I Z A

IZA DP No. 4841

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

Sergio Firpo

March 2010

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

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

Sergio Firpo
*São Paulo School of Economics, FGV
and IZA*

Discussion Paper No. 4841
March 2010

ABSTRACT

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

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 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 good relative performance in small samples is robust to changes in the distribution of latent selection variables. 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, using several popular inequality measures to capture distributional impacts of the program.

JEL Classification: C1, C3

Keywords: inequality measures, treatment effects, semiparametric efficiency, reweighting estimator

Corresponding author:

Sergio Firpo
Escola de Economia de São Paulo – FGV-SP
R. Itapeva, 474/1215
São Paulo, SP 01332-000
Brazil
E-mail: sergio.firpo@fgv.br

1 Introduction

For the evaluation of a social program, the policy-maker may want to learn about the distributional effects of the program going beyond the program's 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 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 impact 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 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 (2007).

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.

A two step estimation procedure is proposed. 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. Note that unlike previous works, as Firpo (2007), estimation of all quantiles to recover differences in a given inequality measure is not necessary. All that is needed is direct calculation of weighted inequality measures, which may or may not involve computation of quantiles, depending on its formula.

---

[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).

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, 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.

The key methodological contribution of this paper is to provide a weighted estimator that can be expressed as a functional of the empirical weighted distribution. Therefore, we generalize the usage of weighted estimators to statistics that may not be represented as solutions to moment conditions. 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. Inference based on the bootstrap is shown to be a valid procedure for testing using the estimators developed in this paper.

Under failure of unconfoundedness, we may not have a causal interpretation. Nevertheless, the 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, of 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 identify how systematic differences in the pay structure of the two countries affect the Gini coefficient, fixing the distribution of covariates to be the same.

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 elegant 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 consistent 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 derives the large sample properties of the inequality treatment effects estimators. Section 5 discusses finite-sample behavior through a Monte Carlo exercise. We present in section 6 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 7 concludes. Proofs of results are left to the Appendix.

In both sections 5 and 6, 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). Somewhat surprisingly, evidence from Monte Carlo exercises reveals 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

We start by assuming that there is an available random sample of N individuals (units). 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.[4] We define the inequality measure ν as a functional of the distribution, $\nu : \mathcal{F}_\nu \to \mathbb{R}$. 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 distributions 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.[5]

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^ν, defined as:

$$
\begin{aligned}
\Delta^\nu &= \nu\left(F_{Y(1)}\right) - \nu\left(F_{Y(0)}\right) \\
&= \nu_1 - \nu_0
\end{aligned}
$$

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

$$
\begin{aligned}
\Delta_T^\nu &= \nu\left(F_{Y(1)|T=1}\right) - \nu\left(F_{Y(0)|T=1}\right) \\
&= \nu_{11} - \nu_{01}
\end{aligned}
$$

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):[6]

$$
\begin{aligned}
\Delta_C^\nu &= \nu(F_Y) - \nu\left(F_{Y(0)}\right) \\
&= \nu_Y - \nu_0
\end{aligned}
$$

3 Identification of Inequality Treatment Effects

This section is divided up into four 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, while in subsection 3.3 we present 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.

---

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

[5]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.

[6]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.

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, T, X). Let the data be defined by the sequence $\{Y_i, T_i, X_i\}_{i=1}^N$ where each element (Y_i, T_i, X_i) is a random draw from $F_{Y,T,X}$, the joint distribution of $(Y, T, X) \in \mathcal{Y} \times \{0,1\} \times \mathcal{X}$, 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 those 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 at y that uses proper functions of T and X as weights is

$$F_Y^\omega(y) = E[\omega \cdot \mathbb{I}\{Y \le y\}], \tag{1}$$

where $\mathbb{I}\{\cdot\}$ is the indicator function and ω is the weighting function. 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 $\omega = 1$.

Let us define the *propensity-score*, $p(x)$, as the probability that given a value $x \in \mathcal{X}$ an individual will be in the treatment group, that is, $p(x) \equiv \Pr[T = 1|X = x]$. The unconditional probability, $\Pr[T = 1]$, is p, which is assumed to be positive. Let $\mathcal{P} \subset [0,1]$ be the image set of the mapping $p(\cdot)$, $p: \mathcal{X} \to \mathcal{P}$. A restriction on $\mathcal{P}$ will be made later on Assumption 2.

Next, we define the following four "weighting functions", generally written as ω, such that $\omega: \{0,1\} \times \mathcal{P} \to \mathbb{R}$:

$$
\begin{aligned}
\omega_1(t, p(x)) &= t/p(x) \\
\omega_0(t, p(x)) &= (1-t)/(1-p(x)) \\
\omega_{11}(t, p(x)) &= t/p \\
\omega_{01}(t, p(x)) &= ((1-t)/(1-p(x))) \cdot (p(x)/p).
\end{aligned}
$$

Therefore, a weighted c.d.f. of Y using ω_A, which is a general representation of the above weights, is

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

These weighting functions will be used to identify the marginal c.d.f.'s of distributions of $Y(1)$ and $Y(0)$ as we show later.

3.2 Identifying Assumptions

Here we invoke the set of identifying restrictions that will permit that we write the distribution of the unobserved potential outcomes in terms of observable data. Moreover, those distributions will actually fall into the category of the weighted distributions just defined.

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, $(Y(1), Y(0)) \perp\!\!\!\perp T|X = x$.*

Assumption 1 is sometimes a strong assumption and its plausibility has to be analyzed in a case by case basis. It 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 overlap assumption:

ASSUMPTION 2 [**Common Support**] *For all x in $\mathcal{X}$, $0 < p(x) < 1$.*

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.*

3.3 Identification Results

Finally, the main identification result will follow as a corollary of the next theorem. We therefore write the ITE parameters as functions of the observable variables (Y,T,X).

LEMMA 1 [**Firpo (2007), Lemma 1**]:*Let $Y(1) \sim F_{Y(1)}$, $Y(0) \sim F_{Y(0)}$, $Y(1)|T = 1 \sim F_{Y(1)|T=1}$, $Y(0)|T = 1 \sim F_{Y(0)|T=1}$ and $Y \sim F_Y$. Under Assumptions 1 and 2, for every $y \in \mathcal{Y}$, the c.d.f.'s associated with these distributions can be written respectively as $F_{Y(1)}(y) = F_Y^{\omega=\omega_1}(y)$; $F_{Y(0)}(y) = F_Y^{\omega=\omega_0}(y)$; $F_{Y(1)|T=1}(y) = F_Y^{\omega=\omega_{11}}(y)$; and $F_{Y(0)|T=1}(y) = F_Y^{\omega=\omega_{01}}(y)$.*

COROLLARY 1 *Under Assumptions 1 and 2 Δ^ν, Δ_T^ν, and Δ_C^ν are identifiable.*

Once we know that the inequality treatment effects are identifiable, 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.

3.4 Some Inequality Measures

We now turn our attention to 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 for inequality measures. 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).[7]

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

---

[7]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.

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:[8]

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

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

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

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

3. **Theil Index** (TI):[9]

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

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

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

4 Estimation and Large Sample Inference

We now focus our attention to estimation of $\nu\left(F_Y^\omega\right)$, the inequality measure of a weighted outcome distribution. We first show how to estimate and derive the asymptotic distribution of the estimator of $\nu\left(F_Y^\omega\right)$ with a general ω, and later show how to use these results to estimation and inference regarding Δ^ν, Δ_T^ν and Δ_C^ν.

4.1 Estimation

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, $\widehat{F}_Y^{\omega=\widehat{\omega}}$, and plug it into the functional ν. The estimator will therefore be:

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

---

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

[9]The Theil index requires that the support of the outcome variable be restricted to the positive real numbers.

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

$$\widehat{F}_Y^{\omega=\widehat{\omega}}(y) = N^{-1} \sum_{i=1}^{N} \widehat{\omega}_i \cdot \mathbb{I}\{Y_i \leq y\}.$$

It is clear that we have to consider carefully the estimation of weights ω by $\widehat{\omega}$.

