Escola de ECONOMIA de São Paulo

Textos para Discussão

214

Outubro de 2009

IDENTIFICATION AND ESTIMATION OF DISTRIBUTIONAL IMPACTS OF INTERVENTIONS USING CHANGES IN INEQUALITY MEASURES

SERGIO FIRPO

FGV EESP

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

Sergio Firpo†

August 2008

Abstract

This paper presents semiparametric estimators of changes in inequality measures of a dependent variable distribution taking into account the possible changes on the distributions of covariates. When we do not impose parametric assumptions on the conditional distribution of the dependent variable given covariates, this problem becomes equivalent to estimation of distributional impacts of interventions (treatment) when selection to the program is based on observable characteristics. The distributional impacts of a treatment will be calculated as differences in inequality measures of the potential outcomes of receiving and not receiving the treatment. These differences are called here 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. Using the inverse probability weighting method to estimate parameters of the marginal distribution of potential outcomes, in the second step weighted sample versions of inequality measures are computed. Root-N consistency, asymptotic normality and semiparametric efficiency are shown for the semiparametric estimators proposed. A Monte Carlo exercise is performed to investigate the behavior in finite samples of the estimator derived in the paper. We also apply our method to the evaluation of a job training program.

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

*An earlier version of this paper circulated under the names "Treatment Effects on Some Inequality Measures" and "Inequality Treatment Effects". I have benefited from comments from David Green, Jim Heckman, Guido Imbens, Jeff Smith and participants at the 2004 Meeting of the Sociedade Brasileira de Econometria, 2005 World Congress of the Econometric Society, seminars at UBC, USC, and EPGE-FGV. Rafael Cayres and Breno Braga helped me with superb research assistance. Financial support from CNPq-Brazil is acknowledged. All errors are mine.

†Escola de Economia de São Paulo, Fundação Getúlio Vargas. Mailing address: R. Itapeva 474/1215, São Paulo - SP, Brazil, 01332-000. E-mail: sergio.firpo@fgv.br. Webpage: http://www.fgv.br/eesp/sergiofirpo

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 here *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. Some recent examples are the papers by Heckman (1992), Imbens and Rubin (1997), Heckman, Smith and Clements (1997), Heckman and Smith (1998), Abadie, Angrist and Imbens (2002), Carneiro, Hansen and Heckman (2001, 2003), Cunha, Heckman and Navarro (2005), Aakvik, Heckman and Vytlacil (2005), Dehejia (2005) and Firpo (2007). In the applied literature a recent paper that focuses on the distributional effects of a social program is Bitler, Gelbach and Hoynes (2006).

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. Important examples where this assumption has been used are Barnow, Cain and Goldberger (1980), Rosenbaum and Rubin (1983), Heckman, Ichimura, Smith and Todd (1998), Dehejia and Wahba (1999) and Hirano, Imbens and Ridder (2003). 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. Under unconfoundedness assumption we show that those estimators are consistent for the ITE parameters. For the class of estimators of inequality measures that are asymptotically linear we also show that the ITE estimators are asymptotically normal and semiparametrically efficient.

The key methodological contribution of this paper is to provide a detailed estimation procedure (with asymptotically valid inference) for treatment effects on inequality measures. We consider four popular inequality measures: the coefficient of variation, the interquartile range, the Theil index and the Gini coefficient. Our discussion is made on a very general level treating those measures as functionals of the distribution. In this sense, our approach can also be seen as a generalization of the discussion of identification and estimation of average and quantile treatment effects as both are differences in functionals of the distribution.

Recently, Tarozzi (2007) and Chen, Hong and Tarozzi (2008) have shown how to generalize treatment effects identification and estimation under unconfoundedness for a class of parameters that satisfy certain moment conditions. Their discussion encompasses a broader class of

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

problems, such as missing data and non-classical errors in variables. On inequality treatment effects see also the recent work of Thuysbaert (2007).

Under failure of unconfoundedness, the estimation method we develop here can be seen as a way to compare inequality measures taking into account the role of the covariates (observables). Applied researchers are very often interested in comparing features of two or more outcome distributions. For example, we might be interested in comparing the Gini coefficient 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 will 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. Of course, we could make parametric and functional form assumptions in order to relate covariates to the wage distribution in each economy. However, if we are not willing to impose restrictive parametric assumptions, it is not clear how to compare Gini coefficients fixing the distribution of observed covariates.

Note that even though in the example of a cross-country comparison of Gini coefficients we do not have a clear "treatment" involved, the problem of estimation of changes in distributions of a dependent variable when the distribution of covariates is fixed can be equivalently stated as the estimation problem of distributional impacts of social programs (treatments) when selection to the program is based on observable characteristics. Although we lose causal interpretation, our method is easily implementable and robust to misspecification of the conditional wage distributions.

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, in fact it serves as an interesting example where there is failure in the randomization to treatment and control groups. Finally, section 7 concludes. Proofs of results are left to the Appendix.

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.[2] 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.[3]

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):[4]

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

3 Identification of Inequality Treatment Effects

This section is divided up into four subsections. In the first one, we introduce some weighting functions and define the concept of weighted distributions. Subsection 2 presents the identification assumptions, while in subsection 3 we present the main identification results. Finally, subsection 4 brings some examples of inequality measures and shows how they fit into the framework just presented.

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

[3] 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.

[4] 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 Weighted Distributions

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 $(Y,\ T,\ X)$. Thus, 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 defining the *propensity-score* (or simply, *p-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 will be 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, define the following four "weighting functions", generally written as ω, such that $\omega : \{0,1\} \times \mathcal{P} \to \mathbb{R}$:

$$\begin{array}{cc} \omega_1(t,p(x)) = \frac{t}{p(x)} & \omega_0(t,p(x)) = \frac{1-t}{1-p(x)} \\ \omega_{11}(t,p(x)) = \frac{t}{p} & \omega_{01}(t,p(x)) = \left(\frac{1-t}{1-p(x)}\right)\cdot\left(\frac{p(x)}{p}\right) \end{array} \tag{1}$$

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 $(Y,T,X) \in \mathbb{R}\times\{0,1\}\times\mathcal{X}$. Assuming that $F_{Y,T,X}$ is absolutely continuous, the joint density function of (Y,T,X) is:

$$f_{Y,T,X}(y,t,x) = \left(p(x)\cdot dF_{Y|T,X}(y;1,x)\right)^t \cdot \left((1-p(x))\cdot dF_{Y|T,X}(y;0,x)\right)^{(1-t)}\cdot dF_X(x) \tag{2}$$

where $F_{Y|T,X}(y;t,x)$ is the cumulative distribution function (c.d.f.) of Y given $T=t$ and $X=x$ evaluated at y; and $F_X(x)$ is the c.d.f. of X at x. For simplicity, assume that those two distributions are absolutely continuous, which allows their respective densities to be well defined as $f_{Y|T,X}(y;t,x)$ and $f_X(x)$. After straight integration we obtain the c.d.f. of Y evaluated at a real number a:

$$F_Y(a) = \int_{y\in\mathbb{R}}^{a}\int_{x\in\mathcal{X}}\left(f_{Y,T,X}(y,1,x)+f_{Y,T,X}(y,0,x)\right)\cdot dx$$

$$= \int_{x\in\mathcal{X}}\left(p(x)\cdot\int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;1,x)\cdot dy + (1-p(x))\cdot\int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;0,x)\cdot dy\right)\cdot f_X(x)\cdot dx$$

Introduction of proper weights that depend on t and $p(x)$ only is straightforward. Assuming that the following integrals exist, we can write the weighted density and the weighted c.d.f. of Y (at a) as:

$$f_Y^\omega(y) = \int_{x\in\mathcal{X}}\{\omega(1,p(x))\cdot p(x)\cdot f_{Y|T,X}(y;1,x) + \omega(0,p(x))\cdot(1-p(x))\cdot f_{Y|T,X}(y;0,x)\}\cdot f_X(x)\cdot dx$$

$$F_Y^\omega(a) = \int_{x\in\mathcal{X}}\{\omega(1,p(x))\cdot p(x)\cdot\int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;1,x)\cdot dy + \omega(0,p(x))\cdot(1-p(x))\cdot\int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;0,x)\cdot dy\}\cdot f_X(x)\cdot dx$$

where $\omega(t, p(x)) = \omega(1, p(x))$ if $t = 1$ and $\omega(t, p(x)) = \omega(0, p(x))$ if $t = 0$. Simple algebra allows us to show that:

$$F_Y^{\omega}(a) = E[\omega(T, p(X)) \cdot \mathbb{1}\{Y \leq a\}] \tag{3}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function. Note that the definitions of weighted c.d.f and p.d.f. of Y encompass the case of the simple c.d.f. and p.d.f. of Y by making $\omega = 1$. It is worthwhile also noting that the c.d.f. may be seen as a particular case of a weighted expectation. In general, for an integrable function $\varsigma(Y)$ of the random variable Y, we write its weighted mean using weights ω as:

$$E[\omega(T, p(X)) \cdot \varsigma(Y)] = \int \varsigma(Y) \cdot f_Y^{\omega}(y) \cdot dy. \tag{4}$$

3.2 Identifying Assumptions

We now 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 and 1984), 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 there is evidence that 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).

Theorem 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 $a \in \mathbb{R}$, the c.d.f.s associated with those distributions can be written as follows:*

$$\begin{aligned} F_{Y(1)}(a) &= \int_{x \in \mathcal{X}} \int_{y \in \mathbb{R}}^{a} f_{Y|T,X}(y; 1, x) \cdot dy \cdot f_X(x) \cdot dx \\ &= F_Y^{\omega=\omega_1}(a) \\ &= E[\omega_1(T, p(X)) \cdot \mathbb{1}\{Y \leq a\}] \end{aligned}$$

$$
\begin{aligned}
F_{Y(0)}(a) &= \int_{x\in\mathcal{X}}\int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;0,x)\cdot dy\cdot f_X(x)\cdot dx \\
&= F_Y^{\omega=\omega_0}(a) \\
&= E\left[\omega_0\left(T,p\left(X\right)\right)\cdot 1\!\!1\{Y\le a\}\right]
\end{aligned}
$$

$$
\begin{aligned}
F_{Y(1)|T=1}(a) &= \int_{x\in\mathcal{X}}\frac{p(x)}{p}\cdot\int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;1,x)\cdot dy\cdot f_X(x)\cdot dx \\
&= F_Y^{\omega=\omega_{11}}(a) \\
&= E\left[\omega_{11}\left(T,p\left(X\right)\right)\cdot 1\!\!1\{Y\le a\}\right]
\end{aligned}
$$

$$
\begin{aligned}
F_{Y(0)|T=1}(a) &= \int_{x\in\mathcal{X}}\frac{p(x)}{p}\cdot\int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;0,x)\cdot dy\cdot f_X(x)\cdot dx \\
&= F_Y^{\omega=\omega_{01}}(a) \\
&= E\left[\omega_{01}\left(T,p\left(X\right)\right)\cdot 1\!\!1\{Y\le a\}\right]
\end{aligned}
$$

Corollary 1 *Under Assumptions 1 and 2 Δ^{ν}, Δ^{ν}_{T}, and Δ^{ν}_{C} are identifiable.*

Once we know that the inequality treatment effects are identifiable, we can now turn our attention to estimation and inference. Before doing so, let us be explicit about the 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).[5]

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

We proceed treating those four measures as functionals of a weighted outcome distribution. By doing that, we gain the flexibility necessary to further define the treatment effects as differences in functionals of weighted distributions:[6]

[5] 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.

[6] In what follows we assume that $\mu^{\omega}_{Y} \equiv \int y\cdot dF^{W}_{Y}(y;\omega)\neq 0$.

