JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. (2015)
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.2448

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

SERGIO FIRPO* AND CRISTINE PINTO
Sao Paulo School of Economics, Fundação Getulio Vargas, Brazil

SUMMARY

This paper presents estimators of distributional impacts of interventions when selection to the program is based on observable characteristics. Distributional impacts are calculated as differences in inequality measures of the marginal distributions of potential outcomes of receiving and not receiving the treatment. The estimation procedure involves a first non-parametric estimation of the propensity score. In the second step weighted versions of inequality measures are computed using weights based on the estimated propensity score. Consistency, semi-parametric efficiency and validity of inference based on the percentile bootstrap are shown for the estimators. Results from Monte Carlo exercises show its good performance in small samples. Copyright © 2015 John Wiley & Sons, Ltd.

Received 15 February 2012; Revised 30 December 2014

All supporting information may be found in the online version of this article.

1. INTRODUCTION

When evaluating social programs, a policymaker may wish to determine not only their average impacts but also their distributional effects. For example, it is reasonable to assume that the policymaker is interested in the effect of the intervention 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, such as those belonging to the generalized entropy class.[1]

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

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

The second branch, to which this paper contributes, defines the 'distributional impacts of a treatment' as the treatment's impact on distributions. In that case, one is interested in determining how a program changes the distribution of the outcome under two scenarios: with and without the program.

---

* Correspondence to: Sergio Firpo, EESP-FGV, Rua Itapeva 474/1215, Sao Paulo, SP, 01332-000, Brazil. E-mail: sergio.firpo@fgv.br

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

[2] Contributions to this branch of the literature include the papers by Heckman (1992), Heckman *et al.* (1997b), Heckman and Smith (1998), Carneiro *et al.* (2001, 2003), Cunha *et al.* (2006), Aakvik *et al.* (2005), Firpo and Ridder (2008) and Fan and Park (2010).

Copyright © 2015 John Wiley & Sons, Ltd.

Those are potential outcomes because, in general, we do not observe the entire population receiving the treatment or not receiving it. To evaluate the impact on the overall distribution, one may examine either the entire cumulative distribution function (CDF) or at all quantiles, as done by Imbens and Rubin (1997), Abadie (2002), Abadie *et al.* (2002), Firpo (2007), Frölich and Melly (2013) and Donald and Hsu (2014).

We discuss the identification of the inequality treatment effects parameters under the assumption termed *treatment unconfoundedness*, which is also known as the *selection on observables* assumption.[3] The unconfoundedness assumption is a conditional independence assumption: given observable characteristics, the decision to be treated is independent of the potential outcomes of being treated or not being treated. This assumption is crucial because it allows potential outcome distributions to be identified from the observed data as weighted distributions of the outcome. Because we are interested in specific functionals of potential outcome distributions, those are also identified from the observed data.

We propose a two-step estimation procedure. In the first step, weighting functions are non-parametrically estimated; in the second step, inequality measures are calculated using the weighted data. The effect of the program is thus estimated as a simple difference in weighted inequality measures. Our estimator therefore belongs to the class of weighted or, more precisely, inverse probability weighted (IPW) estimators, because the weights will be inversely proportional to the propensity score, i.e. the probability of being treated given the covariates.

IPW estimators have been widely used in the missing data and treatment effects literature. Leading examples using the IPW estimator are Robins and Rotnitzky (1995) and Wooldridge (2007) in the missing data literature and Hirano *et al.* (2003) in the treatment effects literature. Recently, Tarozzi (2007), Chen *et al.* (2008) and Cattaneo (2010) have demonstrated how to generalize the identification of treatment effects and semi-parametric estimation under unconfoundedness for a class of parameters that satisfies certain moment conditions. All of these papers have presented IPW estimators and used them in the context of M-estimation because the parameters of interest solve some moment condition.

Our approach contributes to the treatment effects literature by considering parameters that can be written more generally as differences in the functionals of the CDF of the potential outcomes. The identification conditions ensure that those parameters can be represented as differences in the functionals of the weighted marginal distribution of the outcome.

Semi-parametric reweighting methods applied to M-estimators are two-step methods: in the first step the propensity score is estimated in a non-parametric manner and used as an argument of the weighting function; in the second step, one computes weighted averages. In our case, we have a semi-parametric reweighting method applied to plug-in estimators. The difference between the two approaches therefore lies in the second step. We compute functionals of the weighted empirical distributions, which in our case correspond to empirically weighted inequality measures. The generalization in the class of parameters and the use of plug-in estimators are important differences between our method and those presented by Chen *et al.* (2008) and Cattaneo (2010).

Despite the differences, our inference theory is closely related to that presented in Chen *et al.* (2008) for IPW estimators. This is because making inferences from plug-in estimators typically requires some degree of smoothness in the statistical functional. If the functional admits some linearization, its lead approximation argument will asymptotically play the same role as that of a moment condition. Clearly, one must deal with the remaining terms of this linearization, and we demonstrate that these terms do not contribute to the asymptotic behavior of our estimator.

In our paper, we consider the class of Hadamard differentiable functionals of the CDF. That class encompasses numerous interesting inequality measures, which, despite being highly nonlinear

[3] Important examples of contributions adopting this assumption are, among others, Rubin (1977), Rosenbaum and Rubin (1983), Heckman *et al.* (1998), Dehejia and Wahba (1999) and Hirano *et al.* (2003).

Copyright © 2015 John Wiley & Sons, Ltd.

functionals of the distribution, may admit a linear functional derivative. We demonstrate that four of the most popular and relevant inequality measures belong to that class: the coefficient of variation, the interquartile range, the Theil index and the Gini coefficient. Because of the relevance of these inequality indices to the literature, we apply our proposed method to each one of them.

Donald and Hsu (2014) apply a similar semi-parametric reweighting argument to estimate the CDF of potential outcomes in a binary treatment approach. The CDF and its inverse, the quantile process, are either their final objects of interest or necessary intermediate arguments to construct feasible Kolmogorov–Smirnoff test statistics for differences in distributions. In their case, one needs to estimate the CDFs. In this paper, however, we directly estimate the four selected inequality measures using their weighted empirical counterparts. The difference lies in the implementation: our estimators of inequality treatment effects are differences in empirically weighted inequality measures. Thus, unlike Donald and Hsu (2014), we do not need to estimate weighted empirical distributions at each point in the support of the potential outcomes, but solely at the observed values of the actual outcome. The computational simplicity of our approach possibly makes it attractive to the applied researcher, because all it requires is a first-step estimation of weights and a second-step calculation of inequality measures using these weights.

The estimation method proposed is also related to the literature of decomposition methods. As noted by Fortin *et al.* (2011), potential outcome distributions do fall into the class of 'counterfactual' distributions. Moreover, our method can serve the goal of comparing inequality measures while controlling for the distribution of covariates (observables). Specifically, our method can be used to compare the actual inequality measure with a counterfactual one, based on the hypothetical outcome distribution that would leave the covariate distribution unchanged. Applied researchers are often interested in comparing the features of two or more outcome distributions while maintaining the distribution of observables fixed. For example, one might be interested in comparing the Gini coefficient computed for two different wage distributions, as would be the case in a comparison between two different countries. Acknowledging that there are many observed factors that differ across countries, such as schooling and job experience, leads us to attempt to control for these factors when comparing Gini coefficients. By doing so, we would be able to identify how differences in the distribution of covariates explain differences in Gini coefficients between the countries.

Some papers in the literature of decomposition methods employ reweighting and/or imputation methods to recover 'counterfactual' distributions. DiNardo *et al.* (1996) analyze the changes in wage densities over time, while controlling for covariates through a weighting scheme. Juhn *et al.* (1993) construct counterfactual distributions using fitted values and residuals from linear regressions, which is a parametric procedure falling within the class of imputation methods for the counterfactual distribution. These regression/imputation methods have been generalized in numerous ways, and recent contributions on the subject include the papers by Gosling *et al.* (2000), Donald *et al.* (2000) and Machado and Mata (2005).

Recently, Chernozhukov *et al.* (2013) and Rothe (2010) extended the analysis based on counterfactuals to situations in which one may be interested in determining the features of the entire marginal distribution of outcomes using a completely new distribution of covariates, namely the one that could prevail after a policy intervention that exclusively affects the distribution of covariates. Chernozhukov *et al.* (2013) estimate the effects of a policy intervention on the marginal distribution of the variable of interest. Their approach is semi-parametric: in a first stage they non-parametrically estimate the conditional CDF of the outcome, treating the covariates as given. In the second stage, using the new distribution of covariates, they construct marginal counterfactual distributions and recover features from those distributions. Rothe (2010) proposes a non-parametric approach to estimate the effect of a change in the distribution of a covariate on the unconditional distribution of the variable of interest.

Those approaches differ from ours in two important respects. First, we are estimating the propensity score, not the conditional CDF in the first step. A second important difference is that we do not need

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

to calculate the counterfactual CDF, as mentioned above in the comparison of our paper to Donald and Hsu (2014). We can directly compute weighted inequality measures using the weights that will be inversely proportional to the estimated propensity score.

This paper has seven sections besides this introduction. In Section 2 we formally present the ITE class of parameters. Section 3 presents the main identification result. Section 4 discusses estimation and Section 5 derives the large-sample properties of the inequality treatment effects estimators; that is, we demonstrate that under the unconfoundedness assumption and mild regularity conditions our estimators are consistent, asymptotically normal and semi-parametrically efficient. In that section, we also provide a consistent estimator of the asymptotic variance and further demonstrate that inference based on the percentile bootstrap is a valid procedure. Section 6 discusses finite-sample behavior using a Monte Carlo exercise. In Section 7 we present a brief empirical exercise using data on a Brazilian job training program from the late 1990s. Although the training program had been designed to be as a randomized experiment, randomization was performed at the strata (class) level with different proportions of treated units across strata. Thus controlling for strata is crucial in obtaining consistent estimates of the program impact. Finally, Section 8 concludes.[4]

We believe it relevant to emphasize that, in both Sections 6 and 7, we compare our estimation procedure with three other methods: a naive procedure, which computes simple differences in inequality measures with no attempt to control for selection; a regression-based method inspired by Juhn *et al.* (1993); and a method based on non-parametric estimation of the conditional distribution of the outcome proposed by Chernozhukov *et al.* (2013). Evidence from the Monte Carlo exercises reveals that although we may have a much less cumbersome estimation procedure, the relative costs of our method—when they exist (in terms of bias, variance and coverage rate)—seem to be negligible, even at small sample sizes.

2. INEQUALITY TREATMENT EFFECT PARAMETERS

Suppose that a random sample of N individuals (units) is available. For each unit i, let X_i be a random vector of observed covariates with support $\mathcal{X} \subset \mathbb{R}^r$. Define $Y_i(1)$ as the potential outcome for individual i if she enters the program and $Y_i(0)$ as 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 zero otherwise. As we only observe each unit in one treatment status, we say that the unobserved outcome is the counterfactual outcome. Thus the observed outcome can be expressed as

$$Y_i = T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0), \quad \forall i \tag{1}$$

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:

$$W(F) = \Omega(\mu(F), \quad \nu(F)) \tag{2}$$

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

[4] We have two supplementary appendices to this paper, provided as supporting information. The first one, entitled 'Supplement to "Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures", Part I: Proofs, derivations and some additional technical material', provides an extensive discussion of all theoretical results presented in this paper. The second one, whose title is 'Supplement to "Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures", Part II: Monte Carlo experiments and the empirical application', presents additional material that due to space constraints was not included in the paper.