4.1.1 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. We follow the sieve ML approach proposed 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.[10] Stacking all these polynomials in a vector, we end up with $H_K(x) = [H_{K,j}(x)] \quad (j = 1, ..., K)$, a vector of length K of polynomial functions of $x \in \mathcal{X}$. The estimation procedure will therefore involve computation of the vector of length K of coefficients $\hat{\pi}_K$:

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

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:

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

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.[11]

We propose an estimator $\widehat{\omega}$ for the weighting function ω:

$$\widehat{\omega}_i = \omega(T_i, \widehat{p}(X_i)).$$

Specifically for the weighting functions ω_1, ω_0, ω_{11}, and ω_{01} we have

$$\widehat{\omega}_{1,i} = \omega_1(T_i, \widehat{p}(X_i)) = \frac{T_i}{\widehat{p}(X_i)}$$

---

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

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

$$\widehat{\omega}_{0,i} = \omega_0\left(T_i, \widehat{p}\left(X_i\right)\right) = \left(\frac{1 - T_i}{1 - \widehat{p}\left(X_i\right)}\right)$$

$$\widehat{\omega}_{11,i} = \omega_{11}\left(T_i, \widehat{p}\left(X_i\right)\right) = \frac{T_i}{\widehat{p}}$$

$$\widehat{\omega}_{01,i} = \omega_{01}\left(T_i, \widehat{p}\left(X_i\right)\right) = \frac{\widehat{p}\left(X_i\right)}{\widehat{p}} \cdot \left(\frac{1 - T_i}{1 - \widehat{p}\left(X_i\right)}\right).$$

4.1.2 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^ν, Δ_T^ν, and Δ_C^ν as

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

4.2 Large Sample Inference

We now devote our attention to the asymptotic behavior of our estimators. 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 of interest could be considered beyond the inequality measures here studied.[12]

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 establish efficiency of our estimators.[13] Finally, we argue that the inference based on bootstrap is a valid method for the weighted estimators considered in this paper.

4.2.1 Hadamard Differentiability

We invoke a smoothness condition in order to be able to derive the asymptotic normality of the inequality estimators just proposed. We restrict the discussion to the class of inequality measures that are Hadamard differentiable functionals of the distribution.

A functional ν is a map $\nu : \mathcal{F}_\nu \to \mathbb{R}$ defined on a subset $\mathcal{F}_\nu$ of the normed space of c.d.f.'s $\mathcal{F}$ that contains ν. Following van der Vaart (1998, chapter 20), we say that ν is *Hadamard*

---

[12] For examples of some estimands and hypotheses tests that could involve, for instance, the quantile process, first- and second-order stochastic dominance, and Kolmogorov-Smirnov tests, see Abadie (2002) and Chernozhukov, Fernandez-Val and Melly (2009).

[13] 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.

differentiable at $F \in \mathcal{F}_\nu$ if there exists a continuous, linear map $\psi^\nu : \mathcal{F} \to \mathbb{R}$ such that

$$\lim_{s \downarrow 0} \left\| \frac{\nu\left(F + s \cdot \delta_s\right) - \nu\left(F\right)}{s} - \psi^\nu\left(\delta\right) \right\| = 0$$

for all functions $\delta_s \to \delta$ such that $F + s \cdot \delta_s \in \mathcal{F}_\nu$, for all small $s > 0$.

That smoothness condition is used here to help us establishing (i) asymptotic normality, (ii) efficiency and (iii) validity of the inference procedure for our estimators. We therefore write it as an assumption.

ASSUMPTION 3 [**Hadamard**] *The inequality measure ν defined over the marginal distribution of potential outcomes is Hadamard differentiable.*

Under some mild additional conditions on the c.d.f., Assumption 3 can be verified to hold for all four inequality measures considered in subsection 3.4. The coefficient of variation and the Theil index are known functions of expectations, which are already linear functionals, and therefore satisfy Assumption 3 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.[14] 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.[15]

4.2.2 Asymptotic Normality, Efficiency and Valid Inference

We now derive the limiting distribution of estimators of inequality measures for weighted distributions. We first show that under an additional regularity condition, they will be asymptotically equivalent to a sum of terms that do not depend on the estimated weights, but instead, on the true ones.

ASSUMPTION 4 [**Smoothness**] *For all weighting functions ω considered, $\Pr\left[Y \leq y | X = x, T = t\right]$ is continuously differentiable for all x in $\mathcal{X}$, and at all $(t, y) \in \{0, 1\} \times \mathcal{Y}$.*

Assumption 4 is analogous to the more technical requirement imposed by Hirano, Imbens and Ridder (2003, assumption 3) that the conditional expectation of Y be continuously differentiable. As we have a more general framework, we need that the conditional weighted probability be continuously differentiable.

We are now able to obtain the limiting distribution of our estimators. For that goal, we first consider a proposition that establishes uniform root-N consistency for $\widehat{F}_Y^{\omega=\widehat{\omega}}$ and presents 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 $\widehat{\nu}$.

The assumptions required for the following proposition have been stated along the text; we also invoke an assumption presented in the appendix (Assumption A.1) that guarantees uniform convergence of the estimated propensity-score to the true one.

---

[14]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 ψ^ν 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.

[15]Some of the inequality measures of subsection 3.4 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.

PROPOSITION 1 *Let $h\left(\cdot, p\left(x\right)\right) \equiv \partial\omega\left(\cdot, p\left(x\right)\right)/\partial p\left(x\right)$. Then, under Assumptions 1, 2, 4 and A.1:*

$$\sqrt{N}\left(\widehat{F}_Y^{\omega=\widehat{\omega}} - F_Y^{\omega}\right) = \Psi_N^{\omega} + o_p\left(1\right) \qquad uniformly\ in \quad \ell^{\infty}\left(\mathcal{Y} \times \{0,1\} \times \mathcal{X}\right),$$

where at a given $y \in \mathcal{Y}$

$$
\begin{aligned}
\Psi_N^{\omega}\left(y\right) \;=\; & N^{-1/2}\sum_{i=1}^{N}\omega\left(T_i, p\left(X_i\right)\right)\mathbb{I}\{Y_i \leq y\} \\
& + E\left[h\left(T, p\left(X_i\right)\right)\cdot\mathbb{I}\{Y \leq y\}|X = X_i\right]\left(T_i - p\left(X_i\right)\right) \\
& - E\left[\omega\left(T, p\left(X\right)\right)\cdot\mathbb{I}\{Y \leq y\}\right].
\end{aligned}
$$

Proposition 1 follows after we specialize the results in Newey (1994) to the missing data case and Hirano, Imbens and Ridder (2003) and Chen, Hong and Tarozzi (2008) to the indicator function. Proposition 1 is a general result that can be used along with the four weighting functions considered here.

One can check that the derivatives of the four weighting functions with respect to $p\left(x\right)$ are

$$
\begin{aligned}
h_1\left(t, p\left(x\right)\right) & \equiv \partial\omega_1\left(t, p\left(x\right)\right)/\partial p\left(x\right) = -t/p^2\left(x\right) \\
h_0\left(t, p\left(x\right)\right) & \equiv \partial\omega_0\left(t, p\left(x\right)\right)/\partial p\left(x\right) = \left(1 - t\right)/\left(1 - p\left(x\right)\right)^2 \\
h_{11}\left(t, p\left(x\right)\right) & \equiv \partial\omega_{11}\left(t, p\left(x\right)\right)/\partial p\left(x\right) = 0 \\
h_{01}\left(t, p\left(x\right)\right) & \equiv \partial\omega_{01}\left(t, p\left(x\right)\right)/\partial p\left(x\right) = \left(1 - t\right)/\left[p\cdot\left(1 - p\left(x\right)\right)^2\right].
\end{aligned}
$$

As the estimators of the inequality treatment effect parameters are simply differences in estimators of inequality measures of weighted distributions, we can establish the following result as a direct consequence of Proposition 1:

PROPOSITION 2 *Under Assumptions 1-4 and A.1:*

$$
\begin{aligned}
& \sqrt{N}\cdot\left(\widehat{\Delta}^{\nu} - \Delta^{\nu}\right) \\
& = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_1\left(T_i, p\left(X_i\right)\right)\cdot\psi^{\nu}\left(Y_i; F_Y^{\omega=\omega_1}\right) - \omega_0\left(T_i, p\left(X_i\right)\right)\cdot\psi^{\nu}\left(Y_i; F_Y^{\omega=\omega_0}\right) \\
& \qquad + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(E\left[h_1\left(T, p\left(X\right)\right)\cdot\psi^{\nu}\left(Y; F_Y^{\omega=\omega_1}\right)|X = X_i\right]\right. \\
& \qquad \left. - E\left[h_0\left(T, p\left(X\right)\right)\cdot\psi^{\nu}\left(Y; F_Y^{\omega=\omega_0}\right)|X = X_i\right]\right)\cdot\left(T_i - p\left(X_i\right)\right) + o_p(1) \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \xrightarrow{D} N\left(0, V\right)
\end{aligned}
$$

$$
\begin{aligned}
& \sqrt{N}\cdot\left(\widehat{\Delta}_T^{\nu} - \Delta_T^{\nu}\right) \\
& = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_{11}\left(T_i, p\left(X_i\right)\right)\cdot\psi^{\nu}\left(Y_i; F_Y^{\omega=\omega_{11}}\right) - \omega_{01}\left(T_i, p\left(X_i\right)\right)\cdot\psi^{\nu}\left(Y_i; F_Y^{\omega=\omega_{01}}\right) \\
& \qquad - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}E\left[h_{01}\left(T, p\left(X\right)\right)\cdot\psi^{\nu}\left(Y; F_Y^{\omega=\omega_{01}}\right)|X = X_i\right]\cdot\left(T_i - p\left(X_i\right)\right) + o_p(1) \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \xrightarrow{D} N\left(0, V_T\right)
\end{aligned}
$$

$$\sqrt{N} \cdot \left(\widehat{\Delta}_C^\nu - \Delta_C^\nu\right) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \psi^\nu \left(Y_i; F_Y\right) - \omega_0 \left(T_i, p\left(X_i\right)\right) \cdot \psi^\nu \left(Y_i; F_Y^{\omega=\omega_0}\right)$$

$$- \frac{1}{\sqrt{N}} \sum_{i=1}^N E\left[h_0\left(T, p\left(X\right)\right) \cdot \psi^\nu \left(Y; F_Y^{\omega=\omega_0}\right) | X = X_i\right] \cdot \left(T_i - p\left(X_i\right)\right) + o_p(1)$$

$$\xrightarrow{D} N\left(0, V_C\right)$$

where V, V_T and V_C , whose formulas are given below, are the semiparametric efficiency bounds for, respectively, Δ^ν, Δ_T^ν, and Δ_C^ν.

