{GC}\left(F_Y^{\omega}\right) = 1 - 2\frac{\int_0^1 \int^{\nu^{Q_\tau}\left(F_Y^{\omega}\right)} y \cdot dF_Y^{\omega}(y) \cdot d\tau}{\int y \cdot dF_Y^{\omega}(y)}.$$

4 Estimation

We first show how to estimate $\nu\left(F_Y^{\omega}\right)$ with a general ω, and later show how to use these results to estimate Δ_O^{ν}, Δ_{TT}^{ν} and Δ_{CIT}^{ν}.

Estimation of $\nu\left(F_Y^{\omega}\right)$ follows from the sample analogy principle. We replace the population distribution F_Y^{ω}, by its empirical distribution counterpart with estimated weights, $\widehat{F}_Y^{\widehat{\omega}}$, and plug it into the functional ν. The estimator will therefore be:

$$\widehat{\nu}_Y^{\widehat{\omega}} = \nu\left(\widehat{F}_Y^{\widehat{\omega}}\right).$$

Note that we take advantage of the fact that the weighted c.d.f. is expressed as $F_Y^{\omega}(y) = E\left[\omega\left(T, p(X)\right) \cdot \mathbb{I}\{Y \leq y\}\right]$, and we write its sample analog as:

$$\widehat{F}_Y^{\widehat{\omega}}(y) = N^{-1}\sum_{i=1}^{N} \omega\left(T, \widehat{p}\left(X_i\right)\right) \cdot \mathbb{I}\{Y_i \leq y\}.$$

It is clear that in order to have a feasible estimation procedure, we first have to carefully address the estimation of the weighting functions $\omega = \omega\left(T, p(X)\right)$ by $\widehat{\omega} = \omega\left(T, \widehat{p}(X)\right)$.

[10] The Theil index requires that the support of the outcome variable be restricted to the positive real numbers. For the Gini coefficient to be well defined, that is, to be between 0 and 1, it is also required the same support restriction.

4.1 Weights Estimation

We have four weighting functions to consider: ω_1, ω_0, ω_{11}, and ω_{01}. Three of them depend on the propensity-score $p(x)$, the exception being ω_{11}.

For the propensity-score estimation we do not impose any parametric assumption on the conditional distribution of T given X nor assume that the propensity-score has a given functional form. In this paper we use a nonparametric logistic approach (Hastie, 1983) that has also been used by Hirano, Imbens and Ridder (2003). They approximate the log odds ratio of the propensity score, $L(p(x))$ by a series of polynomial functions of x.[11] Stacking all these polynomials in a vector, we end up with $H_K(x) = [H_{K,j}(x)]$ $(j = 1, ..., K)$, a vector of length K of polynomial functions of $x \in \mathcal{X}$. The estimation procedure will therefore involve computation of the length K vector of coefficients $\hat{\pi}_K$:

$$\begin{aligned} L(\widehat{p}(x)) &= H_K(x)^\top \widehat{\pi}_K \\ \widehat{p}(x) &= L^{-1}\left(H_K(x)^\top \widehat{\pi}_K\right) = \Lambda\left(H_K(x)' \widehat{\pi}_K\right) \end{aligned}$$

where $\Lambda : \mathbb{R} \to \mathbb{R}$, $\Lambda(z) = (1 + exp(-z))^{-1}$ is the the c.d.f. of a logistic distribution evaluated at z. The nonparametric flavor of such procedure comes from the fact that K is a function of the sample size N such that $K(N) \to \infty$ as $N \to \infty$. Therefore, the vector $\hat{\pi}_K$ increases in length as the sample size increases. The actual calculation of $\hat{\pi}_K$ follows by a pseudo-maximum likelihood approach:

$$\hat{\pi}_K = \arg\max_{\pi_K \in \mathbb{R}^K} \sum_{i=1}^{N} \left(T_i \cdot \log(\Lambda(H_K(X_i)^\top \pi_K)) + (1 - T_i) \cdot \log(1 - \Lambda(H_K(X_i)^\top \pi_K))\right).$$

In the implementation of this procedure, following Hirano, Imbens and Ridder (2003), we restrict the choice of $H_K(\cdot)$ to the class of polynomial vectors satisfying at least the following three properties: (i) $H_K : \mathcal{X} \to \mathbb{R}^K$; (ii) $H_{K,1}(x) = 1$, and (iii) if $K > (n+1)^r$, then $H_K(x)$ includes all polynomials up order n.[12] In addition to unconfoundedness and overlap, we need to control how K, the size of the polynomial H_K, increases with N. The rate at which additional terms are added to the polynomial depends on degree of smoothness of the propensity score and the dimension of X.

Assumption 3 [***Propensity Score***] *For all $x \in \mathcal{X}$, the propensity score $p(x)$ is s_p times continuously differentiable with $s_p \geq 7r$.*

Assumption 4 [***Series***] *The order of $H_K(x)$, K, is of the form $K = C \cdot N^{c_p}$ where C is a constant and $c_p \in \left(\frac{1}{4\left(\frac{s_p}{r}-1\right)}, \frac{1}{9}\right)$.*

After we have estimated the propensity score, the estimated weights are simply $\widehat{\omega} = \omega(T, \widehat{p}(X))$.[13]

[11] The log odds ratio of z, $L(z)$, is $L(z) = \log(z/(1-z))$.

[12] Further details regarding the choice of $H_K(x)$ and its asymptotic properties can be found in the supplemental appendix to this paper and in Hirano, Imbens and Ridder (2003).

[13] Notice that under assumptions 2 and 4 the weighting functions $\omega(T, p(X))$ are measurable bounded differentiable functions with bounded derivatives. This property of the weighted functions are important to derive the large sample propeties of our estimators.

4.2 Estimation of inequality treatment effects

Once the weights have been computed, the three ITE parameters are easily estimated by the plug-in method. We write the corresponding estimators of Δ_O^ν, Δ_{TT}^ν, and Δ_{CIT}^ν as

$$\begin{aligned}
\widehat{\Delta}_O^\nu &= \widehat{\nu}_1 - \widehat{\nu}_0 = \nu\left(\widehat{F}_Y^{\widehat{\omega}_1}\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_0}\right) \\
\widehat{\Delta}_{TT}^\nu &= \widehat{\nu}_{11} - \widehat{\nu}_{01} = \nu\left(\widehat{F}_Y^{\widehat{\omega}_{11}}\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_{01}}\right) \\
\widehat{\Delta}_{CIT}^\nu &= \widehat{\nu}_Y - \widehat{\nu}_0 = \nu\left(\widehat{F}_Y\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_0}\right).
\end{aligned}$$

5 Large Sample Inference

In this section we derive the asymptotic distribution for inequality treatment effect parameters based on inequality measures that are Hadamard differentiable functionals of the distribution of potential outcomes. Although we use four inequality measures as concrete examples, our analysis is more general and could be extended to other functionals of the distribution that satisfy the same differentiability property, as for example, inequality measures that belong to the Generalized Entropy Class. In fact, many other estimands and hypothesis tests could be considered beyond the inequality measures here studied.[14]

We then use results from the semiparametric efficiency literature and treatment effects literature (e.g. Hahn, 1998 and Hirano, Imbens and Ridder, 2003, Cattaneo, 2009) to show efficiency of our estimators.[15] Finally, we provide a consistent estimator for the asymptotic variance of the ITE estimators considered in this paper and show that the inference based on bootstrap is also a valid procedure.

We first impose some smoothness conditions on the joint distribution of $(Y(1), Y(0), X, T)$.

Assumption 5 [***Smoothness***]

- *The support of $\mathcal{X}$ of X is a compact subset of $\mathbb{R}^r$.*
- *The density of X is bounded and bounded away from 0 on $\mathcal{X}$.*
- *$Y(1), Y(0)$ are distributed respectively as F_1 and F_0, which are defined over a common compact support $\mathcal{Y}$.*
- *F_1 and F_0 are continuos differentiable functions on $\mathcal{Y}$ with $F_1(0) = F_0(0) = 0$.*
- *For any given x in $\mathcal{X}$, $F_0(y|x)$ and $F_1(y|x)$ are continuous in $y \in \mathcal{Y}$.*

[14] For examples of some estimands and hypothesis tests that could involve, for instance, quantile processes, first- and second-order stochastic dominance, and Kolmogorov-Smirnov tests, see Abadie (2002), Chernozhukov, Fernandez-Val and Melly (2009) and Donald and Hsu (2010). The latter proposes a test for stochastic dominance between the potential outcome distribtuions. Their test statistics is another example of estimand that is in the class studied in this paper. In order to obtain large sample properties of their estimator, one could use inference results presented in this section.

[15] Similar efficiency results can be found in the missing data literature. Robins, Rotnitzky, and Zhao (1994), Robins and Rotnitzky (1995) and Rotnitzky and Robins (1995) provide calculations of the semiparametric efficiency bounds for nonlinear regression models.

Under Assumption 5 we have that the density of X is bounded in its entire support, and that all covariates are continuous. In addition, we have that $Y(1), Y(0)$ have uniformly continuously differentiable cdf's.

In addition to Assumption 5, we also impose a smoothness condition that enables us to derive the asymptotic normality of the proposed inequality estimators. We restrict the discussion to the class of inequality measures that are Hadamard differentiable functionals of the distribution. Imposing such smoothness condition is useful to help us establishing (i) asymptotic normality, (ii) efficiency and (iii) validity of the inference procedure for our estimators. Although all four inequality measures previously discussed satisfy that condition, we write it as an assumption to make it generalizable to other functionals.

ASSUMPTION 6 [***Hadamard***] *The inequality measure ν defined over the marginal distribution of potential outcomes is Hadamard differentiable.*[16]

Under some additional mild conditions on the c.d.f., Assumption 6 can be verified to hold for all four inequality measures considered in subsection 3.3. The coefficient of variation and the Theil index are known functions of expectations, which are already linear functionals, and therefore satisfy Assumption 6 by definition. The interquartile range is a known function of quantiles, which are Hadamard differentiable if the c.d.f. F is continuously differentiable with positive derivative f at the quantile.[17] Finally, as shown in Bhattacharya (2007, Proposition 2), if in addition to the continuous differentiability of the c.d.f. and the existence of positive density we impose a tail restriction that makes the density not to go to 0 too slowly at the tails, then the Gini coefficient will also be Hadamard differentiable.[18]

We are now able to obtain the limiting distribution of our estimators. For that goal, we first establish uniform root-N consistency for $\widehat{F}_Y^{\widehat{\omega}}$ and present a uniform asymptotically linear representation for that estimator. In possession of these results we can later apply the Functional Delta-Method to get the limiting distribution of $\widehat{\nu}$.

LEMMA 2 *Under assumptions 1, 2, 3, 4 and 5*

$$\sup_{y\in\mathcal{Y}}\left|\sqrt{N}\left(\widehat{F}_Y^{\widehat{\omega}}(y)-F_Y^{\omega}(y)\right)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\psi\left(Y_i,X_i,T_i,y\right)-F_Y^{\omega}(y)\right)\right|=o_p(1)$$

where $\psi\left(Y_i,X_i,T_i,y\right)=\omega\left(T_i,p\left(X_i\right)\right)\cdot\mathbb{I}\left\{Y_i\leq y\right\}+\mathbb{E}\left[\frac{\partial\omega(T_i,p(X_i))}{\partial p(X_i)}\cdot\mathbb{I}\left\{Y_i\leq y\right\}\middle|X_i\right]\left(T_i-p\left(X_i\right)\right).$

Lemma 2 provides a general result that can be applied to the four weighted marginal distribution functions considered in the paper.[19] One can easily check that the derivatives to the

[16] See Ren and Sen (1991)'s Equation (2.9) for a formal definition of Hadamard (compact) differentiable functionals.

[17] To be precise and following van der Vaart (1998, chapters 20 and 21), the $\tau-$th quantile of the c.d.f. will be Hadamard differentiable *tangentially* to the subset of c.d.f.s where the derivative is well defined. In fact, as discussed in van der Vaart (1998, Theorem 20.8) one only needs tangential Hadamard differentiability (and not necessarily Hadamard differentiability) for functional delta-method applications.

[18] Some of the inequality measures of subsection 3.3 also need restrictions on $\mathcal{Y}$, the support of Y_1 and Y_0, to be well defined. For example, the Gini and the Theil are defined for $(0,\infty)$. The coefficient of variation and the Gini also require that mean of the distribution is different from zero.

[19] As it is clear in the proof, this result holds for any estimator of the propensity score that satisfies the restriction $\sup_{x\in\mathcal{X}}\left|\widehat{p}(x)-p(x)\right|=o_p\left(N^{-1/4}\right)$. Thus, we could use other estimators of the propensity score that satify that restriction, but that are not necessarily series-type estimators.

four weighting functions with respect to $p(X)$ are

$$\partial\omega_1(T,p(X))/\partial p(X) = -T/p^2(X), \qquad \partial\omega_0(T,p(X))/\partial p(X) = (1-T)/(1-p(X))^2,$$
$$\partial\omega_{11}(T,p(X))/\partial p(X) = 0, \quad \text{and} \quad \partial\omega_{01}(T,p(X))/\partial p(X) = (1-T)/\left[p\cdot(1-p(X))^2\right].$$

Let us consider two weighting functions ω_A and ω_B out of the four presented so far. One consequence of the linear representation of the weighted distribution function is the following result.

THEOREM 1 *Suppose assumptions 1, 2,, 3 ,4 and 5 hold, then*

$$\sqrt{N}\left[\begin{array}{c}\widehat{F}_Y^{\widehat{\omega}_A}(y) - F_Y^{\omega_A}(y)\\ \widehat{F}_Y^{\widehat{\omega}_B}(y) - F_Y^{\omega_B}(y)\end{array}\right] = \mathbb{G}_N^{\omega_{A,B}} + o_p(1) \Rightarrow \mathbb{G}^{\omega_{A,B}}$$

where
(i) $\mathbb{G}_N^{\omega_{A,B}} = \left[\begin{array}{c}\mathbb{G}_N^{\omega_A}\\ \mathbb{G}_N^{\omega_B}\end{array}\right]$, *and* $\mathbb{G}_N^{\omega_j}$ *is a empirical process such that at a given* $y\in\mathcal{Y}$, *for j=A,B*

$$\mathbb{G}_N^{\omega_j}(y) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\psi_j(Y_i,X_i,T_i,y) - F_Y^{\omega_j}(y)\right);$$

(ii) $\Rightarrow$ *denotes weak convergence;*
(iii) $\mathbb{G}^{\omega_{A,B}}$ *is a Gaussian process with variance-covariance matrix given by*

$$\mathbb{E}\left[\mathbb{G}^{\omega_j}(s)\,\mathbb{G}^{\omega_k}(t)\right] = \mathbb{E}\left[\left(\psi_j(Y,X,T,s) - F_Y^{\omega_j}(s)\right)\cdot\left(\psi_k(Y,X,T,t) - F_Y^{\omega_k}(t)\right)\right]$$

for $(s,t)\in\mathcal{Y}\times\mathcal{Y}$, *j and k=A,B*
(iv) $\psi_j(Y,X,T,y) = \omega_j(T,p(X))\cdot\mathbb{I}\{Y\le y\} + \mathbb{E}\left[\frac{\partial\omega_j(T,p(X))}{\partial p(X)}\cdot\mathbb{I}\{Y\le y\}\middle| X\right](T-p(X))$, *for j=A,B.*

The proof of this theorem follows from establishing that the class of measurable functions from $(\mathcal{Y}\times\mathcal{X}\times\{0,1\})\to\mathbb{R}$, $\mathcal{H} = \{\psi(Y_i,X_i,T_i,y)|\, y\in\mathcal{Y}\}$ is P-Donsker, and then by applying the Donsker's Theorem (page 266 of Van der Vaart,1998). Interestingly, under the same assumptions of Theorem 1, one can show that the maximal asymptotic precision with which we can estimate $F_Y^{\omega}(y)$ is given by

$$\mathbb{E}\left[\left(\omega(T,p(X))\cdot\mathbb{I}\{Y\le y\} + \mathbb{E}\left[\frac{\partial\omega(T,p(X))}{\partial p(X)}\cdot\mathbb{I}\{Y\le y\}\middle| X\right](T-p(X)) - F_Y^{\omega}(y)\right)^2\right].$$

The characterization of this bound follows the work of Newey (1990) and Bickell et al (1993).[20]

Once we have established joint uniform convergence for two empirical weighted marginal distributions, we can establish that estimators $\nu\left(\widehat{F}_Y^{\widehat{\omega}_A}\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_B}\right)$ will be asymptotically normal. Moreover, as estimators of the inequality treatment effect parameters are simply differences in estimators of inequality measures of weighted distributions, we can show that our estimators will be semiparametrically efficient. These results are direct a consequence of Functional Delta Method (Theorem 20.6 of Van der Vaart, 1998).

[20] A detailed proof of this result can be found in the supplemental appendix.

Theorem 2 *Under assumptions 1, 2, 3, 4, 5 and 6*

$$\begin{aligned}&\sqrt{N}\left(\nu\left(\widehat{F}_Y^{\widehat{\omega}_A}\right)-\nu\left(\widehat{F}_Y^{\widehat{\omega}_B}\right)-\left(\nu\left(F_Y^{\omega_A}\right)-\nu\left(F_Y^{\omega_B}\right)\right)\right)\\ =\;&\psi^{\nu}\left(\mathbb{G}_N^{\omega_A};F_Y^{\omega_A}\right)-\psi^{\nu}\left(\mathbb{G}_N^{\omega_B};F_Y^{\omega_B}\right)+o_p(1)\end{aligned}$$

where for j=A,B

$$\begin{aligned}\psi^{\nu}\left(\mathbb{G}_N^{\omega_j};F_Y^{\omega_j}\right) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_j\left(T_i,p\left(X_i\right)\right)\cdot\phi^{\nu}\left(Y_i;F_Y^{\omega_j}\right)\\ &\quad+\mathbb{E}\left[\left.\frac{\partial\omega_j\left(T,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\cdot\phi^{\nu}\left(Y;F_Y^{\omega_j}\right)\right|X_i\right]\left(T_i-p\left(X_i\right)\right),\end{aligned}$$

ψ^{ν} is the functional ν's Hadamard derivative, $\phi^{\nu}\left(Y_i;\cdot\right)=\psi^{\nu}\left(\delta_{Y_i};\cdot\right)$ and δ_{Y_i} is the Dirac measure at observation i. Moreover, $\nu\left(\widehat{F}_Y^{\widehat{\omega}_A}\right)-\nu\left(\widehat{F}_Y^{\widehat{\omega}_B}\right)$ is asymptotically efficient.

We can apply the general result in Theorem 2 to our ITE estimators. In fact,

$$\begin{aligned}\sqrt{N}\left(\widehat{\Delta}_O^{\nu}-\Delta_O^{\nu}\right) &\rightarrow {}_D\mathcal{N}\left(0,V_O\right)\\ \sqrt{N}\left(\widehat{\Delta}_{TT}^{\nu}-\Delta_{TT}^{\nu}\right) &\rightarrow {}_D\mathcal{N}\left(0,V_{TT}\right)\\ \sqrt{N}\left(\widehat{\Delta}_{CIT}^{\nu}-\Delta_{CIT}^{\nu}\right) &\rightarrow {}_D\mathcal{N}\left(0,V_{CIT}\right)\end{aligned}$$

where analytical expressions for V_O, V_{TT}, and V_{CIT} are, respectively

$$V_O=\mathbb{E}[(\omega_1\left(T,p\left(X\right)\right)\cdot\phi^{\nu}\left(Y;F_Y^{\omega_1}\right)-\omega_0\left(T,p\left(X\right)\right)\cdot\phi^{\nu}\left(Y;F_Y^{\omega_0}\right)+\left(g_1\left(X\right)-g_0\left(X\right)\right)\left(T-p\left(X\right)\right))^2]$$

$$V_{TT}=\mathbb{E}[(\omega_{11}\left(T,p\left(X\right)\right)\cdot\phi^{\nu}\left(Y;F_Y^{\omega_{11}}\right)-\omega_{01}\left(T,p\left(X\right)\right)\cdot\phi^{\nu}\left(Y;F_Y^{\omega_{01}}\right)-g_{01}\left(X\right)\left(T-p\left(X\right)\right))^2]$$

and

$$V_{CIT}=\mathbb{E}[(\phi^{\nu}\left(Y;F_Y\right)-\omega_0\left(T,p\left(X\right)\right)\cdot\phi^{\nu}\left(Y;F_Y^{\omega_0}\right)-g_0\left(X\right)\left(T-p\left(X\right)\right))^2],$$

and where for $j=0,1$

$$g_j\left(x\right)\equiv\mathbb{E}\left[\left.\frac{\partial\omega_j\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\cdot\phi^{\nu}\left(Y;F_Y^{\omega_j}\right)\right|X=x\right]$$

and

$$g_{01}\left(x\right)\equiv\mathbb{E}\left[\left.\frac{\partial\omega_{01}\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\cdot\phi^{\nu}\left(Y;F_Y^{\omega_{01}}\right)\right|X=x\right].$$

Valid inference for estimators of inequality treatment effect parameters can be implemented either by estimation of the analytical expressions for the variance terms or by resampling methods, such as the bootstrapping. In the next Theorems we show that both methods are consistent for variance estimation. Again, we use a general form and then specialize for the case of interest.

Estimators for the variance based on its analytical expression have to deal with estimation of functions $g\left(x\right)$. Notice that these functions depend on the propensity-score and, therefore,

we again follow Hirano, Imbens and Ridder (2003) and replace $p(X)$ by $\widehat{p}(X)$. In addition, for $j = A, B$ our estimator of $g_j(\cdot)$ is $\widehat{g}_j(\cdot)$, a nonparametric regression of $\left.\frac{\partial \omega_j(T,p(X))}{\partial p(X)}\right|_{p(X)=\widehat{p}(X)} \cdot \phi^\nu\left(Y; \widehat{F}_Y^{\widehat{\omega}_j}\right)$ on X using series. The series estimator is written as

$$\widehat{g}_j(\cdot) = H_{K_j}(\cdot)^\top \widehat{\gamma}_K^{\omega_j},$$

where

$$\widehat{\gamma}_K^{\omega_j} = \arg\min_\gamma \sum_{i=1}^N \left(\left.\frac{\partial \omega_j(T_i, p(X))}{\partial p(X)}\right|_{p(X)=\widehat{p}(X_i)} \cdot \phi^\nu\left(Y_i; \widehat{F}_Y^{\widehat{\omega}_j}\right) - H_{K_j}(X_i)^\top \gamma \right)^2.$$

In this case, we use a series of orthonormal polynomials such that

$$\sup_{x\in\mathcal{X}} \left\| H_{K_j}(x) \right\| = \zeta(K_j) \leq CK_j$$

where $H_{K_j}(\cdot)$ needs to satisfy the following properties: (i) $H_{K_j}(\cdot) : \mathcal{X} \to \mathbb{R}^{K_j}$; (ii) $H_{K_j,1}(\cdot) = 1$ and (iii) if $K_j > (n_j + 1)^r$, $H_{K_j}(\cdot)$ includes all the polynomials up to order n_j. In order to derive the large sample properties of the estimator of the conditional expectation, we need to control how K_j increases with N. We impose the following assumptions.

Assumption 7 [***Series Estimator***] *For $j = A, B$, and all $x \in \mathcal{X}$:*
(i) $g_j(x)$ is bounded and s_j times continuously differentiable;
(ii) The order of $H_{K_j}(x)$, K_j, is of the form $K_j = C \cdot N^{c_j}$ where C is a constant and $c_{jl} \in (0, \frac{1}{2}\left(\frac{s_p}{2r} - 1\right) c_p)$;
(iii) $\sup_y \left| \left.\frac{\partial \phi^\nu(y;z)}{\partial z}\right|_{z=F_Y^\omega} \right| \leq M$.

Using series estimators for $g(\cdot)$ functions, we propose a consistent estimator for the asymptotic variance of $\sqrt{N}\left(\nu\left(\widehat{F}_Y^{\widehat{\omega}_A}\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_B}\right)\right)$, which we represent as

$$\begin{aligned} V_{AB} \quad = \quad & \mathbb{E}[(\omega_A(T, p(X)) \cdot \phi^\nu\left(Y; F_Y^{\omega_A}\right) - \omega_B(T, p(X)) \cdot \phi^\nu\left(Y; F_Y^{\omega_B}\right) \\ & + (g_A(X) - g_B(X))(T - p(X)))^2] \end{aligned}$$

Theorem 3 *Under assumptions 1-7,*

$$\widehat{V}_{AB} \to_P V_{AB}$$

where

$$\begin{aligned} \widehat{V}_{AB} \quad = \quad & \frac{1}{N}\sum_{i=1}^N (\omega_A(T_i, \widehat{p}(X_i)) \cdot \phi^\nu\left(Y_i; F_Y^{\widehat{\omega}_A}\right) - \omega_B(T_i, \widehat{p}(X_i)) \cdot \phi^\nu\left(Y_i; F_Y^{\widehat{\omega}_B}\right) \\ & + (\widehat{g}_A(X_i) - \widehat{g}_B(X_i))(T_i - \widehat{p}(X_i)))^2. \end{aligned}$$

Consistency of $\widehat{V}_{AB}$ follows from uniformly consistency of $\widehat{g}_j(x)$ to $g_j(x)$ and proposition 2. Calculation of $\widehat{V}_{AB}$ may be computationally demanding since we need to estimate two series

estimators. Next, we show that in applied work, calculations via bootstrap of standard errors for $\widehat{\Delta}$ are a valid procedure for inference.

Consider a random sample $Z = \{(Y_i, X_i, T_i) : i = 1, ..., N\}$ from $\mathcal{P}$, the joint distribution of (Y, X, T). The estimator $\widehat{\Delta} = \nu\left(\widehat{F}_Y^{\widehat{\omega}_A}\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_B}\right)$ is a function of the original sample Z.

Suppose that we construct B bootstrap samples, $Z^* = \{Z_b : b = 1, ..., \text{B}\}$, where for each Z_b we randomly draw N observations from Z with replacement, that is, $Z_b = \{(Y_i^*, X_i^*, T_i^*) : i = 1, ..., N\}$. The bootstrap weighted empirical distribution is the empirical measure

$$\widehat{F}_{Yb}^{\widehat{\omega}_j}(y) = N^{-1} \sum_{i=1}^{N} \omega_j\left(T_i^*, \widehat{p}\left(X_i^*\right)\right) \mathbb{1}\{Y_i^* \leq y\}$$

and the bootstrap empirical process is defined as

$$\mathbb{G}_N^{\omega_j *} = \sqrt{N}\left(\widehat{F}_{Yb}^{\widehat{\omega}_j} - \widehat{F}_Y^{\widehat{\omega}_j}\right).$$

The bootstrap estimator is $\widehat{\Delta}_b = \nu\left(\widehat{F}_{Yb}^{\widehat{\omega}_{Ab}}\right) - \nu\left(\widehat{F}_{Yb}^{\widehat{\omega}_{Bb}}\right)$ and the bootstrap variance is

$$V_{AB}^{\mathfrak{B}} = \mathbb{E}\left[\left(\widehat{\Delta}_b - \widehat{\Delta}\right)^2 \middle| Z\right].$$

Given the B replica samples, an unbiased estimator for this variance is

$$\widehat{V}_{AB}^{\mathfrak{B}} = \frac{1}{\text{B}} \sum_{b=1}^{\text{B}} \left(\widehat{\Delta}_b - \overline{\widehat{\Delta}}\right)^2$$

where $\overline{\widehat{\Delta}} = \text{B}^{-1} \sum_{b=1}^{\text{B}} \widehat{\Delta}_b$.

Theorem 4 *Suppose that assumptions 1-7 hold, then*

$$\widehat{V}_{AB}^{\mathfrak{B}} \rightarrow_P V_{AB}.$$

In fact, the bootstrap is surely an easier alternative than calculation of the analytical standard errors. In the next sections, we present a Monte Carlo exercise and an empirical application that use bootstrapped standard errors.

6 A Monte Carlo Exercise

In this section we report the results of Monte Carlo exercises. We are interested in learning how the estimators for the inequality treatment effect behave in small samples. One thousand (1,000) replications of the experiment with sample sizes of 250, 1,000 and 4,000 observations were considered.

We designed the data generation process (d.g.p.) to produce "selection on observables", that is, the conditional distribution of X given T will differ from the marginal distribution of X, but marginal distributions of the potential outcomes will be independent of T given X. Note that as $Y(1)$ and $Y(0)$ are known for each observation i, we can compute "unfeasible" estimators of functionals of the marginal distributions of $Y(1)$ and $Y(0)$. If we restrict our attention

to subpopulations, for example, the treated, we can still compute "unfeasible" statistics of estimators $Y(1)|T=1$ and $Y(0)|T=1$.

The generated data follows a very simple specification. Starting with $X=[X_1,X_2]^\intercal$ we set $X_1 \sim \text{Unif}\left[\mu_{X_1}-\frac{\sqrt{12}}{2},\mu_{X_1}+\frac{\sqrt{12}}{2}\right]$ and $X_2 \sim \text{Unif}\left[\mu_{X_2}-\frac{\sqrt{12}}{2},\mu_{X_2}+\frac{\sqrt{12}}{2}\right]$, which will be independent random variables with the following means and variances: $E\left[X_1\right]=\mu_{X_1}$, $E\left[X_2\right]=\mu_{X_2}$ and $V\left[X_1\right]=V\left[X_2\right]=1$. The treatment indicator is set to be

$$T=\mathbb{I}\{\delta_0+\delta_1X_1+\delta_2X_2+\delta_3X_1^2+\delta_4X_2^2+\delta_5X_1X_2+\eta>0\}.$$

We consider two possible distributions for η: (i) logistic, $\eta \sim F_\eta(n)=\left(1+\exp\left(-\frac{\pi n}{10\sqrt{3}}\right)\right)^{-1}$; (ii) normal, $\eta \sim F_\eta(n)=\int_{-\infty}^{\frac{n}{10}}(2\pi)^{-1/2}\exp\left(-z^2/2\right)dz$. In all cases, $\eta \sim (0,100)$, that is, η has mean zero and standard deviation 10.

The potential outcomes are

$$\begin{aligned} Y(0) &= \exp\left(\beta_{00}+\beta_{01}X_1+\beta_{02}X_2+\beta_{03}X_1^2+\beta_{04}X_2^2+\beta_{05}X_1X_2+\epsilon_0\right) \\ Y(1) &= \exp\left(\beta_{10}+\beta_{11}X_1+\beta_{12}X_2+\beta_{13}X_1^2+\beta_{14}X_2^2+\beta_{15}X_1X_2+\epsilon_1\right) \end{aligned}$$

where

$$\begin{aligned} \epsilon_0 &= \left(\beta_{00}^s+\beta_{01}^sX_1+\beta_{02}^sX_2+\beta_{03}^sX_1^2+\beta_{04}^sX_2^2+\beta_{05}^sX_1X_2\right)\cdot\kappa_0 \\ \epsilon_1 &= \left(\beta_{10}^s+\beta_{11}^sX_1+\beta_{12}^sX_2+\beta_{13}^sX_1^2+\beta_{14}^sX_2^2+\beta_{15}^sX_1X_2\right)\cdot\kappa_1 \end{aligned}$$

and where κ_0 and κ_1 are distributed as standard normals. The variables X, η, κ_0 and κ_1 are mutually independent. Under this specification, $Y(1)$ and $Y(0)$ will not have a closed form distribution. We compute target functionals using median values from 100 simulations of size 100,000 for the "unfeasible estimator", which is presented below.

The parameters were chosen to be $\mu_{X_1}=1$, $\mu_{X_2}=5$ and those in the table below.[21]

Table 1: Parameter specification for
Monte Carlo Exercise

$coeff.\backslash j$	0	1	2	3	4	5
δ_j	-0.5	1.35	-0.2	0.15	-0.1	0.5
β_{0j}	0.01	-0.01	0.01	0.01	-0.01	-0.02
β_{1j}	0.1	0.01	0.01	0.01	0.01	0.01
β_{0j}^s	0.01	-0.01	0.01	0.01	-0.01	-0.02
β_{1j}^s	0.01	0.01	0.01	0.01	0.01	0.01

We will compute inequality treatment effects on the treated. For that purpose, it is important to have the values of some functionals of the distributions of potential outcomes for the treated. These are listed below.

[21] Under this parameter specification, the propensity score satisfies the common support assumption. For example, in the logistic case, $p(X)$ attains values between 0.80 and 0.15. In all designs, $p=0.5$.

Table 2: Features of the Distributions of Potential Outcomes for the Treated (conditional on $T = 1$), scaled by 100.