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

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

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

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

$$\Delta_{\mathrm{O}}^\nu = \nu\left(F_{Y(1)}\right) - \nu\left(F_{Y(0)}\right) = \nu_1 - \nu_0 \tag{3}$$

Other parameters could be defined for subpopulations. In particular, consider the *inequality treatment effect on the treated* (ITT), Δ_{T}^ν:

$$\Delta_{\mathrm{TT}}^\nu = \nu\left(F_{Y(1)|T=1}\right) - \nu\left(F_{Y(0)|T=1}\right) \tag{4}$$

$$= \nu_{11} - \nu_{01} \tag{5}$$

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

We finally consider a parameter that serves to compare the current inequality $\nu(F_Y)$ with the one that we would encounter if there were no program $\nu\left(F_{Y(0)}\right)$. We call this parameter the *current inequality treatment effect* (CIT): [7]

$$\Delta_{\mathrm{CIT}}^\nu = \nu(F_Y) - \nu\left(F_{Y(0)}\right) \tag{6}$$

$$= \nu_Y - \nu_0 \tag{7}$$

3. IDENTIFICATION OF INEQUALITY TREATMENT EFFECTS

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

3.1. The Set-Up

We now establish assumptions for the identification of Δ^ν. Recall that because $Y(1)$ and $Y(0)$ are never fully observable, we need to impose identifying assumptions to be able to express functionals of their marginal distributions as functionals of the joint distribution of observable variables (Y, T, X).

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

[7] If ν is not decomposable, we cannot write the CIT as a 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, such as the difference in inequality measures between treated and control subpopulations that where formed following a rule that is a function of pretreatment covariates X.

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Let the data be defined by the sequence $\{Y_i, T_i, X_i\}_{i=1}^N$ where each element (Y_i, T_i, X_i) is a random draw from $F_{Y,T,X}$, the joint distribution of $(Y, T, X) \in \mathcal{Y} \times \{0, 1\} \times \mathcal{X}$, where $\mathcal{Y} \subset \mathbb{R}$.

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

We begin by writing the weighted marginal distribution of Y, which is a key tool in our identification strategy. The weighted marginal distribution of Y at y that uses proper functions of T and X as weights is

$$F_Y^\omega(y) = E\left[\omega\left(T, p(X)\right) \cdot \mathbb{I}\{Y \leq y\}\right] \tag{8}$$

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

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

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

$$\omega_1\left(t, p\left(x\right)\right) = t / p\left(x\right) \tag{9}$$

$$\omega_0\left(t, p\left(x\right)\right) = \left(1 - t\right) / \left(1 - p\left(x\right)\right) \tag{10}$$

$$\omega_{11}\left(t, p\left(x\right)\right) = t / p \tag{11}$$

$$\omega_{01}\left(t, p\left(x\right)\right) = \left(\left(1 - t\right) / \left(1 - p\left(x\right)\right)\right) \cdot \left(p\left(x\right) / p\right) \tag{12}$$

These weighting functions will be used to identify the marginal, and those conditional on $T = 1$, CDFs of the distributions of $Y(1)$ and $Y(0)$.

3.2. Identification

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

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

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

Assumption 1 is occasionally a strong assumption, and its plausibility must be analyzed on a case-by-case basis. This assumption has been used, however, in several studies on the effect of treatments or programs. Prominent examples are Rosenbaum and Rubin (1983), Heckman and Robb (1986), LaLonde (1986), Card and Sullivan (1988), Heckman *et al.* (1997a, 1998), Hahn (1998), Lechner (1999), Dehejia and Wahba (1999) and Becker and Ichino (2002). In the empirical section, we present an example for which, by design, Assumption 1 is valid.

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

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Assumption 2. (common support) For all x in $\mathcal{X}$, there is a real number c, such that $0 < c \leq p(x) \leq 1 - c < 1$.

Assumption 2 states that with probability one there will be no particular value x in $\mathcal{X}$ that belongs to either the treated group or the control group. This assumption is important because it allows the groups ($T = 1$ and $T = 0$) to be fully comparable in terms of X.[8] Assumptions 1 and 2 are jointly termed *strong unconfoundedness*.

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

Lemma 1. (identification) Under Assumptions 1 and 2 $\Delta_{\mathrm{O}}^{\nu}$, $\Delta_{\mathrm{TT}}^{\nu}$, and $\Delta_{\mathrm{CIT}}^{\nu}$ are identifiable from observable data (Y, T, X).

The proof of this lemma is based on the fact that, under Assumptions 1 and 2, we can write the marginal distribution function of $Y(1)$ and $Y(0)$ as a weighted CDF of Y. Having identified the marginal (and conditional on $T = 1$) distributions of potential outcomes, the functionals of these distributions will also be identified. Thus our inequality treatment effects are identifiable from data on (Y, X, T).

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

3.3. Some Inequality Measures

We now turn our attention to concrete examples of inequality measures and express them as functionals of a weighted distribution of Y.

Comparisons of inequality measures are often performed based on 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 determine which of the usual properties an inequality measure possesses. Among those properties, the most common and important are the *principle of transfers*, *invariance*, *decomposability* and *anonymity*. For a detailed discussion of this topic, see Cowell (2000, 2003).[9]

We consider four popular inequality measures: the coefficient of variation, the interquartile range, the Theil index and the Gini coefficient. As discussed by Cowell (2000), the coefficient of variation will satisfy all properties listed above, except invariance. The interquartile range will not satisfy any of those properties other than anonymity. The Theil index, being a member of the generalized entropy class, will satisfy all four properties, whereas the Gini coefficient, likely the most commonly used inequality measure, is known to be non-decomposable.

We proceed treating those four measures as functionals of a weighted outcome distribution. By doing so, we gain the flexibility necessary to further define the treatment effects as differences in the functionals of weighted distributions:[10]

[8] The assumption that the propensity score is bounded away from 0 and 1 is also employed subsequently to obtain the typical convergence results for our first-stage nonparametric estimator.

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

[10] In what follows we assume that $\mu_Y^{\omega} \equiv \int y \cdot \mathrm{d}F_Y^{\omega}(y) \neq 0$.

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

1. *Coefficient of variation* (CV):

$$\nu^{\mathrm{CV}}\left(F_Y^\omega\right) = \frac{\left(\int \left(y - \int z \cdot \mathrm{d}F_Y^\omega(z)\right)^2 \cdot \mathrm{d}F_Y^\omega(y)\right)^{1/2}}{\int y \cdot \mathrm{d}F_Y^\omega(y)} \tag{13}$$

2. *Interquartile range* (IQR):

$$\nu^{\mathrm{IQR}}\left(F_Y^\omega\right) = \nu^{Q.75}\left(F_Y^\omega\right) - \nu^{Q.25}\left(F_Y^\omega\right) \tag{14}$$

$$= \inf_q \left\{\int_{-\infty}^q \mathrm{d}F_Y^\omega(y) \geq \frac{3}{4}\right\} - \inf_q \left\{\int_{-\infty}^q \mathrm{d}F_Y^\omega(y) \geq \frac{1}{4}\right\} \tag{15}$$

3. *Theil index* (TI):[11]

$$\nu^{\mathrm{TI}}\left(F_Y^\omega\right) = \frac{\int y \cdot \left(\log(y) - \log\left(\int z \cdot \mathrm{d}F_Y^\omega(z)\right)\right) \cdot \mathrm{d}F_Y^\omega(y)}{\int y \cdot \mathrm{d}F_Y^\omega(y)} \tag{16}$$

4. *Gini coefficient* (GC):

$$\nu^{\mathrm{GC}}\left(F_Y^\omega\right) = 1 - 2\frac{\int_0^1 \int^{\nu^{Q\tau}\left(F_Y^\omega\right)} y \cdot \mathrm{d}F_Y^\omega(y) \cdot \mathrm{d}\tau}{\int y \cdot \mathrm{d}F_Y^\omega(y)} \tag{17}$$

4. ESTIMATION

We first demonstrate how to estimate $\nu\left(F_Y^\omega\right)$ with a general ω, and subsequently explain how to use these results to estimate Δ_{O}^ν, Δ_{TT}^ν and $\Delta_{\mathrm{CIT}}^\nu$.

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

$$\widehat{\nu}_Y^{\widehat{\omega}} = \nu\left(\widehat{F}_Y^{\widehat{\omega}}\right) \tag{18}$$

Note that we exploit that the weighted CDF is expressed as $F_Y^\omega(y) = E\left[\omega(T, p(X)) \cdot \mathbb{1}\{Y \leq y\}\right]$, and we write its sample analog as

$$\widehat{F}_Y^{\widehat{\omega}}(y) = N^{-1} \sum_{i=1}^N \omega\left(T, \widehat{p}(X_i)\right) \cdot \mathbb{1}\{Y_i \leq y\} \tag{19}$$

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

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

It is clear that to arrive at a feasible estimation procedure we must first address the estimation of the weighting function, $\omega = \omega\left(T, p\left(X\right)\right)$ by $\widehat{\omega} = \omega\left(T, \widehat{p}\left(X\right)\right)$

4.1. Weights Estimation

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

In the propensity score estimation, we do not impose any parametric assumption regarding the conditional distribution of T given X or assume that the propensity score has a given functional form. In this paper we adopt the non-parametric logistic approach (Hastie and Tibshirani, 1990), which has also been used by Hirano *et al.* (2003). They approximate the log odds ratio of the propensity score, $L\left(p\left(x\right)\right)$, using a series of polynomial functions of x.[12] Stacking all of these polynomials in a vector, we obtain $H_K(x) = [H_{K,j}(x)] \quad (j = 1, \ldots, K)$, a vector of length K of polynomial functions of $x \in \mathcal{X}$. The estimation procedure therefore involves computing the length K vector of coefficients $\hat{\pi}_K$:

$$L\left(\widehat{p}\left(x\right)\right) = H_K\left(x\right)' \widehat{\pi}_K \tag{20}$$

$$\widehat{p}\left(x\right) = L^{-1}\left(H_K\left(x\right)' \widehat{\pi}_K\right) = \Lambda\left(H_K\left(x\right)' \widehat{\pi}_K\right) \tag{21}$$

where $\Lambda : \mathbb{R} \to \mathbb{R},\ \Lambda(z) = (1 + \exp(-z))^{-1}$ is the CDF of a logistic distribution evaluated at z. The non-parametric aspect of this procedure is a result of 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:

$$\hat{\pi}_K = \underset{\pi_K \in \mathbb{R}^K}{\arg\max} \sum_{i=1}^{N} \left(T_i \cdot \log(\Lambda(H_K(X_i)'\pi_K)) + (1 - T_i) \cdot \log(1 - \Lambda(H_K(X_i)'\pi_K))\right) \tag{22}$$

In implementing this procedure, following Hirano *et al.* (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 to order n.[13] In addition to unconfoundedness and overlap, we need to control how the size of the polynomial H_K increases with N. The rate at which additional terms are added to the polynomial depends on the degree of smoothness of the propensity score and the dimension of X.

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

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

Assumptions 3 and 4 are invoked to guarantee that the non-parametric series estimator for propensity score has a negligible bias and converges sufficiently fast uniformly in probability to the true

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

[13] Further details regarding the choice of $H_K(x)$ and its asymptotic properties can be found in supporting information to this paper and in the work of Hirano *et al.* (2003).

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

propensity score. Those assumptions might be weakened if one uses other basis functions such as B-splines. See, for example, Newey (1997).