$$V = E[(\omega_1\left(T, p\left(X\right)\right) \cdot \psi^\nu \left(Y; F_Y^{\omega=\omega_1}\right) - \omega_0\left(T, p\left(X\right)\right) \cdot \psi^\nu \left(Y; F_Y^{\omega=\omega_0}\right)$$
$$+ (E\left[h_1\left(T, p\left(X\right)\right) \cdot \psi^\nu \left(Y; F_Y^{\omega=\omega_1}\right) | X\right]$$
$$- E\left[h_0\left(T, p\left(X\right)\right) \cdot \psi^\nu \left(Y; F_Y^{\omega=\omega_0}\right) | X\right]) \cdot \left(T - p\left(X\right)\right))^2]$$

$$V_T = E[(\omega_{11}\left(T, p\left(X\right)\right) \cdot \psi^\nu \left(Y; F_Y^{\omega=\omega_{11}}\right) - \omega_{01}\left(T, p\left(X\right)\right) \cdot \psi^\nu \left(Y; F_Y^{\omega=\omega_{01}}\right)$$
$$- E\left[h_{01}\left(T, p\left(X\right)\right) \cdot \psi^\nu \left(Y; F_Y^{\omega=\omega_{01}}\right) | X\right] \cdot \left(T - p\left(X\right)\right))^2]$$

$$V_C = E[(\psi^\nu \left(Y; F_Y\right) - \omega_0\left(T, X\right) \cdot \psi^\nu \left(Y; F_Y^{\omega=\omega_0}\right)$$
$$- E\left[h_0\left(T, p\left(X\right)\right) \cdot \psi^\nu \left(Y; F_Y^{\omega=\omega_0}\right) | X\right] \cdot \left(T - p\left(X\right)\right))^2]$$

Furthermore, inference based on the bootstrap will be valid as bootstrapped variances will be consistent to V, V_T and V_C.

Valid inference for inequality treatment effect parameters can be implemented either by estimation of the analytical expressions for the variance terms presented in Proposition 2 or by resampling methods, such as the bootstrapping as stated in that same proposition. Recently, Abadie and Imbens (2008) gave an example of a root-N consistent and asymptotically normal estimator for which the bootstrap does not work, so it is worth checking if usual conditions for inference based on the bootstrap are valid in our setting. 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.

5 A Monte Carlo Exercise

In this section we report the results of Monte Carlo exercises. The interest is in learning how the estimators for the overall inequality treatment effect (ITE). One thousand (1,000) replications of the experiment with sample sizes of 250, 1,000 and 4,000 observations were considered.

We design 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 the 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 $X = [X_1, X_2]^\mathsf{T}$ 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: $E[X_1] = \mu_{X_1}$, $E[X_2] = \mu_{X_2}$ 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 three possible distributions for η: (i) logistic, $\eta \sim F_\eta(n) = 10\left(1 + \exp\left(-\pi n/\sqrt{3}\right)\right)^{-1}$; (ii) normal, $\eta \sim F_\eta(n) = \int_{-\infty}^n 10\left(2\pi\right)^{-1/2}\exp\left(-z^2/2\right)dz$; (iii) uniform, $\eta \sim F_\eta(n) = \sqrt{12}n/120 + 1/2$. 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

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

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 1: Parameter specification for Monte Carlo Exercise

$coeff.\backslash\jmath$	0	1	2	3	4	5
δ_j	-1	10	2	-10	-3	10
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.

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

η	Logistic		Normal		Uniform	
ν\Distribution	$Y(0)$	$Y(1)$	$Y(0)$	$Y(1)$	$Y(0)$	$Y(1)$
Mean	76.59	193.33	76.61	193.19	76.61	193.12
Standard Deviation (s.d.)	28.17	123.33	28.21	123.23	28.12	122.90
Mean of Logarithm	-33.58	52.92	-33.58	52.88	-33.54	52.87
S.D. of Logarithm	38.72	48.39	38.68	48.34	38.58	48.29
10th Percentile	43.50	97.68	43.52	97.68	43.58	97.73
1st Quartile	58.64	124.23	58.66	124.18	58.69	124.17
Median	75.04	162.57	75.04	162.50	75.07	162.55
3rd Quartile	91.30	222.28	91.31	222.25	91.31	222.10
90th Percentile	108.66	315.46	108.65	314.93	108.57	314.68

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

η	Logistic		Normal		Uniform	
ν\Distribution	$Y(0)$	$Y(1)$	$Y(0)$	$Y(1)$	$Y(0)$	$Y(1)$
Coefficient of Variation	.3677	.6389	.3681	.6377	.3671	.6363
Interquartile Range	.3264	.9806	.3262	.9804	.3259	.9789
Theil Index	.0651	.1466	.0651	.1464	.0647	.1458
Gini Coefficient	.1961	.2817	.1960	.2814	.1956	.2810

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.1 and A.2 with logistic and uniform d.g.p.'s are left to the Appendix.[16]

We provide in Tables 4 and 5 results for the unfeasible estimator and also for five estimators that do not use information from $Y(1)$ and $Y(0)$ but instead use information from usually available data (Y, T, X). The first one is the estimator proposed here and labeled "weighted estimator". In order to simplify the estimation procedure, we considered a parametric first step, in which we computed the propensity-score by a logit using the correct quadratic specification.

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 to 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 $X_1 X_2$, one for each group ($T = 0$ and $T = 1$). Save residuals $\hat{u}_j$ and compute s_j^2 where $j = 0, 1$ indexes treatment groups, $s_j^2 = (N_j - 6)^{-1} \sum_i^{N_j} (\hat{u}_{ji})^2$ and $N_1 = \sum_i^N T_i$ and $N_0 = N - N_1$. Save coefficient estimates for group $T = 0$, $\hat{\gamma}_{00}, \hat{\gamma}_{10}, \hat{\gamma}_{20}, \hat{\gamma}_{30}, \hat{\gamma}_{40}, \hat{\gamma}_{50}$. Then, let Y_i^* be counterfactual outcome of treated observation i:

$$Y_i^* = \hat{\gamma}_{00} + \hat{\gamma}_{10}X_{1i} + \hat{\gamma}_{20}X_{2i} + \hat{\gamma}_{30}X_{1i}^2 + \hat{\gamma}_{40}X_{2i}^2 + \hat{\gamma}_{50}X_{1i}X_{2i} + \sqrt{s_0^2/s_1^2} \cdot \hat{u}_{1i}$$

---

[16]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.

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.

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.[17] These 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 $D_y = \mathbb{I}\{Y \le y\}$ 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, for each y we used the full quadratic 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.1 and A.2 show distribution features for each one of the estimators of inequality treatment effect parameters. We report average, standard deviation and quantiles (10th percentile, median, and 90th percentile) for the four types of treatment effects on inequality measures here considered (coefficient of variation, interquartile range, Theil index and Gini coefficient). 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 estimator, 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. Relatively to other estimators also being analyzed, the weighted estimator clearly dominates in all considered criteria the naive estimator and the two estimators based on Juhn, Murphy and Pierce (1993).

---

[17]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 $N = 4,000$.

6   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[18]. 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.[19] 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.

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).[20]

---

[18]According to the Brazilian Ministry of Labor, during January 01 1996 and October 27 2007, exactly R\$ 4,312,426,625.55 or US\$2,661,991,699.67 (using June 2008 exchange rate) have been spent on PLANFOR all over the country.