η	Logistic		Normal	
ν\Distribution	$Y(0)$	$Y(1)$	$Y(0)$	$Y(1)$
Mean	0.7787	1.8871	0.7802	1.8831
Standard Deviation (s.d.)	0.2751	1.1730	0.2737	1.1644
Mean of Logarithm	−0.3155	0.5110	−0.3129	0.5098
S.D. of Logarithm	0.3780	0.4714	0.3760	0.4697
10th Percentile	0.4449	0.9840	0.4477	0.9849
1st Quartile	0.6022	1.2293	0.6046	1.2293
Median	0.7749	1.5834	0.7773	1.5811
3rd Quartile	0.9312	2.1633	0.9325	2.1586
90th Percentile	1.0749	3.0708	1.0746	3.0601

Table 3: Inequality Measures of Potential Outcomes for the Treated (conditional on $T = 1$)

η	Logistic		Normal	
ν\Distribution	$Y(0)$	$Y(1)$	$Y(0)$	$Y(1)$
Coefficient of Variation	0.3532	0.6216	0.3508	0.6183
Interquartile Range	0.9340	0.3290	0.9293	0.3279
Theil Index	0.0608	0.1401	0.0601	0.1390
Gini Coefficient	0.1889	0.2755	0.1877	0.2745

Naked eye inspection of Tables 2 and 3 reveal that target functionals are little affected by the distribution underlying the selection model. Thus, consistent semiparametric estimators of functionals of these distributions should not be affected by the nature of the d.g.p.. In Tables 4 and 5 we present results for the d.g.p. based on the normal specification. Tables A.2 and A.3 with logistic are left to the Monte Carlo Supplemental Appendix.[22]

We provide in Tables 4 and 5 results for the unfeasible estimator and for five feasible estimators. The first one is the estimator proposed here and labeled "weighted estimator". In order to simplify the estimation procedure, we have considered a parametric first step. In this step, we computed the propensity-score by a logit using the correct quadratic specification and a logit using a misspecified linear model.[23]

The second estimator is the one based on the empirical distributions of $Y|T = 1$ and $Y|T = 0$. We call that estimator the "naive estimator". Given that there is selection into treatment based on observables, the naive estimator will be inconsistent for the ITE parameters.

We then consider what we call the "location shift estimator". This is constructed in the following way. We first run two linear regressions (with intercept) of Y on X_1, X_2, X_1^2, X_2^2 and X_1X_2 , one for each group ($T = 0$ and $T = 1$). Save residuals $\widehat{u}_j$ and compute S_j^2 where $j = 0, 1$ indexes treatment groups, $S_j^2 = (N_j - 6)^{-1} \sum_i^{N_j} (\widehat{u}_{ji})^2$ and $N_1 = \sum_i^N T_i$ and $N_0 = N - N_1$.

[22] Tables of results of Monte Carlo experiments with 1,000 observations are not presented as results are very similar to those with sample size 4,000 presented in the paper.

[23] In the tables, 'weighted' corresponds to the estimator whose first stage uses a quadratic specification for the propensity score and 'weighted-linear' corresponds to the estimator whose first stage uses a linear specification for the propensity score. The same corresponding notation applies to other estimators.

Save coefficient estimates for group $T = 0$, $\widehat{\gamma}_{00}$, $\widehat{\gamma}_{10}$, $\widehat{\gamma}_{20}$, $\widehat{\gamma}_{30}$, $\widehat{\gamma}_{40}$, $\widehat{\gamma}_{50}$. Then, let Y_i^* be counterfactual outcome of treated observation i:

$$Y_i^* = \widehat{\gamma}_{00} + \widehat{\gamma}_{10} X_{1i} + \widehat{\gamma}_{20} X_{2i} + \widehat{\gamma}_{30} X_{1i}^2 + \widehat{\gamma}_{40} X_{2i}^2 + \widehat{\gamma}_{50} X_{1i} X_{2i} + \sqrt{S_0^2 / S_1^2} \cdot \widehat{u}_{1i}$$

and since Y_i^* is well defined for all treated i, we compute the inequality measures for two distributions: $Y|T = 1$ and $Y^*|T = 1$. From the empirical $Y|T = 1$ we estimate functionals of $Y(1)|T = 1$, whereas with $Y^*|T = 1$ we estimate functionals of $Y(0)|T = 1$. Note that this is a way of "controlling" for covariates. We also consider "location shift estimator" in which in the first step we misspecify the conditional distribution function of $Y(1)$ and $Y(0)$ by running linear regressions (with intercept) of Y on X_1, X_2 for each group.

By noticing that Y is distributed over the positive real numbers, an alternative way to implement the same idea is to take the logarithm first. We call this estimator "log-location shift estimator". We proceed by following the same steps for the location shift estimator. The difference is that we apply the logarithm on Y first. Then, after we complete all steps described for the location shift estimator, we exponentiate the counterfactual logarithm of the outcome. By proceeding this way we guarantee that the counterfactual outcome will always be defined over the positive reals, something that we cannot guarantee for the location-shift estimator.[24] Location shift estimators correspond indeed to the procedure proposed by Juhn, Murphy and Pierce (1993).

Our final estimator is the one proposed by Chernozhukov, Fernandez-Val and Melly (2009). We estimate the conditional distribution function of $Y|X, T = 0$ by using logit estimators. To be more precise, let $D_y = \mathbb{1}\{Y \leq y\}$ and $F_{Y|T,X}(y|0, x) = \Pr[D_y = 1|T = 0, X = x]$. For fixed y, the conditional probability can be estimated by a flexible logit. Exactly as we did for all other estimators, we consider two situations. In the first one, for each y we use the full quadratic model in the logit, and in the second one for each y we use a linear model in the logit. The number of points y considered was dependent on the sample size. For $N = 250$, we used 100 points; for $N = 1,000$, we used 500 points; and for $N = 4,000$, we used $1,000$ points from the support of Y. Once we have an estimate of the conditional c.d.f. of $Y|T = 0, X$, we can integrate it using the empirical distribution of $X|T = 1$. We call that estimator the "CFM estimator".

Results in Tables 4, 5, A.2 and A.3 show distribution features for each one of the estimators of inequality treatment effect parameters. We report average, standard deviation and quantiles (10^{th} percentile, median, and 90^{th} percentile) for the four types of treatment effects on inequality measures here considered (coefficient of variation, interquartile range, Theil index and Gini coefficient) in the case when we use the correct form of the propensity score or the conditional expectation (quadratic) and in the case when we misspecify by using a linear function. Besides those inequality treatment effects, we also report results for average treatment effects. Finally, we present results that compare the estimates with the population target. Those are reported by the bias, root mean squared error, mean absolute error and the coverage rate of 90% confidence intervals.

In Tables 4 and 5 we present results using a normal d.g.p. for the latent variable in the selection model. Among all three d.g.p.s, this is the least favorable one to the weighted esti-

[24] In fact, in most Monte Carlo replications we obtained negative counterfactual outcomes for the location shift estimator for very few observations. Typically, in $1,000$ replications, around 900 had at least one negative counterfactual and among these replications only less than 10 percent had more than one negative value. Interestingly, we obtained more negative counterfactual outcomes for $N = 250$ than for $N = 4,000$.

mator, which has been constructed using a fixed polynomial model for the logit. The results in Tables 4 and 5 point out, however, that the weighted estimator is a competitive estimator for distributional impacts, when compared to a more elaborated and computationally more demanding estimator as the CFM estimator. The weighted estimator performs well according to the MSE criteria and its variance shrinks as expected as the sample size increases. In addition, the weighted estimator is competitive to the CFM estimator even when we misspecify the propensity score. Relatively to other estimators also being analyzed, when we look at the bias criteria, the weighted estimator dominates the naive estimator and the two estimators based on Juhn, Murphy and Pierce (1993).[25]

7 Empirical Application

The empirical application is on a Brazilian public-sponsored job training program, also known as PLANFOR (*Plano Nacional de Qualificação Profissional*). That program, which started in 1996, has provided classroom training for the formation of the basic skills necessary for certain occupations (e.g. waiters, hairdressers, administrative jobs). The program operates on a continuous basis throughout the year, with new cohorts of participants starting every month. Although funding comes from the federal government, the program was decentralized at the State level. Each state subcontracted for classroom training with vocational proprietary schools and local community colleges. The target population consists of disadvantaged workers, who have been defined as the unemployed, and individuals with low level of schooling and/or income. Enrollment of individuals in the program is voluntary, but its scale in 1998 was relatively small, being around 1.5% of the labor force in all metropolitan areas in Brazil.

The evaluation of PLANFOR involved the first attempt in the country to perform a randomized study designed to measure impacts of a social program. In the years of 1998-99, the Brazilian Ministry of Labor financed an experimental evaluation of the program impact on earnings and employment.[26] Experimental data were collected in two metropolitan areas of the country, namely Rio de Janeiro and Fortaleza. The process of randomization of individuals in and out of the program was performed at the class level and took place in August 1998, and almost all individuals that were selected in attended the training courses in September 1998. In that month, participants in both cities were interviewed through the application of the same questionnaire, and retrospective questions were asked about their labor market history. A follow-up survey took place in November 1999, and retrospective questions were asked going back to September 1998.

The total available sample size from the baseline interview was 5,249 individuals. Given that randomization was performed at class level, for the sake of our analysis we dropped all classes with either only one treated or one control unit, remaining with 5,222 individuals, out of which 2,616 were from Rio de Janeiro. They were distributed in along 237 classes (74 in Rio) that had a median size of 18 students.

[25] The Monte Carlo supplemental appendix available at authors' webpage presents a design in which the propensity score attains values very close to 0 and 1. The results of this design are very similar to the ones presented in the main paper. The weighted estimator is still competitive wiht the CFM estimator and dominates the other estimators. As we expected, the weighted estimator is more sensitive to misspecification of the propensity score in this design in which the common support assumption is almost violated.

[26] This data set has also been used by Foguel (2006), in which further details on the impact evaluation study can be found.

Because of the stratified randomization, we have by design that, conditional on the class (stratum), treatment status is independent of potential outcomes. Thus we can infer causality by applying the proposed method discussed in this paper using class dummies as confounding variables.

We first check whether randomization was properly performed. Because randomization occurred within class, we check whether randomization was well performed in each class through t-tests of differences in means between treated and control groups. We have decided to drop classes that, for at least in one covariate, presented imbalances detected by t-tests at 1% significance level. After we apply that filter 17 classes were dropped (5 in Rio), and we remained with 4,864 observations (2,469 in Rio) out of which 2,298 in the treated group (1,258 from Rio) and 2,566 in the control group (1,211 from Rio).[27]

A summary statistics table, Table 6, shows that for some covariates there are statistically significant differences in means between treated and control groups for pooled data. We therefore applied the weighting function ω_{01} to the control group to recover a counterfactual distribution of covariates that would have prevailed if the control group were distributed across classes exactly as the treated group. By doing so, we expect to 'undo' the problem induced by having different proportions of treated units across classes. Table 6 shows that after applying weights differences by treatment status become non-significant. We interpret this as evidence that randomization was well performed at the class level.

A few interesting features emerge from Table 6, revealing that the target population in those two sites, Rio de Janeiro and Fortaleza are intrinsically different: People in our Fortaleza sample are older (average age of 27 years old) than people in our Rio de Janeiro sample, which consists basically of teenagers/young adults (average age of 18 years old). Average schooling is about 8 years in Rio and 9 in Fortaleza, perhaps reflecting age differences between sites. For the same reason it is no surprise that in Rio de Janeiro, about 50% of sample had worked before, whereas in Fortaleza that number was around 80%.

Using the follow-up survey we constructed two outcome variables: The hourly wage rate at the first job in the 12-month interval after treatment period; and the sum of all monthly salaries received during 12 months after treatment period. Since hourly wage rate at the first job is only well defined for those who obtained a job, a condition that may have been affected by treatment itself, we also consider the second outcome variable, the sum of all earnings. In the construction of that variable we did not drop individuals who remained unemployed after the 12 month period after treatment; instead we assigned them zero earnings.

Our sample size decreased from the baseline to the follow-up stage to 3,783 individuals (2,071 in Rio), 1,884 belonging to the treated group (1,106 in Rio) and 1,899 to the control group (965 from Rio). We checked whether attrition could be explained by treatment status but found no statistical evidence supporting it.[28] We also had a sample size reduction when

[27] Although we found imbalances in 17 out of 237 classes (7.2% of classes), that is not necessarily evidence that controlling for class dummies is insufficient to remove bias in this case. We call attention for the fact that our criteria of dropping classes that presented detectable imbalances at 1% level was applied for 8 covariates. There were no classes that presented more than one unbalanced covariate at that significance level. Thus, we performed 8*237=1,896 tests and rejected the null 17 out of 1,896 tests, that is, we rejected the null at 1% significance level in 0.9% of the tests. Finally, for robustness we present a summary statistics table (Table A.1) in the Appendix with the sample before we dropped classes removed on the t-tests criteria. The features of data remain almost identical after we dropped the 17 classes.

[28] We ran a regression of an indicator of missingness on the treatment dummy, class dummies and interactions between treatment and class dummies; and obtained a non-significant at 5% coefficient for the treatment dummy. We interpret this as evidence that there was no within-class differential attrition between treated and controls.

using the variable hourly wage rate at the first job because that variable is defined only for those who were obtained a job after the program.[29]

In Table 7 we report average and inequality treatment effects estimates. We report point estimates and bootstrapped standard errors (100 replications) for all five feasible estimators that have been defined in the Monte Carlo section. For all of them, except for the naive estimator, we use as controlling or confounding variables classroom dummies. It turns out that because we use a fully saturated model for the propensity-score, first stage of our weighted estimator will be nonparametric. In order to have comparability with all other estimators we also used class dummies as regressors in all of them. For CFM estimator, given the sample sizes in the control group, we estimated the conditional c.d.f. for 100 points.

A problem that may emerge with the location shift estimator is that it might create negative earnings, as predicted values from the linear regression are not necessarily bounded above zero. Having a variable with negative values creates an asymmetry between that estimator and other estimators since some inequality measures are defined only for positive values. We do not attempt to make samples comparable, and interpret that asymmetry as another source of bias for the location shift estimator.

Results are that on average the program is either ineffective or has a small negative effect on earnings. However, we see that for Rio de Janeiro, although the program does seem to induce no average gains, it does reduce inequality among treated according, for example, to the weighted estimator applied to the first hourly wage. One possible interpretation is that the program reduces signaling costs, allowing employers to set similar wages for entering workers that have program certificates. For the sum of all earnings results are less clear. In most part of the results, we obtain no difference in inequality. For Fortaleza, results using weighted estimator are that the program is ineffective in reducing inequality. Finally, log-linear location shift, linear location shift and naive estimators detected significant positive effects. However, as we know, from Monte Carlo exercises, results coming from those estimators are in general biased ones.

8 Conclusion

In this paper we proposed a method that may be useful for applied researchers that are interested in comparing inequality measures of two or more outcome distributions. When comparing Gini coefficients between two groups (for example, treated and non-treated groups), it is important to acknowledge for the fact that there are many observed factors whose distributions differ across groups. Our method allows applied researchers to identify the impact of the treatment through properly weighted differences in Gini coefficients between these two groups of workers. That allows decomposing differences in Gini coefficients into two parts: one that fixes the distribution of covariates and under a selection on observables assumption captures the impact of the treatment on the Gini coefficient; and another that is a composition effect, induced by different distribution in covariates.

In fact, our estimation strategy is mostly useful when the individual decision to participate in the program (the treatment) depends on observable characteristics. If such identification restriction holds, then the reweighing method allows identifying the distribution of potential

[29] For that variable, we have 2,443 non-missing observations for the combined sample, 1,241 in the treated group. In Rio de Janeiro we have 577 treated and 456 control units.

outcomes and, therefore, the causal impact of the program on many functionals of interest for policy analysis, such as inequality indices.

Results from our Monte Carlo study suggest that, when looking at distributional aspects, some alternatives to the weighted estimator would require a very flexible estimation of the conditional distribution of Y given covariates to be competitive. Because the weighted estimator only requires estimation of a single conditional expectation, it is certainly a readily implementable alternative; which clearly contrasts to nonparametric estimators of the conditional distribution. Finally, in addition to its computational simplicity the weighted method also has desired large sample properties and behave relatively well in small samples.

Possible ways to extend the work presented would be to characterize semiparametric estimation of inequality treatment effects using alternative efficient estimators. A natural alternative procedure is the "efficient influence function estimator", also known as the "double robust estimator" after Scharfstein, Rotnitzky and Robins (1999), that was recently proposed by Cattaneo (2009) for the multivalued case in the GMM context. Although such estimator may not be as simple to compute as the weighted estimator, it may represent an interesting mixture, between the estimator here presented and those proposed by Chernozhukov, Fernandez-Val and Melly (2009) and Rothe (2010).

APPENDIX

Proof of Lemma 1: A proof for this Lemma would have two parts. The first part, concerned with identification of marginal (and conditional on $T = 1$) cdf's of potential outcomes is omitted as it follows trivially from Firpo (2007), Lemma 1. Second part follows by own definition of ITE parameters: They are differences in functionals of these marginal (conditional on $T = 1$) identified distributions and therefore they can be expressed as functions of the observable data (Y,X,T).

Q.E.D.

Proof of Lemma 2: We fix notation to what had been used by Hirano, Imbens and Ridder (2003), so our estimator of the propensity score is $\widehat{p}(\cdot) = \widehat{p}_K(\cdot) \equiv \Lambda\left(H_K(\cdot)^\top \widehat{\pi}_K\right)$. In fact, we set notation as in HIR to simplify the comparison of the decomposition presented below with results presented in their appendix. We break down the difference $\sqrt{N}\left(\widehat{F}_Y^{\widehat{\omega}}(y) - F_Y^{\omega}(y)\right)$ into six components:

$$
\begin{aligned}
&\sqrt{N}\left(\widehat{F}_Y^{\widehat{\omega}}(y) - F_Y^{\omega}(y)\right) \\
= \; & \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\omega\left(T_i, \widehat{p}_K(X_i)\right)\mathbb{I}\{Y_i \le y\} - \omega\left(T_i, p(X_i)\right)\mathbb{I}\{Y_i \le y\}\right. && \text{(A-1)} \\
&\left. + \frac{\partial\omega\left(T_i, p(X_i)\right)}{\partial p(X_i)}\mathbb{I}\{Y_i \le y\}\left(\widehat{p}_K(X_i) - p(X_i)\right)\right) \\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(-\frac{\partial\omega\left(T_i, p(X_i)\right)}{\partial p(X_i)}\mathbb{I}\{Y_i \le y\}\left(\widehat{p}_K(X_i) - p(X_i)\right)\right. \\
&\left. + \int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T, p(x)\right)}{\partial p(x)}\cdot\mathbb{I}\{Y \le y\}\middle| x\right]\left(\widehat{p}_K(x) - p(x)\right)dF(x)\right) && \text{(A-2)} \\
&-\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T, p(x)\right)}{\partial p(x)}\cdot\mathbb{I}\{Y \le y\}\middle| x\right]\left(\widehat{p}_K(x) - p(x)\right)dF(x) \\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{\Psi}_K(X_i)\left(\frac{T_i - p_K(X_i)}{\sqrt{p_K(X_i)\left(1 - p_K(X_i)\right)}}\right) && \text{(A-3)} \\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\widetilde{\Psi}_K(X_i) - \Psi_K(X_i)\right)\left(\frac{T_i - p_K(X_i)}{\sqrt{p_K(X_i)\left(1 - p_K(X_i)\right)}}\right) && \text{(A-4)} \\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\Psi_K(X_i)\left(\frac{T_i - p_K(X_i)}{\sqrt{p_K(X_i)\left(1 - p_K(X_i)\right)}}\right) - \Psi_0(X_i)\left(\frac{T_i - p(X_i)}{\sqrt{p(X_i)\left(1 - p(X_i)\right)}}\right)\right) && \text{(A-5)} \\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\omega\left(T_i, p(X_i)\right)\cdot\mathbb{I}\{Y_i \le y\} - F_Y^{\omega}(y)\right) + \Psi_0(X_i)\left(\frac{T_i - p(X_i)}{\sqrt{p(X_i)\left(1 - p(X_i)\right)}}\right) && \text{(A-6)}
\end{aligned}
$$

with

$$\widetilde{\Psi}_K(x) = -\int_{\mathcal{X}} \mathbb{E}\left[\frac{\partial\omega(T_i, p(z))}{\partial p(z)} \cdot \mathbb{I}\{Y_i \leq y\} \middle| z\right] \Lambda'\left(H_K(z)^\top \widetilde{\pi}_K\right) H_K(z)^\top dF(z) \tag{A-7}$$
$$\cdot\, \widetilde{\Sigma}_K^{-1} \sqrt{\Lambda'\left(H_K(x)^\top \pi_K\right)} H_K(x)$$

$$\Psi_K(x) = -\int_{\mathcal{X}} \mathbb{E}\left[\frac{\partial\omega(T_i, p(z))}{\partial p(z)} \cdot \mathbb{I}\{Y_i \leq y\} \middle| z\right] \Lambda'\left(H_K(z)^\top \pi_K\right) H_K(z)^\top dF(z) \tag{A-8}$$
$$\cdot\, \Sigma_K^{-1} \sqrt{\Lambda'\left(H_K(x)^\top \pi_K\right)} H_K(x)$$

$$\Psi_0(X_i) = -\mathbb{E}\left[\frac{\partial\omega(T_i, p(x))}{\partial p(x)} \cdot \mathbb{I}\{Y_i \leq y\} \middle| x\right] \sqrt{p(X_i)(1-p(X_i))} \tag{A-9}$$

and

$$\Sigma_K = \mathbb{E}\left[\Lambda'\left(H_K(X)^\top \pi_K\right) H_K(X) H_K(X)^\top\right]$$
$$\widetilde{\Sigma}_K = \frac{1}{N}\sum_{i=1}^N \Lambda'\left(H_K(X_i)^\top \widetilde{\pi}_K\right) H_K(X_i) H_K(X_i)^\top,$$

where $\widetilde{\pi}_K$ lies between $\widehat{\pi}_K$ and π_K. Now we show that each term can be bounded uniformly in y.

$$\sup_{y\in\mathcal{Y}}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N \left(\omega(T_i, \widehat{p}_K(X_i))\mathbb{I}\{Y_i \leq y\} - \omega(T_i, p(X_i))\mathbb{I}\{Y_i \leq y\}\right.\right.$$
$$\left.\left. + \frac{\partial\omega(T_i, p(X_i))}{\partial p(X_i)}\mathbb{I}\{Y_i \leq y\}(\widehat{p}_K(X_i) - p(X_i))\right)\right|$$
$$= O_p\left(\frac{\zeta(K(N))^3}{\sqrt{N}}\right) + O_p\left(\sqrt{N}\zeta(K(N))^2 K^{-\frac{sp}{r}}\right) + O_p\left(\zeta(K(N))^{\frac{5}{2}} K^{-\frac{sp}{2r}}\right)$$

$$\sup_{y\in\mathcal{Y}}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N \left(-\frac{\partial\omega(T_i, p(X_i))}{\partial p(X_i)}\mathbb{I}\{Y_i \leq y\}(\widehat{p}_K(X_i) - p(X_i))\right.\right.$$
$$\left.\left. + \int_{\mathcal{X}} \mathbb{E}\left[\frac{\partial\omega(T, p(x))}{\partial p(x)} \cdot \mathbb{I}\{Y \leq y\} \middle| x\right](\widehat{p}_K(x) - p(x))\, dF(x)\right)\right|$$
$$= O_p\left(\zeta(K(N)) K^{-\frac{sp}{2r}}\right) + O_p\left(\frac{\zeta(K(N))^2}{\sqrt{N}}\right)$$

$$\sup_{y\in\mathcal{Y}}\left|-\sqrt{N}\int_{\mathcal{X}} \mathbb{E}\left[\frac{\partial\omega(T, p(x))}{\partial p(x)}\mathbb{I}\{Y \leq y\} \middle| x\right](\widehat{p}_K(x) - p(x))\, dF(x)\right.$$
$$\left. + \frac{1}{\sqrt{N}}\sum_{i=1}^N \widetilde{\Psi}_K(X_i) \cdot \frac{(T_i - p_K(X_i))}{\sqrt{p_K(X_i)(1 - p_K(X_i))}}\right|$$
$$= O\left(\sqrt{N}\zeta(K(N)) K^{-\frac{sp}{2r}}\right).$$

$$\sup_{y\in\mathcal{Y}}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\widetilde{\Psi}_K(X_i)-\Psi_K(X_i)\right)\left(\frac{T_i-p_K(X_i)}{\sqrt{p_K(X_i)(1-p_K(X_i))}}\right)\right|=O_p\left(\frac{\zeta(K(N))^{\frac{11}{2}}}{\sqrt{N}}\right)$$

$$\sup_{y\in\mathcal{Y}}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\Psi_K(X_i)\left(\frac{T_i-p_K(X_i)}{\sqrt{p_K(X_i)(1-p_K(X_i))}}\right)-\Psi_0(X_i)\left(\frac{T_i-p(X_i)}{\sqrt{p(X_i)(1-p(X_i))}}\right)\right)\right|$$
$$=O_p\left(\max\left(K^{-\frac{t}{2r}},\zeta(K(N))K^{-\frac{s_p}{2r}}\right)\right).$$

Combining the bounds, we have

$$\sup_{y\in\mathcal{Y}}\left|\sqrt{N}\left(\widehat{F}_Y^{\widehat{\omega}}(y)-F_Y^{\omega}(y)\right)\right.$$
$$\left.-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{(\omega(T_i,p(X_i))\cdot\mathbb{I}\{Y_i\leq y\}-F_Y^{\omega}(y))-\Psi_0(X_i)\left(\frac{T_i-p(X_i)}{\sqrt{p(X_i)(1-p(X_i))}}\right)\right\}\right|$$
$$=O_p\left(\sqrt{N}\zeta(K(N))^2K^{-\frac{s_p}{r}}\right)+O_p\left(\zeta(K(N))^{\frac{5}{2}}K^{-\frac{s_p}{2r}}\right)+O_p\left(\frac{\zeta(K(N))^{\frac{11}{2}}}{\sqrt{N}}\right).$$

And under the assumptions on $\zeta(K(N))$ and s_p, this sum is $o_p(1)$.

Q.E.D.

Proof of Theorem 1: The proof is divided in two parts. First we show that the collection of measurable functions from $(\mathcal{Y}\times\mathcal{X}\times\{0,1\})\to\mathbb{R}$, $\mathcal{H}=\{\psi(Y,X,T,y)|\,y\in\mathcal{Y}\}$ is P-Donsker using the following steps:

(i) The measurable collection of functions $\mathcal{W}=\{\mathbb{I}\{Y\leq y\}|\,y\in\mathcal{Y}\}$ is Donsker since the bracketing number $N_{[]}(\sqrt{\varepsilon},\mathcal{W},L_2(P))\leq\frac{2}{\varepsilon}$ are of the polynomial order $\left(\frac{1}{\varepsilon}\right)^2$. The bracketing integral is finite since it converges at a slower rate than $\left(\frac{1}{\varepsilon}\right)^2$.

(ii) Define the measurable functions $\mathcal{K}=\{F_Y(y|\,X)|\,y\in\mathcal{Y}\}$. Using the proof of Lemma A2 in Donald and Hsu (2010), $\mathcal{K}$ is Donsker. In that proof, Donald and Hsu (2010) show that $N_{[]}(\varepsilon,\mathcal{K},L_2(P))\leq 1+\left(\frac{1}{\varepsilon}\right)^2$, and the bracketing integral is finite since it converges at a slower rate than $\left(\frac{1}{\varepsilon}\right)^2$.

(iii) Since $p(x)$ is bounded away from zero and one in $\mathcal{X}$, $d_1(T,X)\equiv\omega(T,p(X))$ is a uniformly bounded measurable function, and $d_1(T,X)\cdot\mathcal{W}$ is Donsker. Similarly, define $d_2(T,X)\equiv\mathbb{E}\left[\left.\frac{\partial\omega(T,p(X))}{\partial p(X)}\right|X\right](T-p(X))$, $d_2(T,X)$ is a uniformly bounded measurable function, and $d_2(T,X)\cdot\mathcal{K}$ is Donsker.

(iv) $\mathcal{H}=\{d_1(T,X)\cdot\mathcal{W}+d_2(T,X)\cdot\mathcal{K}|\,y\in\mathcal{Y}\}$ is Donsker.

For the second part of the proof, we let $\mathbb{P}_N$ be the empirical measure of the sample (Y,X,T). Using the fact that $\mathcal{H}$ is P-Donsker and Donsker's Theorem (Theorem 19.3 at page 266 of Van der Vaart,1998), $\sqrt{N}\begin{pmatrix}\widehat{F}_Y^{\widehat{\omega}_A}-F_Y^{\omega_A}\\ \widehat{F}_Y^{\widehat{\omega}_B}-F_Y^{\omega_B}\end{pmatrix}$ converges to a zero mean Gaussian process, defined by $\mathbb{G}^{\omega_{A,B}}$, with variance-covariance matrix given by

$$\mathbb{E}\left[\mathbb{G}^{\omega_j}(s)\,\mathbb{G}^{\omega_k}(t)\right]=\mathbb{E}\left[\left(\psi_j(Y,X,T,s)-F_Y^{\omega_j}(s)\right)\cdot\left(\psi_k(Y,X,T,t)-F_Y^{\omega_k}(t)\right)\right]$$

for $(s,t)\in\mathcal{Y}\times\mathcal{Y}$, j and $k=A,B$.

Q.E.D.

Proof of Theorem 2: The proof is divided into three parts. In the first part (Part I), we fix $j = A, B$ and derive the asymptotic distribution of $\nu\left(\widehat{F}_Y^{\widehat{\omega}}\right)$. In the second part (Part II), we show that for a fixed y, $\widehat{F}_Y^{\widehat{\omega}}(y)$ is efficient for $F_Y^{\omega}(y)$ and, therefore, $\nu\left(\widehat{F}_Y^{\widehat{\omega}}\right)$ will be efficient since ν is Hadamard. Finally (in Part III), we derive the asymptotic distribution of the difference $\nu\left(\widehat{F}_Y^{\widehat{\omega}_A}\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_B}\right)$ and show that it is efficient.

Part I. We fix $j = A, B$ and therefore drop the subscript. Using results in Theorem 1,

$$\sqrt{N}\left(\widehat{F}_Y^{\widehat{\omega}} - F_Y^{\omega}\right) = \mathbb{G}_N^{\omega} + o_p(1) \Rightarrow \mathbb{G}^{\omega}$$

where $\mathbb{G}^{\omega}$ is a Gaussian process with variance-covariance matrix given by

$$\mathbb{E}\left[\mathbb{G}^{\omega}\mathbb{G}^{\omega}\right](s,t) = \mathbb{E}\left[\left(\psi(Y,X,T,s) - F_Y^{\omega}(s)\right)\cdot\left(\psi(Y,X,T,t) - F_Y^{\omega}(t)\right)\right]$$

for $(s,t) \in \mathcal{Y}\times\mathcal{Y}$. Because the map $\nu : \mathcal{F}_\nu \to \mathbb{R}$ is Hadamard differentiable at $F_Y^{\omega} \in \mathcal{F}_\nu$ we can apply van der Vaart's (1998) Theorem 20.8:

$$\sqrt{N}\left(\nu\left(\widehat{F}_Y^{\widehat{\omega}}\right) - \nu\left(F_Y^{\omega}\right)\right) = \psi^{\nu}\left(\mathbb{G}_N^{\omega}; F_Y^{\omega}\right) + o_p(1).$$

And since ν is Hadamard differentiable, its functional derivative $\psi^{\nu}\left(\cdot; F_Y^{\omega}\right)$ is linear, implying that

$$\psi^{\nu}\left(\mathbb{G}_N^{\omega}; F_Y^{\omega}\right) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega\left(T_i, p(X_i)\right)\cdot\phi^{\nu}\left(Y_i; F_Y^{\omega}\right) + \mathbb{E}\left[\frac{\partial\omega\left(T, p(X_i)\right)}{\partial p(X_i)}\cdot\phi^{\nu}\left(Y; F_Y^{\omega}\right)\middle| X_i\right]\left(T_i - p(X_i)\right).$$

Part II. Consider a (regular) parametric submodel of the joint distribution of (Y, X, T) with cdf $F(y,x,t;\theta)$. The log-likelihood is

$$\ln f(y,x,t|\theta) = t\ln f_1(y|x,\theta) + (1-t)\ln f_0(y|x,\theta) + t\ln p(x|\theta) + (1-t)\ln\left(1 - p(x|\theta)\right) + \ln f(x|\theta)$$

where for $j = 0, 1$ we use the ignorability assumption to write $f(y|x, T = j;\theta)$ as $f_j(y|x,\theta)$, which is the conditional density of $Y(j)$ given x for parameter value θ. Following results in Hahn (1998), we have that the corresponding score function:

$$S(y,x,t|\theta) = ts_1(y|x;\theta) + (1-t)s_0(y|x;\theta) + \frac{dp(x|\theta)}{d\theta}\frac{\left(t - p(x|\theta)\right)}{p(x|\theta)\left(1 - p(x|\theta)\right)} + s(x|\theta)$$

where for $j = 0, 1$, $s_j(y|x;\theta) \equiv d\ln f_j(y|x;\theta)/d\theta$ and $s(x|\theta) \equiv d\ln f(x|\theta)/d\theta$. The tangent space for this model is:

$$\mathcal{L} = \left\{S(y,x,t) : S(y,x,t) = ts_1(y|x) + (1-t)s_0(y|x) + a(x)\left(t - p(x)\right) + s(x)\right\}$$

where $a(x)$ is a square-integrable function of x,

$$\int s_j(y|x) f_j(y|x;\theta_0) = 0,\ \forall x,\ j = 0,1, \qquad \int s(x|\theta) f(x|\theta_0) = 0,$$

and for notational simplicity, $p(\cdot|\theta_0) = p(\cdot)$. The derivative of $F_Y^{\omega}(y;\theta)$ with respect to θ evaluated at θ_0 is

$$\begin{aligned}
\left.\frac{\partial F_Y^{\omega}(y;\theta)}{\partial\theta}\right|_{\theta=\theta_0} &= \mathbb{E}\left[\omega(T,p(X))\,\mathbb{1}\{Y\le y\}S(Y,X,T|\theta_0)\right] \\
&\quad +\mathbb{E}\left[\mathbb{E}\left[\left.\frac{\partial\omega(T,p(X))}{\partial p(X)}\cdot\mathbb{1}\{Y\le y\}\right| X\right](T-p(X))\,S(Y,X,T|\theta_0)\right] \\
&= \mathbb{E}\left[\left(\psi(Y,X,T,y)-F_Y^{\omega}(y;\theta_0)\right)S(Y,X,T|\theta_0)\right]
\end{aligned}$$

and a quick inspection of $\psi(Y,X,T,y) - F_Y^{\omega}(y;\theta_0)$ yields that it is in $\mathcal{L}$. According to Bickel, Klaassen, Ritov and Wellner (1993), because the estimator for $F_Y^{\omega}(y)$, $\widehat{F}_Y^{\widehat{\omega}}(y)$, is asymptotically linear with influence function $\psi(Y,X,T,y) - F_Y^{\omega}(y;\theta_0) \in \mathcal{L}$, $\widehat{F}_Y^{\widehat{\omega}}(y)$ is also asymptotically efficient at $\mathcal{P}$, the joint distribution of (Y,X,T). Therefore, since ν is Hadamard differentiable, by theorem 25.47 of van der Vaart (1998) $\nu\left(\widehat{F}_Y^{\widehat{\omega}}\right)$ is asymptotically efficient at $\mathcal{P}$ for estimating $\nu\left(F_Y^{\omega}\right)$.