After we have estimated the propensity score, the estimated weights are $\widehat{\omega}_i = \omega\left(T_i, \widehat{p}\left(X_i\right)\right)$.[14]

4.2. Estimation of Inequality Treatment Effects

Once the weights have been computed, the three ITE parameters are easily estimated by using the plug-in method. We write the corresponding estimators of $\Delta_{\mathrm{O}}^{\nu}$, $\Delta_{\mathrm{TT}}^{\nu}$, and $\Delta_{\mathrm{CIT}}^{\nu}$ as

$$\widehat{\Delta}_{\mathrm{O}}^{\nu} = \widehat{\nu}_1 - \widehat{\nu}_0 = \nu\left(\widehat{F}_Y^{\widehat{\omega}_1}\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_0}\right) \tag{23}$$

$$\widehat{\Delta}_{\mathrm{TT}}^{\nu} = \widehat{\nu}_{11} - \widehat{\nu}_{01} = \nu\left(\widehat{F}_Y^{\widehat{\omega}_{11}}\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_{01}}\right) \tag{24}$$

$$\widehat{\Delta}_{\mathrm{CIT}}^{\nu} = \widehat{\nu}_Y - \widehat{\nu}_0 = \nu\left(\widehat{F}_Y\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_0}\right) \tag{25}$$

As an illustration, we present the estimators of three inequality treatment effect parameters: (i) the coefficient of variation CIT; (ii) the Theil index O; and (iii) the Gini coefficient ITT.[15]

Example 1. Coefficient of variation CIT

$$\widehat{\Delta}_{\mathrm{CIT}}^{\mathrm{CV}} = \frac{\left(\frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \sum_{i=1}^{N}\frac{Y_i}{N}\right)^2\right)^{1/2}}{\sum_{i=1}^{N}\frac{Y_i}{N}} - \frac{\left(\sum_{i=1}^{N}\widehat{\omega}_{0i}\left(Y_i - \sum_{i=1}^{N}\widehat{\omega}_{0i}\cdot Y_i\right)^2\right)^{1/2}}{\sum_{i=1}^{N}\widehat{\omega}_{0i}\cdot Y_i} \tag{26}$$

Example 2. Theil index O

$$\begin{aligned}\widehat{\Delta}_{\mathrm{O}}^{\mathrm{TI}} = {} & \frac{\sum_{i=1}^{N}\widehat{\omega}_{1i}\cdot Y_i\cdot\left(\log\left(Y_i\right) - \log\left(\sum_{i=1}^{N}\widehat{\omega}_{1i}\cdot Y_i\right)\right)}{\sum_{i=1}^{N}\widehat{\omega}_{1i}\cdot Y_i} \\ & - \frac{\sum_{i=1}^{N}\widehat{\omega}_{0i}\cdot Y_i\cdot\left(\log\left(Y_i\right) - \log\left(\sum_{i=1}^{N}\widehat{\omega}_{0i}\cdot Y_i\right)\right)}{\sum_{i=1}^{N}\widehat{\omega}_{0i}\cdot Y_i}\end{aligned} \tag{27}$$

[14] Note that under Assumptions 2, 3 and 4 the weighted functions $\omega\left(T_i, p\left(X_i\right)\right)$ are measurable, bounded, differentiable functions with bounded derivatives. This property of the weighted functions is important in deriving the large-sample properties of our estimators.

[15] The implementation of IQR follows from Firpo (2007) and is suppressed here for the sake of brevity.

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Example 3. Gini coefficient ITT

$$\widehat{\Delta}_{\mathrm{ITT}}^{\mathrm{GC}} = \frac{\sum_{i=1}^{N} \widehat{\omega}_{11i} \cdot \left(\sum_{j=1}^{N} \widehat{\omega}_{11j} \left| Y_i - Y_j \right| \right)}{2 \cdot \sum_{i=1}^{N} \widehat{\omega}_{11i} \cdot Y_i} - \frac{\sum_{i=1}^{N} \widehat{\omega}_{01i} \cdot \left(\sum_{j=1}^{N} \widehat{\omega}_{01j} \left| Y_i - Y_j \right| \right)}{2 \cdot \sum_{i=1}^{N} \widehat{\omega}_{01i} \cdot Y_i} \tag{28}$$

5. LARGE-SAMPLE INFERENCE

In this section, we derive the asymptotic distribution of the inequality treatment effect parameters based on inequality measures that are Hadamard differentiable functionals of the distribution of potential outcomes. Although we use four inequality measures as concrete examples, our analysis is more general and could be extended to other functionals of the distribution that satisfy the same differentiability property, such as inequality measures that belong to the generalized entropy class. Numerous other estimands and hypothesis tests of interest could be considered beyond the inequality measures studied here.[16]

We then use results from the semi-parametric efficiency literature and treatment effects literature (e.g. Hahn 1998; Hirano *et al.*, 2003; Cattaneo, 2010) to demonstrate the efficiency of our estimators.[17] Finally, we provide a consistent estimator of the asymptotic variance of the weighted estimators considered in this paper and demonstrate that we could use the percentile bootstrap approach to do inference.

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

Assumption 5. (smoothness)

(i) The support $\mathcal{X}$ of X is a compact subset of $\mathbb{R}^r$.
(ii) The density of X is bounded and bounded away from 0 on $\mathcal{X}$.
(iii) $Y(1), Y(0)$ are distributed as F_1 and F_0,respectively, which are defined over a common support $\mathcal{Y}$.
(iv) F_0 and F_1 are continuously differentiable functions on $\mathcal{Y}$ with $F_1(0) = F_0(0) = 0$.
(v) For any given x in $\mathcal{X}$, $F_0(y|x)$ and $F_1(y|x)$ are continuous in $y \in \mathcal{Y}$.
(vi) For any given $y \in \mathcal{Y}$, $F_0(y|x)$ and $F_1(y|x)$ are s_f times continuously differentiable on $x \in \mathcal{X}$, with $\frac{s_f}{r} \geq 8$.

Under Assumption 5, the density of X is bounded in its entire support, and all covariates are continuous. [18] In addition, $Y(1), Y(0)$ have uniformly continuously differentiable CDFs.

In addition to Assumption 5, we also impose a smoothness condition that enables us to derive the asymptotic normality of our estimators. We restrict the discussion to the class of inequality measures that are Hadamard differentiable functionals of the distribution. This smoothness condition is

[16] For examples of some estimands and hypotheses tests that could involve, for instance the quantile process, first- and second-order stochastic dominance and Kolmogorov–Smirnov tests, see Abadie (2002), Chernozhukov *et al.* (2013) and Donald and Hsu (2014).
[17] Similar efficiency results can be found in the missing data literature. Robins *et al.* (1994), Robins and Rotnitzky (1995) and Rotnitzky and Robins (1995) provide calculations of the semi-parametric efficiency bounds for nonlinear regression models.
[18] As in Hirano *et al.* (2003), we can generalize the discussion to the case of both continuous and discrete covariates. As in their case, we can use the estimator for the continuous covariates case in samples that are homogeneous in discrete covariates.

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

important in establishing the following results: (i) asymptotic normality; (ii) efficiency; and (iii) the validity of the percentile bootstrap for our estimators.

Assumption 6. (Hadamard) The inequality measure $\nu : \mathcal{F}_\nu \to \mathbb{R}$ defined over the marginal distribution of potential outcomes is Hadamard differentiable at F_Y^ω tangentially to the subset $\mathcal{F}_\nu^\omega$.

Under some mild additional conditions on the CDF, Assumption 6 can be verified to hold for all four inequality measures considered in Section 3.3. The coefficient of variation and the Theil index are known functions of expectations, which are already linear functionals, and therefore satisfy Assumption 6 by definition. The interquartile range is a known function of quantiles, which are Hadamard differentiable if the F is continuously differentiable with a positive derivative f at both quartiles.[19] Finally, as shown by Bhattacharya (2007, Proposition 2), if, in addition to the continuous differentiability of the CDF and the existence of positive density, we impose a tail restriction that prevents the density from going to 0 too slowly at the tails, then the Gini coefficient is also Hadamard differentiable.[20]

We are now able to obtain the limiting distribution of our estimators. To do so, we first establish uniform root-N consistency for $\widehat{F}_Y^{\omega=\widehat{\omega}}$ and present a uniform asymptotically linear representation for that estimator. Having these results, we can then apply the functional delta method to obtain the limiting distribution of $\widehat{\nu}$.

Lemma 2. Under Assumptions 1–5:

$$\sup_{y \in \mathcal{Y}} \left| \sqrt{N} \left(\widehat{F}_Y^{\widehat{\omega}}(y) - F_Y^\omega(y) \right) - \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\psi(Y_i, X_i, T_i, y) - F_Y^\omega(y) \right) \right| = o_p(1) \qquad (29)$$

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

The proof of Lemma 2 follows from a modification of the decomposition employed in Theorem 1 in Hirano *et al.* to consider a weighted indicator function instead of conditional expectations. Lemma 2 provides a general result that can be applied to the four weighted marginal distribution functions considered in this paper. One can ensure that the derivatives of the four weighting functions with respect to $p(x)$ are

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

Let us consider two weighting functions ω_A and ω_B of the four presented this far. One consequence of the linear representation of the weighted distribution function is the following result. In what follows, we use the symbol $\Rightarrow$ to denote weak convergence.

[19] To be precise and following van der Vaart (1998, chapters 20 and 21), the τth quantile of the CDF is Hadamard differentiable *tangentially* to the subset of CDFs for which ψ^ν is well defined. Moreover, as discussed by van der Vaart (1998, Theorem 20.8) one only needs tangential Hadamard differentiability (and not necessarily Hadamard differentiability) for functional delta method applications.

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

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Theorem 1. Suppose Assumptions 1–5 hold, then

$$\sqrt{N}\begin{pmatrix} \widehat{F}_Y^{\widehat{\omega}_A}(y) - F_Y^{\omega_A}(y) \\ \widehat{F}_Y^{\widehat{\omega}_B}(y) - F_Y^{\omega_B}(y) \end{pmatrix} = \mathbb{G}_N^{\omega_{A,B}} + o_p(1) \Rightarrow \mathbb{G}^{A,B} \tag{30}$$

where

(i) $\mathbb{G}_N^{\omega_{A,B}} = \begin{bmatrix} \mathbb{G}_N^{\omega_A} \\ \mathbb{G}_N^{\omega_B} \end{bmatrix}$, $\mathbb{G}_N^{\omega_j}$ is an empirical process such that at a given $y \in \mathcal{Y}$, for $j = A, B$

$$\mathbb{G}_N^{\omega_j}(y) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\psi_j(Y_i, X_i, T_i, y) - F_Y^{\omega_j}(y)\right) \tag{31}$$

(ii) $\mathbb{G}^{A,B}$ is a Gaussian process with variance–covariance matrix given by

$$\mathbb{E}\left[\mathbb{G}^{\omega_j}(s)\,\mathbb{G}^{\omega_k}(t)\right] = \mathbb{E}\left[\left(\psi_j(Y_i, X_i, T_i, s) - F_Y^{\omega_j}(s)\right) \cdot \left(\psi_k(Y_i, X_i, T_i, t) - F_Y^{\omega_k}(t)\right)\right] \tag{32}$$

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

(iii) $\psi_j(Y_i, X_i, T_i, s) = \omega_j(T_i, p(X_i)) \cdot \mathbb{I}\{Y \le y\} + \mathbb{E}\left[\frac{\partial \omega_j(T, p(X))}{\partial p(X)} \cdot \mathbb{I}\{Y \le y\} \middle| X\right](T - p(X))$, for $j = A, B$.

The proof of this theorem follows from establishing that the class of measurable functions from $(\mathcal{Y} \times \{0,1\} \times \mathcal{X}) \to \mathbb{R}$, $\mathcal{H} = \{\psi(Y_i, X_i, T_i, y) | \, y \in \mathcal{Y}\}$ is P-Donsker, and then by applying the Donsker's theorem (Van der Vaart, 1998, p. 266).