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

[20]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

Because within each class we have few observations to rely solely on t-tests, which are only asymptotically valid,[21] we pool the data.[22] However, the proportion of treated units within class greatly varied across classes, from 0.11 to 0.86 and, therefore, unless covariates were mean independent of classes, t-tests with pooled data would not be a valid procedure for testing imbalance in covariates. Indeed, there is evidence that covariates distributions do vary by class, as regressions of covariates on class dummies always produced significant F-stats. This is explained by the fact that there were many different courses being offered attracting people with different backgrounds.

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 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 never worked before, whereas in Fortaleza that number was around 20%.

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.[23] We also had a sample size reduction when

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.

[21] For robustness we present a summary statistics table (Table A.3) 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.

[22] Fisher's permutation tests are commonly used as an alternative procedure when we cannot rely on asymptotic approximations. Their validity rely, however, on an exchangeability assumption. Also, given the typical class size we have, they will not be very powerful tests for the difference in means. For a detailed survey on exact inference see Agresti (1992).

[23] 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.[24]

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 1,000 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 we obtain no difference in inequality, which reinforces that interpretation. 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, all detected significant positive effects. We are agnostic regarding these results as we know, from the Monte Carlo exercises, that these estimators are in general biased ones.

7   Conclusion

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 in a part fixing the distribution of covariates and another that is merely a composition effect, induced by different distribution in covariates.

The estimation strategy adopted here is useful when the individual decision to participate in the program (the treatment) depends on observable characteristics. If the identification restrictions hold, then the reweighing method allows identifying the distribution of potential outcomes and, therefore, the causal impact of the program on many functionals of interest for policy analysis, such as inequality indices.

---

[24]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.

Results from our Monte Carlo study suggest that, when looking at distributional aspects, consistent alternatives to the weighted estimator would probably require a very complete model for the conditional distribution of Y given covariates. Because the weighted estimator only requires estimation of a single conditional expectation, it is a easy and readily implementable alternative; which clearly contrasts to non-parametric estimators of the conditional distribution, which may also suffer from additional dimensionality difficulties. 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 that estimator and those proposed by Chernozhukov, Fernandez-Val and Melly (2009) and Rothe (2010).

APPENDIX

Proofs

Proof of Lemma 1: See Lemma 1 in Firpo (2007).

Proof of Corollary 1:

By definition of ITE parameters, we have that they are the following differences in functionals of the distributions:

$$\Delta^\nu = \nu\left(F_{Y(1)}\right) - \nu\left(F_{Y(0)}\right) = \nu\left(F_Y^{\omega=\omega_1}\right) - \nu\left(F_Y^{\omega=\omega_0}\right)$$

$$\Delta^\nu_{T=1} = \nu\left(F_{Y(1)|T=1}\right) - \nu\left(F_{Y(0)|T=1}\right) = \nu\left(F_Y^{\omega=\omega_{11}}\right) - \nu\left(F_Y^{\omega=\omega_{01}}\right)$$

$$\Delta^\nu_C = \nu\left(F_Y\right) - \nu\left(F_{Y(0)}\right) = \nu\left(F_Y\right) - \nu\left(F_Y^{\omega=\omega_0}\right)$$

and therefore those three parameters can be expressed as functions of the observable data (Y,T,X).

$$Q.E.D.$$

Proof of Proposition 1

In order to be able to prove Proposition 1, we need first to guarantee that the propensity-score estimated here as proposed by Hirano, Imbens and Ridder (2003) is uniformly consistent for the true propensity-score. They show that with an extra assumption uniform consistency is achieved. For sake of completeness we state such assumption and the desired result:

ASSUMPTION A.1 [**First Step**]:

(i) $\mathcal{X}$ is a compact subset of $\mathbb{R}^r$;

(ii) the density of X, $f(x)$, satisfies $0 < \inf_{x\in\mathcal{X}} f(x) \leq \sup_{x\in\mathcal{X}} f(x) < \infty$

(iii) $p(x)$ is s-times continuously differentiable, where $s \geq 7r$ and r is the dimension of X;

(iv) the order of $H_K(x)$, K, is of the form $K = C\cdot N^c$ where C is a constant and $c \in \left(\frac{1}{4\left(\frac{s}{r}-1\right)}, \frac{1}{9}\right)$

Newey (1995, 1997) has established that for orthogonal polynomials $H_K(x)$ and compact $\mathcal{X}$:

$$\Gamma(K) = \sup_{x\in\mathcal{X}} \|H_K(x)\| \leq C \cdot K \tag{A-1}$$

where C is a generic constant. Note then that because of part (iv) of Assumption A.1, Γ will be a function of N since K is assumed to be a function of N.

Uniform consistency of the estimated propensity-score is guaranteed by the following lemma:

LEMMA A.1 [**First Step**]: Under Assumptions 2 and A.1, the following results hold:

(I) $\sup_{x\in\mathcal{X}} |p(x) - p_K(x)| \leq C \cdot \Gamma(K) \cdot K^{-s/2r} \leq C \cdot K^{1-s/2r} \leq C \cdot N^{(1-s/2r)c} = o(1)$; where $p_K(x) = L(H_K(x)'\pi_K)$ and $\pi_K = \arg\max_{\pi\in\mathbb{R}^K} E\Big\{ p(X)\cdot\log(L(H_K(X)'\pi)) + (1-p(X))\cdot\log(1-L(H_K(X)'\pi)) \Big\}$;

(II) $\|\hat{\pi}_K - \pi_K\| = O_p\left(\sqrt{\frac{K(N)}{N}}\right) \leq C \cdot O_p\left(\sqrt{\frac{N^c}{N}}\right) \leq C \cdot O_p\left(N^{\frac{c-1}{2}}\right) = o_p(1)$;

(III) $\sup_{x\in\mathcal{X}} |p(x) - \hat{p}(x)| \leq C_1 \cdot N^{(1-s/2r)c} + O_p\left(\Gamma(K) \cdot \sqrt{\frac{K(N)}{N}}\right) \leq C_1 \cdot N^{(1-s/2r)c} + C_2 \cdot N^{(3c-1)/2} \cdot O_p(1) = o_p(1)$;

(IV) There is $\varepsilon > 0$: $\lim_{N\to\infty} \Pr[\varepsilon < \inf_{X\in\mathcal{X}} \hat{p}(X) \leq \sup_{X\in\mathcal{X}} \hat{p}(X) < 1 - \varepsilon] = 1$.

Proof of Lemma A.1: See Hirano, Imbens and Ridder (2003), Lemmas 1 and 2.

Let us now consider $\sqrt{N}\left(\widehat{F}_Y^{\omega=\widehat{\omega}}(y) - F_Y^\omega(y)\right)$:

$$\sqrt{N}\left(\widehat{F}_Y^{\omega=\widehat{\omega}}(y) - F_Y^\omega(y)\right)$$

$$= N^{-1/2}\sum_{i=1}^N \left(\omega\left(T_i, \widehat{p}(X_i)\right) \mathbb{I}\{Y_i \le y\} - E\left[\omega\left(T, p(X)\right) \cdot \mathbb{I}\{Y \le y\}\right]\right)$$

$$= N^{-1/2}\sum_{i=1}^N \left(\omega\left(T_i, \widehat{p}(X_i)\right) - \omega\left(T_i, p(X_i)\right)\right) \mathbb{I}\{Y_i \le y\}$$

$$+ N^{-1/2}\sum_{i=1}^N \left(\omega\left(T_i, p(X_i)\right) \mathbb{I}\{Y_i \le y\} - E\left[\omega\left(T, p(X)\right) \cdot \mathbb{I}\{Y \le y\}\right]\right)$$

and

$$N^{-1/2}\sum_{i=1}^N \left(\omega\left(T_i, \widehat{p}(X_i)\right) - \omega\left(T_i, p(X_i)\right)\right) \mathbb{I}\{Y_i \le y\}$$

$$= N^{-1/2}\sum_{i=1}^N \left(\omega\left(T_i, \widehat{p}(X_i)\right) - \omega\left(T_i, p(X_i)\right) - h\left(T_i, p(X_i)\right)\left(\widehat{p}(X_i) - p(X_i)\right)\right) \mathbb{I}\{Y_i \le y\} \quad \text{(A-2)}$$

$$+ N^{-1/2}\sum_{i=1}^N \left(h\left(T_i, p(X_i)\right) \mathbb{I}\{Y_i \le y\}\right)\left(\widehat{p}(X_i) - p(X_i)\right) \quad \text{(A-3)}$$

$$- N^{-1/2}\sum_{i=1}^N E\left[E\left[h\left(T, p(X)\right) \cdot \mathbb{I}\{Y \le y\}|X\right]\left(\widehat{p}(X) - p(X)\right)\right]$$

$$+ N^{-1/2}\sum_{i=1}^N E\left[E\left[h\left(T, p(X)\right) \cdot \mathbb{I}\{Y_i \le y\}|X\right]\left(\widehat{p}(X) - p(X)\right)\right] \quad \text{(A-4)}$$

$$- N^{-1/2}\sum_{i=1}^N E\left[h\left(T, p(X_i)\right) \cdot \mathbb{I}\{Y_i \le y\}|X = X_i\right]\left(T_i - p(X_i)\right)$$

$$+ N^{-1/2}\sum_{i=1}^N E\left[h\left(T, p(X_i)\right) \cdot \mathbb{I}\{Y \le y\}|X = X_i\right]\left(T_i - p(X_i)\right).$$