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

$$\nu^{CV}\left(F_Y^{\omega}\right) = \frac{\left(\int\left(y-\int z\cdot dF_Y^{\omega}\left(z\right)\right)^2\cdot dF_Y^{\omega}\left(y\right)\right)^{1/2}}{\int y\cdot dF_Y^{\omega}\left(y\right)} = \frac{\sqrt{E\left[\omega\left(T,p\left(X\right)\right)\cdot\left(Y-E\left[\omega\left(T,p\left(X\right)\right)\cdot Y\right]\right)^2\right]}}{E\left[\omega\left(T,p\left(X\right)\right)\cdot Y\right]}$$

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

$$\nu^{IQR}\left(F_Y^{\omega}\right) = \nu^{Q.75}\left(F_Y^{\omega}\right)-\nu^{Q.25}\left(F_Y^{\omega}\right) = \inf_q\left\{\int_{-\infty}^{q} f_Y^{\omega}\left(y\right)\cdot dy\geq\frac{3}{4}\right\}-\inf_q\left\{\int_{-\infty}^{q} f_Y^{\omega}\left(y\right)\cdot dy\geq\frac{1}{4}\right\}$$

3. **Theil Index** (TI):[7]

$$\nu^{TI}\left(F_Y^{\omega}\right) = \frac{\int y\cdot\left(\log\left(y\right)-\log\left(\int z\cdot dF_Y^{\omega}\left(z\right)\right)\right)\cdot dF_Y^{\omega}\left(y\right)}{\int y\cdot dF_Y^{\omega}\left(y\right)} = E\left[\omega\left(T,p\left(X\right)\right)\cdot\left(\frac{Y}{E\left[\omega\left(T,p\left(X\right)\right)\cdot Y\right]}\right)\cdot\log\left(\frac{Y}{E\left[\omega\left(T,p\left(X\right)\right)\cdot Y\right]}\right)\right]$$

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

$$\nu^{GC}\left(F_Y^{\omega}\right)=1-2\frac{R\left(F_Y^{\omega}\right)}{\int y\cdot dF_Y^{\omega}\left(y;\omega\right)}=1-2\frac{R\left(F_Y^{\omega}\right)}{E\left[\omega\left(T,p\left(X\right)\right)\cdot Y\right]}$$

where

$$R\left(F_Y^{\omega}\right) = \int_0^1\int_{-\infty}^{\nu^{Q\tau}\left(F_Y^{\omega}\right)} y\cdot dF_Y^{\omega}\left(y;\omega\right)\cdot d\tau = \int_0^1 E\left[\omega\left(T,p\left(X\right)\right)\cdot\mathbb{1}\{Y\leq\inf_q\left\{\int_{-\infty}^{q} f_Y^{\omega}\left(z\right)\cdot dz\geq\tau\right\}\}\cdot Y\right]\cdot d\tau$$

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}.

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

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:

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

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

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

and it is clear that we have to consider carefully the estimation of weights $\omega\left(t, p\left(x\right)\right)$ by $\widehat{\omega}\left(t, \widehat{p}\left(x\right)\right)$.

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 $p\left(x\right)$, 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, $L\left(p\left(x\right)\right)$ by a series of polynomial functions of x.[8] 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 $\hat{\pi}_K$:

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

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 $H_K\left(\cdot\right)$ 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.[9]

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

[9] 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).

We propose an estimator $\widehat{\omega}$ (normalized to sum up exactly one) for the weighting function ω:

$$\widehat{\omega}_i = \frac{1}{N_\omega} \cdot \omega\left(T_i, \widehat{p}\left(X_i\right)\right)$$

where N_ω, which is chosen to converge in probability to one, equals

$$N_\omega = \frac{1}{N} \cdot \sum_{i=1}^{N} \omega\left(T_i, \widehat{p}\left(X_i\right)\right)$$

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

$$\begin{aligned}
\widehat{\omega}_{1,i} &= \frac{1}{N_1} \cdot \omega_1\left(T_i, \widehat{p}\left(X_i\right)\right) \\
&= \frac{1}{N_1} \cdot \frac{T_i}{\widehat{p}\left(X_i\right)} \\
&= \left(\frac{1}{N} \cdot \sum_{j=1}^{N} \frac{T_j}{\widehat{p}\left(X_j\right)}\right)^{-1} \cdot \frac{T_i}{\widehat{p}\left(X_i\right)}
\end{aligned}$$

$$\begin{aligned}
\widehat{\omega}_{0,i} &= \frac{1}{N_0} \cdot \omega_0\left(T_i, \widehat{p}\left(X_i\right)\right) \\
&= \frac{1}{N_0} \cdot \left(\frac{1-T_i}{1-\widehat{p}\left(X_i\right)}\right) \\
&= \left(\sum_{j=1}^{N} \frac{1-T_j}{1-\widehat{p}\left(X_j\right)}\right)^{-1} \cdot \left(\frac{1-T_i}{1-\widehat{p}\left(X_i\right)}\right)
\end{aligned}$$

$$\begin{aligned}
\widehat{\omega}_{11,i} &= \frac{1}{N_{11}} \cdot \omega_{11}\left(T_i, \widehat{p}\left(X_i\right)\right) \\
&= \frac{1}{N_{11}} \cdot \frac{T_i}{p} \\
&= \left(\frac{1}{N} \cdot \sum_{j=1}^{N} T_j\right)^{-1} \cdot T_i
\end{aligned}$$

$$\begin{aligned}
\widehat{\omega}_{01,i} &= \frac{1}{N_{01}} \cdot \omega_{01}\left(T_i, \widehat{p}\left(X_i\right)\right) \\
&= \frac{1}{N_{01}} \cdot \frac{\widehat{p}\left(X_i\right)}{p} \cdot \left(\frac{1-T_i}{1-\widehat{p}\left(X_i\right)}\right) \\
&= \left(\frac{1}{N} \cdot \sum_{j=1}^{N} \widehat{p}\left(X_j\right) \cdot \frac{1-T_j}{1-\widehat{p}\left(X_j\right)}\right)^{-1} \cdot \widehat{p}\left(X_i\right) \cdot \left(\frac{1-T_i}{1-\widehat{p}\left(X_i\right)}\right)
\end{aligned}$$

where

$$N_1 = \tfrac{1}{N} \cdot \textstyle\sum_{i=1}^{N} \frac{T_i}{\widehat{p}(X_i)} \qquad N_0 = \tfrac{1}{N} \cdot \textstyle\sum_{i=1}^{N} \frac{1-T_i}{1-\widehat{p}(X_i)}$$

$$N_{11} = \tfrac{1}{N} \cdot \textstyle\sum_{i=1}^{N} \frac{T_i}{p} \qquad N_{01} = \tfrac{1}{N} \cdot \textstyle\sum_{i=1}^{N} \frac{\widehat{p}(X_i)}{p} \cdot \left(\frac{1-T_i}{1-\widehat{p}(X_i)}\right)$$

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. Define the corresponding estimators of Δ^{ν}, Δ^{ν}_{T}, and Δ^{ν}_{C} as:

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

As an illustration, we consider in details the estimation of three inequality treatment effect parameters: (i) the coefficient of variation CIT, (ii) the Theil index ITE and (iii) the Gini coefficient ITT:

Example 1: The estimator for the coefficient of variation CIT:

$$\begin{aligned}
\widehat{\Delta}^{CV}_{C} &= \frac{\left(\sum_{i=1}^{N}\frac{1}{N}\cdot\left(Y_i - \sum_{i=1}^{N}\frac{1}{N}\cdot Y_i\right)^2\right)^{1/2}}{\sum_{i=1}^{N}\frac{1}{N}\cdot Y_i}\\
&\quad - \frac{\left(\sum_{i=1}^{N}\widehat{\omega}_{0,i}\cdot\left(Y_i - \sum_{i=1}^{N}\widehat{\omega}_{0,i}\cdot Y_i\right)^2\right)^{1/2}}{\sum_{i=1}^{N}\widehat{\omega}_{0,i}\cdot Y_i}
\end{aligned}$$

Example 2: The estimator for the Theil index ITE:

$$\begin{aligned}
\widehat{\Delta}^{TI} &= \frac{\sum_{i=1}^{N}\widehat{\omega}_{1,i}\cdot Y_i\cdot\left(\log(Y_i) - \log\left(\sum_{i=1}^{N}\widehat{\omega}_{1,i}\cdot Y_i\right)\right)}{\sum_{i=1}^{N}\widehat{\omega}_{1,i}\cdot Y_i}\\
&\quad - \frac{\sum_{i=1}^{N}\widehat{\omega}_{0,i}\cdot Y_i\cdot\left(\log(Y_i) - \log\left(\sum_{i=1}^{N}\widehat{\omega}_{0,i}\cdot Y_i\right)\right)}{\sum_{i=1}^{N}\widehat{\omega}_{0,i}\cdot Y_i}
\end{aligned}$$

Example 3: The estimator for the Gini coefficient ITT:

$$\begin{aligned}
\widehat{\Delta}^{GC}_{T} &= 2\int_0^1 \{\frac{\sum_{i=1}^{N}\widehat{\omega}_{01,i}\cdot \mathbb{1}\{Y_i \le \widehat{q}_{\tau}^{01}\}\cdot Y_i}{\sum_{i=1}^{N}\widehat{\omega}_{01,i}\cdot Y_i}\\
&\quad - \frac{\sum_{i=1}^{N}\widehat{\omega}_{11,i}\cdot \mathbb{1}\{Y_i \le \widehat{q}_{\tau}^{11}\}\cdot Y_i}{\sum_{i=1}^{N}\widehat{\omega}_{11,i}\cdot Y_i}\}\cdot d\tau
\end{aligned}$$

where

$$\begin{aligned}
\widehat{q}_{\tau}^{01} &= \nu^{Q\tau}\left(\widehat{F}_Y^{\omega=\widehat{\omega}_{01}}\right) = \arg\min_q \sum_{i=1}^{N}\widehat{\omega}_{01,i}\cdot\rho_{\tau}(Y_i - q)\\
\widehat{q}_{\tau}^{11} &= \nu^{Q\tau}\left(\widehat{F}_Y^{\omega=\widehat{\omega}_{11}}\right) = \arg\min_q \sum_{i=1}^{N}\widehat{\omega}_{11,i}\cdot\rho_{\tau}(Y_i - q)
\end{aligned}$$

and, following Koenker and Bassett (1978), $\rho_{\tau}(u) = u\cdot(\tau - \mathbb{1}\{u \le 0\})$ is the check function evaluated at a real number u. The integral over τ can be computed by numerical integration.

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 asymptotically linear. We then invoke known results established by Hahn (1998), Hirano, Imbens and Ridder (2003) and Chen, Hong and Tarozzi (2008) to consider efficiency of our estimators.[10]

4.2.1 Asymptotic Linearity

We will need to establish some extra conditions to be able to derive the asymptotic normality of the inequality estimators just proposed. We restrict the discussion to the class of asymptotically linear estimators.

First, we recall the usual definition of the influence function ψ^{ν} of a general functional ν of the distribution (Hampel, 1974). Letting that functional ν be an inequality measure ν, its influence function ψ^{ν} is the directional derivative of ν evaluated in F in the direction of δ_y, the degenerate distribution that put mass only at a single point y:

$$\psi^{\nu}(y; F) = \lim_{s \to 0} \frac{\nu(s \cdot (F - \delta_y) + F) - \nu(F)}{s} \tag{5}$$

For an alternative distribution G, close enough to F, a von Mises type expansion of the functional ν is allowable (Hampel, Ronchetti, Rousseeuw, and Stahel, 1986):

$$\nu(G) = \nu(F) + \int \psi^{\nu}(y; F) \cdot dG(y) + r(F, G) \tag{6}$$

where $r(F, G)$ is a remainder term that depends on the distributions F and G.
Definition 1 [***Asymptotic Linearity***]: Using the general weighted distribution framework, we define a functional estimate as being asymptotically linear if, after expanding the functional at the empirical distribution $\widehat{F}_Y^{\omega=\widehat{\omega}}$ around the population distribution F_Y^{ω} we get a remainder term that converges at the parametric rate $N^{-1/2}$.

Another way to say that the estimator $\widehat{\nu}_Y^{\widehat{\omega}} = \nu\left(\widehat{F}_Y^{\omega=\widehat{\omega}}\right)$ is asymptotically linear is by using directly Equation (6), replacing G by $\widehat{F}_Y^{\omega=\widehat{\omega}}$ and F by $\widehat{F}_Y^{\omega}$. The estimator $\widehat{\nu}_Y^{\widehat{\omega}}$ will be asymptotically linear if the following assumption holds.

Assumption 3 [***Asymptotic Linearity of Inequality Estimators***]*:*

$$\nu\left(\widehat{F}_Y^{\omega=\widehat{\omega}}\right) - \nu\left(F_Y^{\omega}\right) = \sum_{i=1}^{N} \widehat{\omega}_i \cdot \psi^{\nu}\left(Y_i; F_Y^{\omega}\right) + o_p(1/\sqrt{N})$$

This assumption holds for all four inequality measures considered in this article. Treating the weights as fixed, we can verify that three out of four of our estimators are non-linear functions of exactly linear functionals (expectations) evaluated at the empirical distribution; therefore, a simple application of the so-called delta method suffices to show asymptotic linearity. The only one that does not fall into that category, the interquartile range, is however just a difference

[10] 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.

in quantiles; a proof of asymptotic linearity of the sample quantile is found in van der Vaart (1998), for example.