Part III. A mechanical application of Part I gives us that

$$\begin{aligned}
&\sqrt{N}\left(\nu\left(\widehat{F}_Y^{\widehat{\omega}_A}\right)-\nu\left(\widehat{F}_Y^{\widehat{\omega}_B}\right)-\left(\nu\left(F_Y^{\omega_A}\right)-\nu\left(F_Y^{\omega_B}\right)\right)\right) \\
= \;&\sqrt{N}\left(\nu\left(\widehat{F}_Y^{\widehat{\omega}_A}\right)-\nu\left(F_Y^{\omega_A}\right)-\left(\nu\left(\widehat{F}_Y^{\widehat{\omega}_B}\right)-\nu\left(F_Y^{\omega_B}\right)\right)\right) \\
= \;&\psi^{\nu}\left(\mathbb{G}_N^{\omega 1l};F_Y^{\omega 1l}\right)-\psi^{\nu}\left(\mathbb{G}_N^{\omega 0l};F_Y^{\omega 0l}\right)+o_p(1).
\end{aligned}$$

Now, for efficiency, let us define the functional vector $\boldsymbol{\nu}:\mathcal{F}_v\to\mathbb{R}^2$. Thus,

$$\nu\left(\widehat{F}_Y^{\widehat{\omega}_A}\right)-\nu\left(\widehat{F}_Y^{\widehat{\omega}_B}\right)=\begin{bmatrix}1\\-1\end{bmatrix}^{\top}\boldsymbol{\nu}\left(\begin{bmatrix}\widehat{F}_Y^{\widehat{\omega}_A}\\ \widehat{F}_Y^{\widehat{\omega}_B}\end{bmatrix}\right),$$

and because $[1,-1]\,\boldsymbol{\nu}$ is Hadamard we can apply results from Part II.

Q.E.D.

Proof of Theorem 3: We divide the proof into two parts. We first show that for $j = A, B$, $\sup_{x\in\mathcal{X}}\left|g_j(x) - H_{K_j}(x)^{\top}\widetilde{\gamma}_K^{\omega_j}\right| = o_p(1)$. Then we show that since all nonparametric components of the variance estimator converges uniformly in probability to their population counterparts, the variance estimator will be consistent for the asymptotic variance.

Fix $j = A, B$ and drop the subscript. Define $\widetilde{\gamma}_K^{\omega}$, $\overline{\gamma}_K^{\omega}$ and γ_K^{ω}

$$\widetilde{\gamma}_K^{\omega}=\arg\min_{\gamma}\sum_{i=1}^{N}\left(\left.\frac{\partial\omega(T_i,p(X))}{\partial p(X)}\right|_{p(X)=p(X_i)}\cdot\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)-H_K(X_i)^{\top}\gamma\right)^2$$

$$\overline{\gamma}_K^{\omega}=\arg\min_{\gamma}\mathbb{E}\left[\left(\frac{\partial\omega(T,p(X))}{\partial p(X)}\cdot\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right)-H_K(X)^{\top}\gamma\right)^2\right]$$

$$\gamma_K^{\omega}=\arg\min_{\gamma}\mathbb{E}\left[\left(\frac{\partial\omega(T,p(X))}{\partial p(X)}\cdot\phi^{\nu}\left(Y;F_Y^{\omega}\right)-H_K(X)^{\top}\gamma\right)^2\right].$$

Using triangle inequality,

$$\begin{aligned}\sup_{x\in\mathcal{X}}\left|g(x)-H_K(x)^\top\widehat{\gamma}_K^\omega\right| &\leq \sup_{x\in\mathcal{X}}\left|g(x)-H_K(x)^\top\gamma_K^\omega\right| \\ &+\zeta(K)\left(\|\gamma_K^\omega-\overline{\gamma}_K^\omega\|+\|\overline{\gamma}_K^\omega-\widetilde{\gamma}_K^\omega\|+\|\widetilde{\gamma}_K^\omega-\widehat{\gamma}_K^\omega\|\right)\end{aligned}$$

where $\zeta(K)=\sup_x\|H_K(x)\|$. First, under the assumption that the function $g(\cdot)$ is s times continuously differentiable we have that for a fixed K:

$$\sup_{x\in\mathcal{X}}\left|g(x)-H_K(x)^\top\gamma_K^\omega\right|\leq CK^{-\frac{s}{r}}.$$

We then work with differences in the coefficients:

$$\begin{aligned}\|\gamma_K^\omega-\overline{\gamma}_K^\omega\| &= O_p\left(\zeta(K)N^{-1/2}\right),\ \|\overline{\gamma}_K^\omega-\widetilde{\gamma}_K^\omega\|=O_p\left(N^{-1}\zeta(K)\right) \\ \|\widetilde{\gamma}_K^\omega-\widehat{\gamma}_K^\omega\| &= O_p\left(\zeta(K)\zeta(K_\pi)\left(K_\pi^{1/2}N^{-1/2}+K_\pi^{-s_p/2r}\right)\right)\end{aligned}$$

where K_π and s_p refer to the propensity-score. K_π is the order of the polynomial used to estimate the propensity-score, which is assumed to be s_p-differentiable. Therefore, we have that

$$\begin{aligned}\sup_{x\in\mathcal{X}}\left|g(x)-H_K(x)^\top\widehat{\gamma}_K^\omega\right| &= O_p\left(K^{-\frac{s}{r}}(N)\right)+O_p\left(\zeta^2(K)N^{-1/2}\right)+O_p\left(\zeta^2(K(N))N^{-1}\right) \\ &+O_p\left(\zeta^2(K(N))\zeta(K_\pi(N))K_\pi^{1/2}N^{-1/2}\right)+O_p\left(\zeta^2(K(N))\zeta(K_\pi(N))K_\pi^{-s_p/2r}\right) \\ &= o_p(1).\end{aligned}$$

We now have that $\sup_x|\widehat{p}(x)-p(x)|=o_p(1)$, and for $j=A,B$, $\sup_x|\widehat{g}_j(x)-g_j(x)|=o_p(1)$ and $\sup_{y\in\mathcal{Y}}\left|\widehat{F}_Y^{\widehat{\omega}_j}(y)-F_Y^{\omega_j}(y)\right|=o_p(1)$. We can rewrite $\widehat{V}_{AB}$ as

$$\widehat{V}_{AB}=\frac{1}{N}\sum_{i=1}^N h\left(T_i,Y_i,\widehat{p}(X_i),\widehat{g}_A(X_i),\widehat{g}_B(X_i);\widehat{F}_Y^{\widehat{\omega}_A},\widehat{F}_Y^{\widehat{\omega}_B}\right)$$

where h is a continuously differentiable function with respect to $W=\left[p(X),g_A(X),g_B(X),F_Y^{\omega_A},F_Y^{\omega_B}\right]^\top$. For convenience, define $\widehat{W}=\left[\widehat{p}(X),\widehat{g}_A(X),\widehat{g}_B(X),\widehat{F}_Y^{\widehat{\omega}_A},\widehat{F}_Y^{\widehat{\omega}_B}\right]^\top$. Thus, a simple linearization of $\widehat{V}$ yields

$$\begin{aligned}\left|\widehat{V}_l-V_l\right|\leq\left\|\mathbb{E}\frac{\partial h}{\partial Z}(T_i,Y_i,W_i)\right\| &\cdot\sup_{x\in\mathcal{X}}|\widehat{p}(x)-p(x)|\sup_{x\in\mathcal{X}}|\widehat{g}_A(x)-g_A(x)|\sup_{x\in\mathcal{X}}|\widehat{g}_B(x)-g_B(x)| \\ &\cdot\sup_{y\in\mathcal{Y}}\left|\widehat{F}_Y^{\widehat{\omega}_A}(y)-F_Y^{\omega_A}(y)\right|\sup_{y\in\mathcal{Y}}\left|\widehat{F}_Y^{\widehat{\omega}_B}(y)-F_Y^{\omega_B}(y)\right|+o_p(1)=o_p(1).\end{aligned}$$

Q.E.D.

Proof of Proposition 4: Consider a random sample $Z=\{(Y_i,X_i,T_i):i=1,...,N\}$ from $\mathcal{P}$, the joint distribution of (Y,X,T). The estimator $\widehat{\Delta}=\nu\left(\widehat{F}_Y^{\widehat{\omega}_A}\right)-\nu\left(\widehat{F}_Y^{\widehat{\omega}_B}\right)$ is a function of the original sample Z. Suppose that we construct B^* bootstrap samples, $Z^*=\{Z_b:b=1,...,B^*\}$,

where for each Z_b we randomly draw N observations from Z with replacement, that is, $Z_b = \{(Y_i^*, X_i^*, T_i^*) : i = 1, ..., N\}$. The bootstrap weighted empirical distribution is the empirical measure

$$\widehat{F}_{Yb}^{\widehat{\omega}_j}(y) = N^{-1} \sum_{i=1}^{N} \omega_j(T_i^*, \widehat{p}(X_i^*)) \mathbb{1}\{Y_i^* \leq y\}$$

and the bootstrap empirical process is defined as

$$\mathbb{G}_N^{\omega_j *} = \sqrt{N}\left(\widehat{F}_{Yb}^{\widehat{\omega}_j} - \widehat{F}_Y^{\widehat{\omega}_j}\right).$$

The bootstrap estimator is $\widehat{\Delta}_b = \nu\left(\widehat{F}_{Yb}^{\widehat{\omega}_{Ab}}\right) - \nu\left(\widehat{F}_{Yb}^{\widehat{\omega}_{Bb}}\right)$ and the bootstrap variance is

$$V_{AB}^{\mathfrak{B}} = \mathbb{E}\left[\left(\widehat{\Delta}_b - \widehat{\Delta}\right)^2 \middle| Z\right].$$

Given the B^* replicate samples, an unbiased estimator for this variance is

$$\widehat{V}_{AB}^{\mathfrak{B}} = \frac{1}{B^*} \sum_{b=1}^{B^*} \left(\widehat{\Delta}_b - \overline{\widehat{\Delta}}\right)^2$$

where $\overline{\widehat{\Delta}} = B^{*-1} \sum_{b=1}^{B^*} \widehat{\Delta}_b$.

From Theorem 1, $\mathbb{G}_N^{\omega_{AB}}$ is a sequence of maps with values into the normed space, $\ell^\infty(\mathcal{H})$, converging in distribution to the Gaussian Process $\mathbb{G}^{\omega_{AB}}$. Following van der Vaart (1998), section 23.2.1, $\mathbb{G}_N^{\omega_{AB}*} = \left[\mathbb{G}_N^{\omega_A *}, \mathbb{G}_N^{\omega_B *}\right]$ is a sequence of maps with values into the normed space, $\ell^\infty(\mathcal{H})$, converging (conditionally on Z) in distribution to $\mathbb{G}^{\omega_{AB}}$. Or putting more formally, we use van der Vaart's theorem 23.7 to write

$$\sup_{h \in \mathrm{B}L_1(\ell^\infty(\mathcal{H}))} \left|\mathbb{E}_Z\left[h\left(\mathbb{G}_N^{\omega_{AB}*}\right)\right] - \mathbb{E}\left[h\left(\mathbb{G}^{\omega_{AB}}\right)\right]\right| \to_p 0$$

where $\mathbb{E}_Z$ denotes the expectation conditionally on $Z = \{(Y_1, X_1, T_1), (Y_2, X_2, T_2), ..., (Y_N, X_N, T_N)\}$.

Since $\boldsymbol{\nu}(\cdot)$ $\boldsymbol{\nu} : \mathcal{F}_\nu \to \mathbb{R}^2$ is Hadamard differentiable, by the Delta-method for bootstrap in probability (Theorem 23.9 van der Vaart (1998)), $\boldsymbol{\nu}\left(\mathbb{G}_N^{\omega_{AB}*}\right)$ should converge in distribution to $\boldsymbol{\psi}^\nu\left(\mathbb{G}_N^{\omega_{AB}}; F_Y^{\omega_{AB}}\right) = \left[\psi^\nu\left(\mathbb{G}_N^{\omega_A}; F_Y^{\omega_A}\right), \psi^\nu\left(\mathbb{G}_N^{\omega_B}; F_Y^{\omega_B}\right)\right]^\top$, given Z in probability. Put more formally, for every $h \in \mathrm{B}L_1\left(\mathbb{R}^2\right)$, the function $h \circ \boldsymbol{\psi}^\nu$ is contained in $\mathrm{B}L_{\|\boldsymbol{\psi}^\nu\|}(\mathcal{F}_\nu)$

$$\sup_{h \in \mathrm{B}L_1(\mathbb{R}^2)} \left|\mathbb{E}_Z\left[h\left(\boldsymbol{\nu}\left(\mathbb{G}_N^{\omega_{AB}*}\right)\right)\right] - \mathbb{E}\left[h\left(\boldsymbol{\psi}^\nu\left(\mathbb{G}_N^{\omega_{AB}}; F_Y^{\omega_{AB}}\right)\right)\right]\right| \to_p 0.$$

We have therefore that

$$[1, -1]^\top \boldsymbol{\nu}\left(\mathbb{G}_N^{\omega_{AB}*}\right) = \mathrm{B}^{-1/2} \sum_{b=1}^{\mathrm{B}} \left(\widehat{\Delta}_b - \widehat{\Delta}\right)$$

and

$$\begin{aligned} [1, -1]^\top \boldsymbol{\psi}^\nu\left(\mathbb{G}_N^{\omega_{AB}}; F_Y^{\omega_{AB}}\right) \quad = \quad & N^{-1/2} \sum_{i=1}^{N} \left(\omega_A(T_i, \widehat{p}(X_i)) \cdot \phi^\nu\left(Y_i; \widehat{F}_Y^{\widehat{\omega}_A}\right)\right. \\ & -\omega_B(T, \widehat{p}(X_i)) \cdot \phi^\nu\left(Y_i; \widehat{F}_Y^{\widehat{\omega}_B}\right) \\ & \left. + \left(\widehat{g}_A(X_i) - \widehat{g}_B(X_i)\right)(T_i - \widehat{p}(X_i))\right) \end{aligned}$$

Since $\widehat{V}_{AB}^{\mathfrak{B}}$ is a consistent estimator for the variance of $\widehat{\Delta}_b - \widehat{\Delta}$ given Z, it must be consistent for V_{AB}, since $\boldsymbol{\nu}\left(\mathbb{G}_N^{\omega_{AB}*}\right)$ and $\boldsymbol{\psi}^{\nu}\left(\mathbb{G}_N^{\omega_{AB}}; F_Y^{\omega_{AB}}\right)$ are asymptotically equivalent.

Q.E.D.

REFERENCES

Aakvik, A., J. Heckman, and E. Vytlacil, (2005), "Estimating Treatment Effects for Discrete Outcomes When Responses to Treatment Vary: An application to Norwegian vocational rehabilitation programs", *Journal of Econometrics*, 125, 15-51.

Abadie, A., J. Angrist, and G. Imbens, (2002), "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284-292.

Abadie, A., J. Angrist, and G. Imbens, (2002), "Instrumental Variables Estimation of Quantile Treatment Effects," *Econometrica*, 70, 91-117.

Abadie, A. and G. Imbens, (2008) "On the Failure of the Bootstrap for Matching Estimators," *Econometrica*, 76(6), 1537-1557.

Agresti, A., (1992), "A Survey of Exact Inference for Contingency Tables," *Statistical Science*, 7 (1), 131-153.

Bhattacharya D., (2007), "Inference on inequality from household survey data," *Journal of Econometrics*, 137, 674–707.

Becker, S., and A. Ichino, (2002), "Estimation of Average Treatment Effects Based on Propensity Scores," *The Stata Journal*, 2, 358-377.

Card, D., and D. Sullivan, (1988), "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment," *Econometrica*, 56, 497-530.

Carneiro, P., K. Hansen, and J. Heckman, (2001), Removing the Veil of Ignorance in assessing the distributional impacts of social policies," *Swedish Economic Policy Review*, 8, 273-301.

Carneiro, P., K. Hansen, and J. Heckman, (2003), "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, 44, 361-422.

Cattaneo, M., (2009), "Efficient semiparametric estimation of multi-valued treatment effects under ignorability," *Journal of Econometrics*, forthcoming.

Champernowne, D., and F. Cowell, (1999), *Economic Inequality and Income Distribution*, Cambridge University Press.

Chen, X., H. Hong, and A. Tarozzi, (2008), "Semiparametric Efficiency in GMM Models with Auxiliary Data," *The Annals of Statistics*, 36(2), 808-843.

Chernozhukov, V. I. Fernandez-Val and B. Melly, (2009), "Inference on Counterfactual Distributions," cemmap working paper CWP09/09.

Cowell, F., (2000), "Measurement of Inequality," in *Handbook of Income Distribution*, Ed. by A. Atkinson, and F. Bourguignon, North Holland, Amsterdam.

Cowell, F., (2003), "Theil, Inequality and the Structure of Income Distribution," *Distributional Analysis Discussion Paper, STICERD, LSE*, No. 67.

Cowell, F. and M. Victoria-Feser, (1996), "Robustness properties of inequality measures," *Econometrica*, 64, 77-101.

Cunha, F., J. Heckman, and S. Navarro, (2005), "Counterfactual Analysis of Inequality and Social Mobility", in G. Fields et al. (eds.), *Mobility and Inequality: Frontiers of Research from Sociology and Economics*, Stanford University Press, Palo Alto, CA.

Dehejia, R., and S. Wahba, (1999), "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053-1062.

DiNardo, J., N. Fortin and T. Lemieux, (1996), "Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach," *Econometrica*, 65, 1001-1046.

Donald, S. D. Green and H. Paarsch, (2000), "Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates," *Review of Economic Studies*, 67, 609-633.

Fan, Y. and S. Park, (2010), "Sharp Bounds on the Distribution of Treatment Effects and their Statistical Inference," *Econometric Theory*, 26, forthcoming.

Firpo, S. (2007), "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75, 259-276.

Firpo, S. and G. Ridder, (2008), "Bounds on Functionals of the Distribution of Treatment Effects," IEPR Working Paper No. 08-09.

Foguel, M. (2006), "Time Effects, Economic Effects, and Programme Impacts", typescript, available at url: http://www.webmeets.com/files/papers/LACEA-LAMES/2006/468/Paper_Prog_Impacts.pdf.

Frölich, M. and B. Melly., (2008), "Unconditional quantile treatment effects under endogeneity," IZA discussion paper 3288.

Gosling A., Machin S. and C. Meghir, (2000), "The changing distribution of male wages in the UK," *Review of Economic Studies*, 67, 635-686.

Hahn, J., (1998), "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66, 315-331.

Hastie, T., (1983), "Non-parametric Logistic Regression," Stanford University, Dept. of Statistics.

Heckman, J., (1992), “Randomization and Social Program Evaluation,”in Manski, C., Garfinkle, I. (eds.), *Evaluating Welfare and Training Program.* Harvard University Press, Cambridge, MA, 201–230.

Heckman, J., H. Ichimura, and P. Todd, (1997), “Matching as an Econometric Evaluation Estimator,” *Review of Economic Studies*, 65(2), 261-294.

Heckman, J., H. Ichimura, J. Smith, and P. Todd, (1998), “Characterizing Selection Bias Using Experimental Data,” *Econometrica*, 66, 1017-1098.

Heckman, J., R. Robb, (1986), “Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcome,” in *Drawing Inference from Self-Selected Samples*, ed. by H. Wainer, pp. 63-107. New York, Springer-Verlag.

Heckman, J. and J. Smith, (1998), “Evaluating the Welfare State,” in Strom, S. (ed.), *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial*, Econometric Society Monograph Series, Cambridge University Press, Cambridge.

Heckman, J., J. Smith, and N. Clements, (1997), “Making the Most out of Programme Evaluations and Social Experiments Accounting for Heterogeneity in Programme Impacts,” *Review of Economic Studies*, 64(4), 487-535.

Hirano, K., G. Imbens, and G. Ridder, (2003), “Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score,” *Econometrica*, 71, 1161-1189.

Imbens, G., and D. Rubin, (1997), “Estimating Outcome Distributions for Compliers in Instrumental Variable Models,” *Review of Economic Studies*, October, 555-574.

Juhn, C., K. Murphy, and B. Pierce, (1993), “Wage Inequality and the Rise in Returns to Skill,” *Journal of Political Economy*, 101, 410-442.

LaLonde, R., (1986), “Evaluating the Econometric Evaluations of Training Programs with Experimental Data,” *American Economic Review*, 76, 604-620.

Lechner, M., (1999), “Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification,” *Journal of Business & Economic Statistics*, 17, 74-90.

Lehmann, E., (1999), *Elements of Large-Sample Theory*, Springer.

Machado, J. and J. Mata, (2005), “Counterfactual Decomposition of Changes in Wage Distributions using Quantile Regression,” *Journal of Applied Econometrics*, 20, 445-465.

Manski, C. (1997), “The Mixing Problem in Programme Evaluation,” *Review of Economic Studies*, October, 537-554.

Melly, B. (2006), “Estimation of counterfactual distributions using quantile regression,” mimeo.

Newey, W., (1994), “The Asymptotic Variance of Semiparametric Estimators,” *Econometrica*, 62, 1349-1382.

Newey, W., (1995), "Convergence Rates for Series Estimators," in *Advances in Econometrics and Qualitative Economics: Essays in Honor of C.R. Rao*, G. Maddala, P.C. Phillips, and T.N. Srinivasan, eds., Cambridge US, Basil-Blackwell.

Newey, W., (1997), "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147-168.

Ren, J. and P.Kumar, (1991) "On Hadamard Differentiability of Extended Statistical Functional," *Journal of Multivariate Analysis,* 39, 30-43.

Robins, J., and A. Rotnitzky, (1995) "Semiparametric Efficiency in Multivariate Regression Models with Missing Data," *Journal of the American Statistical Association*, 90, 122-129.

Robins, J., A. Rotnitzky and L. P. Zhao, (1994), "Estimation of regression coefficients when some regressions are not always observed," *Journal of the American Statistical Association*, 89, 846-866.

Rosenbaum, P., and D. Rubin, (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.

Rothe, C., (2010), "Nonparametric Estimation of Distributional Policy Effects," *Journal of Econometrics*, 155, 56-70.

Rotnitzky, A. and J. Robins, (1995), "Semiparametric Regression Estimation in the Presence of Dependent Censoring," *Biometrika*, 82, 805-820.

Rubin, D., (1977), "Assignment to Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 2(1), 1-26.

Scharfstein D., A. Rotnitzky and J. Robins, (1999), "Adjusting for Non-ignorable Drop-out Using Semiparametric Non-response Models," *Journal of American Statistical Association*, 94, 1096-1146.

Schluter, C. and M. Trede, (2003), "Local Versus Global Assessment of Mobility," *International Economic Review*, 44(4), 1313-1335.

Tarozzi, A., (2007), "Calculating Comparable Statistics from Incomparable Surveys, with an Application to Poverty in India," *Journal of Business and Economic Statistics*, 25(3), 314-336.

van der Vaart, A., (1991), "Efficiency and Hadamard Differentiability," *Scandinavian Journal of Statistics*, 18 (1), 63-75.

van der Vaart, A., (1998), *Asymptotic Statistics.* Cambridge, Cambridge University Press.

Wooldridge, J., (2007), "Inverse probability weighted estimation for general missing data problems," *Journal of Econometrics*, 141(2), 1281-1301.

Table 4: Results of Monte Carlo Exercise (Sample Size 250. Replications 1000. Normal Selection Model)

Treatment on the Treated Parameters	Estimators	Target	Average	Lower 10th percentile	Median	Upper 10th percentile	Standard Deviation	Bias	Root Mean Squared Error	Mean Absolute Error	Median Absolute Error	90% C.I. Coverage Rate
Mean Treatment Effects		1.103										
	Unfeasible		1.101	0.972	1.095	1.234	0.103	-0.002	0.103	0.082	0.069	0.906
	Naive		1.075	0.953	1.072	1.209	0.100	-0.027	0.104	0.084	0.073	0.898
	Weighted		1.100	0.971	1.097	1.238	0.106	-0.003	0.106	0.084	0.072	0.907
	Weighted (Linear)		1.100	0.967	1.095	1.234	0.105	-0.005	0.105	0.083	0.072	0.907
	Location Shift		1.100	0.969	1.097	1.239	0.106	-0.003	0.106	0.084	0.072	0.905
	Location Shift (Linear)		1.102	0.972	1.099	1.240	0.105	-0.001	0.105	0.083	0.071	0.906
	Log Location Shift		1.109	0.977	1.107	1.248	0.107	0.006	0.107	0.084	0.072	0.906
	Log Location Shift (linear)		1.111	0.981	1.108	1.252	0.105	0.008	0.106	0.083	0.069	0.905
	CFM		1.101	0.970	1.098	1.240	0.106	-0.002	0.106	0.084	0.071	0.905
	CFM (Linear)		1.102	0.974	1.100	1.243	0.106	0.106	0.106	0.084	0.072	0.904
CV		0.268										
	Unfeasible		0.249	0.127	0.230	0.385	0.121	-0.019	0.122	0.090	0.075	0.934
	Naive		0.296	0.178	0.278	0.439	0.121	0.029	0.124	0.084	0.063	0.927
	Weighted		0.254	0.127	0.241	0.403	0.934	-0.014	0.126	0.090	0.071	0.934
	Weighted (Linear)		0.258	0.130	0.243	0.405	0.125	-0.009	0.126	0.089	0.070	0.934
	Location Shift		0.283	0.161	0.266	0.425	0.121	0.016	0.122	0.084	0.065	0.932
	Location Shift (Linear)		0.284	0.167	0.265	0.425	0.120	0.017	0.122	0.083	0.064	0.933
	Log Location Shift		0.284	0.167	0.267	0.422	0.116	0.016	0.117	0.080	0.062	0.933
	Log Location Shift (linear)		0.285	0.172	0.268	0.421	0.113	0.017	0.114	0.079	0.061	0.933
	CFM		0.253	0.127	0.241	0.396	0.127	-0.014	0.127	0.092	0.074	0.929
	CFM (Linear)		0.255	0.131	0.241	0.402	0.125	-0.013	0.126	0.090	0.070	0.930
Interquartile Range		0.601										
	Unfeasible		0.605	0.459	0.601	0.756	0.122	0.003	0.122	0.094	0.076	0.890
	Naive		0.638	0.481	0.633	0.796	0.127	0.037	0.132	0.101	0.081	0.892
	Weighted		0.603	0.452	0.598	0.763	0.128	0.002	0.128	0.099	0.081	0.907
	Weighted (Linear)		0.610	0.459	0.603	0.767	0.128	0.008	0.128	0.099	0.081	0.904
	Location Shift		0.660	0.507	0.656	0.816	0.126	0.058	0.139	0.107	0.088	0.873
	Location Shift (Linear)		0.650	0.499	0.645	0.797	0.124	0.048	0.133	0.102	0.082	0.894
	Log Location Shift		0.615	0.459	0.610	0.773	0.126	0.013	0.127	0.097	0.077	0.908
	Log Location Shift (linear)		0.613	0.459	0.606	0.766	0.126	0.012	0.127	0.097	0.079	0.904
	CFM		0.602	0.452	0.597	0.761	0.128	0.001	0.128	0.099	0.081	0.901
	CFM (Linear)		0.601	0.444	0.595	0.760	0.128	0.000	0.128	0.099	0.082	0.909
Theil Index		0.079										
	Unfeasible		0.077	0.038	0.071	0.119	0.036	-0.002	0.036	0.027	0.022	0.930
	Naive		0.091	0.054	0.085	0.136	0.036	0.012	0.038	0.026	0.019	0.905
	Weighted		0.077	0.038	0.073	0.123	0.037	-0.002	0.037	0.027	0.021	0.926
	Weighted (Linear)		0.079	0.041	0.075	0.124	0.037	0.000	0.037	0.027	0.021	0.928
	Location Shift		0.087	0.048	0.082	0.133	0.038	0.009	0.039	0.027	0.021	0.911
	Location Shift (Linear)		0.089	0.049	0.082	0.135	0.038	0.010	0.039	0.028	0.021	0.909
	Log Location Shift		0.087	0.050	0.082	0.131	0.035	0.008	0.036	0.026	0.020	0.914
	Log Location Shift (linear)		0.088	0.051	0.083	0.132	0.035	0.009	0.036	0.025	0.019	0.911
	CFM		0.077	0.135	0.039	0.072	0.038	-0.002	0.038	0.028	0.022	0.923
	CFM (Linear)		0.078	0.040	0.074	0.123	0.037	-0.001	0.037	0.027	0.021	0.922
Gini Coefficient		0.087										
	Unfeasible		0.085	0.051	0.085	0.120	0.027	-0.002	0.027	0.022	0.019	0.904
	Naive		0.110	0.076	0.108	0.145	0.027	0.023	0.036	0.028	0.024	0.797
	Weighted		0.087	0.050	0.086	0.123	0.030	0.000	0.030	0.024	0.019	0.910
	Weighted (Linear)		0.090	0.052	0.088	0.125	0.030	0.003	0.030	0.023	0.019	0.906
	Location Shift		0.111	0.068	0.109	0.156	0.034	0.024	0.042	0.033	0.027	0.825
	Location Shift (Linear)		0.108	0.067	0.106	0.151	0.032	0.021	0.039	0.030	0.025	0.846
	Log Location Shift		0.099	0.063	0.098	0.135	0.029	0.012	0.031	0.025	0.021	0.887
	Log Location Shift (linear)		0.099	0.064	0.099	0.135	0.028	0.013	0.031	0.024	0.021	0.880
	CFM		0.087	0.050	0.086	0.124	0.031	0.000	0.031	0.024	0.020	0.905
	CFM (Linear)		0.087	0.051	0.087	0.124	0.030	0.000	0.030	0.024	0.020	0.910

Table 5: Results of Monte Carlo Exercise (Sample Size 4,000, Replications 1000, Normal Selection Model)