Therefore:

$$\sqrt{N}\left(\widehat{F}_Y^{\omega=\widehat{\omega}}(y) - F_Y^\omega(y)\right) = \Psi_N^\omega(y) + \text{(A-1.)} + \text{(A-2.)} + \text{(A-3.)}$$

where

$$\Psi_N^\omega(y) = N^{-1/2}\sum_{i=1}^N \psi_i^\omega(y)$$

$$\begin{aligned}
\psi_i^\omega(y) &= \psi_y^\omega(Y_i, T_i, X_i) = \omega(T_i, p(X_i)) \, \mathbb{I}\{Y_i \le y\} \\
&\quad + E[h(T, p(X_i)) \cdot \mathbb{I}\{Y \le y\} | X = X_i](T_i - p(X_i)) \\
&\quad - E[\omega(T, p(X)) \cdot \mathbb{I}\{Y \le y\}].
\end{aligned}$$

We now investigate whether $\|(\text{A-1.}) + (\text{A-2.}) + (\text{A-3.})\|_\infty$ is $o_p(1)$. We notice that we for all choices of weights, ω (ω_1, ω_1, ω_{11}, and ω_{01}), these expressions have already been analyzed by Hirano, Imbens and Ridder (2003), except that they have considered Y, while we replace that random variable by the indicator variable $\mathbb{I}\{Y \le y\}$ and take the sup over y in $\mathcal{Y}$, the support of Y. See for example, Theorem 1 in Hirano, Imbens and Ridder (2003), in which they consider ω_1 as weights and work in details the algebra. Under the same conditions considered here, but adding a bounding restriction to the variance of Y, they obtain that their analogs to terms (A-1.), (A-2.) and (A-3.) are all $o_p(1)$. Given that the the indicator function is bounded (and so is its variance), we can adapt their results to ours, obtaining a uniform root-N consistency result:

$$\sqrt{N}\left(\widehat{F}_Y^{\omega = \widehat{\omega}} - F_Y^\omega\right) = \Psi_N^\omega + o_p(1) \qquad \text{in} \quad \ell^\infty\left(\mathcal{Y} \times \{0,1\} \times \mathcal{X}\right).$$

We now check whether the empirical process Ψ_N^ω converges to a tight random element in the collection of bounded functions $\mathcal{Y} \times \{0,1\} \times \mathcal{X} \to \mathbb{R}$. Consider the i.i.d. sequence $\{Y_i, T_i, X_i^\intercal\}_{i=1}^N$ of random vectors and write $\psi_y^\omega(Y_i, T_i, X_i)$ as

$$\begin{aligned}
\psi_y^\omega(Y_i, T_i, X_i) &= \omega(T_i, p(X_i)) \, \mathbb{I}\{Y_i \le y\} \\
&\quad + \alpha_y^\omega(T_i, p(X_i))
\end{aligned}$$

where

$$\alpha_y^\omega(t, p(x)) = E[h(t, p(x)) \cdot \mathbb{I}\{Y \le y\} | X = x](t - p(x))$$

is the contribution of estimation of $p(x)$ to the influence function ψ_y. Now, note that ψ_y^ω is a weighted indicator function plus a zero mean term, with bounded variance. Thus,

$$N^{-1} \sum_{i=1}^N \psi_y^\omega(Y_i, T_i, X_i)$$

is weighted empirical distribution plus a zero mean $O_p\left(N^{-1/2}\right)$. Therefore, because the weighting function satisfies the conditions in Example 19.12 of van der Vaart (1998), we will have that the sequence of empirical processes Ψ_N^ω is in a Donsker class and converges in $\ell^\infty\left(\mathcal{Y} \times \{0,1\} \times \mathcal{X}\right)$ to a tight random element term with mean zero and finite variance.

<div style="text-align: right">Q.E.D.</div>

Proof of Proposition 2:

With Proposition 1 at hand, we can then apply a functional Delta-method to the inequality measures here considered, as those will be Hadamard differentiable. Thus, according to Theorem

20.8 and 20.9 in van der Vaart (1998), we have that

$$\sqrt{N} \cdot \left(\nu \left(\widehat{F}_Y^{\omega = \widehat{\omega}} \right) - \nu \left(F_Y^{\omega} \right) \right) = \psi^{\nu} \left(\Psi_N^{\omega} \right) + o_p \left(1 \right) = N^{-1/2} \sum_{i=1}^{N} \psi_i^{\nu} \left(F_Y^{\omega} \right) \circ \psi^{\omega} \left(Y_i, T_i, X_i \right) + o_p \left(1 \right)$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \omega \left(T_i, p \left(X_i \right) \right) \cdot \psi^{\nu} \left(Y_i; F_Y^{\omega} \right)$$

$$+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} E \left[h \left(T, p \left(X \right) \right) \cdot \psi^{\nu} \left(Y; F_Y^{\omega} \right) | X = X_i \right] \cdot \left(T_i - p \left(X_i \right) \right) + o_p(1)$$

$$\xrightarrow{D} N \left(0, V_{\nu}^{\omega} \right)$$

where $\psi^{\nu} \left(y; F_Y^{\omega} \right)$ is the *influence function* of the functional ν. The fact that it shares the same notation with the Hadamard derivative is not mere coincidence: If ν is Hadamard differentiable, then $\lim_{s \downarrow 0} \mathrm{d}\nu \left((1 - s) F_Y^{\omega} + s\delta_y \right) / \mathrm{d}s = \psi^{\nu} \left(y; F_Y^{\omega} \right)$, where δ_y is the (Dirac) probability measure that put mass 1 at value y. Conveniently enough, $\int \psi^{\nu} \left(y; F_Y^{\omega} \right) \mathrm{d}F_Y^{\omega} \left(y \right) = 0$.

For the four inequality measures here considered, the influence functions are well known or easily derivable. Their expressions can be found either directly or constructed after a simple delta-method argument from the following references: Lehmann (1999, chapter 6) for the coefficient of variation, interquartile range and Theil index; and the Gini coefficient in Cowell and Victoria-Feser (1996) and Schluter and Trede (2003).

A straightforward application to the four weighting functions separately to allows us to obtain a asymptotic normality result for the estimators $\widehat{\Delta}^{\nu}$, $\widehat{\Delta}_T^{\nu}$ and $\widehat{\Delta}_C^{\nu}$ of the inequality treatment effects Δ^{ν}, Δ_T^{ν} and Δ_C^{ν}:

$$\sqrt{N} \cdot \left(\widehat{\Delta}^{\nu} - \Delta^{\nu} \right) = \psi^{\nu} \left(\Psi_N^{\omega_1} \right) - \psi^{\nu} \left(\Psi_N^{\omega_0} \right) + o_p \left(1 \right)$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \omega_1 \left(T_i, p \left(X_i \right) \right) \cdot \psi^{\nu} \left(Y_i; F_Y^{\omega = \omega_1} \right) - \omega_0 \left(T_i, p \left(X_i \right) \right) \cdot \psi^{\nu} \left(Y_i; F_Y^{\omega = \omega_0} \right)$$

$$+ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(E \left[h_1 \left(T, p \left(X \right) \right) \cdot \psi^{\nu} \left(Y; F_Y^{\omega = \omega_1} \right) | X = X_i \right] \right.$$

$$\left. - E \left[h_0 \left(T, p \left(X \right) \right) \cdot \psi^{\nu} \left(Y; F_Y^{\omega = \omega_0} \right) | X = X_i \right] \right) \cdot \left(T_i - p \left(X_i \right) \right) + o_p(1)$$

$$\xrightarrow{D} N \left(0, V \right)$$

$$\sqrt{N} \cdot \left(\widehat{\Delta}_T^{\nu} - \Delta_T^{\nu} \right) = \psi^{\nu} \left(\Psi_N^{\omega_{11}} \right) - \psi^{\nu} \left(\Psi_N^{\omega_{01}} \right) + o_p \left(1 \right)$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \omega_{11} \left(T_i, p \left(X_i \right) \right) \cdot \psi^{\nu} \left(Y_i; F_Y^{\omega = \omega_{11}} \right) - \omega_{01} \left(T_i, p \left(X_i \right) \right) \cdot \psi^{\nu} \left(Y_i; F_Y^{\omega = \omega_{01}} \right)$$