Under the assumptions employed in Theorem 1, we can demonstrate the maximal asymptotic precision with which we can estimate $F_Y^{\omega}(y)$ is given by

$$\mathbb{E}\left[\left(\omega(T_i, p(X_i)) \cdot \mathbb{I}\{Y \le y\} + \mathbb{E}\left[\frac{\partial \omega(T, p(X))}{\partial p(X)} \cdot \mathbb{I}\{Y \le y\} \middle| X\right](T - p(X)) - F_Y^{\omega}(y)\right)^2\right] \tag{33}$$

The characterization of this bound follows the work of Newey (1990) and Bickel *et al.* (1993).[21]

Once we have established joint uniform convergence for two empirical weighted marginal distributions, we can establish that the estimators $\nu\left(\widehat{F}_Y^{\widehat{\omega}_A}\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_B}\right)$ is asymptotically normal. Moreover, because the estimators for the inequality treatment effect parameters are estimators of the difference in inequality measures of weighted distributions, we can demonstrate that our estimators will be semi-parametrically efficient. These results are direct consequences of the functional delta method (Theorem 20.6 of Van der Vaart, 1998).[22]

---

[21] A detailed proof of this result can be found in the supporting information to this paper.

[22] In what follows. we use $\rightsquigarrow$ to indicate convergence in law.

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Theorem 2. Under Assumptions 1–6:

$$\begin{aligned}&\sqrt{N}\left(\left(\nu\left(\widehat{F_Y^{\omega_A}}\right)-\nu\left(\widehat{F_Y^{\omega_B}}\right)\right)-\left(\nu\left(F_Y^{\omega_A}\right)-\nu\left(F_Y^{\omega_B}\right)\right)\right)\\&\rightsquigarrow \psi^{\nu}\left(\mathbb{G}_N^{\omega_A};F_Y^{\omega_A}\right)-\psi^{\nu}\left(\mathbb{G}_N^{\omega_B};F_Y^{\omega_B}\right)\end{aligned} \tag{34}$$

and for $j = A, B$:

$$\begin{aligned}\psi^{\nu}\left(\mathbb{G}_N^{\omega_j};F_Y^{\omega_j}\right)=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\omega_j\left(T_i,p\left(X_i\right)\right)\cdot\phi^{\nu}\left(Y_i;F_Y^{\omega_j}\right)\\&+\mathbb{E}\left[\frac{\partial\omega_j\left(T_i,p\left(X_i\right)\right)}{\partial p\left(X_i\right)}\cdot\phi^{\nu}\left(Y_i;F_Y^{\omega_j}\right)\right]\left(T_i-p\left(X_i\right)\right)\end{aligned} \tag{35}$$

where $\phi^{\nu}\left(Y_i;F_Y^{\omega_j}\right)$ is the functional $\nu's$ Hadamard derivative, $\phi^{\nu}\left(Y_i;.\right) = \psi^{\nu}\left(_{Y_i};.\right)$ and $_{Y_i}$ is the Dirac measure at observation i. Moreover, $\nu\left(\widehat{F_Y^{\omega_A}}\right)-\nu\left(\widehat{F_Y^{\omega_B}}\right)$ is asymptotically efficient.

We can apply the general result in Theorem 2 to our ITE estimators. For h = O, TT and CIT:

$$\sqrt{N}\left(\widehat{\Delta}_h^{\nu}-\Delta_h^{\nu}\right)\underset{d}{\rightarrow}\mathcal{N}\left(0,V_h\right) \tag{36}$$

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

$$\begin{aligned}V_{\mathrm{O}}=\mathbb{E}\Big[\big(&\omega_1\left(T,p\left(X\right)\right)\cdot\phi^{\nu}\left(Y;F_Y^{\omega_1}\right)-\omega_0\left(T,p\left(X\right)\right)\cdot\phi^{\nu}\left(Y;F_Y^{\omega_0}\right)\\&+\left(g_1\left(X\right)-g_0\left(X\right)\right)\left(T-p\left(X\right)\right)\big)^2\Big]\end{aligned} \tag{37}$$

$$V_{\mathrm{TT}}=\mathbb{E}\left[\left(\omega_{11}\left(T,p\left(X\right)\right)\cdot\phi^{\nu}\left(Y;F_Y^{\omega_{11}}\right)-\omega_{01}\left(T,p\left(X\right)\right)\cdot\phi^{\nu}\left(Y;F_Y^{\omega_0}\right)-g_{01}\left(X\right)\left(T-p\left(X\right)\right)\right)^2\right] \tag{38}$$

$$V_{\mathrm{CIT}}=\mathbb{E}\left[\left(\phi^{\nu}\left(Y;F_Y^{\omega_{11}}\right)-\omega_0\left(T,p\left(X\right)\right)\cdot\phi^{\nu}\left(Y;F_Y^{\omega_0}\right)-g_0\left(X\right)\left(T-p\left(X\right)\right)\right)^2\right] \tag{39}$$

where for $j = 0, 1$:

$$g_j\left(x\right)\equiv\mathbb{E}\left[\frac{\partial\omega_j\left(T,p\left(X\right)\right)}{\partial p\left(X\right)}\cdot\phi^{\nu}\left(Y;F_Y^{\omega_j}\right)\middle|X=x\right] \tag{40}$$

Valid inference for the inequality treatment effect can be approached by either estimating the analytical expressions of the variance terms or applying resampling methods, such as the percentile bootstrap. In the next theorem, we present a consistent estimator for the asymptotic variance of the general form of the estimator. Again, we use a general form and then adapt it to the case of interest.

Estimators of the variance based on analytical expressions must be able to estimate the functions $g\left(x\right)$. Note that these functions depend on the propensity score; we therefore use the series

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

estimator proposed by Hirano *et al.* (2003) and replace $p(X)$ with $\widehat{p}(X)$. In addition, for $j = A, B$ we use a non-parametric estimator for $g_j(\cdot)$. We call $\widehat{g}_j(.)$ the non-parametric regression of $\left.\frac{\partial \omega_j(T,p(X))}{\partial p(X)}\right|_{p(X)=\widehat{p}(X)} \cdot \phi^\nu\left(Y; \widehat{F}_Y^{\omega_j}\right)$ on X using series. The series estimator is written as

$$\widehat{g}_j(.) = H_{K_j}(.)^T \,\widehat{\gamma}_K^{\omega_j} \tag{41}$$

where

$$\widehat{\gamma}_K^{\omega_j} = \arg\min_\gamma \sum_{i=1}^N \left(\left.\frac{\partial \omega_j(T, p(X))}{\partial p(X)}\right|_{p(X)=\widehat{p}(X)} \cdot \phi^\nu\left(Y; \widehat{F}_Y^{\omega_j}\right) - H_{Kj}(.)^T \gamma \right)^2 \tag{42}$$

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

$$\sup_{x\in\mathcal{X}} \left\| H_{K_j}(X_i) \right\| = \zeta\left(K_j\right) \le C K_j$$

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

Assumption 7. (series estimator) For $j = A, B$

(i) $g_j(x)$ is bounded and s_j times continuous differentiable.
(ii) $\frac{s_j}{r} \ge 4$.
(iii) The order of $H_{K_j}(x)$, K_j, is of the form $K_j = C \cdot N^{c_j}$, where C is a constant and $c_j \in \left(0, \frac{1}{2}\left(\frac{s_p}{2r} - 1\right) c_p\right)$.
(iv) $\sup_y \left|\frac{\partial \phi^\nu(y;z)}{\partial z}\right|_{z=F_Y^\omega} \le M$.

Parts (i) and (ii) of Assumption 7 establish that the functions $g_j(\cdot)$ are sufficiently smooth for their non-parametric estimators to be accurate. Part (iii) of the same assumption is analogous to Assumption 4, but it is weaker than that and is used to guarantee that the variance of the non-parametric estimators of $g_j(\cdot)$ go to zero at an appropriate rate as the sample size increases. Part (iv) of Assumption 7 is necessary to guarantee that $\phi^\nu\left(\cdot; \widehat{F}_Y^{\omega_j}\right)$ converges uniformly in probability to $\phi^\nu\left(\cdot; F_Y^{\omega_j}\right)$.

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

$$\begin{aligned} V_{AB} = \mathbb{E}\Big[&\big(\omega_A(T, p(X)) \cdot \phi^\nu\left(Y; F_Y^{\omega_A}\right) - \omega_B(T, p(X)) \cdot \phi^\nu\left(Y; F_Y^{\omega_B}\right) \\ &+ \left(g_A(X) - g_B(X)\right)(T - p(X))\big)^2\Big] \end{aligned} \tag{43}$$

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Theorem 3. Under Assumptions 1–7:

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

where

$$\widehat{V}_{AB} = \frac{1}{N}\sum_{i=1}^{N}\Big\{\omega_A\left(T_i, \widehat{p}\left(X_i\right)\right)\cdot\phi^{\nu}\left(Y_i; F_Y^{\widehat{\omega}_A}\right) - \omega_B\left(T_i, \widehat{p}\left(X_i\right)\right)\cdot\phi^{\nu}\left(Y_i; F_Y^{\widehat{\omega}_B}\right) + \left(\widehat{g}_A\left(X_i\right) - \widehat{g}_B\left(X_i\right)\right)\left(T_i - \widehat{p}\left(X_i\right)\right)\Big\}^2 \tag{44}$$

Calculating $\widehat{V}_{AB}$ may be computationally demanding because we need to estimate three series estimators ($p(x)$, $g_A(x)$ and $g_B(x)$). Next, we demonstrate that we can use the percentile bootstrap for inference.

Consider a random sample $Z = \{(Y_i, X_i, T_i) : i = 1, \ldots, N\}$ from $\mathcal{P}$ (probability distribution) on a measurable space $\mathcal{Y}\times\mathcal{X}\times\{0,1\}$, where $\mathcal{Y} \subset \mathbb{R}$ and $\mathcal{X} \subset \mathbb{R}^r$. The estimator $\widehat{\Delta} = \nu\left(\widehat{F}_Y^{\widehat{\omega}_A}\right) - \nu\left(\widehat{F}_Y^{\widehat{\omega}_B}\right)$ is a function of the original sample Z.

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

$$\widehat{F}_{Y_b}^{\widehat{\omega}_{j,b}}(y) = N^{-1}\sum_{i=1}^{N}\omega_j\left(T_i^*, \widehat{p}_b\left(X_i^*\right)\right)\mathbb{I}\{Y_i^* \leq y\} \tag{45}$$

where $\widehat{p}_b(\cdot)$ is the estimator of the propensity-score using the bth replica of the bootstrap and $\widehat{\omega}_{j,b,i} = \omega_j\left(T_i^*, \widehat{p}_b\left(X_i^*\right)\right)$. Finally, the bootstrap estimator of the bth replica is

$$\widehat{\Delta}_{AB,b} = \nu\left(\widehat{F}_{Y_b}^{\widehat{\omega}_{A,b}}\right) - \nu\left(\widehat{F}_{Y_b}^{\widehat{\omega}_{B,b}}\right) \tag{46}$$

One can use the percentile bootstrap to construct confidence intervals for Δ. Let $\widehat{\Delta}_{[\alpha\cdot\mathrm{B}]}$ be the $\alpha\cdot$B order statistic of the bootstrap distribution, that is

$$\alpha = \mathrm{B}^{-1}\sum_{b=1}^{N}\mathbf{1}\left\{\widehat{\Delta}_b \leq \widehat{\Delta}_{[\alpha\cdot\mathrm{B}]}\right\} \tag{47}$$

where $0 < \alpha < 1$. Then, an approximate $(1-\alpha)\cdot 100\%$ confidence interval for Δ is

$$\mathrm{CI}^*\left(\Delta, (1-\alpha)\cdot 100\%\right) = \left(2\widehat{\Delta} - \widehat{\Delta}_{[(1-\alpha/2)\cdot\mathrm{B}]}, 2\widehat{\Delta} - \widehat{\Delta}_{[(\alpha/2)\cdot\mathrm{B}]}\right) \tag{48}$$

Theorem 4. Suppose that Assumptions 1–7 hold. Then for h = O, TT and CIT:

$$\Pr\left[\Delta_h \in CI^*\left(\Delta_h, (1-\alpha)\cdot 100\%\right)\right] \rightarrow_p 1-\alpha \tag{49}$$

The percentile bootstrap is an easier alternative for inference than calculating the analytical standard errors. In the next sections, we present a Monte Carlo exercise that compares the behavior of the percentile bootstrap with inference based on the analytical standard errors.