Treatment on the Treated Parameters	Estimators	Target	Average	Lower 10th percentile	Median	Upper 10th percentile	Standard Deviation	Bias	Root Mean Squared Error	Mean Absolute Error	Median Absolute Error	90% C.I. Coverage Rate
Mean Treatment Effects		1.103										
	Unfeasible		1.103	1.066	1.102	1.140	0.028	0.000	0.028	0.022	0.018	0.900
	Naive		0.027	1.044	1.077	1.113	0.027	-0.025	0.037	0.031	0.028	0.760
	Weighted		1.103	1.068	1.103	1.140	0.028	0.000	0.028	0.023	0.019	0.894
	Weighted (Linear)		1.103	1.066	1.101	1.138	0.028	-0.002	0.028	0.022	0.018	0.895
	Location Shift		1.102	1.067	1.102	1.139	0.028	0.000	0.028	0.023	0.018	0.894
	Location Shift (Linear)		1.105	1.070	1.104	1.141	0.028	0.002	0.028	0.023	0.018	0.888
	Log Location Shift		1.112	1.077	1.112	1.150	0.029	0.010	0.030	0.024	0.020	0.876
	Log Location Shift (linear)		1.114	1.079	1.113	1.151	0.028	0.011	0.030	0.024	0.020	0.872
	CFM		1.099	1.064	1.099	1.136	0.028	-0.003	0.028	0.023	0.019	0.893
	CFM (Linear)		1.103	1.069	1.102	1.140	0.028	0.028	0.028	0.023	0.018	0.890
CV		0.268										
	Unfeasible		0.267	0.219	0.263	0.316	0.042	0.000	0.042	0.031	0.026	0.932
	Naive		0.315	0.267	0.311	0.364	0.043	0.047	0.064	0.050	0.043	0.754
	Weighted		0.268	0.219	0.264	0.318	0.925	0.000	0.044	0.032	0.026	0.925
	Weighted (Linear)		0.274	0.225	0.270	0.323	0.044	0.006	0.044	0.032	0.026	0.925
	Location Shift		0.305	0.258	0.301	0.354	0.042	0.038	0.057	0.043	0.035	0.807
	Location Shift (Linear)		0.303	0.255	0.298	0.352	0.043	0.035	0.055	0.041	0.033	0.830
	Log Location Shift		0.307	0.263	0.302	0.354	0.040	0.040	0.057	0.043	0.035	0.784
	Log Location Shift (linear)		0.305	0.262	0.300	0.351	0.040	0.037	0.054	0.041	0.033	0.798
	CFM		0.302	0.255	0.297	0.351	0.042	0.034	0.054	0.040	0.032	0.834
	CFM (Linear)		0.296	0.251	0.292	0.345	0.042	0.029	0.051	0.037	0.028	0.864
Interquartile Range		0.601										
	Unfeasible		0.601	0.560	0.600	0.645	0.033	0.000	0.033	0.026	0.023	0.909
	Naive		0.635	0.594	0.635	0.678	0.033	0.034	0.048	0.039	0.035	0.727
	Weighted		0.601	0.559	0.601	0.644	0.033	0.000	0.033	0.027	0.024	0.910
	Weighted (Linear)		0.607	0.565	0.607	0.650	0.033	0.006	0.034	0.027	0.024	0.900
	Location Shift		0.678	0.633	0.677	0.721	0.035	0.076	0.084	0.076	0.075	0.296
	Location Shift (Linear)		0.660	0.618	0.659	0.704	0.034	0.059	0.068	0.060	0.057	0.488
	Log Location Shift		0.620	0.576	0.618	0.661	0.034	0.018	0.038	0.031	0.027	0.862
	Log Location Shift (linear)		0.614	0.570	0.613	0.657	0.034	0.012	0.036	0.029	0.025	0.892
	CFM		0.602	0.560	0.602	0.644	0.033	0.001	0.033	0.027	0.024	0.911
	CFM (Linear)		0.602	0.559	0.602	0.644	0.033	0.000	0.033	0.027	0.023	0.910
Theil Index		0.079										
	Unfeasible		0.079	0.067	0.078	0.091	0.010	0.000	0.010	0.008	0.006	0.905
	Naive		0.093	0.081	0.092	0.106	0.010	0.014	0.017	0.015	0.013	0.619
	Weighted		0.079	0.067	0.079	0.092	0.010	0.000	0.010	0.008	0.007	0.910
	Weighted (Linear)		0.081	0.069	0.080	0.094	0.010	0.002	0.010	0.008	0.007	0.905
	Location Shift		0.093	0.080	0.092	0.107	0.011	0.014	0.018	0.015	0.013	0.690
	Location Shift (Linear)		0.093	0.079	0.092	0.107	0.011	0.014	0.018	0.015	0.013	0.695
	Log Location Shift		0.091	0.079	0.090	0.103	0.010	0.012	0.015	0.012	0.011	0.692
	Log Location Shift (linear)		0.090	0.078	0.089	0.102	0.010	0.011	0.015	0.012	0.010	0.698
	CFM		0.089	0.107	0.077	0.089	0.010	0.010	0.014	0.012	0.010	0.740
	CFM (Linear)		0.088	0.076	0.087	0.100	0.010	0.009	0.013	0.011	0.009	0.785
Gini Coefficient		0.087										
	Unfeasible		0.087	0.078	0.086	0.096	0.007	0.000	0.007	0.006	0.005	0.892
	Naive		0.111	0.102	0.111	0.121	0.008	0.025	0.026	0.025	0.024	0.053
	Weighted		0.087	0.077	0.087	0.097	0.008	0.000	0.008	0.006	0.005	0.892
	Weighted (Linear)		0.090	0.081	0.090	0.100	0.008	0.003	0.009	0.007	0.006	0.869
	Location Shift		0.117	0.103	0.116	0.130	0.011	0.030	0.032	0.030	0.029	0.125
	Location Shift (Linear)		0.114	0.101	0.113	0.128	0.011	0.027	0.029	0.027	0.027	0.180
	Log Location Shift		0.103	0.093	0.102	0.113	0.008	0.016	0.018	0.016	0.015	0.372
	Log Location Shift (linear)		0.101	0.091	0.101	0.111	0.008	0.014	0.016	0.015	0.014	0.437
	CFM		0.100	0.090	0.100	0.110	0.008	0.013	0.015	0.013	0.013	0.494
	CFM (Linear)		0.097	0.087	0.097	0.107	0.008	0.010	0.013	0.011	0.010	0.607

Table 6: Summary Statistics

	Final Sample									
	Rio de Janeiro					Fortaleza				
Variables	Treatment (A)	Control (B)	Weighted Control (C)	Difference (A)-(B)	Difference (A)-(C)	Treatment (D)	Control (E)	Weighted Control (F)	Difference (D)-(E)	Difference (D)-(F)
Previous Labor Market	0.52	0.47	0.51	0.05	0.01	0.82	0.82	0.83	-0.01	-0.01
Engagement (Dummy)	(0.01)	(0.01)	(0.02)	(0.02)**	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Number of children	0.08	0.07	0.09	0.01	-0.01	0.38	0.37	0.40	0.01	-0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Schooling (Years)	8.58	8.31	8.50	0.26	0.08	9.00	9.27	9.18	-0.27	-0.18
	(0.06)	(0.07)	(0.08)	(0.09)***	(0.10)	(0.10)	(0.07)	(0.09)	(0.13)*	(0.14)
Age	18.55	18.03	18.63	0.52	-0.08	27.10	26.52	26.54	0.58	0.56
	(0.12)	(0.12)	(0.19)	(0.18)***	(0.23)	(0.29)	(0.24)	(0.36)	(0.37)	(0.47)
Dummy for Single	0.93	0.94	0.92	0.00	0.01	0.64	0.64	0.61	0.00	0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)
Household Head Dummy	0.03	0.03	0.04	0.00	-0.01	0.21	0.19	0.18	0.02	0.03
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
White Dummy	0.38	0.40	0.41	-0.03	-0.03	0.34	0.31	0.30	0.03	0.04
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)*	(0.03)
Male Dummy	0.36	0.34	0.38	0.02	-0.02	0.39	0.43	0.41	-0.03	-0.02
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)
Number of Observations	1258	1211	1211			1040	1355	1355		

Bootstrapped standard errors in parenthesis. *: Significant at 10%; **: Significant at 5%; ***: Significant at 1%.

Table 7: Inequality Treatment Effects for the PLANFOR data set

	Rio de Janeiro											
	Sum of all earnings during the 12-month period after treatment						Hourly wage rate at first job in a 12-month period after treatment					
	Treatment Effect Estimators						Treatment Effect Estimators					
	Treated	Naive	Weighted	Linear Shift	Log Linear Shift	CFM	Treated	Naive	Weighted	Linear Shift	Log Linear Shift	CFM
Average	699.34	141.62	30.45	2.35	-971.52	4.85	1.54	-0.20	-0.36	-0.39	-0.51	-0.33
	(35.29)	(54.57)***	(59.70)	(82.33)	(204.52)***	(93.62)	(0.10)	(0.16)	(0.19)*	(0.40)	(0.46)	(0.19)*
Coefficient of Variation	1.82	-0.18	-0.12	-0.06	0.04	-0.09	1.44	-0.30	-0.27	-0.31	-0.71	-0.21
	(0.07)	(0.11)*	(0.13)	(0.14)	(0.45)	(0.16)	(0.55)	(0.58)	(0.59)	(0.58)	(0.77)	(0.60)
Interquartile Range	945.57	300.59	167.65	-141.26	-567.35	105.90	0.90	-0.03	-0.10	-0.80	-0.39	-0.13
	(81.41)	(99.64)***	(107.85)	(107.16)	(183.85)***	(133.74)	(0.05)	(0.08)	(0.11)	(0.58)	(0.15)***	(0.12)
Theil Index	2.34	-0.61	-0.16	1.80	1.62	-0.10	0.36	-0.21	-0.22	-0.16	-0.30	-0.20
	(0.20)	(0.24)**	(0.23)	(0.15)***	(0.20)***	(0.24)	(0.10)	(0.16)	(0.17)	(0.16)	(0.21)	(0.18)
Gini Coefficient	0.75	-0.03	-0.02	1.01	2.26	-0.01	0.38	-0.09	-0.10	0.15	-0.13	-0.09
	(0.01)	(0.01)**	(0.02)	(0.12)***	(0.14)***	(0.02)	(0.04)	(0.05)*	(0.05)*	(0.19)	(0.06)**	(0.06)

	Fortaleza											
	Sum of all earnings during the 12-month period after treatment						Hourly wage rate at first job in a 12-month period after treatment					
	Treatment Effect Estimators						Treatment Effect Estimators					
	Treated	Naive	Weighted	Linear Shift	Log Linear Shift	CFM	Treated	Naive	Weighted	Linear Shift	Log Linear Shift	CFM
Average	1403.41	140.14	130.42	110.58	-1287.86	129.08	1.80	0.06	-0.22	-0.20	-0.20	-0.20
	(94.19)	(113.14)	(125.10)	(177.56)	(590.71)**	(140.24)	(0.13)	(0.14)	(0.29)	(0.40)	(0.45)	(0.31)
Coefficient of Variation	1.82	0.33	0.26	0.21	0.17	0.30	1.62	0.22	-0.01	0.01	0.22	0.02
	(0.21)	(0.21)	(0.24)	(0.24)	(0.49)	(0.25)	(0.31)	(0.38)	(0.54)	(0.46)	(0.49)	(0.60)
Interquartile Range	1885.11	56.00	71.22	219.35	-399.47	30.76	1.02	-0.07	-0.11	-0.38	-0.24	-0.10
	(128.69)	(152.98)	(156.85)	(206.08)	(496.25)	(203.49)	(0.06)	(0.08)	(0.10)	(0.26)	(0.23)	(0.17)
Theil Index	1.40	-0.04	-0.07	0.93	0.74	-0.03	0.49	0.10	-0.04	0.01	0.04	-0.04
	(0.10)	(0.14)	(0.18)	(0.12)***	(0.19)***	(0.20)	(0.08)	(0.12)	(0.21)	(0.18)	(0.18)	(0.23)
Gini Coefficient	0.68	0.01	0.00	0.61	1.10	0.01	0.46	0.05	0.00	0.26	0.13	0.00
	(0.02)	(0.02)	(0.02)	(0.08)***	(0.10)***	(0.03)	(0.03)	(0.04)	(0.07)	(0.08)***	(0.07)**	(0.07)

Bootstrapped standard errors in parenthesis. *: Significant at 10%; **: Significant at 5%; ***: Significant at 1%.

Table A.1: Summary Statistics

Variables	Original Sample									
	Rio de Janeiro					Fortaleza				
	Treatment (A)	Control (B)	Weighted Control (C)	Difference (A)-(B)	Difference (A)-(C)	Treatment (D)	Control (E)	Weighted Control (F)	Difference (D)-(E)	Difference (D)-(F)
Previous Labor Market	0.51	0.46	0.50	0.05	0.01	0.81	0.83	0.84	-0.02	-0.02
Engagement (Dummy)	(0.01)	(0.01)	(0.02)	(0.02)**	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Number of children	0.08	0.06	0.08	0.01	0.00	0.38	0.37	0.42	0.01	-0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)
Schooling (Years)	8.50	8.26	8.38	0.24	0.11	9.03	9.31	9.19	-0.29	-0.16
	(0.06)	(0.07)	(0.08)	(0.09)***	(0.10)	(0.09)	(0.07)	(0.09)	(0.12)**	(0.14)
Age	18.42	17.86	18.32	0.56	0.10	27.34	26.64	26.94	0.70	0.40
	(0.13)	(0.11)	(0.16)	(0.17)***	(0.21)	(0.31)	(0.26)	(0.36)	(0.39)*	(0.45)
Dummy for Single	0.93	0.94	0.93	-0.01	0.00	0.63	0.64	0.60	-0.01	0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)
Household Head Dummy	0.03	0.03	0.04	0.01	0.00	0.21	0.19	0.18	0.02	0.03
	(0.00)	(0.00)	(0.01)	(0.01)*	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
White Dummy	0.37	0.42	0.41	-0.05	-0.04	0.35	0.32	0.33	0.03	0.02
	(0.01)	(0.01)	(0.02)	(0.02)**	(0.02)*	(0.01)	(0.01)	(0.02)	(0.02)*	(0.02)
Male Dummy	0.36	0.34	0.39	0.03	-0.02	0.41	0.42	0.40	-0.01	0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)
Number of Observations	1323	1293	1293			1152	1454	1454		

Bootstrapped standard errors in parenthesis. *: Significant at 10%; **: Significant at 5%; ***: Significant at 1%.

Technical Supplemental Material of "Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures"

Sergio Firpo* Cristine Pinto†

September 2011

Abstract

This supplemental material contains the formal proofs of all theorems and lemmas in the paper "Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality". We start by presenting a formal proof of the asymptotically linear representation of the estimator of the weighted distribution function, $\widehat{F}^{\widehat{w}}$. Then we show that the class of measurable influence functions of $\widehat{F}^{\widehat{w}}$ are P-Donsker, and invoke the Donsker´s Theorem to show the joint uniform convergence of the empirical weighted marginal distributions to a Gaussian Process with mean 0 and certain variance matrix. We calculate the maximal asymptotic precision with which we can estimate F^{w}, and show that $\widehat{F}^{\widehat{w}}$ attains the semiparametric efficient bound. Using the fact that the inequality measure v is Hadamard differentiable, we can demonstrate that the estimator $\widehat{v}\left(F_Y^{\widehat{w}_A}\right)-\widehat{v}\left(F_Y^{\widehat{w}_B}\right)$ is asymptotically normal and semiparametrically efficient. Finally, we show that valid inference for estimator of the inequality treatment effects can be implemented by using a consistent estimator for the asymptotic variance or by standard bootstrap variance.

In summary, this technical appendix contains the proofs of the following theorems and lemmas:

- Lemma 1.1: Asymptotically linear representation of $\widehat{F}^{\widehat{w}}$
- Lemma 1.2: The collection of the influence functions of $\widehat{F}^{\widehat{w}}$ are P-Donskers.
- Theorem 1.1: Joint uniform convergence of the empirical weighted marginal distributions.
- Lemma 2.1: Semiparametric efficient of the $\widehat{v}\left(F_Y^{\widehat{w}_A}\right)-\widehat{v}\left(F_Y^{\widehat{w}_B}\right)$

*Corresponding author. Escola de Economia de São Paulo and C-Micro, Fundação Getulio Vargas (Brazil). Mailing address: EESP-FGV, Rua Itapeva 474/1215, São Paulo - SP, Brazil, 01332-000. E-mail: sergio.firpo@fgv.br. Webpage: http://sites.google.com/site/sergiopfirpo

†Escola de Economia de São Paulo and C-Micro, Fundação Getulio Vargas (Brazil). Mailing address: EESP-FGV, Rua Itapeva 474/1200, São Paulo - SP, Brazil, 01332-000. E-mail: cristine.pinto@fgv.br. Webpage: https://sites.google.com/site/cristinepinto/

- Theorem 2.1: Asymptotically linear representation of $\widehat{v}\left(F_Y^{\widehat{w}_A}\right) - \widehat{v}\left(F_Y^{\widehat{w}_B}\right)$
- Lemma 3.1: A series approximation for one of the terms in the expression of the asymptotic variance of the estimator $\widehat{v}\left(F_Y^{\widehat{w}_A}\right) - \widehat{v}\left(F_Y^{\widehat{w}_B}\right)$.
- Theorem 3.1: Consistent estimator for the asymptotic variance
- Theorem 3.2: Validity of the bootstrap variance

Supplemental material: additional proofs and derivations

1 Notation, Identification and Estimation

This supplemental appendix contains the formal proofs of all the theorems and lemmas in the main paper. All notation is defined as in the main text unless stated otherwise. For completeness, we state the functions of interest and the main assumptions.

We have a random sample of N individuals indexed by i. For each i, we observe a vector of covariates X_i with support $\mathcal{X} \subset \mathbb{R}^r$, a binary variable T_i that equals 1 if the individual was assigned to treatment, and the outcome Y_i. The observed outcome can be expressed as a function of the potential outcome

$$Y_i = T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0)$$

where $Y_i(1)$ is the potential outcome if individual i receives treatment, and $Y_i(0)$ is the potential outcome if individual i does not receive treatment. We need a set of identifying restrictions that allow us to write the distribution of the unobserved outcome in terms of observable data. The first assumption that has been used in many papers is stated as follows.

Assumption 1 [***Unconfoundedness***] *Let $(Y(1), Y(0), T, X)$ have a joint distribution. For all x in $\mathcal{X}$: $(Y(1), Y(0))$ is jointly independent from T given $X = x$, that is, $(Y(1), Y(0)) \perp T | X = x$.*

We define an inequality measure ν, where $\nu : \mathcal{F}_v \to \mathbb{R}$, that is, ν is a functional defined over the set of distribution functions, $\mathcal{F}_\nu$, such that $F \in \mathcal{F}_\nu$ if $\nu < +\infty$. The parameter of interest is simply the difference in the results of a given inequality measure evaluted in two different distribution functions.

Because the elements of the set $\mathcal{F}_\nu$ that we are interested in are marginal (or conditional on T) distribution functions of potential outcomes, in order to identify the parameter of interest we need first to identify these distribution functions. We identify the distribution functions of potential outcomes by writing them as weighted marginal distributions of Y,

$$F_Y^\omega(y) = E[\omega(T, p(X)) \cdot \mathbb{I}\{Y \le y\}]$$

where $\omega(T, p(X))$ is a given weighting function that 'converts' marginal (or conditional on T) distribution functions of potential outcomes into weighted marginal distributions of Y. The second argument in the weighting function is $p : \mathcal{X} \to \mathcal{E} \subset [0, 1]$, the propensity score, defined as $p(x) \equiv \Pr[T = 1|X = x]$, where $X \in \mathcal{X}$. Both $p(\cdot)$ and $\omega(\cdot)$, $\omega : \{0, 1\} \times \mathcal{E} \to \mathbb{R}$, are continuously differentiable bounded functions with bounded derivatives.

For identification, we also need that the treatment and control groups are comparable in terms of X, which is guaranteed by the following assumption.

Assumption 2 [***Common Support***] *For all x in $\mathcal{X}$, there are real numbers c_* and c^* such that $0 < c_* \le p(x) \le c^* < 1$.*

Under assumptions 1 and 2, we invoke a result in Firpo (2007) for quantiles that can be straightforwardly applied to cdfs. For a given $y \in \mathcal{Y}$ and $j = 0, 1$:

$$\begin{aligned} F_j(y) &= F_{Y(j)}(y) \equiv \Pr[Y(j) < y] \\ &= E[\omega_{jU}(T, p(X)) \cdot \mathbb{I}\{Y \le y\}], \end{aligned}$$

where

$$\omega_{jU}\left(T,p\left(X\right)\right)=j\cdot\left(\frac{T}{p(X)}\right)+\left(1-j\right)\cdot\left(\frac{1-T}{1-p(X)}\right)$$

and

$$\begin{aligned}F_{j1}\left(y\right) &= F_{Y(j)|T}\left(y|1\right)\equiv Pr[Y\left(j\right)<y|T=1]\\ &= E\left[\omega_{jC}\left(T,p\left(X\right)\right)\cdot \mathbb{I}\{Y\leq y\}\right],\end{aligned}$$

$$\omega_{jC}\left(T,p\left(X\right)\right)=j\cdot\left(\frac{T}{p}\right)+\left(1-j\right)\cdot\left(\frac{1-T}{1-p(X)}\right)\cdot\left(\frac{p(X)}{p}\right)$$

and $p\equiv E[T]$. Thus, we can identify an inequality treatment effect parameter as a difference in inequality measures of two weighted marginal distributions of Y. For example, consider, weights ω_{1U} and ω_{0U}. *Overall Inequality Treatment Effect*, Δ_O^ν, can thus be written as:

$$\Delta_O^\nu=\nu\left(F_1\right)-\nu\left(F_0\right)=\nu\left(F_Y^{\omega_{1U}}\right)-\nu\left(F_Y^{\omega_{0U}}\right)=\Delta_U.$$

Inequality Treatment Effect on the Treated, Δ_{TT}^ν, is identified similarly as

$$\Delta_{TT}^\nu=\nu\left(F_{11}\right)-\nu\left(F_{01}\right)=\nu\left(F_Y^{\omega_{1C}}\right)-\nu\left(F_Y^{\omega_{0C}}\right)=\Delta_C.$$

And *Current Inequality Treatment Effect* (CIT), Δ_{CIT}^ν, is simply

$$\Delta_{CIT}^\nu=\nu\left(F_Y\right)-\nu\left(F_0\right)=\nu\left(F_Y\right)-\nu\left(F_Y^{\omega_{0U}}\right).$$

In what follows, we specialize to the cases of doing estimation and inference for Δ_O^ν and Δ_{TT}^ν. Extending the analysis to Δ_{CIT}^ν is straightforward as it involves doing the same as for Δ_O^ν, but fixing $\omega_{1U}\left(T,p\left(X\right)\right)=1$ for all (T,X). In fact, to facilitate notation we rename Δ_O^ν and Δ_{TT}^ν after the names given to weights, respectively Δ_U (Δ_O^ν) and Δ_C (Δ_{TT}^ν).

Using the sample analog principle, estimators for Δ_U and Δ_C are obtained by replacing the population weighted distribution by the empirical weighted distribution using estimated weights, as

$$\widehat{\Delta}_U=\nu\left(\widehat{F}_1\right)-\nu\left(\widehat{F}_0\right)$$

and

$$\widehat{\Delta}_C=\nu\left(\widehat{F}_{11}\right)-\nu\left(\widehat{F}_{01}\right),$$

where for a given $y\in\mathcal{Y}$ and $j=0,1$:

$$\widehat{F}_j\left(y\right)=\frac{1}{N}\sum_{i=1}^{N}\left\{j\cdot\left(\frac{T_i}{\widehat{p}\left(X_i\right)}\right)+\left(1-j\right)\cdot\left(\frac{1-T_i}{1-\widehat{p}\left(X_i\right)}\right)\right\}\cdot\mathbb{I}\{Y_i\leq y\}$$

and

$$\widehat{F}_{j1}\left(y\right)=\frac{1}{N_1}\sum_{i=1}^{N}\left\{j\cdot T_i+\left(1-j\right)\cdot\left(1-T_i\right)\cdot\left(\frac{\widehat{p}\left(X_i\right)}{1-\widehat{p}\left(X_i\right)}\right)\right\}\cdot\mathbb{I}\{Y_i\leq y\},$$

where $N_1=\sum_{i=1}^N T_i$.

The weights depend on the propensity score. We estimate the propensity score using the approach proposed by Hirano, Imbens and Ridder (2003), which is based on a series estimator. Formally, the estimator of the propensity score is

$$\widehat{p}\left(X_i\right)=\Lambda\left(H_K\left(X_i\right)'\widehat{\pi}_K\right)$$

and

$$\hat{\pi}_K = \arg\max_{\pi_K \in \mathbb{R}^K} \sum_{i=1}^{N} \left(T_i \cdot \log(\Lambda(H_K(X_i)'\pi_K)) + (1-T_i) \cdot \log(1-\Lambda(H_K(X_i)'\pi_K)) \right).$$

where $\Lambda : \mathbb{R} \to \mathbb{R}$, $\Lambda(z) = (1+exp(-z))^{-1}$ is the the c.d.f. of a logistic distribution evaluated at z. In our case, $H_K(\cdot)$ includes polynomials up to order n, and the size K increases with N, and if $K > (n+1)^r$, then $H_K(x)$ includes all polynomials up order n.

We need to control how the size of the series increases with N, which depends on degree of smoothness of the propensity score and the dimension of X. We need to impose the following assumptions:

Assumption 3 [***Propensity Score***] *For all $x \in \mathcal{X}$, the propensity score, $p(x)$, is s_p times continuously differentiable with $s_p \geq 7r$.*

Assumption 4 [***Series Estimator***] *The order of $H_K(x)$, K, is of the form $K = C \cdot N^{c_p}$ where C is a constant and $c_p \in \left(\frac{1}{4\left(\frac{s_p}{r}-1\right)}, \frac{1}{9} \right)$.*

We also impose some regularity conditions.

Assumption 5 [***Smoothness***]

- *The support of $\mathcal{X}$ of X is a compact subset of $\mathbb{R}^r$.*
- *The density of X is bounded and bounded away from 0 on $\mathcal{X}$.*
- *$Y(1), Y(0)$ are distributed respectively as F_1 and F_0, which are defined over a common compact support $\mathcal{Y}$.*
- *F_1 and F_0 are continuos differentiable functions on $\mathcal{Y}$ with $F_1(0) = F_0(0) = 0$*
- *For any given x in $\mathcal{X}$, $F_0(y|x)$ and $F_1(y|x)$ are continuous in $y \in \mathcal{Y}$.*

We require that the density of X to be bounded in the entire support $\mathcal{X}$, and that all the covariates are continuous. In addition, we require that $Y(1), Y(0)$ to be continuos random variables.

2 Empirical Distribution Function

Next result is a slight generalization of Theorem 1 of Hirano, Imbens and Ridder (2003). We use the same decomposition as the addendum of Hirano, Imbens and Ridder (2003).

Lemma S. 1 *Under assumptions 1, 2, 3, 4 and 5, for j=0,1 and l=U,C*

$$\sup_{y \in \mathcal{Y}} \left| \sqrt{N} \left(\widehat{F}_Y^{\widehat{\omega}_{jl}}(y) - F_Y^{\omega_{jl}}(y) \right) - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\psi_{jl}(Y_i, X_i, T_i, y) - F_Y^{\omega_{jl}}(y) \right) \right| = o_p(1)$$

where $\psi_{jl}(Y_i, X_i, T_i, y) = \omega_{jl}(T_i, p(X_i)) \cdot \mathbb{I}\{Y_i \leq y\} + \mathbb{E}\left[\frac{\partial \omega_{jl}(T_i, p(X_i))}{\partial p(X_i)} \cdot \mathbb{I}\{Y_i \leq y\} \middle| X_i \right] (T_i - p(X_i))$.

Proof. *We fix the subscripts j and l and for notational simplicity drop them. Also, we fix notation to what had been used by Hirano, Imbens and Ridder (2003), so our estimator of the propensity score is $\widehat{p}(\cdot) = \widehat{p}_K(\cdot) \equiv \Lambda\left(H_K(\cdot)^\top \widehat{\pi}_K\right)$. In fact, we set notation as in HIR to simplify the comparison of*

the decomposition presented below with results presented in their appendix. We break down the difference $\sqrt{N}\left(\widehat{F}_Y^{\widehat{\omega}}(y)-F_Y^{\omega}(y)\right)$ *into six components:*

$$
\begin{aligned}
& \sqrt{N}\left(\widehat{F}_Y^{\widehat{\omega}}(y)-F_Y^{\omega}(y)\right) \\
= \quad & \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\omega\left(T_i,\widehat{p}_K\left(X_i\right)\right)\mathbb{I}\left\{Y_i\leq y\right\}-\omega\left(T_i,p\left(X_i\right)\right)\mathbb{I}\left\{Y_i\leq y\right\}\right. \qquad (1)\\
& \left.+\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\left(\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
& +\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(-\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\left(\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right)\right.\\
& \left.+\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T_i,p\left(x\right)\right)}{\partial p\left(x\right)}\cdot\mathbb{I}\left\{Y_i\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p\left(x\right)\right)dF\left(x\right)\right) \qquad (2)
\end{aligned}
$$

$$
\begin{aligned}
& -\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\cdot\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p\left(x\right)\right)dF\left(x\right)\\
& +\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{\Psi}_K\left(X_i\right)\left(\frac{T_i-p_K\left(X_i\right)}{\sqrt{p_K\left(X_i\right)\left(1-p_K\left(X_i\right)\right)}}\right) \qquad (3)
\end{aligned}
$$

$$
\begin{aligned}
& +\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\widetilde{\Psi}_K\left(X_i\right)-\Psi_K\left(X_i\right)\right)\left(\frac{T_i-p_K\left(X_i\right)}{\sqrt{p_K\left(X_i\right)\left(1-p_K\left(X_i\right)\right)}}\right) \qquad (4)\\
& +\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\Psi_K\left(X_i\right)\left(\frac{T_i-p_K\left(X_i\right)}{\sqrt{p_K\left(X_i\right)\left(1-p_K\left(X_i\right)\right)}}\right)-\Psi_0\left(X_i\right)\left(\frac{T_i-p\left(X_i\right)}{\sqrt{p\left(X_i\right)\left(1-p\left(X_i\right)\right)}}\right)\right) \qquad (5)\\
& +\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\omega\left(T_i,p\left(X_i\right)\right)\cdot\mathbb{I}\left\{Y_i\leq y\right\}-F_Y^{\omega}(y)\right)+\Psi_0\left(X_i\right)\left(\frac{T_i-p\left(X_i\right)}{\sqrt{p\left(X_i\right)\left(1-p\left(X_i\right)\right)}}\right) \qquad (6)
\end{aligned}
$$

with

$$\widetilde{\Psi}_K(x)=-\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T_i,p\left(z\right)\right)}{\partial p\left(z\right)}\cdot\mathbb{I}\left\{Y_i\leq y\right\}\middle|z\right]\Lambda'\left(H_K(z)^{\top}\widetilde{\pi}_K\right)H_K(z)^{\top}dF(z)\,\widetilde{\Sigma}_K^{-1}\sqrt{\Lambda'\left(H_K(x)^{\top}\pi_K\right)}H_K(x) \qquad (7)$$

$$\Psi_K(x)=-\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T_i,p\left(z\right)\right)}{\partial p\left(z\right)}\cdot\mathbb{I}\left\{Y_i\leq y\right\}\middle|z\right]\Lambda'\left(H_K(z)^{\top}\pi_K\right)H_K(z)^{\top}dF(z)\,\Sigma_K^{-1}\sqrt{\Lambda'\left(H_K(x)^{\top}\pi_K\right)}H_K(x) \qquad (8)$$

$$\Psi_0\left(X_i\right)=-\mathbb{E}\left[\frac{\partial\omega\left(T_i,p\left(x\right)\right)}{\partial p\left(x\right)}\cdot\mathbb{I}\left\{Y_i\leq y\right\}\middle|x\right]\sqrt{p\left(X_i\right)\left(1-p\left(X_i\right)\right)} \qquad (9)$$

and

$$\Sigma_K=\mathbb{E}\left[\Lambda'\left(H_K(X)^{\top}\pi_K\right)H_K(X)\,H_K(X)^{\top}\right]$$

$$\widetilde{\Sigma}_K=\frac{1}{N}\sum_{i=1}^{N}\Lambda'\left(H_K\left(X_i\right)^{\top}\widetilde{\pi}_K\right)H_K\left(X_i\right)H_K\left(X_i\right)^{\top},$$

where $\widetilde{\pi}_K$ *lies between* $\widehat{\pi}_K$ *and* π_K.