$$- \frac{1}{\sqrt{N}} \sum_{i=1}^{N} E \left[h_{01} \left(T, p \left(X \right) \right) \cdot \psi^{\nu} \left(Y; F_Y^{\omega = \omega_{01}} \right) | X = X_i \right] \cdot \left(T_i - p \left(X_i \right) \right) + o_p(1)$$

$$\xrightarrow{D} N \left(0, V_T \right)$$

$$\sqrt{N} \cdot \left(\widehat{\Delta}_C^\nu - \Delta_C^\nu \right) = \psi^\nu \left(\Psi_N \right) - \psi^\nu \left(\Psi_N^{\omega_0} \right) + o_p \left(1 \right)$$

$$= \frac{1}{\sqrt{N}} \sum_{i=1}^N \psi^\nu \left(Y_i ; F_Y \right) - \omega_0 \left(T_i, p\left(X_i \right) \right) \cdot \psi^\nu \left(Y_i ; F_Y^{\omega = \omega_0} \right)$$

$$- \frac{1}{\sqrt{N}} \sum_{i=1}^N E \left[h_0 \left(T, p\left(X \right) \right) \cdot \psi^\nu \left(Y ; F_Y^{\omega = \omega_0} \right) | X = X_i \right] \cdot \left(T_i - p\left(X_i \right) \right) + o_p(1)$$

$$\overset{D}{\longrightarrow} N \left(0, V_C \right)$$

where V, V_T and V_C, whose formulae were given in the statement of the Theorem are simply

$$V = E \left[\left(\psi^\nu \left(\Psi_N^{\omega_1} \right) - \psi^\nu \left(\Psi_N^{\omega_0} \right) \right)^2 \right]$$

$$V_T = E \left[\left(\psi^\nu \left(\Psi_N^{\omega_{11}} \right) - \psi^\nu \left(\Psi_N^{\omega_{01}} \right) \right)^2 \right]$$

$$V_C = E \left[\left(\psi^\nu \left(\Psi_N \right) - \psi^\nu \left(\Psi_N^{\omega_0} \right) \right)^2 \right].$$

We now prove the efficiency claim in Proposition 2. Efficiency follows from van der Vaart (1991) and van der Vaart (1998, Theorem 25.48) who established that if ν is Hadamard differentiable and the empirical distribution is pointwise efficient for all $y \in Y$, then $\widehat{\nu}$ is efficient. Following Hahn (1998), Hirano, Imbens and Ridder (2003) and Cattaneo (2009), who derived efficiency bounds in the multivariate treatment context, we can again replace Y by $\mathrm{I\!I}\{Y \le y\}$ and obtain for every point y, efficiency bounds for $F_{Y(1)}\left(y \right)$, $F_{Y(0)}\left(y \right)$, $F_{Y(1)|T}\left(y | 1 \right)$, and $F_{Y(0)|T}\left(y | 1 \right)$. Inspection of their formulae for the efficient score and replacement of Y by $\mathrm{I\!I}\{Y \le y\}$ yields the respective efficiency bounds on the c.d.f.'s at y: $V_{FY(1)(y)}^{SEB} = E \left[\left(\Psi_N^{\omega_1} \left(y \right) \right)^2 \right]$, $V_{FY(0)(y)}^{SEB} = E \left[\left(\Psi_N^{\omega_0} \left(y \right) \right)^2 \right]$, $V_{FY(1)|T=1(y)}^{SEB} = E \left[\left(\Psi_N^{\omega_{11}} \left(y \right) \right)^2 \right]$, and $V_{FY(0)|T=1(y)}^{SEB} = E \left[\left(\Psi_N^{\omega_{01}} \left(y \right) \right)^2 \right]$.

Now, because, ν is Hadamard differentiable, we can write the efficiency bounds for ν_1, ν_0, ν_{11} and ν_{01} as: $V_{\nu_1}^{SEB} = E \left[\left(\psi^\nu \left(\Psi_N^{\omega_1} \right) \right)^2 \right]$, $V_{\nu_0}^{SEB} = E \left[\left(\psi^\nu \left(\Psi_N^{\omega_0} \right) \right)^2 \right]$, $V_{\nu_{11}}^{SEB} = E \left[\left(\psi^\nu \left(\Psi_N^{\omega_{11}} \right) \right)^2 \right]$, and $V_{\nu_{01}}^{SEB} = E \left[\left(\psi^\nu \left(\Psi_N^{\omega_{01}} \right) \right)^2 \right]$. Finally, given the linearity of the influence functions, we get that $V_{\Delta^\nu}^{SEB} = V$, $V_{\Delta_T^\nu}^{SEB} = V_T$, $V_{\Delta_C^\nu}^{SEB} = V_C$.

Validity of bootstrap-based inference is a direct consequence of the functional delta-method applied to the bootstrap. We have that our inequality measures are Hadamard differentiable functionals; and that sums of our weighted c.d.f.s estimators converge uniformly in law to tight measurable random elements in $\ell^\infty \left(\mathcal{Y} \times \{0,1\} \times \mathcal{X} \right)$; finally our original sample $\{Y_i, T_i, X_i\}_{i=1}^N$ is i.i.d..

Let $F_{Y,T,X}$ be the joint distribution of (Y, T, X). The weighted marginal distribution of Y

using weights given by $\omega(T, p(X))$ can be written as a linear functional of the joint c.d.f.:

$$
\begin{aligned}
F_Y^\omega(\cdot) &= E\left[\omega(T, p(X)) \, \mathbb{I}\{Y \le \cdot\}\right] \\
&= \int_{\mathcal{Y}} \int_{\mathcal{X}} \omega(0, p(x)) \, \mathbb{I}\{z \le \cdot\} \mathrm{d}F_{Y,T,X}(z, 0, x) \\
&\quad + \int_{\mathcal{Y}} \int_{\mathcal{X}} \omega(1, p(x)) \, \mathbb{I}\{z \le \cdot\} \mathrm{d}F_{Y,T,X}(z, 1, x) \\
&= \nu_\omega(F_{Y,T,X}).
\end{aligned}
$$

The functional ν of the weighted c.d.f. is therefore the composite functional $\nu(F_Y^\omega) = \nu \circ \nu_\omega(F_{Y,T,X})$. Our estimator $\widehat{\nu}$ can be written as $\widehat{\nu} = \nu\left(\widehat{F}_Y^{\omega=\widehat{\omega}}\right) = \nu \circ \nu_\omega\left(\widehat{F}_{Y,T,X}\right)$, where $\widehat{F}_{Y,T,X}$ is the joint empirical distribution.

That representation is useful for applying Theorem 23.9 of van der Vaart (1998) that establishes validity of the bootstrap for Hadamard differentiable functionals. Example 23.11 of van der Vaart (1998) adapts that theorem when the empirical process is the empirical distribution. As a consequence, we obtain the following result:

$$
\begin{aligned}
\sqrt{N}\left(\nu \circ \nu_\omega\left(\widehat{F}_{Y,T,X}^*\right) - \nu \circ \nu_\omega\left(\widehat{F}_{Y,T,X}\right)\right) &= \sqrt{N}\left(\nu\left(\widehat{F}_Y^{\omega=\widehat{\omega},*}\right) - \nu\left(\widehat{F}_Y^{\omega=\widehat{\omega}}\right)\right) \\
&= \psi^\nu(\Psi_N^\omega) + o_p(1).
\end{aligned}
$$

where $\widehat{F}_{Y,T,X}^*$ is the bootstrap empirical distribution of (Y, T, X) and $\widehat{F}_Y^{\omega=\widehat{\omega},*}$ is the bootstrap empirical weighted distribution of Y using, in each replication of the bootstrap, estimated weights in that replication.

Because of the asymptotic equivalence of the previous equation, we know that inference for ν using a bootstrap procedure that draws N observations with replacement in each bootstrap replication from the empirical distribution of (Y, T, X) will be valid and variance estimates obtained by the bootstrap should converge in probability to those presented in Proposition 2.

$Q.E.D.$

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., (2007), "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.

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.

