



FUNDAÇÃO
GETULIO VARGAS

SEMINÁRIOS DE PESQUISA ECONÔMICA DA EPGE

Inequality treatment effects

SERGIO FIRPO

(PUC/RJ)

Data: 05/05/2005 (Quinta-feira)

Horário: 16h



FGV

EPGE

Local:

Praia de Botafogo, 190 – 11º andar
Auditório nº 1

Coordenação:

Prof. Luis Henrique B. Braidoe-mail:
lbraidoe@fgv.br

Inequality Treatment Effects*

Sergio Firpo[†]

First draft: July 2003; This version: February 2005

Abstract

This paper presents semiparametric estimators for treatment effects parameters when selection to treatment is based on observable characteristics. The parameters of interest in this paper are those that capture summarized distributional effects of the treatment. In particular, the focus is on the impact of the treatment calculated by differences in inequality measures of the potential outcomes of receiving and not receiving the treatment. These differences are called here inequality treatment effects. The estimation procedure involves a first non-parametric step in which the probability of receiving treatment given covariates, the propensity-score, is estimated. Using the reweighting method to estimate parameters of the marginal distribution of potential outcomes, in the second step weighted sample versions of inequality measures are computed. Calculations of semiparametric efficiency bounds for inequality treatment effects parameters are presented. Root- N consistency, asymptotic normality, and the achievement of the semiparametric efficiency bound are shown for the semiparametric estimators proposed. A Monte Carlo exercise is performed to investigate the behavior in finite samples of the estimator derived in the paper.

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

*An earlier version of this paper circulated under the name "Treatment Effects on Some Inequality Measures". This paper reflects part of my Ph.D. dissertation at UC Berkeley. I am indebted to Guido Imbens and to Jim Powell for their advice, support and many suggestions. I have also benefited from comments from participants at the 2004 Meeting of the Sociedade Brasileira de Econometria and at the empirical lunch seminar at UBC. Financial support from CAPES - Brazil and UBC's Start-Up Research Grant is acknowledged. All errors are mine.

[†]The University of British Columbia and PUC-Rio. Mailing address: University of British Columbia, Department of Economics, 997-1873 East Mall, Vancouver, BC, V6T 1Z1, Canada. E-mail: firpo@interchange.ubc.ca

1 Introduction

In evaluating a social program the policy-maker may want to learn not only about the mean effect of the program, but also about its distributional effects. For example, it would be reasonable to assume that the policy-maker is interested in the effect of the treatment on the dispersion of the outcome, which can be captured by inequality measures such as the Gini coefficient, the quantile-ratio, the interquartile range and many other inequality indices. One way of computing the distributional effect of the social program is through the *inequality treatment effects* (ITE), defined here as differences in inequality measures between the distribution of potential outcome of joining the program and the distribution of potential outcome of not joining it.

The program evaluation literature has not typically focused on inequality measures, or more generally, on distributional characteristics of program impacts.¹ A good reason for that is because treatment effects have been in most cases associated with *individual treatment effects*, defined as the difference in the individual potential outcome of joining the program and the potential outcome of not joining it.

However, it is very well recognized that the central identification problem in the program evaluation literature is that individual treatment effects are never observed, as an individual either receives the treatment or not. This is true even when the interest relies on the most popular treatment effect parameter, the average treatment effect (ATE), which identifies the individual treatment effect only if there is no heterogeneity in the effects, what may be a very strong assumption in many cases. Other treatment effect parameters, such as the quantile treatment effects (QTE), have received less attention, as their interpretability in terms of individual effects is less clear than it is for the mean case.² Nonetheless, there are many relevant aspects for policy purposes that are related to effects of social programs on the distribution of the outcome, a fact that weakens the sole interest in individual effects of those programs.

Unlike many other fields in economics, the welfare economics literature has had a long standing interest in the measurement of income inequality. Inequality measures can be justified

¹Some exceptions are the papers by Imbens and Rubin (1997), Abadie, Angrist and Imbens (2002), Heckman, Smith and Clements (1997) and, more recently, Firpo (2004). In the applied literature a recent paper that focus on the treatment effect of a social program on the income distribution is Bitler, Gelbach and Hoynes (2004).