Proofs that the four inequality measures considered here are asymptotically linear are not provided since they are simple applications of the delta-method. Nevertheless, we list the actual format of the influence function of each one of the four inequality measures, ν^{CV}, ν^{IQR}, ν^{TI}, and ν^{GC}:[11]

Coefficient of Variation:

We write $\nu^{CV}\left(F_Y^{\omega}\right) = g^{CV}\left(\nu^{V}\left(F_Y^{\omega}\right), \mu_Y^{\omega}\right)$, where $\nu^{V}\left(F_Y^{\omega}\right)$ is the variance and the mean is $\mu_Y^{\omega} = E[\omega\left(T, p\left(X\right)\right) \cdot Y]$. Let $g_1^{CV}\left(\cdot,\cdot\right)$ and $g_2^{CV}\left(\cdot,\cdot\right)$ be the partial derivatives with respect to the first and the second arguments respectively. Then, $g^{CV}\left(a,b\right) = a^{1/2} \cdot b^{-1}$, $g_1^{CV}\left(a,b\right) = \frac{1}{2} \cdot a^{-1/2} \cdot b^{-1}$, $g_2^{CV}\left(a,b\right) = -a^{1/2} \cdot b^{-2}$. Let ψ^{V} be the influence of the function of the variance and ψ^{μ} be the influence of the function of the mean (see, for example, Lehmann, 1999, p. 309):

$$\psi^{\mu}\left(y; F_Y^{\omega}\right) = y - \mu_Y^{\omega}$$

$$\psi^{V}\left(y; F_Y^{\omega}\right) = \left(y - \mu_Y^{\omega}\right)^2 - \nu^{V}\left(F_Y^{\omega}\right)$$

By a simple application of the delta-method, we obtain ψ^{CV}, the influence of the function of the coefficient of variation:

$$\begin{aligned}
& \psi^{CV}\left(y; F_Y^{\omega}\right) \\
= \; & g_1^{CV}\left(\nu^{V}\left(F_Y^{\omega}\right), \mu_Y^{\omega}\right) \cdot \psi^{V}\left(y; F_Y^{\omega}\right) + g_2^{CV}\left(\nu^{V}\left(F_Y^{\omega}\right), \mu_Y^{\omega}\right) \cdot \psi^{\mu}\left(y; F_Y^{\omega}\right) \\
= \; & \frac{1}{2} \cdot \frac{\left(y - E\left[\omega\left(T, p\left(X\right)\right) \cdot Y\right]\right)^2 - \left(E\left[\omega\left(T, p\left(X\right)\right) \cdot Y^2\right] - \left(E\left[\omega\left(T, p\left(X\right)\right) \cdot Y\right]\right)^2\right)}{E\left[\omega\left(T, p\left(X\right)\right) \cdot Y\right] \cdot \sqrt{E\left[\omega\left(T, p\left(X\right)\right) \cdot Y^2\right] - \left(E\left[\omega\left(T, p\left(X\right)\right) \cdot Y\right]\right)^2}} \\
& - \frac{\sqrt{E\left[\omega\left(T, p\left(X\right)\right) \cdot Y^2\right] - \left(E\left[\omega\left(T, p\left(X\right)\right) \cdot Y\right]\right)^2}}{\left(E\left[\omega\left(T, p\left(X\right)\right) \cdot Y\right]\right)^2} \cdot \left(y - E\left[\omega\left(T, p\left(X\right)\right) \cdot Y\right]\right)
\end{aligned}$$

Interquartile Range:

Following, for example, Ferguson (1996, p.91) and van der Vaart, (1998, p.307), we define the influence function of the τ-th quantile of the weighted distribution of Y as:[12]

$$\psi^{Q\tau}\left(y; F_Y^{\omega}\right) = \left(\tau - 1\!\!1\{y \leq q_\tau^{\omega}\}\right) / f_Y^{\omega}(q_\tau^{\omega})$$

where $f_Y^{\omega}(q_\tau^{\omega})$ is the density evaluated at the quantile q_τ^{ω}. The influence function of the interquartile range is simply the difference between two quantile influence functions:

$$\begin{aligned}
\psi^{IQR}\left(y; F_Y^{\omega}\right) &= \psi^{Q.75}\left(y; F_Y^{\omega}\right) - \psi^{Q.25}\left(y; F_Y^{\omega}\right) \\
&= \frac{\left(.75 - 1\!\!1\{y \leq \nu^{IQR} + q_{.25}^{\omega}\}\right)}{f_Y^{\omega}(\nu^{IQR} + q_{.25}^{\omega})} \\
&\quad - \frac{\left(.25 - 1\!\!1\{y \leq q_{.25}^{\omega}\}\right)}{f_Y^{\omega}(q_{.25}^{\omega})}.
\end{aligned}$$

[11] We are assuming in what follows that all relevant integrals exist and denominators are non-zero.

[12] For notational simplicity, we denote $\nu^{Q\tau}\left(y; F_Y^{W}\left(\omega\right)\right)$ by q_τ^{W}. The number τ is a real in the open interval $(0,1)$.

For $\tau \in \{1/4, 3/4\}$ we can write the quantile as the minimizer of the expected check function (Koenker and Basset, 1978):

$$q_\tau^\omega = \nu^{Q\tau}(F_Y^\omega) = \arg\min_q E\left[\omega(T, p(X)) \cdot (Y - q) \cdot (\tau - \mathbb{I}\{Y \leq q\})\right]$$

Theil index:

We write $\nu^{TI}(F_Y^\omega) = g^{TI}\left(\nu^{\mu L}(F_Y^\omega), \mu_Y^\omega\right)$, where $\nu^{\mu L}(F_Y^\omega) = E\left[\omega(T, p(X)) \cdot Y \cdot \log(Y)\right]$. Let $g_1^{TI}(\cdot,\cdot)$ and $g_2^{TI}(\cdot,\cdot)$ be the partial derivatives with respect to the first and the second arguments respectively. Then, $g^{TI}(a,b) = a \cdot b^{-1} - \log(b)$, $g_1^{TI}(a,b) = b^{-1}$, $g_2^{TI}(a,b) = -b^{-1} \cdot (1 + a/b)$. Let $\psi^{\mu L}$ be the influence of the function of the $\nu^{\mu L}$:

$$\psi^{\mu L}(y; F_Y^\omega) = y \cdot \log(y) - \nu^{\mu L}(F_Y^\omega)$$

then by another application of the delta-method, we obtain ψ^{TI}:

$$\begin{aligned}
\psi^{TI}(y; F_Y^\omega) &= g_1^{TI}\left(\nu^{\mu L}(F_Y^\omega), \mu_Y^\omega\right) \cdot \psi^{\mu L}(y; F_Y^\omega) + g_2^{TI}\left(\nu^{\mu L}(F_Y^\omega), \mu_Y^\omega\right) \cdot \psi^{\mu}(y; F_Y^\omega) \\
&= \left(E[\omega(T, p(X)) \cdot Y]\right)^{-1} \cdot \left(y \cdot \log(y) - E\left[\omega(T, p(X)) \cdot Y \cdot \log(Y)\right]\right) \\
&\quad - \left(E[\omega(T, p(X)) \cdot Y]\right)^{-2} \cdot \left(E\left[\omega(T, p(X)) \cdot Y \cdot \log(Y)\right] + E\left[\omega(T, p(X)) \cdot Y\right]\right) \\
&\quad \cdot \left(y - E[\omega(T, p(X)) \cdot Y]\right)
\end{aligned}$$

Gini Coefficient:

The influence function of the Gini coefficient was derived by Hoeffding (1948) as an example of results on the asymptotic distribution of U-statistics. We follow a more recent literature on the statistical properties of inequality measures (see the survey paper by Cowell, 2000). For example, following Schluter and Trede (2003), we write the influence function of the Gini coefficient as:

$$\begin{aligned}
\psi^{GC}(y; F_Y^\omega) &= \frac{1}{\mu_Y^\omega} \cdot \Big\{ \left(\nu^{GC}(F_Y^\omega) - 1\right) \cdot (\mu_Y^\omega + y) \\
&\quad -2 \cdot y \cdot \left(1 - \int_{-\infty}^{y} dF_Y^\omega(y;\omega)\right) - 2 \cdot \int_{-\infty}^{\nu^{Q\tau}(F_Y^\omega)} y \cdot dF_Y^\omega(y;\omega) \Big\} \\
&= \frac{1}{E\left[\omega(T, p(X)) \cdot Y\right]} \Big\{ \left(\nu^{GC}(F_Y^\omega) - 1\right) \cdot \left(E\left[\omega(T, p(X)) \cdot Y\right] + y\right) \\
&\quad -2 \cdot y \cdot \left(1 - E\left[\omega(T, p(X)) \cdot \mathbb{I}\{Y \leq y\}\right]\right) \\
&\quad -2 \cdot E\left[\omega(T, p(X)) \cdot \mathbb{I}\{Y \leq \inf_q \left\{\int_{-\infty}^{q} f_Y^\omega(z) \cdot dz \geq \tau\right\}\} \cdot Y\right] \Big\}
\end{aligned}$$

where

$$\nu^{GC}(F_Y^\omega) = 1 - 2\frac{\int_0^1 E\left[\omega(T, p(X)) \cdot \mathbb{I}\{Y \leq \inf_q \left\{\int_{-\infty}^{q} f_Y^\omega(z) \cdot dz \geq \tau\right\}\} \cdot Y\right] \cdot d\tau}{E\left[\omega(T, p(X)) \cdot Y\right]}$$

4.2.2 Asymptotic Normality and Efficiency

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

The additional regularity conditions that we need to impose are directly on the influence functions of the inequality measures:

Assumption 4 [***Influence Function***] *For all weighting functions* ω *considered,*
(i) $\int \left(\psi^{\nu}\left(y;F_{Y}^{\omega}\right)\right)^{2}\cdot f_{Y}^{\omega}\left(y\right)\cdot dy<\infty$*, and*
(ii) $\int \psi^{\nu}\left(y;F_{Y}^{\omega}\right)\cdot f_{Y|X}^{\omega}\left(y;x\right)\cdot dy$ *is continuously differentiable for all* x *in* $\mathcal{X}$*,*
where

$$f_{Y|X}^{\omega}\left(y;x\right)=\omega\left(1,p\left(x\right)\right)\cdot p\left(x\right)\cdot f_{Y|T,X}\left(y;1,x\right)+\omega\left(0,p\left(x\right)\right)\cdot\left(1-p\left(x\right)\right)\cdot f_{Y|T,X}\left(y;0,x\right) \quad (7)$$

Part (i) of Assumption 4 is a condition of finite variance of the influence function. It is definitely important because it allows for a central limit theorem to be used. Part (ii) 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 expectation of the influence function be continuously differentiable.

We are now able to state the central result. We first consider a proposition that establishes asymptotic normality for the estimators of inequality measures using the empirical distribution $\widehat{F}_{Y}^{\omega=\widehat{\omega}}$. The assumptions required for the 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. Asymptotic properties of our estimators of the inequality treatment effects will follow after that as a direct consequence.