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)

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
		1.166										
	Unfeasible		1.165	1.000	1.155	1.344	0.137	-0.001	0.137	0.109	0.093	0.907
	Naive		1.124	0.965	1.113	1.297	0.133	-0.042	0.139	0.113	0.103	0.895
Mean Treatment Effects	Weighted		1.167	0.932	1.166	1.406	0.184	0.000	0.184	0.143	0.120	0.910
	Location Shift		1.163	0.954	1.163	1.377	0.168	-0.003	0.168	0.133	0.107	0.905
	Log Location Shift		1.170	0.950	1.173	1.382	0.171	0.003	0.171	0.134	0.107	0.904
	CFM		1.207	1.008	1.197	1.407	0.163	0.041	0.168	0.131	0.108	0.901
		0.270										
	Unfeasible		0.243	0.094	0.224	0.406	0.144	-0.027	0.147	0.110	0.092	0.936
	Naive		0.306	0.160	0.281	0.469	0.139	0.036	0.144	0.100	0.074	0.926
CV	Weighted		0.286	0.095	0.277	0.477	0.169	0.017	0.170	0.124	0.098	0.920
	Location Shift		0.286	0.130	0.266	0.464	0.146	0.016	0.147	0.104	0.079	0.923
	Log Location Shift		0.289	0.143	0.270	0.461	0.138	0.020	0.139	0.099	0.077	0.927
	CFM		0.228	0.025	0.219	0.417	0.175	-0.042	0.180	0.133	0.103	0.899
		0.654										
	Unfeasible		0.657	0.459	0.649	0.865	0.161	0.003	0.161	0.127	0.104	0.897
	Naive		0.689	0.485	0.684	0.889	0.160	0.035	0.164	0.128	0.105	0.898
Interquartile Range	Weighted		0.689	0.356	0.699	0.990	0.252	0.035	0.254	0.199	0.164	0.913
	Location Shift		0.715	0.512	0.703	0.931	0.164	0.061	0.175	0.136	0.112	0.871
	Log Location Shift		0.670	0.459	0.661	0.893	0.167	0.016	0.168	0.134	0.113	0.906
	CFM		0.607	0.302	0.618	0.897	0.234	-0.047	0.238	0.185	0.147	0.890
		0.081										
	Unfeasible		0.078	0.030	0.072	0.129	0.045	-0.003	0.045	0.034	0.028	0.940
	Naive		0.097	0.050	0.089	0.151	0.044	0.015	0.047	0.033	0.025	0.917
Theil Index	Weighted		0.087	0.025	0.084	0.148	0.056	0.005	0.056	0.040	0.030	0.917
	Location Shift		0.091	0.040	0.084	0.150	0.049	0.010	0.050	0.035	0.027	0.923
	Log Location Shift		0.092	0.045	0.085	0.146	0.045	0.011	0.046	0.033	0.025	0.924
	CFM		0.065	-0.008	0.068	0.132	0.068	-0.016	0.070	0.047	0.033	0.919
		0.085										
	Unfeasible		0.083	0.042	0.083	0.126	0.034	-0.002	0.034	0.027	0.022	0.903
	Naive		0.115	0.075	0.114	0.157	0.032	0.030	0.044	0.035	0.030	0.776
Gini Coefficient	Weighted		0.112	0.032	0.115	0.193	0.067	0.026	0.072	0.056	0.046	0.885
	Location Shift		0.114	0.056	0.113	0.169	0.044	0.028	0.053	0.042	0.035	0.833
	Log Location Shift		0.103	0.057	0.102	0.150	0.036	0.018	0.040	0.032	0.026	0.864
	CFM		0.076	-0.011	0.085	0.147	0.064	-0.010	0.065	0.048	0.036	0.900

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

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
		1.166										
	Unfeasible		1.167	1.123	1.166	1.214	0.035	0.001	0.035	0.028	0.024	0.897
	Naive		1.127	1.084	1.127	1.172	0.034	-0.040	0.052	0.044	0.041	0.684
Mean Treatment Effects	Weighted		1.168	1.108	1.169	1.226	0.047	0.002	0.047	0.037	0.032	0.899
	Location Shift		1.167	1.112	1.166	1.218	0.041	0.001	0.041	0.033	0.027	0.897
	Log Location Shift		1.181	1.129	1.180	1.235	0.041	0.015	0.044	0.035	0.029	0.876
	CFM		1.169	1.112	1.169	1.224	0.043	0.003	0.043	0.034	0.028	0.908
		0.270										
	Unfeasible		0.267	0.210	0.262	0.329	0.050	-0.003	0.050	0.037	0.029	0.916
	Naive		0.331	0.276	0.325	0.389	0.049	0.061	0.078	0.064	0.056	0.699
CV	Weighted		0.274	0.203	0.272	0.352	0.066	0.005	0.066	0.048	0.036	0.918
	Location Shift		0.322	0.267	0.317	0.380	0.050	0.052	0.072	0.057	0.048	0.757
	Log Location Shift		0.325	0.274	0.319	0.382	0.047	0.056	0.073	0.058	0.050	0.721
	CFM		0.282	0.201	0.286	0.364	0.076	0.012	0.077	0.055	0.041	0.930
		0.654										
	Unfeasible		0.655	0.607	0.654	0.707	0.040	0.001	0.040	0.031	0.027	0.901
	Naive		0.687	0.639	0.687	0.738	0.039	0.033	0.052	0.041	0.037	0.806
Interquartile Range	Weighted		0.656	0.579	0.655	0.736	0.061	0.002	0.061	0.049	0.040	0.901
	Location Shift		0.755	0.703	0.753	0.809	0.042	0.101	0.109	0.101	0.099	0.221
	Log Location Shift		0.687	0.637	0.687	0.739	0.040	0.033	0.052	0.042	0.037	0.803
	CFM		0.669	0.603	0.670	0.735	0.052	0.015	0.054	0.043	0.037	0.893
		0.081										
	Unfeasible		0.081	0.066	0.080	0.097	0.012	0.000	0.012	0.009	0.008	0.895
	Naive		0.100	0.085	0.099	0.115	0.012	0.019	0.022	0.019	0.018	0.562
Theil Index	Weighted		0.082	0.062	0.083	0.103	0.017	0.001	0.017	0.013	0.010	0.915
	Location Shift		0.101	0.084	0.100	0.117	0.014	0.019	0.024	0.020	0.018	0.617
	Log Location Shift		0.098	0.084	0.098	0.113	0.012	0.017	0.021	0.018	0.016	0.605
	CFM		0.087	0.066	0.089	0.108	0.019	0.006	0.020	0.015	0.012	0.915
		0.085										
	Unfeasible		0.085	0.074	0.085	0.097	0.009	0.000	0.009	0.007	0.006	0.896
	Naive		0.118	0.107	0.118	0.129	0.009	0.033	0.034	0.033	0.033	0.017
Gini Coefficient	Weighted		0.089	0.064	0.090	0.113	0.019	0.003	0.019	0.015	0.013	0.895
	Location Shift		0.126	0.109	0.126	0.143	0.014	0.041	0.043	0.041	0.040	0.087
	Log Location Shift		0.111	0.099	0.111	0.123	0.009	0.025	0.027	0.025	0.025	0.142
	CFM		0.095	0.072	0.097	0.117	0.018	0.010	0.021	0.017	0.015	0.849

Table 6: Summary Statistics

| | 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	0.03	0.03	0.04	0.00	-0.01	0.21	0.19	0.18	0.02	0.03
Dummy	(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	701.23	143.45	20.57	19.97	-907.40	-48.67	1.53	-0.20	-0.35	-0.34	-0.43	-0.35
	(37.28)	(47.42)***	(62.04)	(87.75)	(244.09)***	(85.79)	(0.10)	(0.18)	(0.22)	(0.23)	(0.29)	(0.24)
Coefficient of Variation	1.82	-0.19	-0.11	0.02	0.25	-0.11	1.56	-0.25	-0.21	-0.20	-0.83	-0.13
	(0.07)	(0.11)*	(0.13)	(0.13)	(0.39)	(0.16)	(0.55)	(0.59)	(0.58)	(0.58)	(0.88)	(0.63)
Interquartile Range	955.15	311.80	184.30	-49.66	-482.40	20.86	0.92	-0.02	-0.07	-0.59	-0.39	-0.07
	(80.21)	(97.53)***	(122.27)	(109.02)	(214.66)**	(145.84)	(0.05)	(0.07)	(0.11)	(0.60)	(0.15)**	(0.11)
Theil Index	2.34	-0.62	-0.11	1.80	1.69	-0.07	0.36	-0.23	-0.23	-0.15	-0.32	-0.20
	(0.15)	(0.25)**	(0.28)	(0.15)***	(0.20)***	(0.30)	(0.13)	(0.16)	(0.16)	(0.14)	(0.23)	(0.19)
Gini Coefficient	0.75	-0.04	-0.01	0.95	2.28	-0.01	0.38	-0.10	-0.10	0.05	-0.13	-0.10
	(0.01)	(0.01)***	(0.02)	(0.14)***	(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	1250.16	145.12	122.17	122.17	-1034.21	141.68	1.74	0.06	-0.28	-0.24	-0.12	-0.47
	(67.54)	(104.43)	(127.89)	(148.51)	(441.02)**	(144.47)	(0.09)	(0.17)	(0.33)	(0.78)	(0.88)	(0.35)
Coefficient of Variation	1.82	0.32	0.26	0.28	0.26	0.34	1.70	0.28	-0.12	0.14	0.34	-0.33
	(0.21)	(0.23)	(0.24)	(0.25)	(0.42)	(0.26)	(0.31)	(0.43)	(0.57)	(0.51)	(0.53)	(0.63)
Interquartile Range	1874.91	58.71	20.31	340.92	-235.70	10.86	1.02	-0.08	-0.13	-0.30	-0.21	-0.13
	(126.16)	(139.52)	(171.68)	(209.96)	(328.28)	(185.48)	(0.06)	(0.09)	(0.11)	(0.32)	(0.25)	(0.18)
Theil Index	1.38	-0.09	-0.09	0.90	0.73	-0.02	0.51	0.12	-0.08	0.01	0.09	-0.20
	(0.12)	(0.16)	(0.19)	(0.13)***	(0.18)***	(0.20)	(0.10)	(0.13)	(0.22)	(0.22)	(0.23)	(0.25)
Gini Coefficient	0.68	0.00	0.00	0.51	1.08	0.01	0.47	0.06	-0.02	0.24	0.15	-0.06
	(0.02)	(0.02)	(0.02)	(0.09)***	(0.10)***	(0.03)	(0.03)	(0.04)	(0.07)	(0.09)***	(0.08)*	(0.08)**