Now we show that each term can be bounded uniformly in y.
By Taylor expansion,

$$\omega\left(T_i,\widehat{p}_K\left(X_i\right)\right)=\omega\left(T_i,p\left(X_i\right)\right)+\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\left(\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right)+o_p\left(\left\|\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right\|^2\right).$$

We can rewrite (1) as,

$$\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\left(\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right)\right)+o_p\left(\sqrt{N}\left\|\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right\|^2\right)$$

$$= \frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\left(\widehat{p}_K\left(X_i\right)-p_K\left(X_i\right)\right)^2\right) \tag{10}$$

$$+\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\left(p_K\left(X_i\right)-p\left(X_i\right)\right)^2\right) \tag{11}$$

$$+\frac{4}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\left(\widehat{p}_K\left(X_i\right)-p_K\left(X_i\right)\right)\left(p_K\left(X_i\right)-p\left(X_i\right)\right)\right) \tag{12}$$

$$+o_p\left(\sqrt{N}\sup_x\left\|\widehat{p}_K\left(x\right)-p\left(x\right)\right\|^2\right).$$

By the mean value theorem,

$$\widehat{p}_K\left(x\right)-p_K\left(x\right)=\Lambda^{'}\left(H_K\left(x\right)^{\top}\widetilde{\pi}_K\right)H_K\left(x\right)^{\top}\left(\widehat{\pi}_K-\pi_K\right)$$

and we can write (10) as

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\Lambda^{'}\left(H_K\left(x\right)^{\top}\widetilde{\pi}_K\right)^2\left(H_K\left(x\right)^{\top}\left(\widehat{\pi}_K-\pi_K\right)\right)^2\right).$$

Note that because $p\left(x\right)$ is bounded away from zero and is less than one, for all $x\in X$ and $T=0,1$,

$$\left|\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\right|\leq C.$$

Using this result,

$$\begin{aligned}
&\sup_{y\in\mathcal{Y}}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\Lambda^{'}\left(H_K\left(X_i\right)^{\top}\widetilde{\pi}_K\right)^2\left(H_K\left(X_i\right)^{\top}\left(\widehat{\pi}_K-\pi_K\right)\right)^2\right)\right| \\
\leq\ &\sup_{y\in\mathcal{Y}}\left|\frac{C}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{I}\left\{Y_i\leq y\right\}\left(H_K\left(X_i\right)^{\top}\left(\widehat{\pi}_K-\pi_K\right)\right)^2\right| \\
\leq\ &C\sqrt{N}\zeta\left(K\left(N\right)\right)^2\left\|\widehat{\pi}_K-\pi_K\right\|^2
\end{aligned}$$

where $\zeta(K) = \sup_x \|H_K(x)\|$. *Next, using Lemma 1 in the appendix of HIR, (11) is bounded by*

$$\begin{aligned}\sup_{y\in\mathcal{Y}}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\partial\omega(T_i,p(X_i))}{\partial p(X_i)}\mathbb{I}\{Y_i\le y\}(p_K(X_i)-p(X_i))^2\right)\right| &\le \frac{C}{\sqrt{N}}\sup_{y\in\mathcal{Y}}\left|\sum_{i=1}^{N}\mathbb{I}\{Y_i\le y\}\zeta(K(N))^2K^{-\frac{sp}{r}}\right|\\ &\le C\sqrt{N}\zeta(K(N))^2K^{-\frac{sp}{r}}.\end{aligned}$$

Using again Lemma 1 in the appendix of HIR, we can bound the last term in the expansion,

$$\begin{aligned}&\sup_{y\in\mathcal{Y}}\left|\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left(\frac{\partial\omega(T_i,p(X_i))}{\partial p(X_i)}\mathbb{I}\{Y_i\le y\}(\widehat{p}_K(X_i)-p_K(X_i))(p_K(X_i)-p(X_i))\right)\right|\\ \le\;& \frac{C}{\sqrt{N}}\sum_{i=1}^{N}|\widehat{p}_K(X_i)-p_K(X_i)|\,|p_K(X_i)-p(X_i)|\\ \le\;& \frac{C\zeta(K(N))}{\sqrt{N}}K^{-\frac{sp}{2r}}\sum_{i=1}^{N}\left|\Lambda'\left(H_K(X_i)^\top\widetilde{\pi}_K\right)H_K(X_i)^\top(\widehat{\pi}_K-\pi_K)\right|\\ \le\;& \sqrt{N}C\zeta(K(N))^2K^{-\frac{sp}{2r}}\|\widehat{\pi}_K-\pi_K\|.\end{aligned}$$

Note that $o_p\left(\sqrt{N}\sup_x\|\widehat{p}_K(x)-p(x)\|^2\right)$ *converges in probability to zero for all sequences of* $K(N)$ *that satisfy the large sample identification condition in HIR,* $\frac{\zeta(K(N))^4}{N}\to 0$ *as* $N\to\infty$. *Combining the expressions above, by Markov inequality and using the results of Lemma 2 at the appendix of HIR,*

$$\begin{aligned}&\sup_{y\in\mathcal{Y}}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\omega(T_i,\widehat{p}_K(X_i))\mathbb{I}\{Y_i\le y\}-\omega(T_i,p(X_i))\mathbb{I}\{Y_i\le y\}+\frac{\partial\omega(T_i,p(X_i))}{\partial p(X_i)}\mathbb{I}\{Y_i\le y\}(\widehat{p}_K(X_i)-p(X_i))\right)\right|\\ =\;& O_p\left(\frac{\zeta(K(N))^3}{\sqrt{N}}\right)+O_p\left(\sqrt{N}\zeta(K(N))^2K^{-\frac{sp}{r}}\right)+O_p\left(\zeta(K(N))^{\frac{5}{2}}K^{-\frac{sp}{2r}}\right).\end{aligned}$$

Now, we find a bound on (2). First, we rewrite this term as

$$\begin{aligned}&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(-\frac{\partial\omega(T_i,p(X_i))}{\partial p(X_i)}\mathbb{I}\{Y_i\le y\}(\widehat{p}_K(X_i)-p_K(X_i))\right. &(13)\\ &\left.+\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega(T,p(x))}{\partial p(x)}\cdot\mathbb{I}\{Y\le y\}\middle|x\right](\widehat{p}_K(x)-p_K(x))\,dF(x)\right)\\ &+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(-\frac{\partial\omega(T_i,p(X_i))}{\partial p(X_i)}\mathbb{I}\{Y_i\le y\}(p_K(X_i)-p(X_i))\right. &(14)\\ &\left.+\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega(T,p(x))}{\partial p(x)}\cdot\mathbb{I}\{Y\le y\}\middle|x\right](p_K(x)-p(x))\,dF(x)\right)\end{aligned}$$

and denote (14) by V_K*. Note that* $\mathbb{E}[V_K]=0$ *and that*

$$\begin{aligned}
\mathbb{V}ar\left[V_K\right] &= \mathbb{E}\left[\mathbb{V}ar\left[\left(\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\right)\mathbb{I}\left\{Y\leq y\right\}\middle| X\right]\left(p_K\left(X\right)-p\left(X\right)\right)^2\right] \\
&\quad +\mathbb{V}ar\left[\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\mathbb{I}\left\{Y\leq y\right\}\middle| X\right]\left(p_K\left(X\right)-p\left(X\right)\right)\right] \\
&\leq \mathbb{E}\left[\mathbb{E}\left[\left(\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\right)^2\mathbb{I}\left\{Y\leq y\right\}\middle| X\right]\left(p_K\left(X\right)-p\left(X\right)\right)^2\right] \\
&\leq C\zeta\left(K\left(N\right)\right)^2K^{-\frac{s_p}{r}}
\end{aligned}$$

and

$$\mathbb{E}\left[\left|V_K\right|\right]\leq\sqrt{\mathbb{V}ar\left[V_K\right]}\leq C\zeta\left(K\left(N\right)\right)K^{-\frac{s_p}{2r}}$$

and

$$\sup_{y\in\mathcal{Y}}\left|V_K\right|=O_p\left(\zeta\left(K\left(N\right)\right)K^{-\frac{s_p}{2r}}\right).$$

Now consider (13), by the mean value theorem this term can be rewritten as

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[-\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\Lambda'\left(H_K\left(X_i\right)^\top\widetilde{\pi}_K\right)H_K\left(X_i\right)^\top\right. \\
&\left.+\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle| x\right]\Lambda'\left(H_K\left(x\right)^\top\widetilde{\pi}_K\right)H_K\left(x\right)^\top dF\left(x\right)\right]\left(\widehat{\pi}_K-\pi_K\right).
\end{aligned}$$

By a second application of the mean value theorem, we write the first term in the expression above as

$$W_{1K(N)}-W_{2K(N)}+W_{3K(N)}$$

with

$$\begin{aligned}
W_{1K(N)} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[-\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\Lambda'\left(H_K\left(X_i\right)^\top\pi_K\right)H_K\left(X_i\right)^\top\right. \\
&\quad\left.+\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle| x\right]\Lambda'\left(H_K\left(x\right)^\top\pi_K\right)H_K\left(x\right)^\top dF\left(x\right)\right]
\end{aligned}$$

$$W_{2K(N)}=\frac{1}{2\sqrt{N}}\sum_{i=1}^{N}\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\Lambda''\left(H_K\left(X_i\right)^\top\widetilde{\widetilde{\pi}}_K\right)H_K\left(X_i\right)H_K\left(X_i\right)^\top\left(\widetilde{\pi}_K-\pi_K\right)$$

$$W_{3K(N)}=\frac{1}{2\sqrt{N}}\sum_{i=1}^{N}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle| x\right]\Lambda''\left(H_K\left(x\right)^\top\widetilde{\widetilde{\pi}}_K\right)H_K\left(x\right)H_K\left(x\right)^\top dF\left(x\right)\left(\widetilde{\pi}_K-\pi_K\right)$$

where $\widetilde{\widetilde{\pi}}_K$ *lies between* $\widetilde{\pi}_K$ *and* π_K. *First, we calculate the variance of* $W_{1K(N)}$,

$$\begin{aligned}
\mathbb{V}ar\left[W_{1K(N)}\right] &= \mathbb{E}\left[\mathbb{V}ar\left[-\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\mathbb{I}\left\{Y\leq y\right\}\middle| X\right]\left(\Lambda'\left(H_K\left(X\right)^\top\pi_K\right)\right)^2 H_K\left(X\right)H_K\left(X\right)^\top\right] \\
&\quad +\mathbb{V}ar\left[\mathbb{E}\left[-\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\mathbb{I}\left\{Y\leq y\right\}\middle| X\right]\Lambda'\left(H_K\left(X\right)^\top\pi_K\right)H_K\left(X\right)^\top\right] \\
&\leq \mathbb{E}\left[\mathbb{E}\left[\left(\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\right)^2\mathbb{I}\left\{Y\leq y\right\}\middle| X\right]\left(\Lambda'\left(H_K\left(X\right)^\top\pi_K\right)\right)^2 H_K\left(X\right)H_K\left(X\right)^\top\right] \\
&\leq C\mathbb{E}\left[H_K\left(X\right)H_K\left(X\right)^\top\right]
\end{aligned}$$

and hence

$$\begin{aligned}
\mathbb{E}\left[\left\|W_{1K(N)}\right\|\right] &\leq \sqrt{tr\left(\mathbb{V}ar\left[W_{1K(N)}\right]\right)} \\
&\leq C\sqrt{tr\left(\mathbb{E}\left[H_K\left(X\right)H_K\left(X\right)^\top\right]\right)} \\
&\leq C\zeta\left(K\left(N\right)\right). \qquad (15)
\end{aligned}$$

Now, working with $W_{2K(N)}$, *we first notice that for* (Y,X,T) *and all* y,

$$\begin{aligned}
&\left\|\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\mathbb{I}\left\{Y\leq y\right\}\Lambda''\left(H_K\left(X\right)^\top\widetilde{\widetilde{\pi}}_K\right)H_K\left(X\right)H_K\left(X\right)^\top\left(\widetilde{\pi}_K-\pi_K\right)\right\| \\
\leq &\left|\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\mathbb{I}\left\{Y\leq y\right\}\Lambda''\left(H_K\left(X\right)^\top\widetilde{\widetilde{\pi}}_K\right)\right|\left\|H_K\left(X\right)H_K\left(X\right)^\top\right\|\left\|\widetilde{\pi}_K-\pi_K\right\|
\end{aligned}$$

and note that,

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\mathbb{I}\left\{Y\leq y\right\}\Lambda''\left(H_K\left(X\right)^\top\widetilde{\widetilde{\pi}}_K\right)\right|\left\|H_K\left(X\right)H_K\left(X\right)^\top\right\|\left\|\widetilde{\pi}_K-\pi_K\right\|\right] \\
\leq\ & C\frac{\zeta\left(K\left(N\right)\right)^{\frac{3}{2}}}{\sqrt{N}}
\end{aligned}$$

and,

$$\mathbb{E}\left[\left\|W_{2K(N)}\right\|\right]\leq C\zeta\left(K\left(N\right)\right)^{\frac{3}{2}}.$$

By analogy we work with $W_{3K(N)}$,

$$\begin{aligned}
&\left\|\frac{\sqrt{N}}{2}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle| x\right]\Lambda''\left(H_K\left(x\right)'\widetilde{\widetilde{\pi}}_K\right)H_K\left(x\right)H_K\left(x\right)^\top dF\left(x\right)\left(\widetilde{\pi}_K-\pi_K\right)\right\| \\
\leq\ & C\zeta\left(K\left(N\right)\right)^{\frac{3}{2}}.
\end{aligned}$$

Using the triangle and Cauchy-Schwartz inequality,

$$\mathbb{E}\left[\left|\left(W_{1K(N)}-W_{2K(N)}+W_{3K(N)}\right)^\top\left(\widehat{\pi}_K-\pi_K\right)\right|\right]\leq C\frac{\zeta\left(K\left(N\right)\right)^2}{\sqrt{N}}.$$

Combining the bounds above, we have

$$\begin{aligned}
&\sup_{y\in\mathcal{Y}}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(-\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\mathbb{I}\left\{Y_i\leq y\right\}\left(\widehat{p}_K\left(X_i\right)-p\left(X_i\right)\right)\right.\right.\\
&\quad\left.\left.+\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p\left(x\right)\right)dF\left(x\right)\right)\right|\\
=&\quad O_p\left(\zeta\left(K\left(N\right)\right)K^{-\frac{s_p}{2r}}\right)+O_p\left(\frac{\zeta\left(K\left(N\right)\right)^2}{\sqrt{N}}\right).
\end{aligned}$$

Now, we work with (3). Note that

$$\begin{aligned}
&-\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p\left(x\right)\right)dF\left(x\right)\\
=&-\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p_K\left(x\right)\right)dF\left(x\right)\\
&-\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(p_K\left(x\right)-p\left(x\right)\right)dF\left(x\right).
\end{aligned}$$

Using the mean value expansion we obtain,

$$\begin{aligned}
&-\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_K\left(x\right)-p_K\left(x\right)\right)dF\left(x\right)\\
=&-\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\Lambda'\left(H_K\left(x\right)^{\top}\widetilde{\pi}_K\right)H_K\left(x\right)^{\top}dF\left(x\right)\sqrt{N}\left(\widehat{\pi}_K-\pi_K\right)
\end{aligned}$$

and define

$$\widetilde{\Xi}_K\equiv\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\Lambda'\left(H_K\left(x\right)^{\top}\widetilde{\pi}_K\right)H_K\left(x\right)^{\top}dF\left(x\right).$$

Using this definition, we can rewrite (7) as

$$\begin{aligned}
\widetilde{\Psi}_K\left(x\right)&=-\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(z\right)\right)}{\partial p\left(z\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|z\right]\Lambda'\left(H_K\left(z\right)^{\top}\widetilde{\pi}_K\right)H_K\left(z\right)^{\top}dF\left(z\right)\Sigma_K^{-1}\sqrt{\Lambda'\left(H_K\left(x\right)^{\top}\pi_K\right)}H_K\left(x\right)\\
&=-\widetilde{\Xi}_K\widetilde{\Sigma}_K^{-1}\sqrt{\Lambda'\left(H_K\left(x\right)^{\top}\pi_K\right)}H_K\left(x\right).
\end{aligned}$$

The first order condition of the pseudo-maximum likelihood approach to calculate $\widehat{\pi}_K$ is

$$\sum_{i=1}^{N}H_K\left(X_i\right)\cdot\left(T_i-\Lambda\left(H_K\left(X_i\right)^{\top}\widehat{\pi}_K\right)\right)=0$$

and by the mean value theorem,

$$\begin{aligned}
\sqrt{N}\left(\widehat{\pi}_K-\pi_K\right)&=\left(\frac{1}{N}\sum_{i=1}^{N}H_K\left(X_i\right)H_K\left(X_i\right)^{\top}\Lambda'\left(H_K\left(X_i\right)^{\top}\widetilde{\pi}_K\right)\right)^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}H_K\left(X_i\right)\cdot\left(T_i-p_K\left(X_i\right)\right)\right)\\
&=\widetilde{\Sigma}_K^{-1}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}H_K\left(X_i\right)\cdot\left(T_i-p_K\left(X_i\right)\right)\right)
\end{aligned}$$

and

$$\begin{aligned}
&-\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_{K}\left(x\right)-p_{K}\left(x\right)\right)dF\left(x\right)\\
=&\ -\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\Lambda^{'}\left(H_{K}\left(x\right)^{\top}\widetilde{\pi}_{K}\right)H_{K}\left(x\right)^{\top}dF\left(x\right)\sqrt{N}\left(\widehat{\pi}_{K}-\pi_{K}\right)\\
=&\ -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{\Xi}_{K}\cdot\widetilde{\Sigma}_{K}^{-1}\cdot\sqrt{\Lambda'\left(H_{K}\left(X_{i}\right)^{\top}\pi_{K}\right)}\cdot H_{K}\left(X_{i}\right)\cdot\frac{\left(T_{i}-p_{K}\left(X_{i}\right)\right)}{\sqrt{\Lambda'\left(H_{K}\left(X_{i}\right)^{\top}\pi_{K}\right)}}\\
=&\ -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{\Psi}_{K}\left(X_{i}\right)\cdot\frac{\left(T_{i}-p_{K}\left(X_{i}\right)\right)}{\sqrt{\Lambda'\left(H_{K}\left(X_{i}\right)^{\top}\pi_{K}\right)}}\\
=&\ -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{\Psi}_{K}\left(X_{i}\right)\cdot\frac{\left(T_{i}-p_{K}\left(X_{i}\right)\right)}{\sqrt{p_{K}\left(X_{i}\right)\left(1-p_{K}\left(X_{i}\right)\right)}}
\end{aligned}$$

and hence

$$\begin{aligned}
&\left|-\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(\widehat{p}_{K}\left(x\right)-p\left(x\right)\right)dF\left(x\right)+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widetilde{\Psi}_{K}\left(X_{i}\right)\cdot\frac{\left(T_{i}-p_{K}\left(X_{i}\right)\right)}{\sqrt{p_{K}\left(X_{i}\right)\left(1-p_{K}\left(X_{i}\right)\right)}}\right|\\
=&\ \left|-\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\left(p_{K}\left(x\right)-p\left(x\right)\right)dF\left(x\right)\right|.
\end{aligned}$$

Since $\mathbb{E}\left[\frac{\partial\omega(T,p(x))}{\partial p(x)}\middle|x\right]$ *is bounded away on* $\mathcal{X}$ *we have that*

$$\begin{aligned}
&\sup_{y\in\mathcal{Y}}\left|-\sqrt{N}\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|X\right]\left(p_{K}\left(x\right)-p\left(x\right)\right)dF\left(x\right)\right|\\
\leq&\ C\sqrt{N}\zeta\left(K\left(N\right)\right)K^{-\frac{sp}{2r}}=O\left(\sqrt{N}\zeta\left(K\left(N\right)\right)K^{-\frac{sp}{2r}}\right).
\end{aligned}$$

Now we work with (4). Let us define

$$\Xi_{K}\equiv\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(x\right)\right)}{\partial p\left(x\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|x\right]\Lambda^{'}\left(H_{K}\left(x\right)^{\top}\pi_{K}\right)H_{K}\left(x\right)^{\top}dF\left(x\right).$$

Using this definition, we can rewrite (8) as

$$\begin{aligned}
\Psi_{K}\left(x\right)=&\ -\int_{\mathcal{X}}\mathbb{E}\left[\frac{\partial\omega\left(T,p\left(z\right)\right)}{\partial p\left(z\right)}\mathbb{I}\left\{Y\leq y\right\}\middle|z\right]\Lambda^{'}\left(H_{K}\left(z\right)^{\top}\pi_{K}\right)H_{K}\left(z\right)^{\top}dF\left(z\right)\Sigma_{K}^{-1}\sqrt{\Lambda'\left(H_{K}\left(x\right)^{\top}\pi_{K}\right)}H_{K}\left(x\right)\\
=&\ -\Xi_{K}\Sigma_{K}^{-1}\sqrt{\Lambda'\left(H_{K}\left(x\right)^{\top}\pi_{K}\right)}H_{K}\left(x\right).
\end{aligned}$$

Therefore, (4) can be rewritten:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\widetilde{\Psi}_{K}\left(X_{i}\right)-\Psi_{K}\left(X_{i}\right)\right)\left(\frac{T_{i}-p_{K}\left(X_{i}\right)}{\sqrt{p_{K}\left(X_{i}\right)\left(1-p_{K}\left(X_{i}\right)\right)}}\right)=\left(\Xi_{K}\Sigma_{K}^{-1}-\widetilde{\Xi}_{K}\widetilde{\Sigma}_{K}^{-1}\right)\cdot B_{K(N)}$$

with

$$B_{K(N)} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} H_K(X_i)(T_i - p_K(X_i)).$$

Note that

$$\left(\Xi_K \Sigma_K^{-1} - \widetilde{\Xi}_K \widetilde{\Sigma}_K^{-1}\right) \cdot B_{K(N)} = \left(\Xi_K - \widetilde{\Xi}_K\right) \widetilde{\Sigma}_K^{-1} B_{K(N)} + \Xi_K \left(\Sigma_K^{-1} - \widetilde{\Sigma}_K^{-1}\right) B_{K(N)}$$

and working with the first term we obtain

$$\left|\left(\Xi_K - \widetilde{\Xi}_K\right) \widetilde{\Sigma}_K^{-1} B_{K(N)}\right| \leq \frac{1}{\lambda_{\min}\left(\widetilde{\Sigma}_K\right)} \left\|B_{K(N)}\right\| \left\|\Xi_K - \widetilde{\Xi}_K\right\|$$

where $\lambda_{\min}(A)$ *is the smallest eigenvalue of a matrix* A*. By the mean value theorem and using the fact that* $p(x)$ *is bounded away from zero,*

$$\begin{aligned} & \left\|\Xi_K - \widetilde{\Xi}_K\right\| \\ = & \left\|\int_{\mathcal{X}} \mathbb{E}\left[\frac{\partial\omega(T, p(z))}{\partial p(z)} \mathbb{I}\{Y \leq y\} \middle| z\right] \left(\Lambda^{'}\left(H_K(z)^\top \pi_K\right) - \Lambda^{'}\left(H_K(z)^\top \widetilde{\pi}_K\right)\right) H_K(z)^\top dF(z)\right\| \\ \leq & \; C \int_{\mathcal{X}} \left|\acute{\Lambda}\left(H_K(z)^\top \widetilde{\widetilde{\pi}}_K\right)\right| \|H_K(z)\|^2 dF(z) \|\pi_K - \widetilde{\pi}_K\| \\ \leq & \; C \int_{\mathcal{X}} \|H_K(z)\|^2 dF(z) \|\pi_K - \widetilde{\pi}_K\| \leq C\zeta(K(N))^2 \|\pi_K - \widetilde{\pi}_K\| \end{aligned}$$

and

$$\left|\left(\Xi_K - \widetilde{\Xi}_K\right) \widetilde{\Sigma}_K^{-1} B_{K(N)}\right| \leq C\zeta(K(N))^2 \|\pi_K - \widetilde{\pi}_K\| \left\|B_{K(N)}\right\|.$$

Now, we work with the second term

$$\begin{aligned} \left|\Xi_K \left(\Sigma_K^{-1} - \widetilde{\Xi}_K\right) B_{K(N)}\right| &= \left|\Xi_K \Sigma_K^{-1} \left(\widetilde{\Sigma}_K - \Sigma_K\right) \widetilde{\Sigma}_K^{-1} B_{K(N)}\right| \\ &\leq \frac{1}{\lambda_{\min(\widetilde{\Sigma}_K)}} \left\|B_{K(N)}\right\| \left\|\left(\widetilde{\Sigma}_K - \Sigma_K\right) \Sigma_K^{-1} \Xi_K\right\|. \end{aligned}$$

Define $\widehat{\Sigma}_K = \frac{1}{N} \sum_{i=1}^{N} \Lambda^{'}\left(H_K(X_i)^\top \pi_K\right) H_K(X_i) H_K(X_i)^\top$ *and* $W_K = \Sigma_K^{-1} \Xi_K$*,*

$$\begin{aligned} & \left\|\left(\widetilde{\Sigma}_K - \Sigma_K\right) W_K\right\| \\ \leq & \left\|\left(\widetilde{\Sigma}_K - \widehat{\Sigma}_K\right) W_K\right\| + \left\|\left(\widehat{\Sigma}_K - \Sigma_K\right) W_K\right\| \\ \leq & \frac{1}{N} \sum_{i=1}^{N} \left\|H_K(X_i)^\top (\widetilde{\pi}_K - \pi_K) \Lambda^{''}\left(H_K(X_i)^\top \widetilde{\widetilde{\pi}}_K\right) H_K(X_i) H_K(X_i)^\top W_K\right\| + \\ & \frac{1}{\sqrt{N}} \left\|\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\Lambda^{'}\left(H_K(X_i)^\top \pi_K\right) H_K(X_i) H_K(X_i)^\top - \mathbb{E}\left[\Lambda^{'}\left(H_K(X)^\top \pi_K\right) H_K(X) H_K(X)^\top\right]\right) W_K\right\| \\ \leq & \; C\zeta(K(N))^3 \|\pi_K - \widetilde{\pi}_K\| \|W_K\| + \\ & \frac{1}{\sqrt{N}} \left\|\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\Lambda^{'}\left(H_K(X_i)^\top \pi_K\right) H_K(X_i) H_K(X_i)^\top - \mathbb{E}\left[\Lambda^{'}\left(H_K(X)^\top \pi_K\right) H_K(X) H_K(X)^\top\right]\right) W_K\right\|. \end{aligned}$$

Note that

$$\begin{aligned} \|W_K\| &\leq C\|\Xi_K\| \\ &\leq C\zeta(K(N)) \end{aligned}$$

and that

$$\begin{aligned} & \left\| \mathbb{V}ar\left[\Lambda'\left(H_K(X)^\top \pi_K\right) H_K(X) H_K(X)^\top W_K\right]\right\| \\ = \; & W_K^\top \mathbb{V}ar\left[\Lambda'\left(H_K(X)^\top \pi_K\right) H_K(X) H_K(X)^\top\right] W_K \\ \leq \; & W_K^\top \mathbb{E}\left[\left(\Lambda'\left(H_K(X)^\top \pi_K\right) H_K(X) H_K(X)^\top\right)^2\right] W_K \\ \leq \; & C W_K^\top \mathbb{E}\left[\left(H_K(X) H_K(X)^\top\right)^2\right] W_K \\ \leq \; & C\zeta(K(N))^4. \end{aligned}$$

Now, we look at $B_{K(N)}$

$$B_{K(N)} = \frac{1}{\sqrt{N}}\sum_{i=1}^N H_K(X_i)(T_i - p(X_i)) + \frac{1}{\sqrt{N}}\sum_{i=1}^N H_K(X_i)(p(X_i) - p_K(X_i)).$$

Since $p(x)$ *is bounded away from zero and one,*

$$\begin{aligned} \mathbb{V}ar\left[B_{K(N)}\right] &\leq \mathbb{E}\left[p(X)(1-p(X)) H_K(X) H_K(X)^\top\right] + \mathbb{E}\left[(p(X) - p_K(X))^2 H_K(X) H_K(X)^\top\right] \\ &\leq C_1\zeta(K(N))^2 + C_2\zeta(K(N))^4 K^{-\frac{sp}{r}} \end{aligned}$$

and

$$\mathbb{E}\left[\left\|B_{K(N)}\right\|\right] \leq C\zeta(K(N))^2.$$

By Cauchy-Schwartz inequality we have

$$\mathbb{E}\left[\left|\left(\Xi_K - \widetilde{\Xi}_K\right)\widetilde{\Sigma}_K^{-1} B_{K(N)}\right|\right] \leq \frac{C\zeta(K(N))^{\frac{7}{2}}}{\sqrt{N}}$$

$$\mathbb{E}\left[\left|\Xi_K\left(\Sigma_K^{-1} - \widetilde{\Xi}_K\right) B_{K(N)}\right|\right] \leq C_1\frac{\zeta(K(N))^{\frac{11}{2}}}{\sqrt{N}} + C_2\frac{\zeta(K(N))^3}{\sqrt{N}}.$$

At the end,

$$\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N \left(\widetilde{\Psi}_K(X_i) - \Psi_K(X_i)\right)\left(\frac{T_i - p_K(X_i)}{\sqrt{p_K(X_i)(1-p_K(X_i))}}\right)\right| = O_p\left(\frac{\zeta(K(N))^{\frac{11}{2}}}{\sqrt{N}}\right).$$

Now we work with (5). Since (5) is a sum of iid random variables with mean different than zero, we

bound this term by deriving the order of its second moment,

$$\begin{aligned}
&\mathbb{E}\left[\left(\Psi_K(X)\left(\frac{T-p_K(X)}{\sqrt{p_K(X)(1-p_K(X))}}\right)-\Psi_0(X)\left(\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)\right)^2\right]\\
=\;&\mathbb{E}\left[\left(\Psi_K(X)\left(\frac{T-p_K(X)}{\sqrt{p_K(X)(1-p_K(X))}}\right)-\Psi_K(X)\left(\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)\right)^2\right]\\
&+\mathbb{E}\left[\left(\Psi_K(X)\left(\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)-\Psi_0(X)\left(\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)\right)^2\right]\\
&+2\cdot\mathbb{E}\left[\left(\Psi_K(X)\left(\frac{T_i-p_K(X)}{\sqrt{p_K(X)(1-p_K(X))}}\right)-\Psi_K(X)\left(\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)\right)\cdot\right.\\
&\left.\left(\Psi_K(X)\left(\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)-\Psi_0(X)\left(\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)\right)\right]\\
=\;&\mathbb{E}\left[\Psi_K(X)^2\left(\frac{T-p_K(X)}{\sqrt{p_K(X)(1-p_K(X))}}-\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)^2\right]\\
&+\mathbb{E}\left[(\Psi_K(X)-\Psi_0(X))^2\left(\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)^2\right]\\
&+2\mathbb{E}\left[\Psi_K(X)(\Psi_K(X)-\Psi_0(X))\left(\frac{T-p_K(X)}{\sqrt{p_K(X)(1-p_K(X))}}-\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)\left(\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)\right].
\end{aligned}$$

We can approximate $\Psi_K(x)$ as the least squares projection of $\Psi_0(x)$ on $H_K(x)\sqrt{p_K(x)(1-p_K(x))}$. If we assume that $\Psi_0(x)$ is t times continuously differentiable, and using Lemma 1 of HIR we have,

$$\sup_{x\in\mathcal{X}}|\Psi_0(x)-\Psi_K(x)|<CK^{-\frac{t}{r}}$$

and hence

$$\begin{aligned}
\mathbb{E}\left[(\Psi_K(X)-\Psi_0(X))^2\left(\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)^2\right] &= \mathbb{E}\left[(\Psi_K(X)-\Psi_0(X))^2\right]\\
&\le CK^{-\frac{2t}{r}}.
\end{aligned}$$

Now, consider the first term in the expansion,

$$\begin{aligned}
&\mathbb{E}\left[\Psi_K(X)^2\left(\frac{T-p_K(X)}{\sqrt{p_K(X)(1-p_K(X))}}-\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)^2\right]\\
=\;&\mathbb{E}\left[\Psi_K(X)^2\frac{(p_K(X)-p(X))^2}{p_K(X)(1-p_K(X))}\right]+\mathbb{E}\left[\Psi_K(X)^2\left(\frac{\sqrt{p(X)(1-p(X))}}{\sqrt{p_K(X)(1-p_K(X))}}-1\right)^2\right].
\end{aligned}$$

Using the approximation of $\Psi_K(x)$,

$$\Psi_K(x)^2\le\Psi_0(x)^2+|\Psi_0(x)|CK^{-\frac{t}{r}}$$

and therefore

$$
\begin{aligned}
& \mathbb{E}\left[\Psi_K(X)^2 \frac{(p_K(X)-p(X))^2}{p_K(X)(1-p_K(X))}\right] \\
\leq\ & \mathbb{E}\left[\Psi_0(X)^2 \frac{(p_K(X)-p(X))^2}{p_K(X)(1-p_K(X))}\right] + CK^{-\frac{sp}{r}} \mathbb{E}\left[\left|\Psi_0(X)\right| \frac{(p_K(X)-p(X))^2}{p_K(X)(1-p_K(X))}\right] \\
\leq\ & \mathbb{E}\left[\mathbb{E}\left[\frac{\partial \omega(T,p(X))}{\partial p(X)} \cdot \mathbb{I}\{Y \leq y\} \middle| X\right]^2 \frac{p(X)(1-p(X))}{p_K(X)(1-p_K(X))} (p_K(X)-p(X))^2\right] \\
& + CK^{-\frac{sp}{r}} \mathbb{E}\left[\left|\mathbb{E}\left[\frac{\partial \omega(T,p(X))}{\partial p(X)} \cdot \mathbb{I}\{Y \leq y\} \middle| X\right]\right| \frac{\sqrt{p(X)(1-p(X))}(p_K(X)-p(X))^2}{p_K(X)(1-p_K(X))}\right].
\end{aligned}
$$

Since $p(x)$ *is bounded from 0 and 1 on* $\mathcal{X}$ *and using Lemma 1 in the appendix of HIR,*

$$
\begin{aligned}
& \mathbb{E}\left[\left|\mathbb{E}\left[\frac{\partial w(T_i,p(x))}{\partial p(x)} \cdot \mathbb{I}\{Y_i \leq y\} \middle| X\right]\right| \frac{\sqrt{p(X_i)(1-p(X_i))}(p_K(X_i)-p(X_i))^2}{p_K(X_i)(1-p_K(X_i))}\right] \\
\leq\ & C\mathbb{E}\left[\frac{(p_K(X_i)-p(X_i))^2}{p_K(X_i)(1-p_K(X_i))}\right] \leq C\zeta(K(N))^2 K^{-\frac{sp}{r}}
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}\left[\Psi_K(X)^2 \frac{(p_K(X)-p(X))^2}{p_K(X)(1-p_K(X))}\right] &\leq C_1\zeta(K(N))^2 K^{-\frac{sp}{r}} + C_2\zeta(K(N))^2 K^{-\frac{sp}{r}-\frac{t}{r}} \\
&\leq C\zeta(K(N))^2 K^{-\frac{sp}{r}}.
\end{aligned}
$$

By analogy,

$$
\mathbb{E}\left[\Psi_K(X)^2 \left(\frac{\sqrt{p(X)(1-p(X))}}{\sqrt{p_K(X)(1-p_K(X))}} - 1\right)^2\right] \leq C\zeta(K(N))^2 K^{-\frac{sp}{r}}
$$

and since $\Psi_K(x)$ *is bounded and* $p_K(x)$ *and* $p(x)$ *are bounded away from* 0 *and* 1*,*

$$
\begin{aligned}
& \mathbb{E}\left[\Psi_K(X)(\Psi_K(X)-\Psi_0(X))\left(\frac{T-p_K(X)}{\sqrt{p_K(X)(1-p_K(X))}} - \frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)\left(\frac{T-p(X)}{\sqrt{p(X)(1-p(X))}}\right)\right] \\
\leq\ & C \cdot \mathbb{E}\left[\left|\Psi_K(X)-\Psi_0(X)\right|\right] \\
\leq\ & CK^{-\frac{t}{r}}.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
& \sup_{y\in\mathcal{Y}}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\Psi_K(X_i)\left(\frac{T_i-p_K(X_i)}{\sqrt{p_K(X_i)(1-p_K(X_i))}}\right) - \Psi_0(X_i)\left(\frac{T_i-p(X_i)}{\sqrt{p(X_i)(1-p(X_i))}}\right)\right)\right| \\
=\ & O_p\left(\max\left(K^{-\frac{t}{2r}}, \zeta(K(N))K^{-\frac{sp}{2r}}\right)\right).
\end{aligned}
$$

Combining the bounds, we have

$$\begin{aligned}
&\sup_{y\in\mathcal{Y}}\left|\sqrt{N}\left(\widehat{F}_Y^{\widehat{\omega}}(y)-F_Y^{\omega}(y)\right)-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\left(\omega\left(T_i,p\left(X_i\right)\right)\cdot\mathbb{I}\left\{Y_i\le y\right\}-F_Y^{\omega}(y)\right)-\Psi_0\left(X_i\right)\left(\frac{T_i-p\left(X_i\right)}{\sqrt{p\left(X_i\right)\left(1-p\left(X_i\right)\right)}}\right)\right\}\right| \\
=\;& O_p\left(\frac{\zeta\left(K\left(N\right)\right)^3}{\sqrt{N}}\right)+O_p\left(\sqrt{N}\zeta\left(K\left(N\right)\right)^2K^{-\frac{s_p}{r}}\right)+O_p\left(\zeta\left(K\left(N\right)\right)^{\frac{5}{2}}K^{-\frac{s_p}{2r}}\right) \\
&+O_p\left(\zeta\left(K\left(N\right)\right)K^{-\frac{s_p}{2r}}\right)+O_p\left(\frac{\zeta\left(K\left(N\right)\right)^2}{\sqrt{N}}\right) \\
&+O\left(\sqrt{N}\zeta\left(K\left(N\right)\right)K^{-\frac{s_p}{2r}}\right)+O_p\left(\frac{\zeta\left(K\left(N\right)\right)^{\frac{11}{2}}}{\sqrt{N}}\right)+O_p\left(\max\left(K^{-\frac{t}{2r}},\zeta\left(K\left(N\right)\right)K^{-\frac{s_p}{2r}}\right)\right) \\
=\;& O_p\left(\sqrt{N}\zeta\left(K\left(N\right)\right)^2K^{-\frac{s_p}{r}}\right)+O_p\left(\zeta\left(K\left(N\right)\right)^{\frac{5}{2}}K^{-\frac{s_p}{2r}}\right)+O_p\left(\frac{\zeta\left(K\left(N\right)\right)^{\frac{11}{2}}}{\sqrt{N}}\right).
\end{aligned}$$

And under the assumptions on $\zeta\left(K\left(N\right)\right)$ and s_p, this sum is $o_p\left(1\right)$. ■

Let $\mathcal{P}$ be the joint distribution of (Y,X,T). Define the collection of measurable functions from $(\mathcal{Y}\times\mathcal{X}\times\{0,1\})\to\mathbb{R}$, $\mathcal{H}=\left\{\psi\left(Y,X,T,y\right)\mid y\in\mathcal{Y}\right\}$. Next, we show that this class $\mathcal{H}$ of measurable functions is P-Donsker.