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

6. A MONTE CARLO EXERCISE

In this section, we report the results of Monte Carlo exercises. We are interested in learning how the estimators for the inequality treatment effect (ITE) behave in small samples. One thousand replications of the experiment with sample sizes of 250 and 1000 observations were considered.

We design the data generation process (DGP) to produce 'selection on observables'; that is, the conditional distribution of X given T will differ from the marginal distribution of X, but the marginal distributions of the potential outcomes will be independent of T given X. Note that because $Y(1)$ and $Y(0)$ are known for each observation i, we can compute 'unfeasible' estimators of functionals of the marginal distributions of $Y(1)$ and $Y(0)$. If we restrict our attention to subpopulations—for example, the treated—we can still compute 'unfeasible' statistics of estimators $Y(1)|T=1$ and $Y(0)|T=1$.

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

$$T=\mathbb{1}\{\delta_0+\delta_1X_1+\delta_2X_2+\delta_3X_1^2+\delta_4X_2^2+\delta_5X_1X_2+\eta>0\} \tag{50}$$

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

The potential outcomes are

$$Y(0)=\exp\left(\beta_{00}+\beta_{01}X_1+\beta_{02}X_2+\beta_{03}X_1^2+\beta_{04}X_2^2+\beta_{05}X_1X_2+\epsilon_0\right) \tag{51}$$

$$Y(1)=\exp\left(\beta_{10}+\beta_{11}X_1+\beta_{12}X_2+\beta_{13}X_1^2+\beta_{14}X_2^2+\beta_{15}X_1X_2+\epsilon_1\right) \tag{52}$$

where

$$\epsilon_0=\left(\beta_{00}^s+\beta_{01}^sX_1+\beta_{02}^sX_2+\beta_{03}^sX_1^2+\beta_{04}^sX_2^2+\beta_{05}^sX_1X_2\right)\cdot\kappa_0 \tag{53}$$

$$\epsilon_1=\left(\beta_{10}^s+\beta_{11}^sX_1+\beta_{12}^sX_2+\beta_{13}^sX_1^2+\beta_{14}^sX_2^2+\beta_{15}^sX_1X_2\right)\cdot\kappa_1 \tag{54}$$

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, the distributions of $Y(1)$ and $Y(0)$ will not have analytical closed forms that could be easily derived. We therefore computed target functionals using median values from 100 simulations of size 100,000 for the unfeasible estimators that use the empirical marginal distributions of $Y(1)$ and $Y(0)$.[23]

The parameters were specified as $\mu_{X_1}=1$, $\mu_{X_2}=5$ and those in Table I.[24]

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

[23] In the supporting information we provide some descriptive statistics of the 'empirical target' parameters.

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

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Table I. Parameter specification for Monte Carlo exercise

Coeff.\ j	0	1	2	3	4	5
δ_j	−0.5	1.35	−0.2	0.15	−0.1	0.5
β_{0j}	0.01	−0.01	0.01	0.01	−0.01	−0.02
β_{1j}	0.1	0.01	0.01	0.01	0.01	0.01
β_{0j}^s	0.01	−0.01	0.01	0.01	−0.01	−0.02
β_{1j}^s	0.01	0.01	0.01	0.01	0.01	0.01

Table II. Features of the distributions of potential outcomes for the treated (conditional on $T = 1$)

η	Logistic		Normal	
ν\Distribution	$Y(0)$	$Y(1)$	$Y(0)$	$Y(1)$
Mean	0.7787	1.8870	0.7798	1.8814
Standard deviation (SD)	0.2746	1.1731	0.2734	1.1648
Mean of logarithm	−0.3155	0.5109	−0.3133	0.5090
SD of logarithm	0.3777	0.4714	0.3759	0.4694
10th percentile	0.4451	0.9841	0.4474	0.9846
1st quartile	0.6022	1.2291	0.6044	1.2285
Median	0.7752	1.5823	0.7772	1.5805
3rd quartile	0.9312	2.1634	0.9320	2.1562
90th percentile	1.0748	3.0695	1.0737	3.0546

Table III. Inequality measures of potential outcomes for the treated (conditional on $T = 1$)

η	Logistic		Normal	
ν\Distribution	$Y(0)$	$Y(1)$	$Y(0)$	$Y(1)$
Coefficient of variation	0.3526	0.6217	0.3505	0.6191
Interquartile range	0.9343	0.3290	0.9277	0.3276
Theil index	0.0607	0.1402	0.0600	0.1390
Gini coefficient	0.1887	0.2755	0.1876	0.2743

A quick inspection of Tables II and III reveals that the target functionals are little affected by the distribution underlying the selection model. Thus consistent semiparametric estimators of the functionals of these distributions should not be largely affected by the nature of the DGP. In Tables IV–VII we present the results for the DGP based on the normal specification.[25]

Tables IV and V present results for the unfeasible estimator and for five feasible estimators. The first is the estimator proposed here and labeled 'weighted estimator'. To simplify the estimation procedure, we considered a parametric first step, in which we computed the propensity score by a logit model using the correct quadratic specification and a logit model using a misspecified linear model.[26]

The second estimator is the one based on the empirical distributions of $Y|T = 1$ and $Y|T = 0$. We term that estimator the 'naive estimator'. Because there is selection into treatment based on observables, the naive estimator is inconsistent with the ITE parameters.

[25] In the Monte Carlo Supplemental Appendix we report Tables A.1, A.2, A.3 and A.4, which correspond respectively to the results presented in Tables IV–VII. The difference is that we use logistic specifications in the Appendix.

[26] In the tables, 'weighted' corresponds to the estimator with a first stage that employs a quadratic specification for the propensity score and 'weighted-linear' corresponds to the estimator with a first stage that employs a linear specification of the propensity score. The same corresponding notation applies to the other estimators. Finally, in the Monte Carlos and in the empirical application we are normalizing weights to sum one.

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

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

Treatment on the Treated Parameters	Estimators	Target	Average	Lower 10th percentile	Median	Upper 10th percentile	Standard Deviation	Bias	Root Mean Squared Error	Mean Absolute Error	Median Absolute Error	90% C.I. Coverage Rate
Mean Treatment Effects		1.102										
	Unfeasible		1.104	0.969	1.100	1.243	0.113	0.002	0.113	0.089	0.075	0.880
	Naive		1.078	0.947	1.075	1.213	0.109	−0.023	0.112	0.089	0.076	0.800
	Weighted		1.104	0.963	1.100	1.251	0.114	0.002	0.114	0.091	0.077	0.840
	Weighted (Linear)		1.102	0.963	1.098	1.244	0.113	0.000	0.113	0.090	0.075	0.850
	Location Shift		1.104	0.961	1.100	1.246	0.114	0.002	0.114	0.091	0.076	0.840
	Location Shift (Linear)		1.106	0.968	1.101	1.246	0.114	0.010	0.114	0.090	0.076	0.830
	Log Location Shift		1.112	0.969	1.108	1.256	0.115	0.004	0.115	0.091	0.076	0.850
	Log Location Shift (linear)		1.114	0.973	1.109	1.255	0.114	0.012	0.115	0.091	0.077	0.890
	CFM		1.050	0.896	1.083	1.239	0.245	−0.051	0.250	0.136	0.087	0.850
	CFM (Linear)		1.106	0.966	1.103	1.247	0.114	0.005	0.114	0.091	0.077	0.860
CV		0.269										
	Unfeasible		0.255	0.124	0.235	0.401	0.134	−0.014	0.135	0.098	0.082	0.700
	Naive		0.302	0.165	0.282	0.455	0.134	0.033	0.138	0.094	0.068	0.810
	Weighted		0.261	0.119	0.244	0.415	0.138	−0.007	0.138	0.099	0.075	0.740
	Weighted (Linear)		0.266	0.124	0.249	0.423	0.137	−0.003	0.137	0.098	0.074	0.750
	Location Shift		0.289	0.151	0.267	0.440	0.133	0.020	0.135	0.093	0.066	0.800
	Location Shift (Linear)		0.290	0.155	0.267	0.442	0.133	0.021	0.134	0.092	0.067	0.790
	Log Location Shift		0.288	0.152	0.268	0.441	0.129	0.020	0.131	0.090	0.064	0.800
	Log Location Shift (linear)		0.289	0.158	0.270	0.436	0.126	0.020	0.128	0.088	0.065	0.810
	CFM		0.278	0.123	0.259	0.446	0.144	0.009	0.144	0.104	0.081	0.760
	CFM (Linear)		0.262	0.119	0.242	0.417	0.136	−0.007	0.136	0.098	0.079	0.760
Interquartile Range		0.600										
	Unfeasible		0.601	0.437	0.593	0.766	0.128	0.000	0.128	0.103	0.088	0.890
	Naive		0.635	0.467	0.631	0.806	0.130	0.035	0.135	0.106	0.090	0.910
	Weighted		0.603	0.438	0.599	0.778	0.132	0.002	0.132	0.105	0.088	0.890
	Weighted (Linear)		0.609	0.446	0.601	0.780	0.130	0.009	0.131	0.104	0.088	0.900
	Location Shift		0.658	0.494	0.654	0.823	0.130	0.058	0.142	0.112	0.091	0.900
	Location Shift (Linear)		0.647	0.486	0.644	0.812	0.129	0.047	0.137	0.108	0.089	0.910
	Log Location Shift		0.612	0.447	0.609	0.780	0.130	0.012	0.130	0.103	0.087	0.920
	Log Location Shift (linear)		0.610	0.445	0.605	0.779	0.131	0.010	0.131	0.105	0.090	0.890
	CFM		0.634	0.443	0.612	0.857	0.164	0.034	0.168	0.127	0.096	0.900
	CFM (Linear)		0.600	0.436	0.597	0.776	0.131	0.000	0.131	0.104	0.089	0.890

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Table IV. Continued

		0.079										
	Unfeasible		0.078	0.036	0.073	0.124	0.040	−0.001	0.040	0.029	0.023	0.780
	Naive		0.092	0.049	0.086	0.140	0.040	0.013	0.042	0.029	0.021	0.860
	Weighted		0.079	0.034	0.074	0.127	0.041	0.000	0.041	0.030	0.022	0.820
	Weighted (Linear)		0.081	0.037	0.076	0.131	0.041	0.002	0.041	0.029	0.022	0.820
Theil Index	Location Shift		0.089	0.044	0.083	0.139	0.041	0.010	0.043	0.030	0.022	0.840
	Location Shift (Linear)		0.090	0.045	0.084	0.142	0.042	0.011	0.043	0.030	0.022	0.830
	Log Location Shift		0.088	0.044	0.082	0.136	0.039	0.009	0.040	0.028	0.020	0.850
	Log Location Shift (linear)		0.089	0.046	0.083	0.136	0.039	0.010	0.040	0.028	0.020	0.850
	CFM		0.082	0.034	0.078	0.134	0.042	0.003	0.042	0.031	0.023	0.820
	CFM (Linear)		0.080	0.035	0.075	0.128	0.041	0.001	0.041	0.029	0.022	0.820
		0.087										
	Unfeasible		0.085	0.049	0.085	0.124	0.029	−0.001	0.029	0.023	0.020	0.890
	Naive		0.110	0.072	0.109	0.147	0.030	0.023	0.037	0.030	0.025	0.910
	Weighted		0.087	0.046	0.088	0.128	0.032	0.001	0.032	0.025	0.021	0.850
	Weighted (Linear)		0.090	0.050	0.091	0.130	0.031	0.003	0.032	0.025	0.021	0.850
Gini Coefficient	Location Shift		0.111	0.067	0.109	0.157	0.036	0.024	0.043	0.034	0.027	0.910
	Location Shift (Linear)		0.108	0.065	0.107	0.153	0.034	0.022	0.041	0.032	0.026	0.940
	Log Location Shift		0.099	0.060	0.098	0.139	0.031	0.012	0.033	0.027	0.022	0.910
	Log Location Shift (linear)		0.099	0.060	0.099	0.137	0.030	0.012	0.033	0.026	0.022	0.940
	CFM		0.102	0.047	0.095	0.160	0.048	0.015	0.050	0.036	0.026	0.880
	CFM (Linear)		0.088	0.047	0.088	0.128	0.032	0.001	0.032	0.025	0.021	0.850

Note: Coverage rates were computed using 100 bootstrap replications and applying the percentile bootstrap method as discussed in the text.