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 1-4 and A.1:*

$$\sqrt{N}\cdot\left(\nu\left(\widehat{F}_{Y}^{\omega=\widehat{\omega}}\right)-\nu\left(F_{Y}^{\omega}\right)\right)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega\left(T_{i},p\left(X_{i}\right)\right)\cdot\psi^{\nu}\left(Y_{i};F_{Y}^{\omega}\right)$$
$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}E\left[h\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega}\right)|X=X_{i}\right]\cdot\left(T_{i}-p\left(X_{i}\right)\right)+o_{p}(1)\xrightarrow{D}N\left(0,V_{\nu}\right)$$

where

$$V_{\nu}=E\left[\left(\omega\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega}\right)+E\left[h\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega}\right)|X\right]\cdot\left(T-p\left(X\right)\right)\right)^{2}\right]$$

Proposition 1 follows immediately after the results by Newey (1994) and, particularly, those by Hirano, Imbens and Ridder (2003) and Chen, Hong and Tarozzi (2008). Proposition 1 is a general result that specializes for the four weighting functions considered here. It can be easily seen that the derivatives of the four weighting functions with respect to $p\left(x\right)$ are

$$h_{1}\left(t,p\left(x\right)\right)\equiv\frac{\partial\omega_{1}(t,p(x))}{\partial p(x)}=-\frac{t}{p^{2}(x)} \qquad h_{0}\left(t,p\left(x\right)\right)\equiv\frac{\partial\omega_{0}(t,p(x))}{\partial p(x)}=\frac{1-t}{(1-p(x))^{2}}$$
$$h_{11}\left(t,p\left(x\right)\right)\equiv\frac{\partial\omega_{11}(t,p(x))}{\partial p(x)}=0 \qquad h_{01}\left(t,p\left(x\right)\right)\equiv\frac{\partial\omega_{01}(t,p(x))}{\partial p(x)}=\frac{1-t}{p\cdot(1-p(x))^{2}}$$

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:*

$$\sqrt{N}\cdot\left(\widehat{\Delta}^{\nu}-\Delta^{\nu}\right)$$
$$=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_{1}\left(T_{i},p\left(X_{i}\right)\right)\cdot\psi^{\nu}\left(Y_{i};F_{Y}^{\omega=\omega_{1}}\right)-\omega_{0}\left(T_{i},p\left(X_{i}\right)\right)\cdot\psi^{\nu}\left(Y_{i};F_{Y}^{\omega=\omega_{0}}\right)$$
$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(E\left[h_{1}\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_{1}}\right)|X=X_{i}\right]\right.$$
$$\left.-E\left[h_{0}\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_{0}}\right)|X=X_{i}\right]\right)\cdot\left(T_{i}-p\left(X_{i}\right)\right)+o_{p}(1)$$
$$\xrightarrow{D}N\left(0,V\right)$$

$$\sqrt{N}\cdot\left(\widehat{\Delta}_{T}^{\nu}-\Delta_{T}^{\nu}\right)$$
$$=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_{11}\left(T_{i},p\left(X_{i}\right)\right)\cdot\psi^{\nu}\left(Y_{i};F_{Y}^{\omega=\omega_{11}}\right)-\omega_{01}\left(T_{i},p\left(X_{i}\right)\right)\cdot\psi^{\nu}\left(Y_{i};F_{Y}^{\omega=\omega_{01}}\right)$$
$$-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}E\left[h_{01}\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_{01}}\right)|X=X_{i}\right]\cdot\left(T_{i}-p\left(X_{i}\right)\right)+o_{p}(1)$$
$$\xrightarrow{D}N\left(0,V_{T}\right)$$

$$\sqrt{N}\cdot\left(\widehat{\Delta}_{C}^{\nu}-\Delta_{C}^{\nu}\right)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi^{\nu}\left(Y_{i};F_{Y}\right)-\omega_{0}\left(T_{i},p\left(X_{i}\right)\right)\cdot\psi^{\nu}\left(Y_{i};F_{Y}^{\omega=\omega_{0}}\right)$$
$$-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}E\left[h_{0}\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_{0}}\right)|X=X_{i}\right]\cdot\left(T_{i}-p\left(X_{i}\right)\right)+o_{p}(1)$$
$$\xrightarrow{D}N\left(0,V_{C}\right)$$

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}

$$V=E[(\omega_{1}\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_{1}}\right)-\omega_{0}\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_{0}}\right)$$
$$+\left(E\left[h_{1}\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_{1}}\right)|X\right]\right.$$
$$\left.-E\left[h_{0}\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_{0}}\right)|X\right]\right)\cdot\left(T-p\left(X\right)\right)\big)^{2}]$$

$$\begin{aligned}V_{T} &= E[(\omega_{11}\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_{11}}\right)-\omega_{01}\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_{01}}\right)\\ &\quad -E\left[h_{01}\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_{01}}\right)|X\right]\cdot\left(T-p\left(X\right)\right)\big)^{2}]\end{aligned}$$

$$\begin{aligned}V_{C} &= E[(\psi^{\nu}\left(Y;F_{Y}\right)-\omega_{0}\left(T,X\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_{0}}\right)\\ &\quad -E\left[h_{0}\left(T,p\left(X\right)\right)\cdot\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_{0}}\right)|X\right]\cdot\left(T-p\left(X\right)\right)\big)^{2}]\end{aligned}$$

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. Given the asymptotic normality of the estimators, the bootstrap may be a valid and is surely an easier alternative for calculation of 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) and estimators of their asymptotic variances behave in finite samples. One thousand replications of the experiment with sample sizes of 500 and 2,500 observations were considered.

We design the data generation process 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 parameters of the marginal distributions of $Y(1)$ and $Y(0)$.

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

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

where η has a standard logistic c.d.f. $F_\eta(n) = \left(1 + \exp\left(-\pi n/\sqrt{3}\right)\right)^{-1}$. The potential outcomes are

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

where

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

and where κ_0 and κ_1 are distributed as standard normals. The variables X, η, κ_0 and κ_1 are mutually independent. Under this specification, $Y(1)$ and $Y(0)$ will not have a closed form distribution. We present in the Appendix a characterization of their conditional densities given $X_1 = x_1$ and $X_2 = x_2$ as functions of the parameters. The unconditional p.d.f.'s are obtained through numerical integration.

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

$coeff.\backslash j$	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^s_{0j}	0.01	-0.01	0.01	0.01	-0.01	-0.02
β^s_{1j}	0.01	0.01	0.01	0.01	0.01	0.01

The values of some functionals of the distributions of are listed below:

Table 2: Features of the Distributions of Potential Outcomes

$\nu\backslash$Distribution	$Y(0)$	$Y(1)$
Mean	0.79	1.83
Standard Deviation (s.d.)	0.26	1.08
Mean of Logarithm	-0.29	0.49
S.D. of Logarithm	0.35	0.45
10th Percentile	0.47	0.99
1st Quartile	0.63	1.22
Median	0.80	1.55
3rd Quartile	0.94	2.09
90th Percentile	1.07	2.92

Table 3: Inequality Measures of Potential Outcomes

$\nu\backslash$Distribution	$Y(0)$	$Y(1)$
Coefficient of Variation	0.33	0.59
Interquartile Range	0.31	0.87
Theil Index	0.05	0.13
Gini Coefficient	0.18	0.26

We provide in Tables 4-7 results for the unfeasible estimator and also for three types of 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 labelled "feasible estimator". The second is the one based on the empirical distributions of $Y|T = 1$ and $Y|T = 0$. We call this estimator the "naive estimator". Given that there is selection into treatment based on observables, the naive estimator will be inconsistent to the ITE parameters. Finally we consider what we call the "counterfactual distribution (CD) estimator". This is constructed in the following way. We first run two linear regressions (with intercept) of Y on X_1 and X_2 for $T = 1$ and of Y on X_1 and X_2 for $T = 0$. Save residuals $\widehat{u}_j$, coefficient estimates $\widehat{\gamma}_{0j}$, $\widehat{\gamma}_{1j}$, $\widehat{\gamma}_{2j}$, and s_j^2 where $j = 0, 1$ indexes treatment groups, $s_j^2 = (N_j - 3)^{-1} \sum_i^{N_j} (\widehat{u}_{ji})^2$ and $N_1 = \sum_i^N T_i$ and $N_0 = N - N_1$. Then, let Y^*_{ij} be counterfactual of observation i:

$$Y^*_{i1} = \begin{cases} Y_i & \text{if } T_i = 1 \\ \widehat{\gamma}_{01} + \widehat{\gamma}_{11} \cdot X_{1i} + \widehat{\gamma}_{21} \cdot X_{2i} + \sqrt{s_1^2/s_0^2} \cdot \widehat{u}_{0i} & \text{if } T_i = 0 \end{cases}$$

$$Y^*_{i0} = \begin{cases} Y_i & \text{if } T_i = 0 \\ \widehat{\gamma}_{00} + \widehat{\gamma}_{10} \cdot X_{1i} + \widehat{\gamma}_{20} \cdot X_{2i} + \sqrt{s_0^2/s_1^2} \cdot \widehat{u}_{1i} & \text{if } T_i = 1 \end{cases}$$

and since Y_{i1}^* and Y_{i0}^* are well defined for all i, we compute the inequality measures for these two distributions. Note that this is a way of "controlling" for covariates.

Results in Tables 4 and 5 show distribution features for each one of the estimators of inequality treatment effect parameters. We report average, standard deviation and quantiles (10^{th} percentile, lower quartile, median, upper quartile and 90^{th} percentile) for the four types of treatment effects on inequality measures here considered (coefficient of variation, interquartile range, Theil index and Gini coefficient). 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.[13] Tables 6 and 7 show some features of the distribution of bootstrapped standard errors for 1,000 Monte Carlo replications using within each replication a bootstrapping procedure with 100 repetitions.

The results indicate that the "feasible estimator" performs well according to the MSE criteria and that its variance shrinks as expected as the sample size increases. Relatively to other estimators being analyzed, the "feasible estimator" has better properties than the naive and the CD estimators, in terms of bias, variance, mean squared error, absolute error and coverage rate.

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[14]. 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 relativley 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 social programs. In the years of 1998-99, the Brazilian Ministry of Labor financed an experimental evaluation of the program impact on earnings and employment.[15] 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 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

[13] We present coverage rates of two types of 90% confidence intervals. In the first column, each Monte Carlo replication used a bootstrap (100 repetitions) variance estimate of that replication. In the second column we used for all replications the same variance estimate, which was the one given by the sampling variation between replications.

[14] 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.

[15] This data set has been analyzed by Foguel (2007) and in Firpo and Foguel (2008), where more details on the impact evaluation study can be found.

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 was 4,603 individuals, out of which 2,091 were from Rio de Janeiro. We selected individuals who obtained a job between the end of the treatment period and the follow-up survey. We also restricted our sample for individuals between 16 and 50 years old. We ended up with 548 treated individuals and 408 controls in Rio de Janeiro and 756 treated and 767 controls in Fortaleza. Table 8 presents some summary statistics of the sample used in this paper. A few interesting features emerge from that table, 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 19 years old). Average schooling is just above Brazilian elementary schooling of 8 years. In Rio de Janeiro, about 30% of sample were not working during the months before the treatment, whereas in Fortaleza that number is around 40%.

We can see that apparently covariates are unbalanced across treatment status. In Rio de Janeiro, a male worker who was older and unemployed during the months before treatment had a greater chance to be in the treatment group than a young female employed worker. In Fortaleza, race/ethnicity and schooling appear to be unbalanced. The main reason for those differences across groups is that the randomization took place within the class level. There were many different courses being offered and attracting different people. Also classes had uneven proportions of treatment and control observations. We thus checked whether at the class level, covariates are unbalanced. In other words, let C be a variable that indicates which class the workers was enrolled and Y, T and X be defined as before. We test whether $X \perp\!\!\!\perp T | C = c$ as a check for good randomization. For Fortaleza, in the original sample there were 196 classes (160 with common support, that is, with at least one treated and one control). When we perform a t-test for differences in means at the class level, using either the original or the restricted sample, we rarely observe any significant difference. Only 22 out of 160 classes present at least one unbalanced variable. For Rio de Janeiro, the proportion of unbalanced covariates at the class level is of the same order of magnitude (13 of 70). We use that as piece of evidence that the randomization process was done reasonably well.

To avoid having to control for as many as 160 cells, as would be in the case of Fortaleza, we opted for controlling for some pre-treatment covariates. We chose those which seemed to be unbalanced when we performed an unconditional test of difference in means. Thus in Rio de Janeiro we control for age and employment one month before treatment, while in Fortaleza we control for race, age and schooling. We discretized schooling and age by constructing groups.[16] Thus, with all "confounding" variables being discrete, we ended up having cells, but in a smaller number than if we had used classes. For Fortaleza we have 24 cells, whereas for Rio de Janeiro, we ended up with 8 cells. For each cell, we inspected whether we had failures of the common support assumption. All cells had treated and control units, so we did not have a failure of common support. Regarding covariates balancing between treated and control groups, we have that for Rio only 2 out of 8 cells present at least one unbalanced variable, while in Fortaleza we found 7 out of 24 unbalanced cells. That is about the same proportion of unbalancing classes when looking at the class level.

[16] For age we grouped 16-17, 18-19, 20-21, and 22-50, whereas for schooling we grouped 0-7 (incomplete elementary), 8-10 (incomplete high school) and 11-15 (at least high school).