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

Table A.1: 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
		1.167										
	Unfeasible		1.168	1.122	1.167	1.214	0.035	0.000	0.035	0.028	0.025	0.903
	Naive		1.126	1.083	1.126	1.171	0.034	-0.041	0.053	0.045	0.041	0.669
Mean Treatment Effects	Weighted		1.166	1.104	1.168	1.226	0.048	-0.001	0.048	0.038	0.034	0.900
	Location Shift		1.166	1.115	1.166	1.218	0.041	-0.002	0.041	0.033	0.028	0.902
	Log Location Shift		1.181	1.129	1.181	1.232	0.041	0.013	0.044	0.035	0.030	0.887
	CFM		1.167	1.114	1.169	1.222	0.043	0.000	0.043	0.034	0.029	0.903
		0.270										
	Unfeasible		0.267	0.211	0.263	0.325	0.050	-0.002	0.050	0.038	0.031	0.918
	Naive		0.332	0.277	0.327	0.388	0.049	0.062	0.079	0.064	0.057	0.696
CV	Weighted		0.272	0.199	0.270	0.350	0.067	0.002	0.067	0.048	0.036	0.918
	Location Shift		0.323	0.267	0.317	0.383	0.049	0.053	0.072	0.057	0.048	0.751
	Log Location Shift		0.325	0.271	0.320	0.381	0.047	0.056	0.073	0.058	0.051	0.728
	CFM		0.283	0.206	0.288	0.363	0.076	0.013	0.077	0.054	0.040	0.921
		0.654										
	Unfeasible		0.656	0.606	0.655	0.706	0.039	0.001	0.039	0.031	0.026	0.896
	Naive		0.688	0.640	0.688	0.741	0.040	0.034	0.052	0.042	0.036	0.782
Interquartile Range	Weighted		0.655	0.575	0.654	0.737	0.062	0.000	0.062	0.049	0.039	0.896
	Location Shift		0.756	0.703	0.755	0.811	0.042	0.102	0.110	0.102	0.101	0.200
	Log Location Shift		0.688	0.639	0.687	0.741	0.040	0.034	0.052	0.042	0.037	0.792
	CFM		0.670	0.603	0.669	0.737	0.052	0.015	0.054	0.043	0.037	0.887
		0.081										
	Unfeasible		0.081	0.066	0.080	0.096	0.012	0.000	0.012	0.010	0.008	0.905
	Naive		0.100	0.085	0.099	0.116	0.012	0.019	0.022	0.019	0.018	0.570
Theil Index	Weighted		0.082	0.060	0.082	0.103	0.018	0.000	0.018	0.013	0.011	0.929
	Location Shift		0.101	0.084	0.100	0.118	0.014	0.020	0.024	0.020	0.018	0.615
	Log Location Shift		0.098	0.084	0.097	0.114	0.012	0.017	0.021	0.018	0.016	0.619
	CFM		0.087	0.067	0.089	0.108	0.020	0.006	0.020	0.015	0.012	0.907
		0.086										
	Unfeasible		0.086	0.074	0.085	0.097	0.009	0.000	0.009	0.007	0.006	0.889
	Naive		0.118	0.108	0.118	0.130	0.009	0.033	0.034	0.033	0.032	0.015
Gini Coefficient	Weighted		0.088	0.063	0.089	0.111	0.019	0.002	0.019	0.015	0.013	0.913
	Location Shift		0.127	0.110	0.126	0.144	0.014	0.041	0.043	0.041	0.040	0.082
	Log Location Shift		0.111	0.100	0.110	0.123	0.009	0.025	0.027	0.025	0.025	0.123
	CFM		0.095	0.073	0.097	0.116	0.018	0.010	0.021	0.017	0.014	0.868

Table A.2: Results of Monte Carlo Exercise (Sample Size 4,000, Replications 1000, Uniform 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
		1.165										
	Unfeasible		1.164	1.119	1.164	1.211	0.035	0.000	0.035	0.028	0.024	0.911
	Naive		1.124	1.081	1.123	1.168	0.034	-0.041	0.053	0.045	0.043	0.655
Mean Treatment Effects	Weighted		1.164	1.108	1.165	1.219	0.044	0.000	0.044	0.035	0.030	0.885
	Location Shift		1.162	1.106	1.161	1.216	0.043	-0.003	0.043	0.035	0.029	0.886
	Log Location Shift		1.176	1.120	1.176	1.232	0.043	0.011	0.045	0.036	0.030	0.894
	CFM		1.168	1.107	1.170	1.229	0.050	0.004	0.050	0.040	0.032	0.900
		0.270										
	Unfeasible		0.266	0.209	0.260	0.324	0.051	-0.004	0.051	0.038	0.030	0.933
	Naive		0.328	0.274	0.322	0.382	0.049	0.058	0.076	0.060	0.053	0.737
CV	Weighted		0.253	0.180	0.254	0.326	0.063	-0.016	0.065	0.048	0.037	0.902
	Location Shift		0.320	0.265	0.314	0.375	0.050	0.050	0.071	0.054	0.045	0.788
	Log Location Shift		0.322	0.271	0.316	0.375	0.047	0.052	0.070	0.055	0.047	0.766
	CFM		0.253	0.150	0.267	0.345	0.097	-0.017	0.098	0.065	0.045	0.920
		0.653										
	Unfeasible		0.653	0.603	0.652	0.706	0.040	0.000	0.040	0.031	0.027	0.895
	Naive		0.685	0.634	0.684	0.737	0.040	0.032	0.051	0.041	0.035	0.787
Interquartile Range	Weighted		0.646	0.574	0.644	0.715	0.060	-0.006	0.060	0.046	0.038	0.898
	Location Shift		0.752	0.701	0.750	0.808	0.041	0.099	0.107	0.099	0.097	0.227
	Log Location Shift		0.683	0.633	0.682	0.740	0.041	0.031	0.051	0.041	0.035	0.816
	CFM		0.666	0.589	0.672	0.741	0.070	0.013	0.071	0.051	0.040	0.930
		0.081										
	Unfeasible		0.081	0.066	0.080	0.096	0.012	0.000	0.012	0.010	0.008	0.915
	Naive		0.099	0.084	0.098	0.114	0.012	0.018	0.021	0.018	0.017	0.579
Theil Index	Weighted		0.075	0.055	0.076	0.094	0.016	-0.006	0.017	0.013	0.010	0.886
	Location Shift		0.100	0.083	0.099	0.117	0.014	0.019	0.023	0.019	0.018	0.637
	Log Location Shift		0.097	0.082	0.096	0.112	0.012	0.016	0.020	0.167	0.015	0.643
	CFM		0.077	0.048	0.082	0.104	0.030	-0.004	0.030	0.019	0.013	0.941
		0.085										
	Unfeasible		0.085	0.074	0.085	0.097	0.009	0.000	0.009	0.007	0.006	0.895
	Naive		0.117	0.107	0.117	0.129	0.009	0.032	0.033	0.032	0.032	0.017
Gini Coefficient	Weighted		0.080	0.059	0.080	0.102	0.018	-0.005	0.019	0.015	0.012	0.886
	Location Shift		0.126	0.109	0.125	0.143	0.014	0.040	0.042	0.040	0.039	0.088
	Log Location Shift		0.110	0.098	0.109	0.121	0.009	0.024	0.026	0.024	0.024	0.156
	CFM		0.084	0.052	0.089	0.113	0.028	-0.001	0.028	0.020	0.015	0.927

Table A.3: Summary Statistics

	Original 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	0.51	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	0.03	0.03	0.04	0.01	0.00	0.21	0.19	0.18	0.02	0.03
Dummy	(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%.