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Table V. Results of Monte Carlo Exercise (Sample Size 1,000, Replications 1000, Normal Selection Model)

Treatment on the Treated Parameters	Estimators	Target	Average	Lower 10th percentile	Median	Upper 10th percentile	Standard Deviation	Bias	Root Mean Squared Error	Mean Absolute Error	Median Absolute Error	90% C.I. Coverage Rate
		1.102										
	Unfeasible		1.104	1.035	1.101	1.180	0.055	0.002	0.055	0.044	0.035	0.910
	Naive		1.079	1.010	1.077	1.149	0.055	−0.023	0.059	0.048	0.042	0.890
	Weighted		1.104	1.031	1.103	1.176	0.057	0.002	0.057	0.045	0.038	0.900
	Weighted (linear)		1.102	1.029	1.101	1.175	0.056	0.001	0.056	0.045	0.038	0.900
	Location shift		1.104	1.030	1.102	1.176	0.057	0.002	0.057	0.045	0.038	0.900
	Location shift (linear)		1.106	1.034	1.105	1.179	0.057	0.012	0.057	0.045	0.038	0.890
Mean treatment effects	Log location shift		1.113	1.038	1.112	1.187	0.057	0.004	0.058	0.046	0.039	0.900
	Log location shift (linear)		1.115	1.041	1.113	1.189	0.057	0.013	0.058	0.046	0.038	0.890
	CFM		1.100	1.027	1.100	1.173	0.057	−0.001	0.057	0.045	0.038	0.890
	CFM (linear)		1.105	1.031	1.103	1.177	0.057	0.003	0.057	0.045	0.038	0.000
		0.269										
	Unfeasible		0.261	0.185	0.250	0.352	0.070	−0.007	0.070	0.053	0.043	0.810
	Naive		0.309	0.228	0.297	0.399	0.071	0.040	0.081	0.059	0.042	0.850
	Weighted		0.263	0.177	0.252	0.357	0.074	−0.005	0.074	0.057	0.046	0.810
	Weighted (linear)		0.269	0.184	0.258	0.360	0.074	0.001	0.073	0.055	0.044	0.820
	Location shift		0.298	0.219	0.288	0.389	0.071	0.030	0.077	0.055	0.041	0.870
	Location shift (linear)		0.297	0.218	0.285	0.387	0.071	0.028	0.076	0.054	0.040	0.900
CV	Log location shift		0.300	0.226	0.290	0.388	0.068	0.032	0.075	0.054	0.040	0.880
	Log location shift (linear)		0.298	0.226	0.289	0.384	0.066	0.030	0.072	0.052	0.038	0.910
	CFM		0.287	0.204	0.278	0.381	0.075	0.018	0.077	0.056	0.042	0.860
	CFM (linear)		0.286	0.202	0.276	0.378	0.072	0.017	0.074	0.054	0.040	0.870
		0.600										
	Unfeasible		0.599	0.517	0.596	0.683	0.066	−0.001	0.065	0.051	0.042	0.890
	Naive		0.632	0.554	0.628	0.719	0.066	0.032	0.073	0.057	0.045	0.890
	Weighted		0.598	0.516	0.594	0.685	0.066	−0.002	0.066	0.052	0.041	0.920
	Weighted (linear)		0.605	0.522	0.600	0.691	0.066	0.004	0.066	0.051	0.041	0.910
	Location shift		0.670	0.587	0.667	0.757	0.068	0.070	0.098	0.080	0.070	0.780
	Location shift (linear)		0.653	0.573	0.649	0.739	0.067	0.053	0.086	0.069	0.058	0.910
Interquartile range	Log location shift		0.617	0.532	0.615	0.703	0.067	0.017	0.069	0.054	0.044	0.840
	Log location shift (linear)		0.611	0.526	0.610	0.697	0.067	0.011	0.068	0.054	0.044	0.890
	CFM		0.600	0.517	0.596	0.685	0.066	0.000	0.066	0.052	0.041	0.920
	CFM (linear)		0.599	0.515	0.594	0.684	0.066	−0.002	0.066	0.051	0.041	0.920

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Table V. Continued

		0.079										
Theil index	Unfeasible		0.078	0.056	0.077	0.104	0.019	−0.001	0.019	0.015	0.012	0.870
	Naive		0.092	0.069	0.091	0.118	0.019	0.013	0.023	0.018	0.014	0.880
	Weighted		0.079	0.055	0.077	0.105	0.020	0.000	0.020	0.015	0.013	0.840
	Weighted (linear)		0.081	0.056	0.079	0.107	0.020	0.002	0.020	0.015	0.012	0.860
	Location shift		0.091	0.066	0.089	0.119	0.021	0.012	0.024	0.018	0.013	0.900
	Location shift (linear)		0.091	0.066	0.089	0.120	0.021	0.012	0.024	0.018	0.014	0.900
	Log location shift		0.090	0.067	0.088	0.115	0.019	0.011	0.022	0.016	0.013	0.870
	Log location shift (linear)		0.090	0.067	0.088	0.114	0.018	0.011	0.021	0.016	0.012	0.890
	CFM		0.087	0.063	0.085	0.113	0.020	0.008	0.021	0.016	0.012	0.890
	CFM (linear)		0.086	0.062	0.084	0.112	0.019	0.007	0.020	0.016	0.012	0.900
		0.087										
Gini coefficient	Unfeasible		0.087	0.068	0.087	0.105	0.015	0.000	0.015	0.011	0.010	0.920
	Naive		0.111	0.092	0.110	0.131	0.015	0.024	0.028	0.025	0.024	0.550
	Weighted		0.087	0.066	0.087	0.108	0.016	0.000	0.016	0.013	0.010	0.910
	Weighted (linear)		0.090	0.069	0.090	0.110	0.016	0.003	0.016	0.013	0.011	0.910
	Location shift		0.115	0.090	0.113	0.140	0.019	0.028	0.034	0.029	0.027	0.560
	Location shift (linear)		0.112	0.088	0.110	0.138	0.019	0.025	0.032	0.026	0.024	0.780
	Log location shift		0.102	0.082	0.102	0.123	0.016	0.015	0.022	0.018	0.016	0.630
	Log location shift (linear)		0.101	0.081	0.101	0.121	0.015	0.014	0.021	0.017	0.015	0.800
	CFM		0.098	0.077	0.098	0.119	0.016	0.011	0.020	0.016	0.013	0.860
	CFM (linear)		0.096	0.074	0.095	0.116	0.016	0.009	0.018	0.014	0.012	0.880

Note: Coverage rates were computed using 100 bootstrap replications and applying the percentile bootstrap method as discussed in the text.

Copyright © 2015 John Wiley & Sons, Ltd.

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

Treatment on the Treated Parameters	Estimator	Linear		Quadratic		Cubic	
		Standard Error	90% C.I. Coverage Rate	Standard Error	90% C.I. Coverage Rate	Standard Error	90% C.I. Coverage Rate
Mean treatment effects	Weighted	0.110	0.900	0.110	0.900	0.110	0.900
	Weighted (linear)	0.109	0.893	0.109	0.893	0.108	0.893
CV	Weighted	0.089	0.796	0.088	0.795	0.088	0.795
	Weighted (linear)	0.088	0.791	0.087	0.784	0.087	0.783
Interquartile Range	Weighted	0.129	0.886	0.128	0.885	0.128	0.885
	Weighted (linear)	0.131	0.904	0.130	0.904	0.130	0.904
Theil index	Weighted	0.032	0.861	0.032	0.852	0.032	0.852
	Weighted (linear)	0.032	0.862	0.032	0.855	0.032	0.854
Gini coefficient	Weighted	0.037	0.937	0.036	0.931	0.037	0.941
	Weighted (linear)	0.036	0.936	0.036	0.935	0.037	0.948

Note: Standard errors and coverage rates were calculated using analytical expressions provided in the main text.

We then consider what we term the 'location shift estimator'. This is constructed in the following manner. We first estimate two linear regressions (with intercepts) of Y on X_1, X_2, X_1^2, X_2^2 and X_1X_2, one for each group ($T = 0$ and $T = 1$). Save residuals $\widehat{u}_j$ and compute s_j^2 where $j = 0, 1$ indexes treatment groups, $s_j^2 = \left(N_j - 6\right)^{-1} \sum_i^{N_j} \left(\widehat{u}_{ji}\right)^2$ and $N_1 = \sum_i^N T_i$ and $N_0 = N - N_1$. Save coefficient estimates for group $T = 0$, $\widehat{\gamma}_{00}$, $\widehat{\gamma}_{10}$, $\widehat{\gamma}_{20}$, $\widehat{\gamma}_{30}$, $\widehat{\gamma}_{40}$, $\widehat{\gamma}_{50}$. Then, let Y_i^* be counterfactual outcome of treated observation i:

$$Y_i^* = \widehat{\gamma}_{00} + \widehat{\gamma}_{10} X_{1i} + \widehat{\gamma}_{20} X_{2i} + \widehat{\gamma}_{30} X_{1i}^2 + \widehat{\gamma}_{40} X_{2i}^2 + \widehat{\gamma}_{50} X_{1i} X_{2i} + \sqrt{s_0^2/s_1^2} \cdot \widehat{u}_{1i} \tag{55}$$

and because Y_i^* is well defined for all treated i, we compute the inequality measures for two distributions: $Y|T = 1$ and $Y^*|T = 1$. From the empirical $Y|T = 1$ we estimate functionals of $Y(1)|T = 1$, whereas with $Y^*|T = 1$ we estimate functionals of $Y(0)|T = 1$. Note that this is a means of 'controlling' for covariates. We also consider a 'location shift estimator' in which, in the first step, we misspecify the conditional distribution of $Y(1)$ and $Y(0)$ by estimating linear regressions (with intercept) of Y on X_1, X_2 for each group.

By noting that Y is distributed over the positive real numbers, an alternative way to implement the same concept is to take the logarithm first. We term this estimator the 'log-location shift estimator'. We proceed by following the same steps for the location shift estimator. The difference is that we apply the logarithm to Y first. Then, after we complete all steps described for the location shift estimator, we exponentiate the counterfactual logarithm of the outcome. By proceeding in this manner, we guarantee that the counterfactual outcome is always defined over the positive reals, which we cannot guarantee for the location shift estimator.[27] Location shift estimators correspond to the procedure proposed by Juhn *et al.* (1993).