The outcome variable of interest is the first monthly salary received in the first job after treatment period. For the whole treatment effect analysis we dropped observations with zero earnings (2 in Rio de Janeiro and 3 in Fortaleza). In Table 9 we report average and inequality treatment effects estimates. We have three estimators: naive, which is a simple difference between groups, reweighted and counterfactual estimators. Our weights in the reweighted estimators are given by the fully saturated model for the propensity-score, which is feasible as we have discrete covariates. In order to have comparability with the counterfactual estimator, we used cell dummies for the linear regression (see the previous section for details on this estimator). As expected, for average treatment effects, both methods are algebraically identical and we can see that in Table 9. They however differ from the naive estimator, which should be inconsistent as it does not "undo" the mixing of classes with different proportions of treated and controls. We compute bootstrapped standard errors as we do not derive analytical standard errors for the counterfactual estimator.[17] Also, for this particular exercise, bootstrapped standard errors are simpler to generate for all estimators and should be valid.

We note that the program, when it affects expected earnings (as in Rio de Janeiro), it has a negative impact. However, we also see that for Rio de Janeiro, although the program does seem to induce no average gains, it does reduce inequality among treated. One possible interpretation is that the program reduces signaling costs, allowing employers to set similar wages for entering workers that have program certificates. For Fortaleza, results are that program is ineffective in reducing inequality.

We correct for the fact that randomization was clustered by using the reweighted and counterfactual estimators. A problem that may emerge with the counterfactual 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 counterfactual and reweighted estimator since some inequality measures are defined only for positive values. We also compute estimators for a sample that drops negative predicted values from the counterfactual approach, for the inequality measures that require positive support. Results in both sites are that the counterfactual estimators are very different from the reweighted and naive estimators, possibly indicating that such deletion procedure may aggravate the bias problem of the counterfactual estimator.

7 Conclusion

We proposed a method that helps applied researchers who 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 what is the impact of the treatment that explain differences in Gini coefficients between these two groups of workers and separate it from a composition effect, induced by different distribution in covariates. Of course, we could make parametric and functional form assumptions in order to relate covariates to the outcome distribution in each group. However, if we are not willing to impose restrictive parametric assumptions, it is not clear how to compare Gini coefficients fixing the distribution of observed covariates.

[17] Although for ATT counterfactual and reweighted should have the same variance, for other functionals this is not true.

Our method consisted of the following. We first estimate, through a reweighing method, the inequality measures of the potential outcomes, and then take the difference between those estimates. That estimation strategy is useful for policy-making purposes when the individual decision to participate in the social 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, many of their inequality parameters.

We showed that the proposed reweighting inequality treatment effect estimators are root-N consistent, asymptotically normal and efficient. Finally, we performed a series of Monte Carlo exercises and apply the method to a new data set. Overall and in both the Monte Carlo and the empirical exercise, the reweighting estimator proposed in this article performed better than its competitors, the estimator based on estimation of counterfactual distributions and the simple difference (naive) estimator.

APPENDIX

Details of the Monte Carlo Exercise

Define:

$$X = \left[1, X_1, X_2, X_1^2, X_2^2, X_1 \cdot X_2\right]^\intercal$$

so a realization of X is $x = \left[1, x_1, x_2, x_1^2, x_2^2, x_1 \cdot x_2\right]^\intercal$. Define also the vector of parameters:

$$\begin{aligned}
\beta_0 &= [\beta_{00}, \beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta_{05}]^\intercal \\
\beta_1 &= [\beta_{10}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{15}]^\intercal \\
\beta_0^s &= [\beta_{00}^s, \beta_{01}^s, \beta_{02}^s, \beta_{03}^s, \beta_{04}^s, \beta_{05}^s]^\intercal \\
\beta_1^s &= [\beta_{10}^s, \beta_{11}^s, \beta_{12}^s, \beta_{13}^s, \beta_{14}^s, \beta_{15}^s]^\intercal .
\end{aligned}$$

Thus the under the specification given in the section 5, the conditional density of the potential outcomes will be

$$\begin{aligned}
f_{Y(0)|X_1,X_2}\left(y|x_1,x_2\right) &= \left(y \cdot \sqrt{2\pi \cdot \left(x^\intercal \beta_0^s\right)^2}\right)^{-1} \cdot \exp\left\{-\frac{1}{2} \cdot \left(\frac{\ln\left(y\right) - x^\intercal \beta_0}{x^\intercal \beta_0^s}\right)^2\right\} \\
f_{Y(1)|X_1,X_2}\left(y|x_1,x_2\right) &= \left(y \cdot \sqrt{2\pi \cdot \left(x^\intercal \beta_1^s\right)^2}\right)^{-1} \cdot \exp\left\{-\frac{1}{2} \cdot \left(\frac{\ln\left(y\right) - x^\intercal \beta_1}{x^\intercal \beta_1^s}\right)^2\right\}
\end{aligned}$$

thus the unconditional densities are given by:

$$\begin{aligned}
f_{Y(0)}\left(y\right) &= \frac{1}{12} \cdot \int_{\mu_{X_1} - \frac{\sqrt{12}}{2}}^{\mu_{X_1} + \frac{\sqrt{12}}{2}} \left(\int_{\mu_{X_2} - \frac{\sqrt{12}}{2}}^{\mu_{X_2} + \frac{\sqrt{12}}{2}} f_{Y(0)|X_1,X_2}\left(y|x_1,x_2\right) \cdot dx_2\right) \cdot dx_1 \\
f_{Y(1)}\left(y\right) &= \frac{1}{12} \cdot \int_{\mu_{X_1} - \frac{\sqrt{12}}{2}}^{\mu_{X_1} + \frac{\sqrt{12}}{2}} \left(\int_{\mu_{X_2} - \frac{\sqrt{12}}{2}}^{\mu_{X_2} + \frac{\sqrt{12}}{2}} f_{Y(1)|X_1,X_2}\left(y|x_1,x_2\right) \cdot dx_2\right) \cdot dx_1
\end{aligned}$$

We compute those integrals numerically. Once we have the densities of the marginals we are able to evaluate numerically the statistics that we listed in section 5, as the mean, the median and some inequality measures.

Proofs

Proof of Theorem 1:

For the first set of equalities, by the definition of observed outcomes ($Y = Y\left(1\right) \cdot T + Y\left(0\right) \cdot \left(1 - T\right)$) and Assumption 1, we have that the first equality holds as the conditional p.d.f. of Y given $T = 1$ and $X = x$ and evaluated at y is exactly the p.d.f. of $Y\left(1\right)$ given $X = x$ evaluated at y:

$$f_{Y|T,X}\left(y;1,x\right) = f_{Y(1)|T,X}\left(y;1,x\right) = f_{Y(1)|X}\left(y;x\right)$$

thus

$$\begin{aligned}
F_{Y(1)}(a) &= \int_{x \in \mathcal{X}} \int_{y \in \mathbb{R}}^{a} f_{Y(1)|X}\left(y;x\right) \cdot dy \cdot f_X\left(x\right) \cdot dx \\
&= \int_{x \in \mathcal{X}} \int_{y \in \mathbb{R}}^{a} f_{Y|T,X}\left(y;1,x\right) \cdot dy \cdot f_X\left(x\right) \cdot dx
\end{aligned}$$

The second equality holds by the definition of weighted c.d.f.s and by the fact that $p(x) > 0$ (Assumption 2):

$$F_Y^{\omega=\omega_1}(a) = \int_{x\in\mathcal{X}} \left(\omega_1(1,p(x)) \cdot p(x) \cdot \int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;1,x)\cdot dy \right.$$
$$\left. + \omega_1(0,p(x)) \cdot (1-p(x)) \cdot \int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;0,x)\cdot dy\right) \cdot f_X(x)\cdot dx$$
$$= \int_{x\in\mathcal{X}} \frac{1}{p(x)} \cdot p(x) \cdot \int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;1,x)\cdot dy \cdot f_X(x)\cdot dx$$
$$= \int_{x\in\mathcal{X}} \int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;1,x)\cdot dy \cdot f_X(x)\cdot dx$$

and the third is just the definition of a weighted expectation given in Equation 4 with the specialization to $\varsigma(Y) = 1\!\!1\{Y \le a\}$ and $\omega = \omega_1$:

$$E\left[\omega_1(T,p(X)) \cdot 1\!\!1\{Y \le a\}\right]$$
$$= \int_{x\in\mathcal{X}} \left(\omega_1(1,p(x)) \cdot p(x) \cdot \int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;1,x)\cdot dy \right.$$
$$\left. + \omega_1(0,p(x)) \cdot (1-p(x)) \cdot \int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;0,x)\cdot dy\right) \cdot f_X(x)\cdot dx$$
$$= F_Y^{\omega=\omega_1}(a)$$

The second set of equalities follows by analogy. The third is trivial as it involves the c.d.f. of $Y(1)|T=1$. Finally, we consider the fourth set of equalities. We have again, by Assumption 1:

$$f_{Y|T,X}(y;0,x) = f_{Y(0)|T,X}(y;0,x) = f_{Y(0)|T,X}(y;1,x)$$

thus

$$\begin{aligned} F_{Y(0)|T=1}(a) &= \int_{x\in\mathcal{X}} \int_{y\in\mathbb{R}}^{a} f_{Y(0)|T,X}(y;1,x)\cdot dy \cdot f_{X|T}(x;1)\cdot dx \\ &= \int_{x\in\mathcal{X}} \frac{p(x)}{p} \cdot \int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;0,x)\cdot dy \cdot f_X(x)\cdot dx \end{aligned}$$

and because of the Bayes' rule we have that

$$f_{X|T}(x;1) = \frac{p(x)}{p} \cdot f_X(x).$$

The second equality holds once again by the definition of weighted c.d.f.s and by the fact that $1 - p(x) > 0$ (Assumption 2):

$$F_Y^{\omega=\omega_{01}}(a) = \int_{x\in\mathcal{X}} \left(\omega_{01}(1,p(x)) \cdot p(x) \cdot \int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;1,x)\cdot dy \right.$$
$$\left. + \omega_{01}(0,p(x)) \cdot (1-p(x)) \cdot \int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;0,x)\cdot dy\right) \cdot f_X(x)\cdot dx$$
$$= \int_{x\in\mathcal{X}} \left(\frac{1}{1-p(x)}\right) \cdot \left(\frac{p(x)}{p}\right) \cdot (1-p(x)) \cdot \int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;0,x)\cdot dy \cdot f_X(x)\cdot dx$$
$$= \int_{x\in\mathcal{X}} \frac{p(x)}{p} \cdot \int_{y\in\mathbb{R}}^{a} f_{Y|T,X}(y;0,x)\cdot dy \cdot f_X(x)\cdot dx$$

and finally the last equality holds as

$$
\begin{aligned}
& E\left[\omega_{01}\left(T, p\left(X\right)\right) \cdot 1\!\!1\{Y \leq a\}\right] \\
& = \int_{x\in\mathcal{X}} \left(\omega_{01}\left(1, p\left(x\right)\right) \cdot p\left(x\right) \cdot \int_{y\in\mathbb{R}}^{a} f_{Y|T,X}\left(y; 1, x\right) \cdot dy \right.\\
& \left. + \,\omega_{01}\left(0, p\left(x\right)\right) \cdot \left(1 - p\left(x\right)\right) \cdot \int_{y\in\mathbb{R}}^{a} f_{Y|T,X}\left(y; 0, x\right) \cdot dy\right) \cdot f_X\left(x\right) \cdot dx \\
& = F_Y^{\omega=\omega_{01}}\left(a\right)
\end{aligned}
$$

Q.E.D.

Proof of Corollary 1:

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

$$
\begin{aligned}
\Delta^{\nu} &= \nu\left(F_{Y(1)}\right) - \nu\left(F_{Y(0)}\right) = \nu\left(F_Y^{\omega=\omega_1}\right) - \nu\left(F_Y^{\omega=\omega_0}\right) \\
\Delta^{\nu}_{T=1} &= \nu\left(F_{Y(1)|T=1}\right) - \nu\left(F_{Y(0)|T=1}\right) = \nu\left(F_Y^{\omega=\omega_{11}}\right) - \nu\left(F_Y^{\omega=\omega_{01}}\right) \\
\Delta^{\nu}_{C} &= \nu\left(F_Y\right) - \nu\left(F_{Y(0)}\right) = \nu\left(F_Y\right) - \nu\left(F_Y^{\omega=\omega_0}\right)
\end{aligned}
$$

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 $\pi_K = \arg\max_{\pi\in\mathbb{R}^K} E\left\{p(X) \cdot \log(L(H_K(X)'\pi)) + (1 - p(X)) \cdot \log(1 - L(H_K(X)'\pi))\right\}$;

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

(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.