Lemma S. 2 *Under assumptions 2, 3, 5, for j=0,1 and l=U,C, $\mathcal{H}_{jl}=\left\{\psi_{jl}\left(Y,X,T,y\right)\mid y\in\mathcal{Y}\right\}$ is Donsker.*
Proof. *Again we fix j and l. Then, the measurable collection of functions $\mathcal{W}=\left\{\mathbb{I}\left\{Y\le y\right\}\mid y\in\mathcal{Y}\right\}$ is Donsker since the bracketing number $N_{[]}\left(\sqrt{\varepsilon},\mathcal{W},L_2\left(P\right)\right)\le\frac{2}{\varepsilon}$ are of the polynomial order $\left(\frac{1}{\varepsilon}\right)^2$. The bracketing integral is finite since it converges at a slower rate than $\left(\frac{1}{\varepsilon}\right)^2$.*

Now, we work the measurable functions $\mathcal{K}=\left\{F_Y\left(y\mid X\right)\mid y\in\mathcal{Y}\right\}$. Using the proof of Lemma A2 in Donald and Hsu (2010), we can claim that $\mathcal{K}$ is Donsker. In this proof, Donald and Hsu (2010) show that $N_{[]}\left(\varepsilon,\mathcal{K},L_2\left(P\right)\right)\le 1+\left(\frac{1}{\varepsilon}\right)^2$, and the bracketing integral is finite since it converges at a slower rate than $\left(\frac{1}{\varepsilon}\right)^2$.

Since $p\left(x\right)$ is bounded away from zero and one in $\mathcal{X}$, $d_1\left(T,X\right)\equiv\omega\left(T,p\left(X\right)\right)$ is a uniformly bounded measurable function, and $d_1\left(T,X\right)\cdot\mathcal{W}$ is Donsker. Similarly, define $d_2\left(T,X\right)\equiv\mathbb{E}\left[\left.\frac{\partial\omega(T,p(X))}{\partial p(X)}\right|X\right]\left(T-p\left(X\right)\right)$, $d_2\left(T,X\right)$ is a uniformly bounded measurable function, and $d_2\left(T,X\right)\cdot\mathcal{K}$ is Donsker.

Hence, $\mathcal{H}=\left\{d_1\left(T,X\right)\cdot\mathcal{W}+d_2\left(T,X\right)\cdot\mathcal{K}\mid y\in\mathcal{Y}\right\}$ is Donsker. ■

Theorem 1 *Suppose assumptions 1, 2, 3, 4 and 5 hold, then for l=U,C*

$$\sqrt{N}\begin{bmatrix}\widehat{F}_Y^{\widehat{\omega}_{1l}}-F_Y^{\omega_{1l}}\\ \widehat{F}_Y^{\widehat{\omega}_{0l}}-F_Y^{\omega_{0l}}\end{bmatrix}=\mathbb{G}_N^{\omega_l}+o_p\left(1\right)\Rightarrow\mathbb{G}^{\omega_l}$$

where
(i) $\mathbb{G}_N^{\omega_l}=\begin{bmatrix}\mathbb{G}_N^{\omega_{1l}}\\ \mathbb{G}_N^{\omega_{0l}}\end{bmatrix}$, and $\mathbb{G}_N^{\omega_{jl}}$ is a empirical process such that at a given $y\in\mathcal{Y}$, for j=0,1

$$\mathbb{G}_N^{\omega_{jl}}\left(y\right)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\psi_{jl}\left(Y_i,X_i,T_i,y\right)-F_Y^{\omega_{jl}}\left(y\right)\right);$$

(ii) $\Rightarrow$ denotes weak convergence;
(iii) $\mathbb{G}^{\omega_l}$ is a Gaussian process with variance-covariance matrix given by

$$\mathbb{E}\left[\mathbb{G}^{\omega_{jl}}\left(s\right)\mathbb{G}^{\omega_{kl}}\left(t\right)\right]=\mathbb{E}\left[\left(\psi_{jl}\left(Y,X,T,s\right)-F_Y^{\omega_{jl}}\left(s\right)\right)\cdot\left(\psi_{kl}\left(Y,X,T,t\right)-F_Y^{\omega_{kl}}\left(t\right)\right)\right]$$

for $(s,t) \in \mathcal{Y} \times \mathcal{Y}$, j and k=0,1.

Proof. *Let $\mathbb{P}_N$ be the empirical measure of the sample (Y, X, T). Using Lemma 2 and Donsker's Theorem (Theorem 19.3 at page 266 of Van der Vaart,1998), $\sqrt{N}\begin{pmatrix} \widehat{F}_Y^{\widehat{\omega}1l} - F_Y^{\omega 1l} \\ \widehat{F}_Y^{\widehat{\omega}0l} - F_Y^{\omega 0l} \end{pmatrix}$ converges to a zero mean Gaussian process, defined by $\mathbb{G}^{\omega_l}$, with variance-covariance matrix*

$$\mathbb{E}\left[\mathbb{G}^{\omega_{jl}}(s)\,\mathbb{G}^{\omega_{kl}}(t)\right] = \mathbb{E}\left[\left(\psi_{jl}(Y,X,T,s) - F_Y^{\omega_{jl}}(s)\right)\cdot\left(\psi_{kl}(Y,X,T,t) - F_Y^{\omega_{kl}}(t)\right)\right],$$

for $(s,t) \in \mathcal{Y} \times \mathcal{Y}$, where j and $k = 0, 1$. ■

3 Estimator of the Inequality Measures: Large Sample Properties

In this section we derive the large sample properties of the estimators of the inequality treatment effects. We show that our estimators are asymptotically normal and efficient. We work only with inequality treatment effects that differences in inequality measures that are Hadamard differentiable functionals of the distribution.

Assumption 6 [***Hadamard***] *The inequality measure ν defined over $\mathcal{F}_\nu$ is Hadamard differentiable.*

We start by deriving the distribution of a Hadamard differentiable functional at $\widehat{F}_Y^{\widehat{\omega}jl}$ for $j = 0, 1$, and $l = U, C$.

Lemma S. 3 (*Semiparametric Efficiency*) *For $j = 0, 1$, and $l = U, C$, if assumptions 1, 2, 3, 4, 5 and 6 hold, then,*

$$\begin{aligned}
& \sqrt{N}\left(\nu\left(\widehat{F}_Y^{\widehat{\omega}_{jl}}\right) - \nu\left(F_Y^{\omega_{jl}}\right)\right) \\
= \; & \psi^\nu\left(\mathbb{G}_N^{\omega_{jl}}; F_Y^{\omega_{jl}}\right) + o_p(1) \\
= \; & \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega\left(T_i, p(X_i)\right)\cdot\phi^\nu\left(Y_i; F_Y^\omega\right) \\
& + \mathbb{E}\left[\frac{\partial\omega\left(T, p(X_i)\right)}{\partial p(X_i)}\cdot\phi^\nu\left(Y; F_Y^\omega\right)\middle| X_i\right]\left(T_i - p(X_i)\right) + o_p(1)
\end{aligned}$$

where ψ^ν is the functional ν's Hadamard derivative, $\phi^\nu(Y_i;\cdot) = \psi^\nu(\delta_{Y_i};\cdot)$ and δ_{Y_i} is the Dirac measure at observation i. Moreover, $\nu\left(\widehat{F}_Y^{\widehat{\omega}_{jl}}\right)$ is asymptotically efficient.

Proof. *Again, we fix j and l and therefore drop these subscripts. Using results in Theorem 1,*

$$\sqrt{N}\left(\widehat{F}_Y^{\widehat{\omega}} - F_Y^\omega\right) = \mathbb{G}_N^\omega + o_p(1) \Rightarrow \mathbb{G}^\omega$$

where $\mathbb{G}^\omega$ is a Gaussian process with variance-covariance matrix given by

$$\mathbb{E}\left[\mathbb{G}^\omega\mathbb{G}^\omega\right](s,t) = \mathbb{E}\left[\left(\psi(Y,X,T,s) - F_Y^\omega(s)\right)\cdot\left(\psi(Y,X,T,t) - F_Y^\omega(t)\right)\right]$$

for $(s,t) \in \mathcal{Y} \times \mathcal{Y}$. Because the map $\nu : \mathcal{F}_\nu \to \mathbb{R}$ is Hadamard differentiable at $F_Y^\omega \in \mathcal{F}_\nu$ we can apply van der Vaart's (1998) Theorem 20.8:

$$\sqrt{N}\left(\nu\left(\widehat{F}_Y^{\widehat{\omega}}\right) - \nu\left(F_Y^\omega\right)\right) = \psi^\nu\left(\mathbb{G}_N^\omega; F_Y^\omega\right) + o_p(1).$$

And since ν is Hadamard differentiable, its functional derivative $\psi^{\nu}\left(\cdot;F_Y^{\omega}\right)$ is linear, implying that

$$\begin{aligned}\psi^{\nu}\left(\mathbb{G}_N^{\omega};F_Y^{\omega}\right) &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega\left(T_i,p\left(X_i\right)\right)\cdot\phi^{\nu}\left(Y_i;F_Y^{\omega}\right) \\ &\quad +\mathbb{E}\left[\left.\frac{\partial\omega\left(T,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\cdot\phi^{\nu}\left(Y;F_Y^{\omega}\right)\right|X_i\right]\left(T_i-p\left(X_i\right)\right).\end{aligned}$$

We now establish efficiency for $\nu\left(\widehat{F}_Y^{\widehat{\omega}}\right)$ as an estimator for $\nu\left(F_Y^{\omega}\right)$. Consider a (regular) parametric submodel of the joint distribution of (Y,X,T) with cdf $F\left(y,x,t;\theta\right)$. The log-likelihood is

$$\begin{aligned}\ln f\left(y,x,t|\theta\right) &= t\left[\ln f_1\left(y|x,\theta\right)+\ln p\left(x|\theta\right)\right]+\left(1-t\right)\left[\ln f_0\left(y|x,\theta\right)+\ln\left(1-p\left(x|\theta\right)\right)\right]+\ln f\left(x|\theta\right) \\ &= t\ln f_1\left(y|x,\theta\right)+\left(1-t\right)\ln f_0\left(y|x,\theta\right)+t\ln p\left(x|\theta\right)+\left(1-t\right)\ln\left(1-p\left(x|\theta\right)\right)+\ln f\left(x|\theta\right)\end{aligned}$$

where for $j=0,1$ we use the ignorability assumption to write $f\left(y|x,T=j;\theta\right)$ as $f_j\left(y|x,\theta\right)$, which is the conditional density of $Y\left(j\right)$ given x for parameter value θ. Following results in Hahn (1998), we have that the corresponding score function:

$$S\left(y,x,t|\theta\right)=ts_1\left(y|x;\theta\right)+\left(1-t\right)s_0\left(y|x;\theta\right)+\frac{dp\left(x|\theta\right)}{d\theta}\frac{\left(t-p\left(x|\theta\right)\right)}{p\left(x|\theta\right)\left(1-p\left(x|\theta\right)\right)}+s\left(x|\theta\right)$$

where for $j=0,1$, $s_j\left(y|x;\theta\right)\equiv d\ln f_j\left(y|x;\theta\right)/d\theta$ and $s\left(x|\theta\right)\equiv d\ln f\left(x|\theta\right)/d\theta$. The tangent space for this model is:

$$\mathcal{L}=\left\{S\left(y,x,t\right):S\left(y,x,t\right)=ts_1\left(y|x\right)+\left(1-t\right)s_0\left(y|x\right)+a\left(x\right)\left(t-p\left(x\right)\right)+s\left(x\right)\right\}$$

where $a\left(x\right)$ is a square-integrable function of x,

$$\int s_j\left(y|x\right)f_j\left(y|x;\theta_0\right)=0,\ \forall x,\ j=0,1,$$

$$\int s\left(x|\theta\right)f\left(x|\theta_0\right)=0,$$

and for notational simplicity, $p\left(\cdot|\theta_0\right)=p\left(\cdot\right)$.

A parameter $\rho\left(\theta\right)$ is pathwise differentiable if there is a differentiable zero-mean function $F_{\rho}\left(Y,X,T\right)$ such that $F_{\rho}\left(Y,X,T\right)\in\mathcal{L}$ and for all regular parametric models:

$$\left.\frac{\partial\rho\left(\theta\right)}{\partial\theta}\right|_{\theta=\theta_0}=\mathbb{E}\left[F_{\rho}\left(Y,X,T\right)\cdot S\left(Y,X,T|\theta_0\right)\right]$$

We specialize $\rho\left(\theta\right)$ to $F_Y^{\omega}\left(y;\theta\right)$, the weighted distribution of Y at a given $y\in\mathcal{Y}$, which can be written as

$$\begin{aligned}&F_Y^{\omega}\left(y;\theta\right) \\ =\ &\int\left(\int\mathbb{I}\{z\leq y\}f_1\left(z|x,\theta\right)dz\right)\omega\left(1,p\left(x|\theta\right)\right)p\left(x|\theta\right)f\left(x|\theta\right)dx \\ &+\int\left(\int\mathbb{I}\{z\leq y\}f_0\left(z|x,\theta\right)dz\right)\omega\left(0,p\left(x|\theta\right)\right)\left(1-p\left(x|\theta\right)\right)f\left(x|\theta\right)dx\end{aligned}$$

and calculate its derivative with respect to θ, which can be written as

$$
\begin{aligned}
& \frac{\partial F_Y^{\omega}(y;\theta)}{\partial\theta} \\
= & \int\left(\int \mathbb{I}\{z\leq y\} s_1(z|x,\theta)\, f_1(z|x,\theta)\,dz\right)\omega(1,p(x|\theta))\,p(x|\theta)\,f(x|\theta)\,dx \\
& +\int\left(\int \mathbb{I}\{z\leq y\} s_0(z|x,\theta)\, f_0(z|x,\theta)\,dz\right)\omega(0,p(x|\theta))\,(1-p(x|\theta))\,f(x|\theta)\,dx \\
& \\
& +\int\left(\int \mathbb{I}\{z\leq y\} f_1(z|x,\theta)\,dz\right)\omega(1,p(x|\theta))\,\frac{dp(x|\theta)}{d\theta}f(x|\theta)\,dx \\
& -\int\left(\int \mathbb{I}\{z\leq y\} f_0(z|x,\theta)\,dz\right)\omega(0,p(x|\theta))\,\frac{dp(x|\theta)}{d\theta}f(x|\theta)\,dx \\
& \\
& +\int\left(\int \mathbb{I}\{z\leq y\} f_1(z|x,\theta)\,dz\right)\omega(1,p(x|\theta))\,p(x|\theta)\,s(x|\theta)\,f(x|\theta)\,dx \\
& +\int\left(\int \mathbb{I}\{z\leq y\} f_0(z|x,\theta)\,dz\right)\omega(0,p(x|\theta))\,(1-p(x|\theta))\,s(x|\theta)\,f(x|\theta)\,dx \\
& \\
& +\int\left(\int \mathbb{I}\{z\leq y\} f_1(z|x,\theta)\,dz\right)\frac{\partial\omega(1,p(x|\theta))}{\partial p(x|\theta)}\frac{dp(x|\theta)}{d\theta}p(x|\theta)\,f(x|\theta)\,dx \\
& +\int\left(\int \mathbb{I}\{z\leq y\} f_0(z|x,\theta)\,dz\right)\frac{\partial\omega(0,p(x|\theta))}{\partial p(x|\theta)}\frac{dp(x|\theta)}{d\theta}(1-p(x|\theta))\,f(x|\theta)\,dx.
\end{aligned}
$$

Now, evaluating that derivative at θ_0 we have

$$
\begin{aligned}
\left.\frac{\partial F_Y^{\omega}(y;\theta)}{\partial\theta}\right|_{\theta=\theta_0} &= \mathbb{E}\left[\omega(T,p(X))\,\mathbb{I}\{Y\leq y\}S(Y,X,T|\theta_0)\right] \\
&+\mathbb{E}\left[\mathbb{E}\left[\left.\frac{\partial\omega(T,p(X))}{\partial p(X)}\cdot\mathbb{I}\{Y\leq y\}\right|X\right]\left.\frac{dp(X|\theta)}{d\theta}\right|_{\theta=\theta_0}\right]
\end{aligned}
$$

and after some tedious algebra one can show that

$$
\begin{aligned}
& \mathbb{E}\left[\mathbb{E}\left[\left.\frac{\partial\omega(T,p(X))}{\partial p(X)}\cdot\mathbb{I}\{Y\leq y\}\right|X\right]\left.\frac{dp(X|\theta)}{d\theta}\right|_{\theta=\theta_0}\right] \\
= & \mathbb{E}\left[\mathbb{E}\left[\left.\frac{\partial\omega(T,p(X))}{\partial p(X)}\cdot\mathbb{I}\{Y\leq y\}\right|X\right](T-p(X))\,S(Y,X,T|\theta_0)\right]
\end{aligned}
$$

and therefore

$$
\left.\frac{\partial F_Y^{\omega}(y;\theta)}{\partial\theta}\right|_{\theta=\theta_0} = \mathbb{E}\left[F_{\rho}(Y,X,T)\cdot S(Y,X,T|\theta_0)\right]
$$

where, for $\rho(\theta) = F_Y^{\omega}(y;\theta)$, we have that

$$
\begin{aligned}
& F_{\rho}(Y,X,T) \\
= \; & \omega(T,p(X))\, \mathbb{I}\{Y \leq y\} - F_Y^{\omega}(y;\theta_0) \\
& + \mathbb{E}\left[\frac{\partial \omega(T,p(X))}{\partial p(X)} \mathbb{I}\{Y \leq y\} \right| X \bigg] (T - p(X)) \\
= \; & \psi(Y,X,T,y) - F_Y^{\omega}(y;\theta_0).
\end{aligned}
$$

We now show that $F_{\rho}(Y,X,T) \in \mathcal{L}$. We rewrite $\psi(Y,X,T,y)$ as

$$
\begin{aligned}
& \psi(Y,X,T,y) \\
= \; & T\omega(1,p(X)) \left(\mathbb{I}\{Y \leq y\} - F_{Y|X,T}(y|X,1;\theta_0) \right) \\
& + (1-T)\,\omega(0,p(X)) \left(\mathbb{I}\{Y \leq y\} - F_{Y|X,T}(y|X,0;\theta_0) \right) \\
& + (\mathbb{E}\left[\frac{\partial \omega(T,p(X))}{\partial p(X)} \mathbb{I}\{Y \leq y\} \right| X \bigg] \\
& + \omega(1,p(X))\, F_{Y|X,T}(y|X,1;\theta_0) - \omega(0,p(X))\, F_{Y|X,T}(y|X,0;\theta_0)\,)\,(T - p(X)) \\
& + p(X)\,\omega(1,p(X))\, F_{Y|X,T}(y|X,1;\theta_0) + (1-p(X))\,\omega(0,p(X))\, F_{Y|X,T}(y|X,0;\theta_0)
\end{aligned}
$$

and we can easily check that for $j = 0,1$:

$$
\mathbb{E}\left[\omega(j,p(X)) \left(\mathbb{I}\{Y \leq y\} - F_{Y|X,T}(y|X,j;\theta_0) \right) |X, T=j \right] = 0
$$

and

$$
F_Y^{\omega}(y;\theta_0) = \mathbb{E}\left[F_{Y|X,T}(y|X,1;\theta_0)\,\omega(1,p(X))\,p(X) + F_{Y|X,T}(y|X,0;\theta_0)\,\omega(0,p(X))\,(1-p(X)) \right].
$$

According to Bickel, Klaassen, Ritov and Wellner (1993), because the estimator for F_Y^{ω}, $\widehat{F}_Y^{\widehat{\omega}}$, is asymptotically linear with influence function $F_{\rho}(Y,X,T) \in \mathcal{L}$, $\widehat{F}_Y^{\widehat{\omega}}$ is also asymptotically efficient at $\mathcal{P}$, the joint distribution of (Y,X,T). Therefore, since ν is Hadamard differentiable, by theorem 25.47 of van der Vaart (1998) $\nu\left(\widehat{F}_Y^{\widehat{\omega}}\right)$ is asymptotically efficient at $\mathcal{P}$ for estimating $\nu(F_Y^{\omega})$. ■

Theorem 2 *Under assumptions 1, 2, 3, 4, 5 and 6, for l=U,C*

$$
\sqrt{N}\left(\widehat{\Delta}_l - \Delta_l \right) = \psi^{\nu}\left(\mathbb{G}_N^{\omega 1l}; F_Y^{\omega 1l} \right) - \psi^{\nu}\left(\mathbb{G}_N^{\omega 0l}; F_Y^{\omega 0l} \right) + o_p(1).
$$

In addition, $\widehat{\Delta}_l$ is asymptotically efficient for Δ_l.
Proof. *This result follows mechanically from previous lemma. By definition, for $l = U, C$,*

$$
\begin{aligned}
\sqrt{N}\left(\widehat{\Delta}_l - \Delta_l \right) \; & \equiv \; \sqrt{N}\left(\nu\left(\widehat{F}_Y^{\widehat{\omega}1l} \right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}0l} \right) - \left(\nu\left(F_Y^{\omega 1l} \right) - \nu\left(F_Y^{\omega 0l} \right) \right) \right) \\
& = \; \psi^{\nu}\left(\mathbb{G}_N^{\omega 1l}; F_Y^{\omega 1l} \right) - \psi^{\nu}\left(\mathbb{G}_N^{\omega 0l}; F_Y^{\omega 0l} \right) + o_p(1).
\end{aligned}
$$

Now, for efficiency, let us define the functional vector $\boldsymbol{\nu} : \mathcal{F}_v \to \mathbb{R}^2$. Thus,

$$
\begin{aligned}
& \sqrt{N}\left(\nu\left(\widehat{F}_Y^{\widehat{\omega}1l} \right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}0l} \right) - \left(\nu\left(F_Y^{\omega 1l} \right) - \nu\left(F_Y^{\omega 0l} \right) \right) \right) \\
= \; & \sqrt{N} \begin{bmatrix} 1 \\ -1 \end{bmatrix}^{\top} \left(\boldsymbol{\nu}\left(\begin{bmatrix} \widehat{F}_Y^{\widehat{\omega}1l} \\ \widehat{F}_Y^{\widehat{\omega}0l} \end{bmatrix} \right) - \boldsymbol{\nu}\left(\begin{bmatrix} F_Y^{\omega 1l} \\ F_Y^{\omega 0l} \end{bmatrix} \right) \right).
\end{aligned}
$$

Result from previous lemma allows us to write

$$\partial \left[\begin{array}{c} F_Y^{\omega_{1l}}(y;\theta) \\ F_Y^{\omega_{0l}}(y;\theta) \end{array} \right] \Big/ \partial\theta \Big|_{\theta=\theta_0} = \mathbb{E}\left[\left[\begin{array}{c} \left(\psi_{1l}(Y,X,T,y) - F_Y^{\omega_{1l}}(y;\theta_0)\right) \\ \left(\psi_{0l}(Y,X,T,y) - F_Y^{\omega_{0l}}(y;\theta_0)\right) \end{array} \right] S(Y,X,T|\theta_0) \right]$$

and therefore we have that $v\left(\left[\begin{array}{c} \widehat{F}_Y^{\widehat{\omega}_{1l}} \\ \widehat{F}_Y^{\widehat{\omega}_{0l}} \end{array}\right]\right)$ *is efficient for* $v\left(\left[\begin{array}{c} F_Y^{\omega_{1l}} \\ F_Y^{\omega_{0l}} \end{array}\right]\right)$. *Moreover,* $\left[\begin{array}{c} 1 \\ -1 \end{array}\right]^\top \left(v\left(\left[\begin{array}{c} \widehat{F}_Y^{\widehat{\omega}_{1l}} \\ \widehat{F}_Y^{\widehat{\omega}_{0l}} \end{array}\right]\right)\right)$ *is efficient for* $\left[\begin{array}{c} 1 \\ -1 \end{array}\right]^\top \left(v\left(\left[\begin{array}{c} F_Y^{\omega_{1l}} \\ F_Y^{\omega_{0l}} \end{array}\right]\right)\right)$. ■

4 Estimator of ITE Measures: Inference

In this section, we show the validity of some inference procedures for our estimators. We provide a consistent estimator for the asymptotic variance of our inequality treatment effect estimators and show that we can also use bootstrap to estimate that variance.

From theorems 1, 2 and Lemma 3 the analytical expression for the asymptotic variance of $\sqrt{N}\left(\widehat{\Delta}_l - \Delta_l\right)$, $l = U, C$ is

$$\begin{aligned} V_l &= \mathbb{E}[(\omega_{1l}(T,p(X)) \cdot \phi^\nu(Y;F_Y^{\omega_{1l}}) + g_{1l}(X)(T-p(X)) \\ &= -(\omega_{0l}(T,p(X)) \cdot \phi^\nu(Y;F_Y^{\omega_{0l}}) + g_{0l}(X)(T-p(X))))^2] \end{aligned}$$

where for $j = 0,1$,

$$g_{jl}(X) \equiv \mathbb{E}\left[\frac{\partial\omega_{jl}(T,p(X))}{\partial p(X)} \cdot \phi^\nu\left(Y;F_Y^{\omega_{jl}}\right)\Big| X\right].$$

As in Hirano, Imbens and Ridder (2003), we replace $p(X)$ by $\widehat{p}(X)$. In addition, our estimator of $g_{jl}(\cdot)$ is $\widehat{g}_{jl}(\cdot)$, a nonparametric regression of $\frac{\partial\omega_{jl}(T,p(X))}{\partial p(X)}\Big|_{p(X)=\widehat{p}(X)} \cdot \phi^\nu\left(Y;\widehat{F}_Y^{\widehat{\omega}_{jl}}\right)$ on X using series. The series estimator is written as

$$\widehat{g}_{jl}(\cdot) = H_{K_{jl}}(\cdot)^\top \widehat{\gamma}_K^{\omega_{jl}},$$

where

$$\widehat{\gamma}_K^{\omega_{jl}} = \arg\min_\gamma \sum_{i=1}^N \left(\frac{\partial\omega_{jl}(T_i,p(X))}{\partial p(X)}\Big|_{p(X)=\widehat{p}(X_i)} \cdot \phi^\nu\left(Y_i;\widehat{F}_Y^{\widehat{\omega}_{jl}}\right) - H_{K_{jl}}(X_i)^\top\gamma\right)^2.$$

In this case, we use a series of orthonormal polynomials such that

$$\sup_{x\in\mathcal{X}} \left\|H_{K_{jl}}(x)\right\| = \zeta(K_{jl}) \leq CK_{jl}$$

where $H_{K_{jl}}(\cdot)$ needs to satisfy the following properties: (i) $H_{K_{jl}}(\cdot):\mathcal{X}\to\mathbb{R}^{K_{jl}}$; (ii) $H_{K_{jl},1}(\cdot)=1$ and (iii) if $K_{jl} > (n_{jl}+1)^r$, $H_{K_{jl}}(\cdot)$ includes all the polynomials up to order n_{jl}. In order to derive the large sample properties of the estimator of the conditional expectation, we need to control how K_{jl} increases with N. We impose the following assumptions.

Assumption 7 [***Series Estimator***] *For all* $x\in\mathcal{X}$*:*

- $g_{jl}(x)$ *is bounded and* s_{jl} *times continuously differentiable;*
- *The order of* $H_{K_{jl}}(x)$, K_{jl}, *is of the form* $K_{jl} = C\cdot N^{c_{jl}}$ *where* C *is a constant and* $c_{jl}\in\left(0,\frac{1}{2}\left(\frac{s_p}{2r}-1\right)c_p\right)$
- $\sup_y\left|\frac{\partial\phi^\nu(y;z)}{\partial z}\Big|_{z=F_Y^\omega}\right|\leq M$.