Our final estimator is that proposed by Chernozhukov *et al.* (2013). We estimate the conditional distribution function of $Y|X, T = 0$ using logit estimators. To be more precise, let $D_y = \mathbb{I}\{Y \leq y\}$ and $F_{Y|T,X}(y|0, x) = \Pr\left[D_y = 1|T = 0, X = x\right]$. For fixed y, the conditional probability can be estimated using a flexible logit. As we did for all other estimators, we considered two situations. In

[27] In most Monte Carlo replications, we obtained negative counterfactual outcomes for the location shift estimator for very few observations. Typically, in 1000 replications, approximately 900 had at least one negative counterfactual outcome, and among these replications less than 10% had more than one negative value. Interestingly, we obtained more negative counterfactual outcomes for $N = 250$ than for $N = 1000$.

Copyright © 2015 John Wiley & Sons, Ltd.

Table VII. Results of Monte Carlo Exercise (Sample Size 1,000, Replications 1000, Normal Selection Model)

Treatment on the Treated Parameters	Estimator	Linear: Standard Error	Linear: 90% C.I. Coverage Rate	Quadratic: Standard Error	Quadratic: 90% C.I. Coverage Rate	Cubic: Standard Error	Cubic: 90% C.I. Coverage Rate
Mean Treatment Effects	Weighted	0.055	0.893	0.055	0.893	0.055	0.893
	Weighted (linear)	0.055	0.892	0.055	0.892	0.055	0.892
CV	Weighted	0.059	0.840	0.059	0.835	0.058	0.835
	Weighted (linear)	0.058	0.840	0.058	0.838	0.058	0.838
Interquartile range	Weighted	0.065	0.897	0.065	0.897	0.064	0.897
	Weighted (linear)	0.066	0.904	0.066	0.904	0.066	0.904
Theil index	Weighted	0.018	0.876	0.018	0.873	0.018	0.872
	Weighted (linear)	0.018	0.881	0.018	0.874	0.018	0.874
Gini coefficient	Weighted	0.015	0.861	0.014	0.855	0.014	0.855
	Weighted (linear)	0.015	0.868	0.015	0.862	0.015	0.863

Note: Standard errors and coverage rates were calculated using analytical expressions provided in the main text.

the first, for each y, we used the full quadratic model in the logit, and in the second, for each y, we used a linear model in the logit. The number of points y considered depended on the sample size. For $N = 250$, we used 100 points; for $N = 1000$, we used 500 points from the support of Y. Once we have an estimate of the conditional CDF of $Y|T = 0, X$, we can integrate it using the empirical distribution of $X|T = 1$. We term that estimator the 'CFM estimator'.

The results presented in Tables IV and V report distribution features of each estimator of inequality treatment effect parameters. We report average, standard deviation and quantiles (10th percentile, median and 90th percentile) for the four types of treatment effects on the inequality measures here considered (coefficient of variation, interquartile range, Theil index and Gini coefficient) in the cases in which we use the correct form of the propensity score, the conditional expectation (quadratic), and misspecify using a linear function. In addition to those inequality treatment effects, we report the results for the average treatment effects. Finally, we present results that contrast the estimates with the population target: We report bias, root mean squared error, mean absolute error and the coverage rate of 90% confidence intervals. Finally, note that the standard deviation entries in Tables IV and V refer to that quantity calculated using the 1000 estimates from the Monte Carlo replications, whereas the 90% CI coverage rate is obtained using 100 bootstrap replications and applying the percentile bootstrap method.

The normal DGP for the selection model is clearly less favorable to our estimator ('weighted estimator') than the one based on the logistic distribution, as our estimator was constructed using a fixed polynomial model for the logit. Despite that disadvantage, results in Tables IV and V indicate that the weighted estimator is a competitive estimator for distributional impacts, compared with more elaborate and computationally demanding estimators such as the CFM estimator. The weighted estimator performs well according to the MSE criteria and, as expected, its variance declines as the sample size increases. The weighted estimator is competitive with the CFM estimator even when we misspecify the propensity score. Relatively to the other estimators, when we consider the bias criteria, the weighted estimator dominates the naive estimator and the two estimators based on Juhn *et al.* (1993).

In Tables VI and VII we focus only on our estimators and report standard errors that were computed using the asymptotic variance formula proposed in the previous section. We also report the coverage rate of 90% confidence intervals that used these analytical standard errors. In the first two columns, we present the results using a linear model for the $g(x)$s functions. In the third and fourth columns,

Copyright © 2015 John Wiley & Sons, Ltd.

we present the results using a quadratic model for $g(x)$. The last two columns contain the results for a cubic model of the $g(x)$s functions. The results in Table VI and VII indicate that the coverage rate for the Gini coefficient is poor in small samples, but that it improves when we increase the sample size. For the coefficient of variation, the coverage rate is always below the desired level. For other inequality measures, the coverage rate is close to 90%. It is important to note that the results are not sensitive to the model used for $g(x)$.[28] Finally, coverage rates and standard errors computed either using the percentile bootstrap method, as in Tables IV and V, or using the analytical formulae, as in Tables VI and VII produce similar results.

7. EMPIRICAL APPLICATION

The empirical application considers a Brazilian public-sponsored job training program, known as PLANFOR (*Plano Nacional de Qualificaç ão Profissional*). This program, launched in 1996, has provided classroom training for the development 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 beginning every month. Although funding comes from the federal government, the program was decentralized at the state level. Each state subcontracted for classroom training with vocational proprietary schools and local community colleges. The target population consists of disadvantaged workers, who have been defined as the unemployed, and individuals with low levels of schooling and/or income. Enrollment of individuals in the program is voluntary, but its scale in 1998 was relatively small, at approximately 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 the impacts of a social program. In the years 1998–99, the Brazilian Ministry of Labor financed an experimental evaluation of the program's impact on earnings and employment. [29] Experimental data were collected in two metropolitan areas of the country, namely Rio de Janeiro and Fortaleza. The process of randomizing individuals into and out of the program was performed at the class level in August 1998, and nearly all individuals who were selected attended the training courses in September 1998. In that month, participants in both cities were interviewed using the same questionnaire, and retrospective questions were asked concerning their labor market histories. A follow-up survey was conducted in November 1999, and retrospective questions were asked dating back to September 1998.

For the sake of simplicity in our analysis we focus on data from Rio de Janeiro. Because randomization was performed at class level, we dropped all observations that were in classes with either only one treated or control unit. Under this criterion, the total available sample size from the baseline interview was 2616 individuals. They were distributed across 74 classes that had a median size of 18 students.

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

We first assess whether the randomization was properly performed. Because randomization occurred within classes, we determine whether the randomization was well performed in each class using t-tests of differences in means between treated and control groups. We decided to drop classes that, in at least one covariate, presented imbalances detected by t-tests at the 1% significance level.

[28] The Monte Carlo supplemental Appendix presents a design in which the propensity score attains values very close to 0 and 1. The results of this design are very similar to those presented in the main paper. The weighted estimator remains competitive with the CFM estimator and dominates all others. As expected, the weighted estimator is more sensitive to misspecification of the propensity score in the design in which the common support assumption is violated. In addition, the coverage rate of the Gini coefficient computed using the analytical formula for the variance is very unstable.

[29] This dataset was also used by Foguel (2006), in which further details on the impact evaluation study can be found.

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Table VIII. Summary Statistics

Variables	Final Sample: Rio de Janeiro: Treatment (A)		Control (B)		Weighted Control (C)		Difference (A)-(B)		Difference (A)-(C)	
Previous Labor Market	0.52		0.47		0.51		0.05		0.01	
Engagement (Dummy)	(0.53	0.56)	(0.45	0.49)	(0.48	0.55)	(0.01	0.08)	(−0.04	0.04)
Number of childred	0.08		0.07		0.09		0.01		−0.01	
	(0.09	0.08)	(0.06	0.08)	(0.07	0.11)	(0	0.03)	(−0.03	0.01)
Schooling (Years)	8.58		8.31		8.50		0.26		0.08	
	(8.63	8.77)	(8.21	8.42)	(8.38	8.62)	(0.12	0.42)	(−0.08	0.26)
Age	18.55		18.03		18.63		0.52		−0.08	
	(18.36	18.73)	(17.83	18.24)	(18.28	18.9)	(0.19	0.84)	(−0.46	0.33)
Dummy for Single	0.93		0.94		0.92		0.00		0.01	
	(0.92	0.95)	(0.93	0.95)	(0.9	0.94)	(−0.02	0.01)	(−0.01	0.03)
Household Head Dummy	0.03		0.03		0.04		0.00		−0.01	
	(0.02	0.04)	(0.02	0.03)	(0.02	0.05)	(−0.01	0.01)	(−0.02	0.01)
White Dummy	0.38		0.40		0.41		−0.03		−0.03	
	(0.35	0.4)	(0.38	0.43)	(0.37	0.44)	(−0.06	0.01)	(−0.07	0.01)
Male Dummy	0.36		0.34		0.38		0.02		−0.02	
	(0.34	0.38)	(0.31	0.36)	(0.34	0.4)	(−0.01	0.05)	(−0.05	0.02)
Number of Observations	1258		1211		1211					

Notes:
(i) The weighted estimator uses a linear specification for the prospensity score
(ii) We report 90% confidence interval in parentheses.

After applying that filter, five classes were dropped, and we were left with 2469 observations, of which 1258 were in the treated group and 1211 were in the control group.[30]

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

Using the follow-up survey, we constructed our outcome variable: the hourly wage rate at the first job in the 12-month interval after the treatment period. Our sample size decreased from the baseline to the follow-up stage, at 2071 individuals, 1106 belonging to the treated group and 965 to the control group. We assessed whether attrition could be explained by treatment status but found no statistical evidence supporting this.[31] Our sample size also declined when using the variable for the hourly

[30] Although we observed imbalances in five out of the 74 classes (6.8% of classes), that is not necessarily evidence that controlling for class dummies is insufficient to remove bias in this case. We direct attention to the fact that our criteria for dropping classes that presented detectable imbalances at the 1% level were applied to eight covariates. There were no classes that presented more than one unbalanced covariate at that significance level. Thus, we performed $8 \times 74 = 592$ tests and rejected the null in five of the 592 tests, that is, we rejected the null at the 1% significance level in 0.8% of the tests. The features of the data remain nearly identical after we dropped the five classes.

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

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Table IX. Inequality Treatment Effects for the PLANFOR Data Set

Hourly wage rate at first job in a 12-month periods after treatment			
	Treatment Effect Estimators		
	Treated	Naive	Weighted
Average	1.53	−0.20	−0.35
	(1.34 1.67)	(−0.51 0.09)	(−0.68 0.02)
Coefficient of Variation	1.56	−0.25	−0.21
	(0.88 2.37)	(−1.02 0.67)	(−0.99 0.72)
Interquartile Range	0.92	−0.02	−0.07
	(0.86 1.03)	(−0.14 0.12)	(−0.22 0.11)
Theil Index	0.36	−0.23	−0.23
	(0.14 0.52)	(−0.52 0.00)	(−0.52 0.02)
Gini Coefficient	0.38	−0.10	−0.11
	(0.32 0.44)	(−0.19 −0.02)	(−0.21 −0.03)
	Linear Shift	Log Linear Shift	CFM
Average	−0.34	−0.43	−0.30
	(−0.68 0.10)	(−0.77 0.16)	(−1.85 0.90)
Coefficient of Variation	−0.20	−0.83	−0.17
	(−0.94 0.80)	(−2.09 0.47)	(−1.43 18.30)
Interquartile Range	−0.59	−0.39	−0.08
	(−1.05 0.75)	(−0.62 −0.09)	(−0.18 0.17)
Theil Index	−0.15	−0.32	−0.23
	(−0.39 0.07)	(−0.65 0.14)	(−0.57 38.84)
Gini Coefficient	0.05	−0.13	−0.11
	(−0.38 0.13)	(−0.22 −0.02)	(−0.29 3.53)

Notes:
(i) First column in the first panel (Treated) presents results in level. All other columns are differences between that column and respective estimates of the counterfactual.
(ii) We report 90% confidence interval in parentheses.

wage rate at the first job because that variable is only defined for those who obtained a job after the program. [32]

In Table IX, we report the average and inequality treatment effect estimates. We report point estimates and bootstrapped standard errors (using 1000 replications) for all five feasible estimators defined in the Monte Carlo section.[33] For all of them, except the naive estimator, we use classroom dummies as controlling or confounding variables. Because we use a fully saturated model for the propensity score, the first stage of our weighted estimator is non-parametric. To ensure comparability with the other estimators, we also use class dummies as regressors in all of them. For the CFM estimator, given the sample sizes in the control group, we estimated the conditional CDF for 100 points.