Now consider $\sqrt{N} \cdot \left(\nu\left(\widehat{F}_Y^{\omega=\widehat{\omega}}\right) - \nu\left(F_Y^{\omega}\right)\right)$. Under the asymptotic linearity assumption (Assumption 3) we have

$$\sqrt{N} \cdot \left(\nu\left(\widehat{F}_Y^{\omega=\widehat{\omega}}\right) - \nu\left(F_Y^{\omega}\right)\right) \tag{A-2}$$

$$= \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \widehat{\omega}_i \cdot \psi^{\nu}\left(Y_i; F_Y^{\omega}\right) + o_p(1) \tag{A-3}$$

$$= \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\widehat{\omega}_i - \omega\left(T_i, p\left(X_i\right)\right)\right) \cdot \psi^{\nu}\left(Y_i; F_Y^{\omega}\right) \tag{A-4}$$

$$-\frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} h\left(T_i, p\left(X_i\right)\right) \cdot \psi^{\nu}\left(Y_i; F_Y^{\omega}\right) \cdot \left(\widehat{p}\left(X_i\right) - p\left(X_i\right)\right) \tag{A-5}$$

$$+\frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} h\left(T_i, p\left(X_i\right)\right) \cdot \psi^{\nu}\left(Y_i; F_Y^{\omega}\right) \cdot \left(\widehat{p}\left(X_i\right) - p\left(X_i\right)\right) \tag{A-6}$$

$$-\frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} E\left[h\left(T, p\left(X\right)\right) \cdot \psi^{\nu}\left(Y; F_Y^{\omega}\right) | X = X_i\right] \cdot \left(T_i - p\left(X_i\right)\right) \tag{A-7}$$

$$+\frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \omega\left(T_i, p\left(X_i\right)\right) \cdot \psi^{\nu}\left(Y_i; F_Y^{\omega}\right) \tag{A-8}$$

$$+\frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} E\left[h\left(T, p\left(X\right)\right) \cdot \psi^{\nu}\left(Y; F_Y^{\omega}\right) | X = X_i\right] \cdot \left(T_i - p\left(X_i\right)\right) + o_p(1) \tag{A-9}$$

We now find bounds for each term of the above sum.

[*Equation* (A-4)]:

$$
\begin{aligned}
& \left| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\widehat{\omega}_i - \omega\left(T_i, p\left(X_i\right)\right)\right) \cdot \psi^{\nu}\left(Y_i; F_Y^{\omega}\right) \right| \\
= & \left| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \psi^{\nu}\left(Y_i; F_Y^{\omega}\right) \cdot \left(\frac{1}{N_{\omega}} \cdot \omega\left(T_i, \widehat{p}\left(X_i\right)\right) - \omega\left(T_i, p\left(X_i\right)\right)\right) \right| \\
= & \left| \sum_{i=1}^{N} \psi^{\nu}\left(Y_i; F_Y^{\omega}\right) \cdot \left(\frac{\omega\left(T_i, \widehat{p}\left(X_i\right)\right) - N_{\omega} \cdot \omega\left(T_i, p\left(X_i\right)\right)}{N_{\omega}} \right) \right| \\
\leq & \; N_{\omega}^{-1} \cdot \left| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\omega\left(T_i, \widehat{p}\left(X_i\right)\right) - \omega\left(T_i, p\left(X_i\right)\right)\right) \cdot \psi^{\nu}\left(Y_i; F_Y^{\omega}\right) \right| \\
& + N_{\omega}^{-1} \cdot \left|1 - N_{\omega}\right| \cdot \left| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \omega\left(T_i, p\left(X_i\right)\right) \cdot \psi^{\nu}\left(Y_i; F_Y^{\omega}\right) \right|.
\end{aligned}
$$

We show that $N_{\omega} = 1 + o_p(1)$. We construct a proof for N_1 only as the algebra for the other three is very similar:

$$
\begin{aligned}
\left|N_1 - 1\right| = & \left| \frac{1}{N} \cdot \sum_{i=1}^{N} \left(\frac{T_i}{\widehat{p}\left(X_i\right)} \right) - 1 \right| = \left| \frac{1}{N} \cdot \sum_{i=1}^{N} \left(\frac{T_i}{\widehat{p}\left(X_i\right)} - \frac{T_i}{p\left(X_i\right)} + \frac{T_i}{p\left(X_i\right)} \right) - 1 \right| \\
= & \left| \frac{1}{N} \cdot \sum_{i=1}^{N} \omega\left(T, \widehat{p}\left(X_i\right)\right) - \omega\left(T, p\left(X_i\right)\right) + \frac{T_i}{\widehat{p}\left(X_i\right)} - 1 \right| \\
\leq & \left| \frac{1}{N} \cdot \sum_{i=1}^{N} \omega\left(T, \widehat{p}\left(X_i\right)\right) - \omega\left(T, p\left(X_i\right)\right) \right| + \left| \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{T_i - \widehat{p}\left(X_i\right)}{\widehat{p}\left(X_i\right)} \right|
\end{aligned}
$$

but

$$
\begin{aligned}
& \left| \frac{1}{N} \cdot \sum_{i=1}^{N} \omega\left(T, \widehat{p}\left(X_i\right)\right) - \omega\left(T, p\left(X_i\right)\right) \right| \\
= & \left| \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{T_i}{\widehat{p}\left(X_i\right)} - \frac{T_i}{p\left(X_i\right)} \right| \\
= & \left| \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{T_i \cdot \left(p\left(X_i\right) - \widehat{p}\left(X_i\right)\right)}{\widehat{p}\left(X_i\right) \cdot p\left(X_i\right)} \right| \\
\leq & \left(\inf_{x \in \mathcal{X}} \widehat{p}(x) \right)^{-1} \cdot \sup_{x \in \mathcal{X}} \left|p(x) - \widehat{p}(x)\right| \cdot \left| \frac{1}{N} \cdot \sum_{i=1}^{N} \frac{T_i}{p\left(X_i\right)} \right| \\
\leq & \; C \cdot \sup_{x \in \mathcal{X}} \left|p(x) - \widehat{p}(x)\right| \cdot \left(1 + \left| \frac{1}{N} \cdot \sum_{i=1}^{N} \left(\frac{T_i}{p\left(X_i\right)} - 1 \right) \right| \right) \\
= & \; C \cdot o_p(1) \cdot \left(1 + o_p(1)\right) \\
= & \; o_p(1)
\end{aligned}
$$

by the law of large numbers, which can be applied since

$$E\left[\frac{T}{p(X)}\right] = E\left[E\left[\frac{T}{p(X)}|X\right]\right] = E\left[\frac{E[T|X]}{p(X)}\right] = E\left[\frac{p(X)}{p(X)}\right] = 1$$

and because the variance is bounded, that is

$$\begin{aligned} V\left[\frac{T}{p(X)}\right] &= E\left[\frac{T}{p^2(X)}\right] - 1 = E\left[\frac{p(X)}{p^2(X)}\right] - 1 \\ &\leq \left(\inf_{x\in\mathcal{X}} p(x)\right)^{-1} - 1 \leq \infty. \end{aligned}$$

We now investigate the other term of the sum:

$$\begin{aligned} & \left|\frac{1}{N}\cdot\sum_{i=1}^{N}\frac{T_i - \widehat{p}(X_i)}{\widehat{p}(X_i)}\right| \\ \leq & \left|\frac{1}{N}\cdot\sum_{i=1}^{N}\left(\frac{T_i - p(X_i)}{\widehat{p}(X_i)}\right)\right| + \left|\frac{1}{N}\cdot\sum_{i=1}^{N}\left(\frac{p(X_i) - \widehat{p}(X_i)}{\widehat{p}(X_i)}\right)\right| \\ \leq & \left(\inf_{x\in\mathcal{X}} \widehat{p}(x)\right)^{-1}\cdot\left(\left|\frac{1}{N}\cdot\sum_{i=1}^{N}(T_i - p + p - p(X_i))\right| + \sup_{x\in\mathcal{X}}|p(x) - \widehat{p}(x)|\right) \\ \leq & C\cdot\left(\left|\frac{1}{N}\cdot\sum_{i=1}^{N}(T_i - p)\right| + \left|\frac{1}{N}\cdot\sum_{i=1}^{N}(p(X_i) - p)\right| + \sup_{x\in\mathcal{X}}|p(x) - \widehat{p}(x)|\right) \\ = & C\cdot(o_p(1) + o_p(1) + o_p(1)) = o_p(1) \end{aligned}$$

Finally, consider

$$\begin{aligned} & \sup_{x\in\mathcal{X}}|\omega(t,\widehat{p}(x)) - \omega(t,p(x))| \\ = & \sup_{x\in\mathcal{X}}\left|\frac{t\cdot(p(x) - \widehat{p}(x))}{\widehat{p}(x)\cdot p(x)}\right| \\ \leq & \left(\inf_{x\in\mathcal{X}}\widehat{p}(x)\right)^{-1}\cdot\left(\inf_{x\in\mathcal{X}}p(x)\right)^{-1}\cdot\sup_{x\in\mathcal{X}}|\widehat{p}(x) - p(x)| \\ \leq & C\cdot\sup_{x\in\mathcal{X}}|\widehat{p}(x) - p(x)| = o_p(1) \end{aligned}$$

Thus, we have that

$$
\begin{aligned}
& N_\omega^{-1} \cdot \left| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \left(\omega\left(T_i, \widehat{p}\left(X_i\right)\right) - \omega\left(T_i, p\left(X_i\right)\right)\right) \cdot \psi^\nu\left(Y_i; F_Y^\omega\right) \right| \\
& + N_\omega^{-1} \cdot \left|1 - N_\omega\right| \cdot \left| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \omega\left(T_i, p\left(X_i\right)\right) \cdot \psi^\nu\left(Y_i; F_Y^\omega\right) \right| \\
\leq \; & C_1 \cdot \sup_{x \in \mathcal{X}} \left|\widehat{p}(x) - p(x)\right| \cdot \left| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \psi^\nu\left(Y_i; F_Y^\omega\right) \right| \\
& + C_2 \cdot \left|1 - N_\omega\right| \cdot \left| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} \omega\left(T_i, p\left(X_i\right)\right) \cdot \psi^\nu\left(Y_i; F_Y^\omega\right) \right| \\
= \; & C_1 \cdot o_p(1) \cdot O_p(1) + C_2 \cdot o_p(1) \cdot O_p(1) \\
= \; & o_p(1).
\end{aligned}
$$

[*Equation* (A-5)]:

$$
\begin{aligned}
& \left| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} h\left(T_i, p\left(X_i\right)\right) \cdot \psi^\nu\left(Y_i; F_Y^\omega\right) \cdot \left(\widehat{p}\left(X_i\right) - p\left(X_i\right)\right) \right| \\
\leq \; & \sup_{x \in \mathcal{X}} \left|\widehat{p}(x) - p(x)\right| \cdot \left| \frac{1}{\sqrt{N}} \cdot \sum_{i=1}^{N} h\left(T_i, p\left(X_i\right)\right) \cdot \psi^\nu\left(Y_i; F_Y^\omega\right) \right| \\
= \; & o_p(1) \cdot O_p(1) = o_p(1)
\end{aligned}
$$

[*Equations* (A-6-A-7)]: We do not provide a detailed derivation for that expression in order to show that it is $o_p(1)$. The particular case of ω_1 and $\nu = \mu$ corresponds to Equations (40)-(43) of Hirano, Imbens and Ridder (2003). As long as $\psi^\nu\left(y; F_Y^\omega\right)$ satisfies the requirements of Assumption 4, their results for the mean can be generalized for the influence function ψ^ν. Also, the algebra for the other three cases (ω_0, ω_{11} and ω_{01}) is similar to the one of ω_1 and the results follow by analogy.

We now have that Equations (A-4)-(A-7) are all $o_p(1)$ and thus, the sum of Equations (A-8) and (A-9) is:

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

Thus, by an application of central limit theorem, we have that

$$
\sqrt{N} \cdot \left(\nu\left(\widehat{F}_Y^{\omega=\widehat{\omega}}\right) - \nu\left(F_Y^\omega\right)\right) \xrightarrow{D} N\left(0, V_\nu\right)
$$

Q.E.D.