Lemma S. 4 *For j=0,1, and l=U,C, under assumptions 1, 2, 3, 4, 5, 6 and 7*

$$\sup_{x\in\mathcal{X}}\left|g_{jl}\left(x\right)-H_{K_{jl}}\left(x\right)^{\top}\widehat{\gamma}_{K}^{\omega_{jl}}\right|=o_{p}\left(1\right)$$

Proof. *Again, we fix j and l. Let us define $\widetilde{\gamma}_{K}^{\omega}$, $\overline{\gamma}_{K}^{\omega}$ and γ_{K}^{ω}*

$$\widetilde{\gamma}_{K}^{\omega}=\arg\min_{\gamma}\sum_{i=1}^{N}\left(\left.\frac{\partial\omega\left(T_{i},p\left(X\right)\right)}{\partial p\left(X\right)}\right|_{p(X)=p(X_{i})}\cdot\phi^{\nu}\left(Y_{i};\widehat{F}_{Y}^{\widehat{\omega}}\right)-H_{K}\left(X_{i}\right)^{\top}\gamma\right)^{2}$$

$$\overline{\gamma}_{K}^{\omega}=\arg\min_{\gamma}\mathbb{E}\left[\left(\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\cdot\phi^{\nu}\left(Y;\widehat{F}_{Y}^{\widehat{\omega}}\right)-H_{K}\left(X\right)^{\top}\gamma\right)^{2}\right]$$

$$\gamma_{K}^{\omega}=\arg\min_{\gamma}\mathbb{E}\left[\left(\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\cdot\phi^{\nu}\left(Y;F_{Y}^{\omega}\right)-H_{K}\left(X\right)^{\top}\gamma\right)^{2}\right].$$

Using triangle inequality,

$$\begin{aligned}\sup_{x\in\mathcal{X}}\left|g\left(x\right)-H_{K}\left(x\right)^{\top}\widehat{\gamma}_{K}^{\omega}\right| &\leq \sup_{x\in\mathcal{X}}\left|g\left(x\right)-H_{K}\left(x\right)^{\top}\gamma_{K}^{\omega}\right|\\ &\quad+\zeta\left(K\right)\left(\|\gamma_{K}^{\omega}-\overline{\gamma}_{K}^{\omega}\|+\|\overline{\gamma}_{K}^{\omega}-\widetilde{\gamma}_{K}^{\omega}\|+\|\widetilde{\gamma}_{K}^{\omega}-\widehat{\gamma}_{K}^{\omega}\|\right)\end{aligned}$$

where $\zeta\left(K\right)=\sup_{x}\|H_{K}\left(x\right)\|$. First, under the assumption that the function $g\left(\cdot\right)$ is s times continuously differentiable we have that for a fixed K:

$$\sup_{x\in\mathcal{X}}\left|g\left(x\right)-H_{K}\left(x\right)^{\top}\gamma_{K}^{\omega}\right|\leq CK^{-\frac{s}{r}}.$$

We then work with differences in the coefficients:

$$\begin{aligned}\|\gamma_{K}^{\omega}-\overline{\gamma}_{K}^{\omega}\| &= \left\|\mathbb{E}\left[H_{K}\left(X\right)\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\left(\phi^{\nu}\left(Y;F_{Y}^{\omega}\right)-\phi^{\nu}\left(Y;\widehat{F}_{Y}^{\widehat{\omega}}\right)\right)\right]\right\|\\ &\leq C\cdot\zeta\left(K\right)\cdot\left|\mathbb{E}\left[\phi^{\nu}\left(Y;\widehat{F}_{Y}^{\widehat{\omega}}\right)-\phi^{\nu}\left(Y;F_{Y}^{\omega}\right)\right]\right|.\end{aligned}$$

Note that because $p\left(\cdot\right)$ is bounded away from zero and is less than one, for $j=0,1$

$$\sup_{x\in\mathcal{X}}\left|\frac{\partial\omega\left(j,p\left(x\right)\right)}{\partial p\left(x\right)}\right|\leq C_{j}\leq\sup_{j}C_{j}=C$$

and as a result of Theorem 1

$$\begin{aligned}\left|\mathbb{E}\left(\phi^{\nu}\left(Y;\widehat{F}_{Y}^{\widehat{\omega}}\right)-\phi^{\nu}\left(Y;F_{Y}^{\omega}\right)\right)\right| &\leq \mathbb{E}\left[\left|\left.\frac{\partial\phi^{\nu}\left(Y;z\right)}{\partial z}\right|_{z=F_{Y}^{\omega}}\right|\right]\sup_{y\in\mathcal{Y}}\left|\widehat{F}_{Y}^{\widehat{\omega}}\left(y\right)-F_{Y}^{\omega}\left(y\right)\right|+\left|\left(\widehat{F}_{Y}^{\widehat{\omega}}-F_{Y}^{\omega}\right)\right|^{2}\\ &\leq CN^{-1/2}\end{aligned}$$

thus

$$\|\gamma_{K}^{\omega}-\overline{\gamma}_{K}^{\omega}\|\leq C\zeta\left(K\right)N^{-1/2}.$$

Now, the difference

$$
\begin{aligned}
& \left\|\overline{\gamma}_K^{\omega}-\widetilde{\gamma}_K^{\omega}\right\| \\
= & \left\|N^{-1}\sum_{i=1}^{N}\left(H_K\left(X_i\right)\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)\right)-\mathbb{E}\left[H_K\left(X\right)\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right)\right]\right\| \\
\leq & \; CN^{-1/2}\left(\mathbb{V}\left[H_K\left(X\right)\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right)\right]\right)^{1/2} \\
= & \; CN^{-1/2}\left(\mathbb{V}\left[H_K\left(X\right)\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\left(\phi^{\nu}\left(Y;F_Y^{\omega}\right)+\left(\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right)-\phi^{\nu}\left(Y;F_Y^{\omega}\right)\right)\right)\right]\right)^{1/2}
\end{aligned}
$$

and working with the variance

$$
\begin{aligned}
& \mathbb{V}\left[H_K\left(X\right)\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\left(\phi^{\nu}\left(Y;F_Y^{\omega}\right)+\left(\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right)-\phi^{\nu}\left(Y;F_Y^{\omega}\right)\right)\right)\right] \\
\leq & \; \mathbb{V}\left[H_K\left(X\right)\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\phi^{\nu}\left(Y;F_Y^{\omega}\right)\right] \\
& +\mathbb{V}\left[H_K\left(X\right)\frac{\partial\omega\left(T_i,p\left(X\right)\right)}{\partial p\left(X\right)}\left(\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right)-\phi^{\nu}\left(Y;F_Y^{\omega}\right)\right)\right] \\
& +2\left|\mathbb{C}ov\left[H_K\left(X\right)\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\phi^{\nu}\left(Y;F_Y^{\omega}\right),H_K\left(X\right)\frac{\partial\omega\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\left(\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right)-\phi^{\nu}\left(Y;F_Y^{\omega}\right)\right)\right]\right| \\
\leq & \; C\zeta^2\left(K\right)E\left[\left(\phi^{\nu}\left(Y;F_Y^{\omega}\right)\right)^2\right] \\
& +C\zeta^2\left(K\right)E\left[\left(\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right)-\phi^{\nu}\left(Y;F_Y^{\omega}\right)\right)^2\right] \\
& +C\zeta^2\left(K\right)\mathbb{C}ov\left[\phi^{\nu}\left(Y;F_Y^{\omega}\right),\left(\phi^{\nu}\left(Y;\widehat{F}_Y^{\widehat{\omega}}\right)-\phi^{\nu}\left(Y;F_Y^{\omega}\right)\right)\right] \\
= & \; C\zeta^2\left(K\right)N^{-1}.
\end{aligned}
$$

Therefore we have that

$$
\left\|\overline{\gamma}_K^{\omega}-\widetilde{\gamma}_K^{\omega}\right\| = CN^{-1/2}\left(\zeta^2\left(K\right)N^{-1}\right)^{1/2} = CN^{-1}\zeta\left(K\right).
$$

Finally,

$$
\begin{aligned}
& \left\|\widetilde{\gamma}_K^{\omega}-\widehat{\gamma}_K^{\omega}\right\| \\
\leq & \; N^{-1}\left\|\sum_{i=1}^{N}H_K\left(X_i\right)\left(\left(\frac{\partial\omega\left(T_i,\widehat{p}\left(X_i\right)\right)}{\partial p\left(X\right)}-\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X\right)}\right)\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)\right)\right\| \\
\leq & \; C\zeta\left(K\right)N^{-1}\left\|\sum_{i=1}^{N}\left(\left(\frac{\partial\omega\left(T_i,\widehat{p}\left(X_i\right)\right)}{\partial p\left(X\right)}-\frac{\partial\omega\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X\right)}\right)\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)\right)\right\| \\
\leq & \; C\zeta\left(K\right)\sup_{t\in\{0,1\},x\in\mathcal{X}}\left|\frac{\partial^2\omega\left(t,p\left(x\right)\right)}{\partial p^2\left(x\right)}\right|N^{-1}\left\|\sum_{i=1}^{N}\left(\widehat{p}\left(X_i\right)-p\left(X_i\right)\right)\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)\right\|
\end{aligned}
$$

and since the second derivative of with respect to $p(x)$ is bounded,we have that

$$\begin{aligned}\left\|\widetilde{\gamma}_K^{\omega}-\widehat{\gamma}_K^{\omega}\right\| &\leq C\zeta(K)N^{-1}\left\|\sum_{i=1}^{N}\left(\widehat{p}(X_i)-p(X_i)\right)\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)\right\| \\ &\leq C\zeta(K)N^{-1}\left\|\sum_{i=1}^{N}\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)\right\|\left(\sup_{x\in\mathcal{X}}\left(\Lambda^{'}\left(H_{K_\pi}(x)^{\top}\widetilde{\pi}_K\right)H_{K_\pi}(x)^{\top}\right)\left\|\left(\widehat{\pi}_K-\pi_K\right)\right\|+\sup_{x\in\mathcal{X}}\left|p_K(x)-p(x)\right|\right).\end{aligned}$$

Now let us work first with

$$\begin{aligned}&\sup_{x\in\mathcal{X}}\left(\Lambda^{'}\left(H_{K_\pi}(x)^{\top}\widetilde{\pi}_K\right)H_{K_\pi}(x)^{\top}\right)\left\|\left(\widehat{\pi}_K-\pi_K\right)\right\|+\sup_{x\in\mathcal{X}}\left|p_K(x)-p(x)\right| \\ \leq\ &\zeta(K_\pi)\left(K_\pi^{1/2}N^{-1/2}+K_\pi^{-s_p/2r}\right)\end{aligned}$$

and then with

$$\begin{aligned}N^{-1}\left\|\sum_{i=1}^{N}\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)\right\| &\leq N^{-1}\left\|\sum_{i=1}^{N}\phi^{\nu}\left(Y_i;F_Y^{\omega}\right)\right\|+N^{-1}\left\|\sum_{i=1}^{N}\left(\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}}\right)-\phi^{\nu}\left(Y_i;F_Y^{\omega}\right)\right)\right\| \\ &\leq O_p(1)+O_p\left(N^{-1/2}\right).\end{aligned}$$

We reach that

$$\left\|\widetilde{\gamma}_K^{\omega}-\widehat{\gamma}_K^{\omega}\right\|=C\zeta(K)\zeta(K_\pi)\left(K_\pi^{1/2}N^{-1/2}+K_\pi^{-s_p/2r}\right).$$

Therefore, we have that

$$\begin{aligned}&\sup_{x\in\mathcal{X}}\left|g(x)-H_K(x)^{\top}\widetilde{\gamma}_K^{\omega}\right| \\ =\ &O_p\left(K^{-\frac{s}{r}}(N)\right)+O_p\left(\zeta^2(K)N^{-1/2}\right) \\ &+O_p\left(\zeta^2(K(N))N^{-1}\right)+ \\ &+O_p\left(\zeta^2(K(N))\zeta(K_\pi(N))K_\pi^{1/2}N^{-1/2}\right) \\ &+O_p\left(\zeta^2(K(N))\zeta(K_\pi(N))K_\pi^{-s_p/2r}\right) \\ =\ &o_p(1).\end{aligned}$$

■

We use the estimator of the conditional distribution function to propose a estimator for V_l, $l=U,C$:

$$\begin{aligned}\widehat{V}_l &= \frac{1}{N}\sum_{i=1}^{N}\left(\omega_{1l}\left(T_i,\widehat{p}(X_i)\right)\cdot\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}_{1l}}\right)\right. \\ &-\omega_{0l}\left(T,\widehat{p}(X_i)\right)\cdot\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}_{0l}}\right) \\ &+\left(\widehat{g}_{1l}(X_i)-\widehat{g}_{0l}(X_i)\right)\left(T_i-\widehat{p}(X_i)\right)\left.\right)^2].\end{aligned}$$

Theorem 3 *Under assumptions 1, 2, 3, 4, 5, 6 and 7, for $l=U,C$,*

$$\left|\widehat{V}_l-V_l\right|=o_p(1).$$

Proof. *We fix l. By lemmas 1 and 4 we have that*

$$\sup_{x} |\widehat{p}(x) - p(x)| = o_p(1),$$

$$\sup_{x} |\widehat{g}(x) - g(x)| = o_p(1)$$

and for $j = 0, 1$,

$$\sup_{y \in \mathcal{Y}} \left| \widehat{F}_Y^{\widehat{\omega}_j}(y) - F_Y^{\omega_j}(y) \right| = o_p(1).$$

We can rewrite $\widehat{V}$ as

$$\widehat{V} = \frac{1}{N} \sum_{i=1}^{N} h\left(T_i, Y_i, \widehat{p}(X_i), \widehat{g}_1(X_i), \widehat{g}_0(X_i); \widehat{F}_Y^{\widehat{\omega}_1}, \widehat{F}_Y^{\widehat{\omega}_0}\right)$$

where h is a continuously differentiable function with respect to $W = [p(X), g_1(X), g_0(X), F_Y^{\omega_1}, F_Y^{\omega_0}]^\top$. For convenience, define $\widehat{W} = \left[\widehat{p}(X), \widehat{g}_1(X), \widehat{g}_0(X), \widehat{F}_Y^{\widehat{\omega}_1}, \widehat{F}_Y^{\widehat{\omega}_0}\right]^\top$. Thus, a simple linearization of $\widehat{V}$ yields

$$\begin{aligned}
\left|\widehat{V}_l - V_l\right| &\leq \left\|\mathbb{E}\frac{\partial h}{\partial Z}(T_i, Y_i, W_i)\right\| \\
&\quad \cdot \sup_{x \in \mathcal{X}} |\widehat{p}(x) - p(x)| \sup_{x \in \mathcal{X}} |\widehat{g}_1(x) - g_1(x)| \sup_{x \in \mathcal{X}} |\widehat{g}_0(x) - g_0(x)| \\
&\quad \cdot \sup_{y \in \mathcal{Y}} \left|\widehat{F}_Y^{\widehat{\omega}_1}(y) - F_Y^{\omega_1}(y)\right| \sup_{y \in \mathcal{Y}} \left|\widehat{F}_Y^{\widehat{\omega}_0}(y) - F_Y^{\omega_0}(y)\right| + o_p(1) \\
&= o_p(1).
\end{aligned}$$

■

Now we show that in applied work, calculations via bootstrap of standard errors for $\widehat{\Delta}$ are a valid procedure for inference. Consider a random sample $Z = \{(Y_i, X_i, T_i) : i = 1, ..., N\}$ from $\mathcal{P}$, the joint distribution of (Y, X, T). The estimator $\widehat{\Delta}_l = \nu\left(\widehat{F}_Y^{\widehat{\omega}_{1l}}\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_{0l}}\right)$ is a function of the original sample Z.

Suppose that we construct B bootstrap samples, $Z^* = \{Z_b : b = 1, ..., \text{B}\}$, where for each Z_b we randomly draw N observations from Z with replacement, that is, $Z_b = \{(Y_i^*, X_i^*, T_i^*) : i = 1, ..., N\}$. The bootstrap weighted empirical distribution is the empirical measure

$$\widehat{F}_{Yb}^{\widehat{\omega}_{jl}}(y) = N^{-1} \sum_{i=1}^{N} \omega_{jl}(T_i^*, \widehat{p}(X_i^*)) \, \mathbb{I}\{Y_i^* \leq y\}$$

and the bootstrap empirical process is defined as

$$\mathbb{G}_N^{\omega_{jl}*} = \sqrt{N}\left(\widehat{F}_{Yb}^{\widehat{\omega}_{jl}} - \widehat{F}_Y^{\widehat{\omega}_{jl}}\right).$$

The bootstrap estimator of Δ_{lb} is $\widehat{\Delta}_{lb} = \nu\left(\widehat{F}_{Yb}^{\widehat{\omega}_{1lb}}\right) - \nu\left(\widehat{F}_{Yb}^{\widehat{\omega}_{0lb}}\right)$ and the bootstrap variance is

$$V_l^{\mathsf{B}} = \mathbb{E}\left[\left(\widehat{\Delta}_{lb} - \widehat{\Delta}_l\right)^2 \middle| Z\right].$$

Given the B replicate samples, an unbiased estimator for this variance is

$$\widehat{V}_l^{\mathsf{B}} = \frac{1}{\mathrm{B}}\sum_{b=1}^{\mathrm{B}}\left(\widehat{\Delta}_{lb}-\overline{\widehat{\Delta}_l}\right)^2$$

where $\overline{\widehat{\Delta}_l} = \mathrm{B}^{-1}\sum_{b=1}^{\mathrm{B}}\widehat{\Delta}_{lb}$.

Theorem 4 *Suppose that assumptions 1, 2, 3, 4, 5, 6 and 7 hold, then for l=U,C,*

$$\left|\widehat{V}_l^{\mathsf{B}} - V_l\right| \to_p 0$$

Proof. *We fix l. From lemmas 1 and 2, $\mathbb{G}_N^{\omega}$ is a sequence of maps with values into the normed space, $\ell^{\infty}(\mathcal{H})$, converging in distribution to the Gaussian Process $\mathbb{G}^{\omega}$. Following van der Vaart (1998), section 23.2.1, $\mathbb{G}_N^{\omega *}$ is a sequence of maps with values into the normed space, $\ell^{\infty}(\mathcal{H})$, converging (conditionally on Z) in distribution to $\mathbb{G}^{\omega}$. Or putting more formally, we use van der Vaart's theorem 23.7 to write*

$$\sup_{h\in \mathrm{B}L_1(\ell^{\infty}(\mathcal{H}))}\left|\mathbb{E}_Z\left[h\left(\mathbb{G}_N^{\omega *}\right)\right]-\mathbb{E}\left[h\left(\mathbb{G}^{\omega}\right)\right]\right| \to_p 0$$

where $\mathbb{E}_Z$ denotes the expectation conditionally on $Z=\{(Y_1,X_1,T_1),(Y_2,X_2,T_2),...,(Y_N,X_N,T_N)\}$.

Since $\boldsymbol{\nu}(\cdot)$ $\boldsymbol{\nu}:\mathcal{F}_{\nu}\to\mathbb{R}^2$ is Hadamard differentiable, by the Delta-method for bootstrap in probability (Theorem 23.9 van der Vaart (1998)), $\boldsymbol{\nu}(\mathbb{G}_N^{\omega *})$ should converge in distribution to $\boldsymbol{\psi}^{\nu}(\mathbb{G}_N^{\omega};F_Y^{\omega}) = [\psi^{\nu}(\mathbb{G}_N^{\omega_1};F_Y^{\omega_1}),\psi^{\nu}(\mathbb{G}_N^{\omega_0};F_Y^{\omega_1})]^{\top}$, given Z in probability. Put more formally, for every $h\in \mathrm{B}L_1(\mathbb{R}^2)$, the function $h\circ\boldsymbol{\psi}^{\nu}$ is contained in $\mathrm{B}L_{\|\boldsymbol{\psi}^{\nu}\|}(\mathcal{F}_{\nu})$

$$\sup_{h\in \mathrm{B}L_1(\mathbb{R}^2)}\left|\mathbb{E}_Z\left[h\left(\boldsymbol{\nu}\left(\mathbb{G}_N^{\omega *}\right)\right)\right]-\mathbb{E}\left[h\left(\boldsymbol{\psi}^{\nu}\left(\mathbb{G}_N^{\omega};F_Y^{\omega}\right)\right)\right]\right| \to_p 0.$$

We have therefore that

$$[1,-1]^{\top}\boldsymbol{\nu}\left(\mathbb{G}_N^{\omega *}\right) = \mathrm{B}^{-1/2}\sum_{b=1}^{\mathrm{B}}\left(\widehat{\Delta}_b-\widehat{\Delta}\right)$$

and

$$\begin{aligned}[1,-1]^{\top}\boldsymbol{\psi}^{\nu}\left(\mathbb{G}_N^{\omega};F_Y^{\omega}\right) &= N^{-1/2}\sum_{i=1}^{N}\left(\omega_{1l}\left(T_i,\widehat{p}\left(X_i\right)\right)\cdot\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}_{1l}}\right)\right.\\ &\quad -\omega_{0l}\left(T,\widehat{p}\left(X_i\right)\right)\cdot\phi^{\nu}\left(Y_i;\widehat{F}_Y^{\widehat{\omega}_{0l}}\right)\\ &\quad \left.+\left(\widehat{g}_{1l}\left(X_i\right)-\widehat{g}_{0l}\left(X_i\right)\right)\left(T_i-\widehat{p}\left(X_i\right)\right)\right)\end{aligned}$$

Since $\widehat{V}^{\mathsf{B}}$ is a consistent estimator for the variance of $\widehat{\Delta}_b-\widehat{\Delta}$ given Z, it must be consistent for V, since $\boldsymbol{\nu}(\mathbb{G}_N^{\omega *})$ and $\boldsymbol{\psi}^{\nu}(\mathbb{G}_N^{\omega};F_Y^{\omega})$ are asymptotically equivalent. ■

Supplemental Material with Monte Carlo Results for the article "Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures"

Sergio Firpo* Cristine Pinto†

September 2011

Abstract

This supplemental material contains one Monte Carlo design in which the true propensity score attains values very close to zero and one. As the design in the main text, we compute inequality treatment effects on treated. We compare the behavior of the weighted estimator proposed in the article with other four estimators described in the paper, an unfeasible estimator, a naive estimator, the "location shift estimator" proposed by Juhn, Murphy and Pierce (1993) and the estimator proposed by Chernozhukov, Fernandez-Val and Melly (2009). The results show that the weighted estimator is more sensitive to misspecification of the propensity score in this design in which the common support assumption is almost violated. Additionally, the weighted estimator is still competitive with the more elaborate estimator proposed by Chernozhukov, Fernandez-Val and Melly (2009) and dominates the other estimators in terms of bias.

This supplemental material also contains additional tables of the results with the logit model for the design stated in the main paper.

*Corresponding author. Escola de Economia de São Paulo and C-Micro, Fundação Getulio Vargas (Brazil). Mailing address: EESP-FGV, Rua Itapeva 474/1215, São Paulo - SP, Brazil, 01332-000. E-mail: sergio.firpo@fgv.br. Webpage: http://sites.google.com/site/sergiopfirpo

†Escola de Economia de São Paulo and C-Micro, Fundação Getulio Vargas (Brazil). Mailing address: EESP-FGV, Rua Itapeva 474/1200, São Paulo - SP, Brazil, 01332-000. E-mail: cristine.pinto@fgv.br. Webpage: https://sites.google.com/site/cristinepinto/

Supplemental material: Monte Carlo

This supplemental appendix contains one monte carlo design in which the true propensity score has limits very close to 0 and 1. All notation is defined as in the main text unless stated otherwise. For completeness, we state the main data generated process (d.g.p.). The generated data follows a very simple specification. Starting with $X = [X_1, X_2]^\intercal$ we set $X_1 \sim$ Unif $\left[\mu_{X_1} - \frac{\sqrt{12}}{2}, \mu_{X_1} + \frac{\sqrt{12}}{2}\right]$ and $X_2 \sim$ Unif $\left[\mu_{X_2} - \frac{\sqrt{12}}{2}, \mu_{X_2} + \frac{\sqrt{12}}{2}\right]$, which will be independent random variables with the following means and variances: $E[X_1] = \mu_{X_1}$, $E[X_2] = \mu_{X_2}$ and $V[X_1] = V[X_2] = 1$. The treatment indicator is set to be

$$T = \mathbb{1}\{\delta_0 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_1^2 + \delta_4 X_2^2 + \delta_5 X_1 X_2 + \eta > 0\}.$$

We consider two possible distributions for η: (i) logistic, $\eta \sim F_\eta(n) = \left(1 + \exp\left(-\frac{\pi n}{10\sqrt{3}}\right)\right)^{-1}$; (ii) normal, $\eta \sim F_\eta(n) = \int_{-\infty}^{n} 10\,(2\pi)^{-1/2} \exp\left(-z^2/2\right) dz$. In all cases, $\eta \sim (0, 100)$, that is, η has mean zero and standard deviation 10.

The potential outcomes are

$$\begin{aligned} Y(0) &= \exp\left(\beta_{00} + \beta_{01} X_1 + \beta_{02} X_2 + \beta_{03} X_1^2 + \beta_{04} X_2^2 + \beta_{05} X_1 X_2 + \epsilon_0\right) \\ Y(1) &= \exp\left(\beta_{10} + \beta_{11} X_1 + \beta_{12} X_2 + \beta_{13} X_1^2 + \beta_{14} X_2^2 + \beta_{15} X_1 X_2 + \epsilon_1\right) \end{aligned}$$

where

$$\begin{aligned} \epsilon_0 &= \left(\beta^s_{00} + \beta^s_{01} X_1 + \beta^s_{02} X_2 + \beta^s_{03} X_1^2 + \beta^s_{04} X_2^2 + \beta^s_{05} X_1 X_2\right) \cdot \kappa_0 \\ \epsilon_1 &= \left(\beta^s_{10} + \beta^s_{11} X_1 + \beta^s_{12} X_2 + \beta^s_{13} X_1^2 + \beta^s_{14} X_2^2 + \beta^s_{15} X_1 X_2\right) \cdot \kappa_1 \end{aligned}$$

and where κ_0 and κ_1 are distributed as standard normals. The variables X, η, κ_0 and κ_1 are mutually independent. Under this specification, $Y(1)$ and $Y(0)$ will not have a closed form distribution. We compute target functionals using median values from 100 simulations of size 100,000 for the "unfeasible estimator", which is presented below.

The parameters were chosen to be $\mu_{X_1} = 1$, $\mu_{X_2} = 5$ and those in the table below.

Table B1: Parameter specification for Monte Carlo Exercise

$coeff.\backslash j$	0	1	2	3	4	5
δ_j	-0.5	10	-2	0.5	-0.1	0.5
β_{0j}	0.01	-0.01	0.01	0.01	-0.01	-0.02
β_{1j}	0.1	0.01	0.01	0.01	0.01	0.01
β^s_{0j}	0.01	-0.01	0.01	0.01	-0.01	-0.02
β^s_{1j}	0.01	0.01	0.01	0.01	0.01	0.01

With this parameter specification the true propensity score attains values that are very close to zero and one, as shown in table B2.

Table B2: Features of the Propensity Score

η	Logistic	Normal
Propensity Score		
Mean	0.5037	0.5052
Maximum	0.9931	0.9969
Minimum	0.0062	0.0026
Propensity Score ($T = 1$)		
Mean	0.7725	0.7638
Maximum	0.9924	0.9969
Minimum	0.0068	0.0031
Propensity Score ($T = 0$)		
Mean	0.2309	0.2411
Maximum	0.9931	0.9964
Minimum	0.0062	0.0026
Q_0	0.5005	0.5004

As in the design in the main text, we compute inequality treatment effects on the treated. Table B3 states the values of some functionals of the distribution of potential outcomes for the treated, and table B4 presents the inequality measures of the potential outcomes for the treated. The values of the inequality measures presented in table are very similar to the values presented in table 3 of the main text for these measures.

Table B3: Features of the Distributions of Potential Outcomes for the Treated (conditional on $T = 1$)

η	Logistic		Normal	
ν\Distribution	$Y(0)$	$Y(1)$	$Y(0)$	$Y(1)$
Mean	0.7636	1.9491	0.7642	1.9480
Standard Deviation (s.d.)	0.2883	1.2514	0.2879	1.2538
Mean of Logarithm	−0.3424	0.5342	−0.3415	0.5335
S.D. of Logarithm	0.3965	0.4902	0.3960	0.4900
10th Percentile	0.4233	0.9770	0.4239	0.9769
1st Quartile	0.5770	1.2413	0.5780	1.2409
Median	0.7495	1.6278	0.7504	1.6255
3rd Quartile	0.9164	2.2468	0.9169	2.2453
90th Percentile	1.0859	3.2134	1.0856	3.2113

Table B4: Inequality Measures of Potential Outcomes for the Treated (conditional on $T = 1$)

η	Logistic		Normal	
ν\Distribution	$Y(0)$	$Y(1)$	$Y(0)$	$Y(1)$
Coefficient of Variation	0.3776	0.6421	0.3768	0.6436
Interquartile Range	1.0056	0.3395	1.0044	0.3389
Theil Index	0.0683	0.1494	0.0681	0.1496
Gini Coefficient	0.2010	0.2855	0.2007	0.2856

As in the main text, tables B5-B8 provides results for the d.g.p based on the normal specification and in the logistic specification. We present the results for the unfeasible estimator and for other five estimator. The first one is the weighted estimator that is the estimator proposed in the paper. We use a parametric step to estimate the propensity score. We compute the propensity score by a logit regression that uses the correct specification, a quadratic function in X_1 and X_2, and a misspecified specification, a linear function of X_1 and X_2. The second estimator is the naive estimator that is based on the empirical distributions of $Y|T = 1$ and $Y|T = 0$. We consider also the "location shift estimator" and the "log-location shift estimator" described in the main paper. We compute these estimators by estimating a full quadratic model

for the conditional expectation of Y, and by misspecifying this conditional distribution using a linear model. Finally, we compute the "CFM estimator" described in the main paper using both a quadratic and a linear specifications for the conditional distribution of Y.

The results in tables B5-B8 show that the weighted estimator is still competitive with CFM estimator, and dominates in terms of the bias the naive estimator and the two estimators based on Juhn, Murphy and Pierce (1993). Compared to the results in the main text, the weighted estimator is more sensitive to the misspecification of the propensity score in this design in which the propensity score attains values very close to 0 and 1.