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

[32] For that variable, we have 1033 non-missing observations: 577 treated and 456 control units.

[33] In the second supplemental Appendix, we present a set of Monte Carlo estimates in which the DGP matches the application scenario. The results of the Monte Carlo simulation indicates that our estimator is competitive with the others in terms of bias and coverage.

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Our results indicate that, on average, the program is ineffective. Moreover, it does also not appear to reduce inequality among the treated individuals as most of estimators used to measure inequality differences lie within 90% CIs.

8. CONCLUSION

In this paper, we proposed a method that may be useful for applied researchers who are interested in comparing the 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 that there are numerous observed factors with distributions that differ across groups. Our method makes it possible to decompose differences in Gini coefficients into two components: one that fixes the distribution of covariates and another that is a composition effect induced by different distributions of covariates.

We have also situated our method within the treatment effect literature. Our estimation strategy is most useful when the individual decision to participate in the program (the treatment) depends on observable characteristics. If this identification restriction holds, then the discussed reweighing method allows the researcher to identify the distribution of potential outcomes and, therefore, the causal impact of the program on numerous functionals of interest for policy analysis, such as inequality indices.

The results of our Monte Carlo study suggest that, when considering distributional aspects, consistent alternatives to the weighted estimator would most likely require a highly flexible estimation of the conditional distribution of Y given covariates to be competitive. Because the weighted estimator only requires the estimation of a single conditional expectation, it is a readily implementable alternative which clearly contrasts with non-parametric estimators of the conditional distributions. Finally, in addition to its computational simplicity, the weighted method also has desirable large-sample properties and performs relatively well in small samples.

Possible extensions of the work presented would be to characterize semi-parametric estimates of inequality treatment effects using alternative efficient estimators. A natural alternative procedure is the èfficient influence function estimator', also known as the 'double robust estimator' after Scharfstein *et al.* (1999), which was recently proposed by Cattaneo (2010) for the multivalued case in the GMM context and discussed extensively by Rothe and Firpo (2014). While this estimator may not be as simple to compute as the weighted estimator, it may represent an interesting combination of that estimator and those proposed by Chernozhukov *et al.* (2013 and Rothe (2010).

ACKNOWLEDGEMENTS

An earlier version of this paper circulated under the names 'Treatment Effects on Some Inequality Measures' and 'Inequality Treatment Effects'. We have benefited from helpful comments from anonymous referees and the editor on an earlier draft, and from comments from David Green, Jim Heckman, Guido Imbens, Thomas Lemieux, and seminars at UBC, USC, and EPGE-FGV. Hugo Jales, Rafael Cayres and Breno Braga helped us with superb research assistance. Financial support from CNPq-Brazil for Firpo is acknowledged (Grant No. 311085/2006-7). The usual disclaimer applies.

REFERENCES

Aakvik A, Heckman J, Vytlacil E. 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. 2002. Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association* **97**: 284–292.

Copyright © 2015 John Wiley & Sons, Ltd.

Abadie A, Angrist J, Imbens G. 2002. Instrumental variables estimation of quantile treatment effects. *Econometrica* **70**: 91–117.
Becker S, Ichino A. 2002. Estimation of average treatment effects based on propensity scores. *Stata Journal* **2**: 358–377.
Bhattacharya D. 2007. Inference on inequality from household survey data. *Journal of Econometrics* **137**: 674–707.
Bickel P, Klaassen C, Ritov Y, Wellner J. 1993. *Efficient and Adaptive Inference in Semiparametric Models*. Johns Hopkins University Press: Baltimore, MD.
Card D, Sullivan D. 1988. Measuring the effect of subsidized training programs on movements in and out of employment. *Econometrica* **56**: 497–530.
Carneiro P, Hansen K, Heckman J. 2001. Removing the veil of ignorance in assessing the distributional impacts of social policies. *Swedish Economic Policy Review* **8**: 273–301.
Carneiro P, Hansen K, Heckman J. 2003. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review* **44**: 361–422.
Cattaneo M. 2010. Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics* **155**: 138–154.
Champernowne D, Cowell F. 1999. *Economic Inequality and Income Distribution*. Cambridge University Press: Cambridge, UK.
Chen X, Hong H, Tarozzi A. 2008. Semiparametric efficiency in GMM models with auxiliary data. *Annals of Statistics* **36**(2): 808–843.
Chernozhukov V, Fernandez-Val I, Melly B. 2013. Inference on counterfactual distributions. *Econometrica* **81**: 2205–2268.
Cowell F. 2000. Measurement of inequality. In *Handbook of Income Distribution*, Atkinson A, Bourguignon F (eds), Vol. 1. North-Holland: Amsterdam; 87–166.
Cowell F. 2003. Theil, inequality and the structure of income distribution. Distributional analysis discussion paper 67. STICERD, LSE.
Cunha F, Heckman J, Navarro S. 2006. Counterfactual analysis of inequality and social mobility. In *Mobility and Inequality: Frontiers of Research from Sociology and Economics*, Morgan SL, Grusky DB, Fields GS (eds). Stanford University Press: Palo Alto, CA; 290–348.
Dehejia R, Wahba S. 1999. Causal effects in non-experimental studies: re-evaluating the evaluation of training programs. *Journal of the American Statistical Association* **94**: 1053–1062.
DiNardo J, Fortin N, Lemieux T. 1996. Labor market institutions and the distribution of wages, 1973–1992: a semiparametric approach. *Econometrica* **65**: 1001–1046.
Donald S, Hsu Y. 2014. Estimation and inference for distribution functions and quantile functions in treatment effect models. *Journal of Econometrics* **178**: 383–397.
Donald S, Green D, Paarsch H. 2000. Differences in wage distributions between Canada and the United States: an application of a flexible estimator of distribution functions in the presence of covariates. *Review of Economic Studies* **67**: 609–633.
Fan Y, Park S. 2010. Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory* **26**: 931–951.
Firpo S. 2007. Efficient semiparametric estimation of quantile treatment effects. *Econometrica* **75**: 259–276.
Firpo S, Ridder G. 2008. Bounds on functionals of the distribution of treatment effects. IEPR working paper 08–09. USC Institute for Economic Policy Research (IEPR) Research Paper Series.
Foguel M. 2006. Time effects, economic effects, and programme impacts. (Available from: http://www.webmeets.com/files/papers/LACEA-LAMES/2006/468/Paper_Prog_Impacts.pdf) Accessed on Nov. 3, 2006. Latest link is http://www.christophrothe.net/papers/sdre_nov2014.pdf. Paper was accessed on Dec. 1, 2014.
Fortin N, Lemieux T, Firpo S. 2011. Decomposition methods in economics. In *Handbook of Labor Economics*, Ashenfelter O, Card D (eds), Vol. 4A. Elsevier: Amsterdam; 1–102.
Frölich M, Melly B. 2013. Unconditional quantile treatment effects under endogeneity. *Journal of Business and Economic Statistics* **31**: 346–357.
Gosling A, Machin S, Meghir C. 2000. The changing distribution of male wages in the UK. *Review of Economic Studies* **67**: 635–686.
Hahn J. 1998. On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica* **66**: 315–331.
Hastie T, Tibshirani R. 1990. *Generalized Additive Models*. Boca Raton, FL: CRC Press.
Heckman J. 1992. Randomization and social program evaluation. In *Evaluating Welfare and Training Program*, Manski CE, Garfinkel I (eds). Harvard University Press: Cambridge, MA; 201–230.

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae

Heckman J, Robb R. 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*, Wainer H (ed). Springer: New York; 63–107.
Heckman J, Smith J. 1998. Evaluating the welfare state. In *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*, Strøm S (ed), Econometric Society Monograph Series. Cambridge University Press: Cambridge, UK; 241.
Heckman J, Ichimura H, Todd P. 1997a. Matching as an econometric evaluation estimator. *Review of Economic Studies* **65**(2): 261–294.
Heckman J, Smith J, Clements N. 1997b. Making the most out of programme evaluations and social experiments accounting for heterogeneity in programme impacts. *Review of Economic Studies* **64**(4): 487–535.
Heckman J, Ichimura H, Smith J, Todd P. 1998. Characterizing selection bias using experimental data. *Econometrica* **66**: 1017–1098.
Hirano K, Imbens G, Ridder G. 2003. Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica* **71**: 1161–1189.
Imbens G, Rubin D. 1997. Estimating outcome distributions for compliers in instrumental variable models. *Review of Economic Studies* **64**: 555–574.
Juhn C, Murphy K, Pierce B. 1993. Wage inequality and the rise in returns to skill. *Journal of Political Economy* **101**: 410–442.
LaLonde R. 1986. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* **76**: 604–620.
Lechner M. 1999. Earnings and employment effects of continuous off-the-job training in East Germany after unification. *Journal of Business and Economic Statistics* **17**: 74–90.
Machado J, Mata J. 2005. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics* **20**: 445–465.
Manski C. 1997. The mixing problem in programme evaluation. *Review of Economic Studies* **64**: 537–554.
Newey W. 1990. Semiparametric efficiency bounds. *Journal of Applied Econometrics* **5**: 99–135.
Newey W. 1997. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* **79**: 147–168.
Robins J, Rotnitzky A. 1995. Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association* **90**: 122–129.
Robins J, Rotnitzky A, Zhao L. 1994. Estimation of regression coefficients when some regressions are not always observed. *Journal of the American Statistical Association* **89**: 846–866.
Rosenbaum P, Rubin D. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**: 41–55.
Rothe C. 2010. Nonparametric estimation of distributional policy effects. *Journal of Econometrics* **155**: 56–70.
Rothe C, Firpo S. 2014. Semiparametric two-step estimation using doubly robust moment conditions. Working paper. Columbia University. (Available from: http://www.christophrothe.net/papers/sdre_nov2014.pdf) Acessed on Dec. 01, 2014.
Rotnitzky A, Robins J. 1995. Semiparametric regression estimation in the presence of dependent censoring. *Biometrika* **82**: 805–820.
Rubin D. 1977. Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics* **2**(1): 1–26.
Scharfstein D, Rotnitzky A, Robins J. 1999. Adjusting for non-ignorable drop-out using semiparametric non-response models. *Journal of American Statistical Association* **94**: 1096–1146.
Tarozzi A. 2007. Calculating comparable statistics from incomparable surveys, with an application to poverty in India. *Journal of Business and Economic Statistics* **25**(3): 314–336.
Van der Vaart A. 1998. *Asymptotic Statistics*. Cambridge University Press: Cambridge, UK.
Wooldridge J. 2007. Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics* **141**(2): 1281–1301.

Copyright © 2015 John Wiley & Sons, Ltd.

DOI: 10.1002/jae