Proof of Proposition 2

Given Proposition 1, the first part of Proposition 2 follows straightforwardly. The remaining part relates to efficiency. We now show that V_ν attains the efficiency bound. We show it again for the case of the $Y(1)$ distribution. We will invoke the results from Chen, Hong and Tarozzi (2008) that show that for moment conditions of the type

$$E\left[m\left(Y(1);\theta_1\right)\right]=0$$

under strong unconfoundedness (assumptions 1 and 2), the core of the efficiency bound will be given by the expectation of the squared efficient influence function ψ^{θ_1}:[18]

$$\psi^{\theta_1}=\frac{T}{p(X)}\cdot\left(m\left(Y(1);\theta_1\right)-E\left[m\left(Y;\theta_1\right)|X,T=1\right]\right)+E\left[m\left(Y;\theta_1\right)|X,T=1\right]$$

thus the bound is

$$\begin{aligned}V_{\theta_1} &= \left(\frac{\partial E\left[m\left(Y(1);\theta_1\right)\right]}{\partial\theta_1}\right)^{-2}\\ &\cdot E\left[\left(\frac{T}{p(X)}\cdot\left(m\left(Y(1);\theta_1\right)-E\left[m\left(Y;\theta_1\right)|X,T=1\right]\right)+E\left[m\left(Y;\theta_1\right)|X,T=1\right]\right)^2\right]\end{aligned}$$

We now show that by expressing our inequality measures in terms of their influence functions, we are able to use their results. First consider

$$\begin{aligned}& E\left[\psi^{\nu}\left(Y(1);F_{Y(1)}\right)\right]\\ =\ & E\left[E\left[\psi^{\nu}\left(Y(1);F_{Y(1)}\right)|X\right]\right]=E\left[E\left[\psi^{\nu}\left(Y(1);F_{Y(1)}\right)|X,T=1\right]\right]\\ =\ & E\left[E\left[\psi^{\nu}\left(Y;F_{Y(1)}\right)|X,T=1\right]\right]=E\left[E\left[T\cdot\psi^{\nu}\left(Y;F_{Y(1)}\right)|X,T=1\right]\right]\\ =\ & E\left[\frac{E\left[T\cdot\psi^{\nu}\left(Y;F_{Y(1)}\right)|X\right]}{p(X)}\right]=E\left[\frac{T}{p(X)}\cdot\psi^{\nu}\left(Y;F_{Y(1)}\right)\right]\\ =\ & E\left[\omega_1\left(T,p(X)\right)\cdot\psi^{\nu}\left(Y;F_{Y(1)}\right)\right]\\ =\ & E\left[\omega_1\left(T,p(X)\right)\cdot\psi^{\nu}\left(Y;F_Y^{\omega=\omega_1}\right)\right]\\ =\ & 0\end{aligned}$$

Thus, the $m(\cdot;\cdot)$ function can be replaced by $\psi^{\nu}(\cdot;\cdot)$. Note also that by definition of the influence function,

$$\frac{\partial E\left[\psi^{\nu}\left(Y(1);F_{Y(1)}\right)\right]}{\partial\nu_1}=-1$$

Now:

$$\begin{aligned}E\left[\psi^{\nu}\left(Y(1);F_{Y(1)}\right)|X\right] &= E\left[\psi^{\nu}\left(Y;F_{Y(1)}\right)|X,T=1\right]\\ &= \frac{E\left[T\cdot\psi^{\nu}\left(Y;F_{Y(1)}\right)|X\right]}{p(X)}\\ &= E\left[\omega_1\left(T,p(X)\right)\cdot\psi^{\nu}\left(Y;F_Y^{\omega=\omega_1}\right)|X\right]\end{aligned}$$

therefore:

$$\begin{aligned}& V_{\nu_1}\\ =\ & E\left[\left(\omega_1\left(T,p(X)\right)\cdot\psi^{\nu}\left(Y;F_Y^{\omega=\omega_1}\right)+E\left[h_1\left(T,p(X)\right)\cdot\psi^{\nu}\left(Y;F_Y^{\omega=\omega_1}\right)|X\right]\cdot\left(T-p(X)\right)\right)^2\right]\\ =\ & E\left[\left(\frac{T}{p(X)}\cdot\psi^{\nu}\left(Y;F_Y^{\omega=\omega_1}\right)-E\left[\frac{T}{p(X)}\cdot\psi^{\nu}\left(Y;F_Y^{\omega=\omega_1}\right)|X\right]\cdot\frac{\left(T-p(X)\right)}{p(X)}\right)^2\right]\end{aligned}$$

[18] Those results are equivalent to those presented by Chen, Hong and Tarozzi (2008) after a notational adjustment.

which after some algebra can be shown to be equal to:

$$
\begin{aligned}
& E\left[\left(\frac{T}{p(X)} \cdot \psi^{\nu}\left(Y(1);F_{Y}^{\omega=\omega_1}\right) - E\left[\frac{T}{p(X)} \cdot \psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_1}\right)|X\right] \cdot \frac{(T-p(X))}{p(X)}\right)^2\right] \\
= \ & E[(\frac{T}{p(X)} \cdot \left(\psi^{\nu}\left(Y(1);F_{Y}^{\omega=\omega_1}\right) - E\left[\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_1}\right)|X,T=1\right]\right) + E\left[\psi^{\nu}\left(Y;F_{Y}^{\omega=\omega_1}\right)|X,T=1\right])^2] \\
= \ & E\left[\left(\frac{T}{p(X)} \cdot \left(m(Y(1);\theta_1) - E\left[m(Y;\theta_1)|X,T=1\right]\right) + E\left[m(Y;\theta_1)|X,T=1\right]\right)^2\right] \\
= \ & V_{\theta_1}
\end{aligned}
$$

thus the estimator $\widehat{\nu}_1$ is efficient in the class of semiparametric estimators. The same results apply to the quantities $\widehat{\nu}_0$, $\widehat{\nu}_{11}$, $\widehat{\nu}_{01}$ and $\widehat{\nu}_Y$ by analogy.

Efficiency of $\widehat{\Delta}^{\nu} = \widehat{\nu}_1 - \widehat{\nu}_0$, $\widehat{\Delta}^{\nu}_T = \widehat{\nu}_{11} - \widehat{\nu}_{01}$ and $\widehat{\Delta}^{\nu}_C = \widehat{\nu}_Y - \widehat{\nu}_0$ follows after noticing that the influence function of the difference is simply the difference in the influence functions. Thus, the normalized asymptotic variances of $\widehat{\Delta}^{\nu}$, $\widehat{\Delta}^{\nu}_T$ and $\widehat{\Delta}^{\nu}_C$, respectively V, V_T and V_C will achieve the semiparametric efficiency bound.

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), "Instrumental Variables Estimation of Quantile Treatment Effects," *Econometrica*, 70, 91-117.

Barnow, B., G. Cain, and A. Goldberger, (1980), "Issues in the Analysis of the Selectivity Bias," *Evaluation Studies*, 5, 42-59.

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

Bitler, M., J. Gelbach, and H. Hoynes, (2006), "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," *American Economic Review*, 96, 988-1012.

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.

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.

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.

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. (2005), "Program Evaluation as a Decision Problem," *Journal of Econometrics*, 125, 141-173.

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.

Ferguson, T. (1996), *A Course in Large Sample Theory*, Chapman & Hall/CRC.

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

Firpo, S. and M. Foguel (2008), "Evaluating the Impact of a Brazilian Public Training Program Using Experimental Data Set," Department of Economics, PUC-Rio. Typescript.

Foguel, M. (2008), "Ensaios sobre Segregação no Mercado de Trabalho, Avaliação de Programas Sociais e Identificação Não-Paramétrica," Doctoral Thesis. Universidade Federal Fluminense, Departamento de Economia.

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

Hampel, F., (1974), "The Influence Curve and Its Role in Robust Estimation," *Journal of the American Statistical Association*, 69, 383-393.

Hampel, F., E. Ronchetti, P. Rousseeuw, and W. Stahel, (1986), *Robust Statistics: The Approach Based on Influence Functions*, John Wiley & Sons.

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.

Hoeffding, W., (1948), "A Class of Statistics With Asymptotically Normal Distribution," *Annals of Mathematical Statistics*, 19, 293-325.

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

Koenker, R., and G. Bassett, (1978), "Regression Quantiles," *Econometrica*, 46, 33-50.

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.

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

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.

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

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

Rosenbaum, P., and D. Rubin, (1984), "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score," *Journal of the American Statistical Association*, 79, 516-574.

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

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.

Thuysbaert, B., (2007), "Econometric Essays on the Measurement of Poverty," Ph.D. thesis, Department of Economics, K.U. Leuven.

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

TABLE 4: Point Estimates from Monte Carlo Exercise (Sample Size 500, Replications 1000)

			Average	Median	Lower 5th percentile	Upper 5th percentile	Standard Deviation	Bias	Root Mean Squared Error	Mean Absolute Error	Median Absolute Error	90% C.I. Coverage Rate (a)	90% C.I. Coverage Rate (b)
Mean Treatment Effects	Target	1.03											
	Unfeasible		1.03	1.03	0.96	1.12	0.05	0.00	0.05	0.04	0.03	89.70%	90.20%
	Feasible		1.05	1.04	0.88	1.25	0.11	0.02	0.11	0.09	0.08	89.40%	92.70%
	Naive		1.10	1.10	0.96	1.26	0.09	0.07	0.12	0.09	0.08	83.80%	81.70%
	C. D.		1.05	1.05	0.87	1.27	0.12	0.02	0.12	0.09	0.08	89.80%	90.30%
CV	Target	0.26											
	Unfeasible		0.26	0.25	0.17	0.39	0.06	0.00	0.06	0.06	0.05	77.60%	94.20%
	Feasible		0.25	0.24	0.10	0.43	0.08	-0.01	0.08	0.08	0.07	76.20%	93.30%
	Naive		0.31	0.29	0.16	0.51	0.08	0.04	0.09	0.09	0.06	80.60%	91.80%
	C. D.		0.33	0.31	0.17	0.54	0.09	0.07	0.12	0.10	0.07	82.40%	90.40%
Interquartile Range	Target	0.56											
	Unfeasible		0.56	0.55	0.46	0.66	0.06	0.00	0.06	0.05	0.04	88.80%	90.30%
	Feasible		0.57	0.55	0.29	0.92	0.18	0.02	0.19	0.15	0.12	87.60%	93.00%
	Naive		0.66	0.65	0.48	0.87	0.12	0.10	0.16	0.12	0.10	81.80%	79.80%
	C. D.		0.85	0.83	0.58	1.20	0.17	0.30	0.34	0.30	0.27	47.80%	63.30%
Theil Index	Target	0.07											
	Unfeasible		0.07	0.07	0.05	0.11	0.02	0.00	0.02	0.01	0.01	84.90%	92.70%
	Feasible		0.07	0.07	0.03	0.13	0.03	0.00	0.03	0.02	0.02	79.70%	92.40%
	Naive		0.09	0.09	0.05	0.15	0.03	0.02	0.03	0.03	0.02	84.00%	87.60%
	C. D.		0.13	0.11	0.05	0.26	0.06	0.06	0.08	0.06	0.04	86.00%	87.10%
Gini Coefficient	Target	0.09											
	Unfeasible		0.09	0.09	0.07	0.11	0.01	0.00	0.01	0.01	0.01	90.60%	90.40%
	Feasible		0.08	0.08	0.04	0.13	0.03	0.00	0.03	0.02	0.02	85.70%	90.80%
	Naive		0.11	0.11	0.07	0.15	0.02	0.03	0.03	0.03	0.03	68.90%	73.60%
	C. D.		0.15	0.14	0.08	0.25	0.04	0.06	0.08	0.07	0.05	62.30%	79.00%

(a) proportion of replications "i" such that $|\text{estimate}_i$ - target$| < 1.645*$(bootstrapped s.e.$_i$)

(b) proportion of replications "i" such that $|\text{estimate}_i$ - target$| < 1.645*$(s.d. of sample {estimate$_1$, estimate$_2$, ..., estimate$_{\#replications}$})

TABLE 5: Point Estimates from Monte Carlo Exercise (Sample Size 2500, Replications 1000)