Table B5: Results of Monte Carlo Exercise (Sample Size 250, Replications 1000, Normal Selection Model)

Treatment on the Treated Parameters	Estimators	Target	Average	Lower 10th percentile	Median	Upper 10th percentile	Standard Deviation	Bias	Root Mean Squared Error	Mean Absolute Error	Median Absolute Error	90% C.I. Coverage Rate
Mean Treatment Effects		1.184										
	Unfeasible		1.178	1.040	1.168	1.324	0.111	-0.005	0.111	0.090	0.077	0.902
	Naive		1.121	0.985	1.114	1.263	0.109	-0.062	0.126	0.104	0.094	0.848
	Weighted		1.173	0.998	1.172	1.356	0.142	-0.011	0.142	0.113	0.095	0.907
	Weighted (Linear)		1.167	0.998	1.164	1.348	0.141	-0.017	0.142	0.111	0.090	0.903
	Location Shift		1.175	0.996	1.172	1.355	0.148	-0.009	0.148	0.117	0.095	0.903
	Location Shift (Linear)		1.191	1.036	1.184	1.355	0.126	0.007	0.126	0.100	0.085	0.905
	Log Location Shift		1.187	1.004	1.191	1.366	0.147	0.003	0.147	0.115	0.097	0.907
	Log Location Shift (linear)		1.212	1.059	1.206	1.377	0.124	0.029	0.127	0.100	0.081	0.888
	CFM		1.195	1.024	1.188	1.369	0.137	0.011	0.137	0.108	0.089	0.905
	CFM (Linear)		1.207	1.043	1.198	1.384	0.135	0.023	0.136	0.107	0.090	0.898
CV		0.267										
	Unfeasible		0.244	0.116	0.225	0.385	0.121	-0.023	0.124	0.094	0.080	0.926
	Naive		0.346	0.222	0.327	0.491	0.119	0.079	0.143	0.100	0.072	0.858
	Weighted		0.281	0.121	0.274	0.454	0.144	0.015	0.145	0.106	0.081	0.913
	Weighted (Linear)		0.288	0.130	0.278	0.457	0.141	0.021	0.142	0.104	0.079	0.912
	Location Shift		0.305	0.169	0.288	0.471	0.131	0.038	0.136	0.096	0.072	0.898
	Location Shift (Linear)		0.319	0.201	0.300	0.470	0.119	0.052	0.130	0.089	0.060	0.894
	Log Location Shift		0.312	0.188	0.295	0.469	0.120	0.045	0.128	0.089	0.063	0.891
	Log Location Shift (linear)		0.334	0.217	0.315	0.476	0.114	0.067	0.132	0.091	0.065	0.867
	CFM		0.249	0.076	0.241	0.426	0.149	-0.018	0.150	0.113	0.091	0.911
	CFM (Linear)		0.214	0.056	0.209	0.395	0.148	-0.053	0.157	0.120	0.098	0.900
Interquartile Range		0.666										
	Unfeasible		0.667	0.503	0.658	0.843	0.136	0.001	0.136	0.109	0.093	0.903
	Naive		0.735	0.567	0.729	0.903	0.136	0.070	0.153	0.119	0.098	0.877
	Weighted		0.674	0.442	0.674	0.920	0.194	0.008	0.194	0.152	0.126	0.906
	Weighted (Linear)		0.682	0.447	0.687	0.922	0.200	0.017	0.201	0.152	0.121	0.916
	Location Shift		0.734	0.556	0.729	0.916	0.144	0.068	0.159	0.124	0.101	0.868
	Location Shift (Linear)		0.747	0.577	0.743	0.929	0.136	0.081	0.158	0.124	0.102	0.851
	Log Location Shift		0.703	0.526	0.691	0.882	0.143	0.038	0.147	0.116	0.096	0.891
	Log Location Shift (linear)		0.733	0.562	0.725	0.916	0.138	0.067	0.154	0.120	0.102	0.871
	CFM		0.623	0.379	0.629	0.869	0.195	-0.042	0.199	0.157	0.129	0.893
	CFM (Linear)		0.624	0.386	0.636	0.847	0.183	-0.041	0.187	0.145	0.120	0.885
Theil Index		0.082										
	Unfeasible		0.078	0.038	0.073	0.123	0.037	-0.004	0.037	0.028	0.023	0.923
	Naive		0.108	0.069	0.102	0.155	0.036	0.027	0.045	0.032	0.024	0.844
	Weighted		0.087	0.034	0.085	0.141	0.046	0.005	0.046	0.034	0.026	0.908
	Weighted (Linear)		0.089	0.038	0.087	0.143	0.045	0.007	0.045	0.034	0.025	0.907
	Location Shift		0.096	0.050	0.092	0.149	0.042	0.015	0.045	0.032	0.024	0.893
	Location Shift (Linear)		0.102	0.062	0.096	0.151	0.038	0.020	0.043	0.031	0.021	0.875
	Log Location Shift		0.099	0.058	0.093	0.147	0.037	0.017	0.041	0.029	0.021	0.883
	Log Location Shift (linear)		0.105	0.067	0.099	0.152	0.036	0.024	0.043	0.031	0.022	0.856
	CFM		0.073	0.014	0.075	0.132	0.053	-0.008	0.053	0.039	0.030	0.910
	CFM (Linear)		0.062	0.005	0.064	0.122	0.052	-0.019	0.055	0.041	0.032	0.901
Gini Coefficient		0.085										
	Unfeasible		0.082	0.047	0.082	0.117	0.028	-0.002	0.028	0.022	0.019	0.897
	Naive		0.136	0.102	0.135	0.171	0.027	0.051	0.057	0.051	0.050	0.428
	Weighted		0.101	0.035	0.105	0.164	0.053	0.017	0.055	0.044	0.038	0.894
	Weighted (Linear)		0.105	0.038	0.108	0.166	0.050	0.020	0.054	0.043	0.037	0.878
	Location Shift		0.123	0.073	0.124	0.173	0.040	0.038	0.055	0.045	0.041	0.761
	Location Shift (Linear)		0.125	0.086	0.123	0.168	0.033	0.040	0.052	0.043	0.038	0.681
	Log Location Shift		0.113	0.071	0.112	0.158	0.033	0.028	0.044	0.036	0.031	0.786
	Log Location Shift (linear)		0.125	0.091	0.124	0.163	0.028	0.040	0.049	0.042	0.039	0.598

Table B6: Results of Monte Carlo Exercise (Sample Size 4,000, Replications 1000, Normal Selection Model)

Treatment on the Treated Parameters	Estimators	Target	Average	Lower 10th percentile	Median	Upper 10th percentile	Standard Deviation	Bias	Root Mean Squared Error	Mean Absolute Error	Median Absolute Error	90% C.I. Coverage Rate
Mean Treatment Effects		1.184										
	Unfeasible		1.184	1.145	1.183	1.223	0.030	0.000	0.030	0.024	0.020	0.893
	Naive		1.127	1.090	1.126	1.166	0.029	-0.057	0.064	0.058	0.058	0.366
	Weighted		1.184	1.136	1.183	1.233	0.037	0.001	0.037	0.030	0.026	0.914
	Weighted (Linear)		1.184	1.132	1.177	1.225	0.035	-0.005	0.036	0.029	0.025	0.905
	Location Shift		1.183	1.137	1.182	1.231	0.036	-0.001	0.036	0.030	0.026	0.904
	Location Shift (Linear)		1.199	1.158	1.199	1.242	0.032	0.015	0.035	0.028	0.024	0.851
	Log Location Shift		1.205	1.158	1.203	1.253	0.036	0.021	0.042	0.033	0.028	0.829
	Log Location Shift (linear)		1.222	1.182	1.223	1.264	0.032	0.038	0.049	0.041	0.039	0.671
	CFM		1.174	1.125	1.174	1.222	0.039	-0.010	0.040	0.032	0.028	0.897
	CFM (Linear)		1.192	1.149	1.191	1.239	0.035	0.035	0.036	0.028	0.024	0.889
CV		0.267										
	Unfeasible		0.263	0.216	0.259	0.312	0.041	-0.003	0.042	0.031	0.026	0.914
	Naive		0.367	0.321	0.364	0.416	0.042	0.100	0.109	0.100	0.098	0.209
	Weighted		0.271	0.209	0.272	0.335	0.922	0.004	0.056	0.041	0.034	0.922
	Weighted (Linear)		0.284	0.223	0.283	0.344	0.053	0.017	0.056	0.041	0.033	0.900
	Location Shift		0.349	0.301	0.346	0.396	0.042	0.082	0.093	0.083	0.079	0.392
	Location Shift (Linear)		0.340	0.293	0.336	0.387	0.042	0.073	0.084	0.074	0.070	0.491
	Log Location Shift		0.356	0.310	0.352	0.404	0.040	0.089	0.098	0.089	0.085	0.291
	Log Location Shift (linear)		0.357	0.312	0.353	0.403	0.040	0.090	0.098	0.090	0.086	0.275
	CFM		0.286	0.211	0.292	0.358	0.072	0.019	0.074	0.054	0.044	0.926
	CFM (Linear)		0.243	0.167	0.247	0.314	0.069	-0.024	0.073	0.051	0.037	0.913
Interquartile Range		0.666										
	Unfeasible		0.667	0.623	0.666	0.714	0.035	0.001	0.035	0.028	0.024	0.903
	Naive		0.736	0.691	0.735	0.780	0.035	0.071	0.079	0.071	0.069	0.350
	Weighted		0.667	0.607	0.666	0.729	0.048	0.002	0.048	0.038	0.032	0.902
	Weighted (Linear)		0.678	0.620	0.679	0.737	0.045	0.013	0.047	0.038	0.032	0.889
	Location Shift		0.778	0.729	0.778	0.826	0.038	0.113	0.119	0.113	0.113	0.095
	Location Shift (Linear)		0.758	0.712	0.757	0.804	0.035	0.092	0.099	0.092	0.091	0.179
	Log Location Shift		0.728	0.680	0.727	0.777	0.036	0.063	0.072	0.064	0.062	0.472
	Log Location Shift (linear)		0.736	0.690	0.736	0.781	0.035	0.070	0.078	0.070	0.070	0.366
	CFM		0.667	0.607	0.669	0.729	0.047	0.001	0.047	0.037	0.033	0.901
	CFM (Linear)		0.669	0.615	0.669	0.722	0.041	0.004	0.041	0.033	0.028	0.898
Theil Index		0.082										
	Unfeasible		0.081	0.068	0.081	0.094	0.010	0.000	0.010	0.008	0.007	0.904
	Naive		0.112	0.099	0.111	0.124	0.010	0.030	0.032	0.030	0.030	0.082
	Weighted		0.083	0.064	0.084	0.101	0.015	0.001	0.015	0.012	0.010	0.915
	Weighted (Linear)		0.087	0.070	0.088	0.104	0.014	0.005	0.015	0.012	0.010	0.876
	Location Shift		0.110	0.095	0.109	0.124	0.012	0.028	0.030	0.028	0.028	0.213
	Location Shift (Linear)		0.107	0.093	0.107	0.122	0.011	0.026	0.028	0.026	0.026	0.277
	Log Location Shift		0.108	0.096	0.108	0.121	0.010	0.027	0.029	0.027	0.026	0.161
	Log Location Shift (linear)		0.109	0.097	0.109	0.122	0.010	0.027	0.029	0.027	0.027	0.128
	CFM		0.089	0.122	0.070	0.092	0.019	0.008	0.020	0.015	0.013	0.913
	CFM (Linear)		0.077	0.057	0.078	0.095	0.018	-0.005	0.018	0.013	0.010	0.925
Gini Coefficient		0.085										
	Unfeasible		0.085	0.076	0.085	0.094	0.007	0.000	0.007	0.006	0.005	0.902
	Naive		0.138	0.129	0.138	0.148	0.008	0.053	0.054	0.053	0.053	0.000
	Weighted		0.087	0.067	0.088	0.107	0.016	0.003	0.016	0.013	0.011	0.910
	Weighted (Linear)		0.094	0.076	0.095	0.112	0.015	0.009	0.017	0.014	0.013	0.827
	Location Shift		0.139	0.123	0.139	0.154	0.012	0.054	0.055	0.054	0.054	0.005
	Location Shift (Linear)		0.132	0.119	0.132	0.145	0.011	0.047	0.048	0.047	0.047	0.004
	Log Location Shift		0.127	0.116	0.127	0.138	0.009	0.042	0.043	0.042	0.042	0.001
	Log Location Shift (linear)		0.128	0.118	0.128	0.138	0.008	0.043	0.044	0.043	0.043	0.000

Table B7: Results of Monte Carlo Exercise (Sample Size 250, Replications 1000, Logistic Selection Model)

Treatment on the Treated Parameters	Estimators	Target	Average	Lower 10th percentile	Median	Upper 10th percentile	Standard Deviation	Bias	Root Mean Squared Error	Mean Absolute Error	Median Absolute Error	90% C.I. Coverage Rate
		1.185										
Mean Treatment Effects	Unfeasible		1.184	1.046	1.171	1.347	0.115	-0.001	0.115	0.091	0.078	0.894
	Naive		1.125	0.993	1.114	1.276	0.111	-0.060	0.126	0.105	0.098	0.855
	Weighted		1.181	0.995	1.188	1.364	0.146	-0.005	0.146	0.115	0.096	0.904
	Weighted (Linear)		1.181	0.992	1.178	1.360	0.148	-0.008	0.148	0.116	0.096	0.898
	Location Shift		1.180	0.992	1.185	1.361	0.150	-0.006	0.150	0.118	0.098	0.903
	Location Shift (Linear)		1.198	1.038	1.194	1.363	0.126	0.012	0.127	0.101	0.087	0.910
	Log Location Shift		1.192	1.005	1.199	1.376	0.150	0.007	0.150	0.118	0.098	0.904
	Log Location Shift (linear)		1.220	1.063	1.216	1.380	0.125	0.035	0.130	0.103	0.085	0.895
	CFM		1.203	1.027	1.199	1.379	0.138	0.018	0.139	0.111	0.093	0.892
	CFM (Linear)		1.214	1.045	1.210	1.381	0.132	0.132	0.135	0.107	0.089	0.894
		0.264										
CV	Unfeasible		0.243	0.112	0.224	0.390	0.123	-0.022	0.125	0.095	0.080	0.924
	Naive		0.346	0.222	0.327	0.491	0.121	0.082	0.146	0.103	0.074	0.872
	Weighted		0.281	0.122	0.270	0.462	0.906	0.017	0.151	0.107	0.078	0.906
	Weighted (Linear)		0.283	0.126	0.273	0.459	0.148	0.018	0.149	0.106	0.077	0.913
	Location Shift		0.305	0.165	0.289	0.469	0.131	0.040	0.137	0.098	0.073	0.905
	Location Shift (Linear)		0.318	0.194	0.298	0.465	0.122	0.054	0.133	0.092	0.066	0.897
	Log Location Shift		0.312	0.188	0.293	0.463	0.121	0.047	0.130	0.091	0.063	0.898
	Log Location Shift (linear)		0.335	0.215	0.316	0.479	0.116	0.070	0.135	0.094	0.066	0.875
	CFM		0.245	0.073	0.239	0.426	0.149	-0.019	0.150	0.112	0.086	0.906
	CFM (Linear)		0.212	0.047	0.204	0.401	0.148	-0.052	0.157	0.120	0.097	0.900
		0.666										
Interquartile Range	Unfeasible		0.671	0.499	0.663	0.854	0.141	0.005	0.141	0.111	0.092	0.897
	Naive		0.741	0.572	0.731	0.925	0.141	0.074	0.159	0.123	0.098	0.863
	Weighted		0.692	0.453	0.695	0.942	0.202	0.026	0.203	0.159	0.131	0.896
	Weighted (Linear)		0.689	0.454	0.697	0.919	0.199	0.023	0.201	0.156	0.130	0.899
	Location Shift		0.737	0.549	0.729	0.936	0.150	0.071	0.166	0.130	0.107	0.870
	Location Shift (Linear)		0.751	0.580	0.741	0.943	0.141	0.085	0.165	0.128	0.104	0.844
	Log Location Shift		0.707	0.528	0.698	0.907	0.149	0.041	0.155	0.121	0.100	0.887
	Log Location Shift (linear)		0.739	0.562	0.728	0.932	0.142	0.073	0.159	0.124	0.102	0.855
	CFM		0.619	0.358	0.628	0.861	0.202	-0.047	0.207	0.161	0.131	0.894
	CFM (Linear)		0.631	0.403	0.624	0.858	0.181	-0.035	0.184	0.147	0.126	0.895
		0.081										
Theil Index	Unfeasible		0.078	0.037	0.073	0.125	0.037	-0.003	0.037	0.028	0.023	0.922
	Naive		0.108	0.069	0.102	0.155	0.037	0.027	0.046	0.033	0.024	0.847
	Weighted		0.087	0.036	0.084	0.144	0.049	0.006	0.049	0.035	0.025	0.904
	Weighted (Linear)		0.087	0.036	0.087	0.142	0.050	0.006	0.050	0.034	0.025	0.923
	Location Shift		0.096	0.049	0.093	0.149	0.042	0.015	0.045	0.033	0.025	0.896
	Location Shift (Linear)		0.102	0.060	0.096	0.152	0.039	0.021	0.044	0.032	0.023	0.882
	Log Location Shift		0.098	0.058	0.093	0.148	0.038	0.017	0.042	0.030	0.021	0.886
	Log Location Shift (linear)		0.106	0.066	0.100	0.152	0.036	0.024	0.044	0.031	0.023	0.858
	CFM		0.072	0.152	0.009	0.074	0.053	-0.009	0.053	0.039	0.028	0.905
	CFM (Linear)		0.063	0.005	0.064	0.124	0.052	-0.018	0.055	0.041	0.032	0.903
		0.085										
Gini Coefficient	Unfeasible		0.082	0.047	0.082	0.118	0.028	-0.002	0.028	0.022	0.018	0.882
	Naive		0.136	0.102	0.136	0.171	0.028	0.051	0.058	0.052	0.051	0.431
	Weighted		0.104	0.037	0.105	0.172	0.055	0.019	0.059	0.045	0.037	0.880
	Weighted (Linear)		0.104	0.040	0.106	0.171	0.054	0.020	0.057	0.044	0.037	0.881
	Location Shift		0.123	0.070	0.124	0.175	0.040	0.038	0.055	0.046	0.042	0.747
	Location Shift (Linear)		0.125	0.082	0.124	0.167	0.034	0.040	0.053	0.044	0.040	0.688
	Log Location Shift		0.113	0.072	0.113	0.156	0.034	0.029	0.045	0.036	0.032	0.777
	Log Location Shift (linear)		0.126	0.089	0.125	0.162	0.029	0.041	0.050	0.043	0.041	0.586

Table B8: Results of Monte Carlo Exercise (Sample Size 4,000, Replications 1000, Logistic Selection Model)

Treatment on the Treated Parameters	Estimators	Target	Average	Lower 10th percentile	Median	Upper 10th percentile	Standard Deviation	Bias	Root Mean Squared Error	Mean Absolute Error	Median Absolute Error	90% C.I. Coverage Rate
		1.185										
Mean Treatment Effects	Unfeasible		1.186	1.147	1.185	1.225	0.030	0.001	0.030	0.024	0.021	0.913
	Naive		1.128	1.091	1.128	1.166	0.029	-0.057	0.064	0.058	0.058	0.372
	Weighted		1.186	1.135	1.186	1.239	0.041	0.000	0.041	0.033	0.028	0.903
	Weighted (Linear)		1.186	1.130	1.181	1.231	0.039	-0.005	0.039	0.031	0.027	0.897
	Location Shift		1.185	1.136	1.184	1.234	0.037	-0.001	0.037	0.030	0.026	0.904
	Location Shift (Linear)		1.202	1.159	1.202	1.244	0.033	0.016	0.036	0.029	0.024	0.869
	Log Location Shift		1.208	1.160	1.207	1.257	0.037	0.023	0.043	0.035	0.029	0.831
	Log Location Shift (linear)		1.225	1.184	1.225	1.267	0.032	0.040	0.051	0.043	0.040	0.659
	CFM		1.176	1.125	1.176	1.228	0.039	-0.009	0.040	0.032	0.028	0.899
	CFM (Linear)		1.194	1.149	1.195	1.240	0.035	0.035	0.036	0.029	0.025	0.888
		0.264										
CV	Unfeasible		0.263	0.215	0.260	0.311	0.042	-0.001	0.042	0.031	0.025	0.922
	Naive		0.368	0.323	0.365	0.415	0.041	0.104	0.112	0.104	0.100	0.182
	Weighted		0.266	0.200	0.268	0.336	0.930	0.002	0.063	0.044	0.033	0.930
	Weighted (Linear)		0.279	0.216	0.279	0.346	0.058	0.014	0.060	0.043	0.034	0.916
	Location Shift		0.350	0.302	0.348	0.400	0.042	0.086	0.096	0.086	0.083	0.345
	Location Shift (Linear)		0.341	0.294	0.338	0.389	0.041	0.077	0.087	0.077	0.074	0.439
	Log Location Shift		0.357	0.312	0.353	0.403	0.040	0.092	0.100	0.092	0.088	0.251
	Log Location Shift (linear)		0.358	0.314	0.353	0.404	0.039	0.093	0.101	0.093	0.089	0.232
	CFM		0.285	0.208	0.291	0.356	0.068	0.020	0.071	0.053	0.043	0.914
	CFM (Linear)		0.239	0.163	0.244	0.312	0.068	-0.025	0.073	0.051	0.039	0.910
		0.666										
Interquartile Range	Unfeasible		0.669	0.625	0.669	0.713	0.034	0.003	0.034	0.028	0.024	0.898
	Naive		0.739	0.694	0.740	0.784	0.035	0.073	0.081	0.073	0.074	0.333
	Weighted		0.669	0.607	0.669	0.728	0.049	0.003	0.049	0.038	0.031	0.902
	Weighted (Linear)		0.680	0.618	0.681	0.742	0.047	0.014	0.049	0.039	0.032	0.884
	Location Shift		0.781	0.734	0.781	0.830	0.037	0.115	0.121	0.115	0.115	0.075
	Location Shift (Linear)		0.761	0.716	0.761	0.807	0.035	0.095	0.101	0.095	0.095	0.154
	Log Location Shift		0.731	0.685	0.730	0.778	0.036	0.065	0.074	0.066	0.064	0.426
	Log Location Shift (linear)		0.739	0.694	0.740	0.784	0.035	0.073	0.081	0.073	0.074	0.324
	CFM		0.669	0.609	0.669	0.728	0.045	0.002	0.045	0.036	0.030	0.894
	CFM (Linear)		0.672	0.621	0.670	0.723	0.039	0.005	0.040	0.031	0.026	0.898
		0.081										
Theil Index	Unfeasible		0.081	0.068	0.081	0.094	0.010	0.000	0.010	0.008	0.007	0.913
	Naive		0.112	0.099	0.111	0.125	0.010	0.031	0.033	0.031	0.030	0.070
	Weighted		0.081	0.062	0.082	0.101	0.018	0.000	0.018	0.013	0.010	0.939
	Weighted (Linear)		0.086	0.068	0.086	0.104	0.015	0.005	0.016	0.012	0.010	0.903
	Location Shift		0.110	0.096	0.110	0.124	0.012	0.029	0.031	0.029	0.029	0.198
	Location Shift (Linear)		0.108	0.094	0.107	0.122	0.011	0.027	0.029	0.027	0.026	0.254
	Log Location Shift		0.109	0.096	0.108	0.121	0.010	0.028	0.029	0.028	0.027	0.152
	Log Location Shift (linear)		0.110	0.097	0.109	0.122	0.010	0.028	0.030	0.028	0.028	0.117
	CFM		0.089	0.122	0.069	0.091	0.018	0.008	0.019	0.015	0.013	0.896
	CFM (Linear)		0.076	0.057	0.078	0.095	0.017	-0.005	0.018	0.013	0.010	0.919
		0.085										
	Unfeasible		0.085	0.075	0.085	0.095	0.007	0.000	0.007	0.006	0.005	0.897
	Naive		0.139	0.129	0.139	0.149	0.008	0.054	0.055	0.054	0.054	0.000
	Weighted		0.086	0.064	0.086	0.108	0.018	0.001	0.018	0.014	0.011	0.912
	Weighted (Linear)		0.093	0.073	0.093	0.113	0.016	0.008	0.018	0.014	0.012	0.855
	Location Shift		0.139	0.124	0.139	0.154	0.012	0.055	0.056	0.055	0.055	0.002
	Location Shift (Linear)		0.132	0.119	0.132	0.145	0.011	0.048	0.049	0.048	0.048	0.001
	Log Location Shift		0.128	0.116	0.128	0.138	0.009	0.043	0.044	0.043	0.043	0.002
	Log Location Shift (linear)		0.129	0.119	0.129	0.138	0.008	0.044	0.045	0.044	0.044	0.000

At the end of this supplemental material, we present the additional tables with the logit model for the design in the main paper.

Table A.2: Results of Monte Carlo Exercise (Sample Size 250, Replications 1000, Logistic Selection Model)

Treatment on the Treated Parameters	Estimators	Target	Average	Lower 10th percentile	Median	Upper 10th percentile	Standard Deviation	Bias	Root Mean Squared Error	Mean Absolute Error	Median Absolute Error	90% C.I. Coverage Rate
Mean Treatment Effects		1.108										
	Unfeasible		1.108	0.976	1.094	1.261	0.109	0.000	0.109	0.087	0.071	0.905
	Naive		1.079	0.951	1.070	1.217	0.106	-0.029	0.110	0.090	0.079	0.896
	Weighted		1.106	0.971	1.100	1.246	0.112	-0.003	0.112	0.089	0.077	0.906
	Weighted (Linear)		1.106	0.971	1.096	1.244	0.111	-0.004	0.111	0.088	0.077	0.908
	Location Shift		1.106	0.971	1.098	1.247	0.112	-0.003	0.112	0.089	0.076	0.908
	Location Shift (Linear)		1.108	0.972	1.100	1.250	0.111	0.000	0.111	0.088	0.075	0.911
	Log Location Shift		1.115	0.981	1.108	1.259	0.113	0.007	0.113	0.089	0.074	0.908
	Log Location Shift (linear)		1.118	0.984	1.111	1.260	0.112	0.010	0.112	0.088	0.074	0.910
	CFM		1.106	0.971	1.099	1.247	0.113	-0.002	0.113	0.089	0.074	0.906
	CFM (Linear)		1.109	0.972	1.103	1.254	0.113	0.113	0.113	0.089	0.075	0.913
CV		0.268										
	Unfeasible		0.244	0.119	0.226	0.393	0.122	-0.024	0.124	0.093	0.079	0.941
	Naive		0.294	0.170	0.274	0.436	0.122	0.025	0.125	0.086	0.064	0.930
	Weighted		0.248	0.117	0.233	0.395	0.924	-0.021	0.128	0.094	0.074	0.924
	Weighted (Linear)		0.252	0.119	0.235	0.401	0.128	-0.016	0.129	0.094	0.074	0.924
	Location Shift		0.280	0.160	0.262	0.425	0.121	0.012	0.122	0.085	0.066	0.934
	Location Shift (Linear)		0.281	0.157	0.262	0.422	0.121	0.013	0.122	0.085	0.066	0.935
	Log Location Shift		0.282	0.166	0.264	0.418	0.117	0.014	0.118	0.082	0.064	0.929
	Log Location Shift (linear)		0.283	0.169	0.266	0.417	0.114	0.015	0.115	0.080	0.063	0.929
	CFM		0.246	0.113	0.234	0.397	0.130	-0.022	0.131	0.096	0.077	0.924
	CFM (Linear)		0.246	0.116	0.229	0.396	0.127	-0.022	0.129	0.095	0.075	0.923
Interquartile Range		0.605										
	Unfeasible		0.612	0.448	0.600	0.795	0.134	0.007	0.134	0.106	0.090	0.892
	Naive		0.648	0.483	0.636	0.835	0.137	0.043	0.144	0.111	0.090	0.877
	Weighted		0.612	0.440	0.600	0.790	0.138	0.007	0.138	0.110	0.092	0.892
	Weighted (Linear)		0.618	0.444	0.604	0.799	0.139	0.013	0.139	0.110	0.092	0.903
	Location Shift		0.668	0.502	0.657	0.855	0.137	0.063	0.151	0.117	0.094	0.868
	Location Shift (Linear)		0.659	0.492	0.648	0.842	0.137	0.054	0.147	0.114	0.090	0.876
	Log Location Shift		0.625	0.458	0.615	0.812	0.138	0.020	0.139	0.108	0.085	0.898
	Log Location Shift (linear)		0.626	0.453	0.614	0.809	0.139	0.021	0.140	0.109	0.086	0.903
	CFM		0.612	0.440	0.598	0.792	0.137	0.007	0.137	0.109	0.090	0.897
	CFM (Linear)		0.609	0.435	0.596	0.788	0.139	0.004	0.139	0.110	0.089	0.897
Theil Index		0.079										
	Unfeasible		0.075	0.037	0.071	0.120	0.036	-0.004	0.037	0.028	0.023	0.929
	Naive		0.090	0.052	0.086	0.135	0.037	0.011	0.038	0.027	0.020	0.914
	Weighted		0.076	0.036	0.071	0.120	0.038	-0.003	0.038	0.028	0.023	0.921
	Weighted (Linear)		0.078	0.038	0.073	0.122	0.038	-0.002	0.038	0.028	0.022	0.913
	Location Shift		0.087	0.047	0.082	0.134	0.038	0.008	0.039	0.028	0.022	0.918
	Location Shift (Linear)		0.088	0.047	0.083	0.135	0.039	0.009	0.040	0.028	0.021	0.923
	Log Location Shift		0.087	0.050	0.082	0.132	0.036	0.007	0.037	0.026	0.020	0.919
	Log Location Shift (linear)		0.087	0.050	0.083	0.131	0.036	0.008	0.036	0.026	0.020	0.916
	CFM		0.075	0.135	0.033	0.071	0.039	-0.004	0.039	0.029	0.024	0.915
	CFM (Linear)		0.076	0.035	0.071	0.121	0.038	-0.004	0.038	0.028	0.023	0.915
Gini Coefficient		0.087										
	Unfeasible		0.084	0.050	0.084	0.120	0.028	-0.002	0.028	0.022	0.019	0.901
	Naive		0.110	0.075	0.111	0.147	0.029	0.024	0.037	0.030	0.027	0.803
	Weighted		0.086	0.049	0.085	0.125	0.031	-0.001	0.031	0.024	0.020	0.902
	Weighted (Linear)		0.088	0.051	0.087	0.126	0.031	0.002	0.031	0.024	0.019	0.901
	Location Shift		0.110	0.069	0.110	0.154	0.034	0.023	0.042	0.033	0.028	0.830
	Location Shift (Linear)		0.108	0.068	0.107	0.151	0.033	0.021	0.039	0.031	0.026	0.844
	Log Location Shift		0.099	0.064	0.099	0.139	0.030	0.013	0.032	0.025	0.021	0.864
	Log Location Shift (linear)		0.100	0.065	0.100	0.137	0.029	0.014	0.032	0.025	0.020	0.860
	CFM		0.086	0.046	0.086	0.126	0.032	-0.001	0.032	0.025	0.020	0.898
	CFM (Linear)		0.085	0.046	0.085	0.124	0.031	-0.002	0.031	0.025	0.020	0.903

Table A.3: Results of Monte Carlo Exercise (Sample Size 4,000, Replications 1000, Logistic Selection Model)

Treatment on the Treated Parameters	Estimators	Target	Average	Lower 10th percentile	Median	Upper 10th percentile	Standard Deviation	Bias	Root Mean Squared Error	Mean Absolute Error	Median Absolute Error	90% C.I. Coverage Rate
Mean Treatment Effects		1.108										
	Unfeasible		1.109	1.073	1.109	1.146	0.028	0.001	0.028	0.023	0.019	0.901
	Naive		1.082	1.047	1.081	1.118	0.027	-0.027	0.038	0.032	0.029	0.746
	Weighted		1.109	1.072	1.108	1.147	0.029	0.001	0.029	0.023	0.019	0.899
	Weighted (Linear)		1.109	1.071	1.106	1.145	0.028	-0.001	0.028	0.023	0.019	0.898
	Location Shift		1.109	1.072	1.108	1.147	0.029	0.001	0.029	0.023	0.019	0.901
	Location Shift (Linear)		1.112	1.075	1.111	1.149	0.029	0.003	0.029	0.023	0.019	0.893
	Log Location Shift		1.120	1.083	1.119	1.158	0.029	0.012	0.031	0.025	0.020	0.874
	Log Location Shift (linear)		1.122	1.085	1.121	1.159	0.029	0.014	0.032	0.025	0.021	0.865
	CFM		1.106	1.069	1.105	1.144	0.029	-0.002	0.029	0.023	0.019	0.900
	CFM (Linear)		1.110	1.074	1.109	1.147	0.029	0.029	0.029	0.023	0.019	0.900
CV		0.268										
	Unfeasible		0.267	0.220	0.262	0.321	0.042	-0.001	0.042	0.032	0.027	0.927
	Naive		0.319	0.272	0.315	0.370	0.042	0.051	0.066	0.053	0.046	0.719
	Weighted		0.268	0.220	0.263	0.320	0.925	-0.001	0.044	0.033	0.026	0.925
	Weighted (Linear)		0.275	0.228	0.271	0.328	0.044	0.007	0.044	0.033	0.026	0.920
	Location Shift		0.309	0.263	0.304	0.358	0.042	0.041	0.059	0.045	0.037	0.800
	Location Shift (Linear)		0.306	0.260	0.301	0.356	0.042	0.038	0.057	0.043	0.035	0.811
	Log Location Shift		0.311	0.266	0.306	0.360	0.041	0.043	0.059	0.046	0.039	0.766
	Log Location Shift (linear)		0.309	0.265	0.304	0.357	0.040	0.040	0.057	0.044	0.036	0.777
	CFM		0.301	0.254	0.295	0.354	0.043	0.033	0.054	0.040	0.030	0.844
	CFM (Linear)		0.294	0.248	0.289	0.345	0.043	0.026	0.050	0.037	0.027	0.873
Interquartile Range		0.605										
	Unfeasible		0.607	0.565	0.606	0.650	0.033	0.002	0.033	0.026	0.022	0.896
	Naive		0.645	0.601	0.644	0.688	0.033	0.040	0.052	0.044	0.040	0.669
	Weighted		0.607	0.565	0.606	0.649	0.033	0.002	0.033	0.026	0.022	0.894
	Weighted (Linear)		0.614	0.571	0.613	0.656	0.033	0.009	0.034	0.027	0.023	0.892
	Location Shift		0.686	0.640	0.686	0.731	0.034	0.081	0.088	0.081	0.081	0.236
	Location Shift (Linear)		0.669	0.625	0.668	0.712	0.034	0.064	0.072	0.064	0.063	0.405
	Log Location Shift		0.629	0.586	0.628	0.672	0.033	0.024	0.041	0.033	0.028	0.823
	Log Location Shift (linear)		0.624	0.581	0.623	0.666	0.033	0.019	0.038	0.031	0.027	0.852
	CFM		0.609	0.566	0.608	0.651	0.033	0.004	0.033	0.026	0.022	0.896
	CFM (Linear)		0.608	0.566	0.607	0.651	0.033	0.003	0.033	0.026	0.022	0.895
Theil Index		0.079										
	Unfeasible		0.079	0.067	0.079	0.093	0.010	0.000	0.010	0.008	0.007	0.910
	Naive		0.095	0.082	0.094	0.108	0.010	0.016	0.019	0.016	0.015	0.582
	Weighted		0.079	0.066	0.079	0.093	0.011	0.000	0.011	0.008	0.007	0.915
	Weighted (Linear)		0.082	0.069	0.081	0.095	0.011	0.003	0.011	0.008	0.007	0.906
	Location Shift		0.094	0.081	0.094	0.109	0.011	0.015	0.019	0.016	0.014	0.657
	Location Shift (Linear)		0.094	0.081	0.093	0.109	0.012	0.015	0.019	0.016	0.014	0.663
	Log Location Shift		0.092	0.080	0.091	0.105	0.010	0.013	0.016	0.013	0.012	0.671
	Log Location Shift (linear)		0.092	0.080	0.091	0.105	0.010	0.012	0.016	0.013	0.012	0.674
	CFM		0.090	0.109	0.077	0.089	0.010	0.010	0.015	0.012	0.010	0.769
	CFM (Linear)		0.088	0.075	0.087	0.101	0.010	0.008	0.013	0.010	0.009	0.806
Gini Coefficient		0.087										
	Unfeasible		0.087	0.077	0.087	0.096	0.007	0.000	0.007	0.006	0.005	0.899
	Naive		0.114	0.104	0.114	0.123	0.008	0.027	0.028	0.027	0.027	0.030
	Weighted		0.087	0.076	0.087	0.097	0.008	0.000	0.008	0.006	0.005	0.900
	Weighted (Linear)		0.090	0.080	0.090	0.101	0.008	0.004	0.009	0.007	0.006	0.860
	Location Shift		0.119	0.105	0.118	0.132	0.011	0.032	0.034	0.032	0.032	0.095
	Location Shift (Linear)		0.116	0.103	0.115	0.129	0.011	0.029	0.031	0.029	0.029	0.140
	Log Location Shift		0.105	0.094	0.105	0.114	0.008	0.018	0.020	0.018	0.018	0.283
	Log Location Shift (linear)		0.103	0.093	0.103	0.113	0.008	0.017	0.019	0.017	0.017	0.303
	CFM		0.100	0.089	0.100	0.110	0.008	0.013	0.015	0.013	0.013	0.503
	CFM (Linear)		0.096	0.086	0.097	0.107	0.008	0.010	0.013	0.011	0.010	0.659

WORKING PAPER ~~20~~/2011 • SEP, 2011 •