			Average	Median	Lower 5th percentile	Upper 5th percentile	Standard Deviation	Bias	Root Mean Squared Error	Mean Absolute Error	Median Absolute Error	90% C.I. Coverage Rate (a)	90% C.I. Coverage Rate (b)
Mean Treatment Effects	Target	1.03											
	Unfeasible		1.03	1.03	1.00	1.07	0.02	0.00	0.02	0.02	0.02	89.30%	90.30%
	Feasible		1.05	1.05	0.97	1.15	0.05	0.02	0.05	0.05	0.04	86.70%	88.60%
	Naive		1.10	1.10	1.03	1.18	0.04	0.07	0.08	0.07	0.07	50.80%	54.80%
	C. D.		1.05	1.05	0.96	1.14	0.05	0.01	0.06	0.05	0.04	87.80%	88.50%
CV	Target	0.26											
	Unfeasible		0.26	0.26	0.21	0.32	0.03	0.00	0.03	0.03	0.02	82.90%	92.70%
	Feasible		0.27	0.26	0.19	0.36	0.05	0.01	0.05	0.04	0.03	84.10%	91.90%
	Naive		0.32	0.31	0.24	0.42	0.05	0.06	0.07	0.06	0.05	70.90%	80.20%
	C. D.		0.34	0.33	0.25	0.45	0.05	0.08	0.09	0.08	0.07	61.40%	72.00%
Interquartile Range	Target	0.56											
	Unfeasible		0.56	0.55	0.52	0.60	0.03	0.00	0.03	0.02	0.02	90.30%	91.20%
	Feasible		0.57	0.57	0.46	0.70	0.08	0.02	0.08	0.06	0.05	90.50%	91.60%
	Naive		0.65	0.65	0.57	0.73	0.05	0.10	0.11	0.10	0.10	41.20%	41.80%
	C. D.		0.85	0.84	0.71	1.02	0.09	0.29	0.31	0.29	0.29	1.80%	5.90%
Theil Index	Target	0.07											
	Unfeasible		0.07	0.07	0.06	0.09	0.01	0.00	0.01	0.01	0.01	88.80%	91.30%
	Feasible		0.08	0.08	0.06	0.10	0.01	0.00	0.01	0.01	0.01	87.10%	89.80%
	Naive		0.10	0.10	0.07	0.12	0.01	0.02	0.03	0.02	0.02	57.50%	63.50%
	C. D.		0.13	0.12	0.09	0.19	0.03	0.06	0.06	0.06	0.05	34.50%	59.10%
Gini Coefficient	Target	0.09											
	Unfeasible		0.09	0.09	0.08	0.10	0.01	0.00	0.01	0.00	0.00	88.00%	87.60%
	Feasible		0.09	0.09	0.07	0.11	0.01	0.00	0.01	0.01	0.01	86.10%	89.10%
	Naive		0.11	0.11	0.09	0.13	0.01	0.03	0.03	0.03	0.03	18.40%	25.10%
	C. D.		0.15	0.15	0.11	0.20	0.02	0.07	0.07	0.07	0.06	7.80%	28.50%

(a) proportion of replications "i" such that $|\text{estimate}_i - \text{target}| < 1.645*(\text{bootstrapped s.e.}_i)$
(b) proportion of replications "i" such that $|\text{estimate}_i - \text{target}| < 1.645*(\text{s.d. of sample } \{\text{estimate}_1, \text{estimate}_2, ..., \text{estimate}_{\#replications}\})$

TABLE 6: Bootstrapped Standard Errors from Monte Carlo Exercise
(Sample Size: 500, Replications: 1000, Bootstrap Replications: 100)

		Average	Median	Lower 5th percentile	Upper 5th percentile
Mean Treatment Effects	Unfeasible	0.05	0.05	0.04	0.06
	Feasible	0.11	0.10	0.07	0.19
	Naive	0.09	0.09	0.06	0.13
	C. D.	0.12	0.11	0.08	0.17
Coefficient of Variation	Unfeasible	0.06	0.05	0.03	0.11
	Feasible	0.08	0.07	0.04	0.15
	Naive	0.08	0.06	0.04	0.17
	C. D.	0.09	0.08	0.04	0.19
Interquartile Range	Unfeasible	0.06	0.06	0.05	0.08
	Feasible	0.18	0.16	0.10	0.35
	Naive	0.12	0.12	0.08	0.17
	C. D.	0.17	0.16	0.10	0.29
Theil Index	Unfeasible	0.02	0.01	0.01	0.03
	Feasible	0.03	0.02	0.01	0.05
	Naive	0.03	0.02	0.01	0.05
	C. D.	0.06	0.04	0.02	0.13
Gini Coefficient	Unfeasible	0.01	0.01	0.01	0.02
	Feasible	0.03	0.02	0.02	0.04
	Naive	0.02	0.02	0.02	0.03
	C. D.	0.04	0.04	0.02	0.09

TABLE 7: Bootstrapped Standard Errors from Monte Carlo Exercise
(Sample Size: 2500, Replications: 1000, Bootstrap Replications: 100)

		Average	Median	Lower 5th percentile	Upper 5th percentile
Mean Treatment Effects	Unfeasible	0.02	0.02	0.02	0.03
	Feasible	0.05	0.05	0.04	0.08
	Naive	0.04	0.04	0.03	0.05
	C. D.	0.05	0.05	0.04	0.07
Coefficient of Variation	Unfeasible	0.03	0.03	0.02	0.05
	Feasible	0.05	0.04	0.03	0.08
	Naive	0.05	0.04	0.03	0.09
	C. D.	0.05	0.05	0.03	0.10
Interquartile Range	Unfeasible	0.03	0.03	0.02	0.03
	Feasible	0.08	0.07	0.05	0.12
	Naive	0.05	0.05	0.04	0.07
	C. D.	0.09	0.08	0.06	0.13
Theil Index	Unfeasible	0.01	0.01	0.01	0.01
	Feasible	0.01	0.01	0.01	0.02
	Naive	0.01	0.01	0.01	0.02
	C. D.	0.03	0.02	0.01	0.06
Gini Coefficient	Unfeasible	0.01	0.01	0.01	0.01
	Feasible	0.01	0.01	0.01	0.02
	Naive	0.01	0.01	0.01	0.01
	C. D.	0.02	0.02	0.01	0.04

Table 8: *Summary Statistics*

		Rio de Janeiro			Fortaleza		
		Control **(A)**	*Treatment* **(B)**	*Difference* **(A)-(B)**	*Control* **(C)**	*Treatment* **(D)**	*Difference* **(C)-(D)**
Male		0.395	0.454	-0.060	0.449	0.422	0.027
		(0.024)	(0.021)	(0.032)*	(0.018)	(0.018)	(0.025)
White		0.421	0.391	0.030	0.317	0.369	-0.052
		(0.025)	(0.021)	(0.032)	(0.017)	(0.018)	(0.024)**
Average Age		19.255	19.887	-0.632	27.256	27.995	-0.739
		(0.219)	(0.226)	(0.322)*	(0.270)	(0.305)	(0.407)*
Household Head		0.044	0.062	-0.018	0.213	0.237	-0.024
		(0.010)	(0.010)	(0.015)	(0.015)	(0.015)	(0.021)
Single		0.931	0.905	0.026	0.605	0.601	0.004
		(0.013)	(0.013)	(0.018)	(0.018)	(0.018)	(0.025)
Average schooling		8.673	8.905	-0.231	9.708	9.377	0.331
		(0.116)	(0.090)	(0.144)	(0.091)	(0.100)	(0.135)**
Number of children		0.164	0.221	-0.057	0.791	0.893	-0.102
		(0.026)	(0.030)	(0.041)	(0.045)	(0.049)	(0.067)
Employment	*June 1998*	0.287	0.261	0.026	0.424	0.435	-0.011
		(0.022)	(0.019)	(0.029)	(0.018)	(0.018)	(0.026)
	July 1998	0.301	0.245	0.057	0.403	0.415	-0.012
		(0.028)	(0.018)	(0.029)**	(0.018)	(0.018)	(0.025)
	August 1998	0.331	0.239	0.092	0.398	0.405	-0.007
		(0.023)	(0.018)	(0.029)***	(0.018)	(0.018)	(0.025)
First hourly wage after the program		1.809	1.518	0.291	1.813	1.828	-0.014
		(0.155)	(0.051)	(0.147)	(0.094)	(0.127)	(0.158)
Observations		408	548		767	756	

Standard deviations in parenthesis. *: *Significant at* 10%; **: *Significant at* 5%; ***: *Significant at* 1%.

Table 9a: Salary at First Job After the Program (adjusted by worked hours): Rio de Janeiro[£]

	Treated Y(1)|T=1	Control Y(0)|T=0	Reweighted Control Y(0)|T=1	Counterfactual Control Y(0)|T=1	Treatment Effect Estimators: Naive	Treatment Effect Estimators: Reweighted	Treatment Effect Estimators: Counterfactual
	(A)	(B)	(C)	(D)	(A)-(B)	(A)-(C)	(A)-(D)
Average	1.518	1.818	1.844	1.844	-0.300	-0.325	-0.325
	(0.197)	(0.282)	(0.291)	(0.291)	(0.164)*	(0.179)*	(0.179)*
Interquartile Range	0.919	0.904	0.935	2.484	0.016	-0.015	-1.564
	(0.123)	(0.129)	(0.129)	(0.686)	(0.064)	(0.069)	(0.592)***
Coefficient of Variation	0.793	1.730	1.751	1.802	-0.937	-0.958	-1.009
	(0.113)	(0.312)	(0.307)	(0.335)	(0.226)***	(0.218)***	(0.243)***
Theil Inequality Index	0.219	0.547	0.566	-	-0.328	-0.347	-
	(0.034)	(0.128)	(0.131)	-	(0.108)***	(0.110)***	-
Gini Coefficient	0.344	0.460	0.467	-	-0.116	-0.123	-
	(0.026)	(0.049)	(0.052)	-	(0.042)***	(0.044)***	-
Theil Inequality Index[££]	0.360	0.547	0.566	1.269	-0.187	-0.207	-0.910
	(0.045)	(0.109)	(0.112)	(0.339)	(0.079)**	(0.085)**	(0.303)***
Gini Coefficient[££]	0.315	0.460	0.467	0.221	-0.144	-0.151	0.094
	(0.023)	(0.042)	(0.045)	(0.052)	(0.047)***	(0.049)***	(0.049)*

*: Significant at 10%; **: Significant at 5%; ***: Significant at 1%.

£: Sample excludes zero earnings (2 observations)

££: Sample excludes non-positive counterfactual outcome variable (114 observations)

Bootstrapped standard errors in parenthesis.

Table 9b: Salary at First Job After the Program (adjusted by worked hours): Fortaleza[£]

	Treated Y(1)|T=1	Control Y(0)|T=0	Reweighted Control Y(0)|T=1	Counterfactual Control Y(0)|T=1	Treatment Effect Estimators		
					Naive	Reweighted	Counterfactual
	(A)	(B)	(C)	(D)	(A)-(B)	(A)-(C)	(A)-(D)
Average	1.835 (0.126)	1.813 (0.093)	1.759 (0.094)	1.759 (0.094)	0.022 (0.158)	0.076 (0.153)	0.076 (0.153)
Interquartile Range	1.036 (0.057)	1.096 (0.053)	1.087 (0.072)	0.901 (0.257)	-0.059 (0.081)	-0.051 (0.090)	0.136 (0.261)
Coefficient of Variation	1.910 (0.361)	1.440 (0.215)	1.257 (0.214)	1.332 (0.207)	0.470 (0.410)	0.652 (0.410)	0.578 (0.404)
Theil Inequality Index	0.541 (0.118)	0.408 (0.071)	0.359 (0.066)	- -	0.134 (0.137)	0.183 (0.133)	- -
Gini Coefficient	0.463 (0.033)	0.420 (0.024)	0.409 (0.023)	- -	0.043 (0.041)	0.054 (0.040)	- -
Theil Inequality Index[££]	0.542 (0.117)	0.408 (0.071)	0.359 (0.066)	0.324 (0.125)	0.135 (0.128)	0.184 (0.124)	0.218 (0.182)
Gini Coefficient[££]	0.463 (0.035)	0.420 (0.024)	0.409 (0.023)	0.017 (0.064)	0.043 (0.042)	0.053 (0.041)	0.446 (0.045)***

*: Significant at 10%; **: Significant at 5%; ***: Significant at 1%.

£: Sample excludes zero earnings (3 observations)

££: Sample excludes non-positive counterfactual outcome variable (1 observation)

Bootstrapped standard errors in parenthesis.