²Individual interpretation of QTE requires the assumption that individuals would keep their rankings in the both the distributions of the potential outcome of being treated and of not being treated. Such discussion can be found in Firpo (2004), Bitler, Gelbach and Hoynes (2004) and Heckman, Smith and Clements (1997).

either as a parameter in the social welfare function or from a *ad hoc* perspective. Many popular inequality measures do not result from an utilitarian point of view and are instead built based on the satisfaction of some desirable properties. Among those properties, the most common and important ones are the principle of transfers, invariance, decomposability and anonymity.³

Anonymity means that inequality measures should not change if we are able to just re-label individuals, as this operation does not alter the distribution of income. This property is shared by all inequality measures that are based solely on the overall income distribution. Examples are all the usual inequality measures, such as the variance, quantile ratios and ranges, the Gini coefficient and several others.

The typical set of assumptions used primarily to identify treatment effect parameters does not, however, allow identification of the joint distribution of potential outcomes. Therefore, assessing individual impacts of social programs is not generally feasible.⁴ That set of assumptions is called the *ignorability* of the treatment.⁵ This assumption is a conditional independence assumption: Given observable characteristics, the decision to be treated is independent of the potential outcomes associated with being treated and with not being treated. This is also known as the *selection on observables* assumption and it makes possible to identify the marginal distributions of potential outcomes.

It is therefore clear that we can combine the genuine interest of the welfare economics with the “limitations” of the program evaluation literature. In this paper, the welfare economics literature and the program evaluation literature are combined and we establish conditions for identification and show how to estimate treatment effects on summary parameters of the outcome distribution.

This paper is divided as follows: In the next section we present a simple economic model of treatment effects on distributions that provides a theoretical justification for the interest in the inequality treatment effects. Section 3 presents the main identification result of this paper. Section 4 discusses estimation and derive the large sample properties of the inequality treatment effects estimators. In that section we show that “natural” estimators of the treatment effects

³See Cowell (2000) and Cowell (2003).

⁴An important exception happens when we are interested in a parameter that is a linear operation of the treatment effect distribution. This is exactly the case of ATE: the mean of the difference in potential outcomes is the simple difference in the means of the potential outcomes.

⁵See, for example, Rosenbaum and Rubin (1983).

on inequality measures are two-step type estimators. We show that these are root- N consistent and asymptotically normal. We also calculate the semiparametric efficiency bounds and show that the proposed estimators achieve them. Section 5 discusses finite-sample behavior through a Monte Carlo exercise. Finally, section 6 concludes. Proofs of results are left to the Appendix.

Along part of this paper we apply the general theory developed here to two particular cases: the variance and the 75-25 inter-quartile range (the difference between the 75th and the 25th percentiles). Note first that these two examples were chosen because they are very illustrative of how broad the methodology presented here can be: while one is an expectation parameter, the other is a function of quantiles. Other inequality measures fit the framework developed here, like the Gini coefficient, those in the Generalized Entropy Class and many others. Also, the approach developed here is broad enough to subsume other parameters of the distribution of the outcome, such as the mean and quantiles. In that sense, ATE and QTE could be seen as particular cases of this method.⁶

2 A Simple Model of Treatment Effects on Distributions

In this section we present a simple model that have two important features that justify the interest in the treatment effects on distributions: First, the individual decision to be in the treatment group depends on a vector of observable covariates and, second, the policy-maker aims to learn features of the marginal distributions of potential outcomes.

We start by assuming that there is an available random sample of N individuals (units). For each unit i , let X_i be a random vector of observed covariates with support $\mathcal{X} \subset \mathbb{R}^r$. Define $Y_i(1)$ as the potential outcome for individual i under treatment, and $Y_i(0)$ the potential outcome for the same individual without the treatment. Let the treatment assignment be defined as T_i , which equals one if individual i is exposed to treatment and equals zero otherwise. As we only observe each unit at one treatment status, we say that the unobserved outcome is the counterfactual outcome. Thus, the observed outcome can be expressed as:

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

⁶Recently, Chen, Hong and Tarozzi (2004) have independently showed that using a similar set of identifying assumptions, the approach of this paper could be generalized to a broader class of problems, such as missing data and non-classical errors in variables.

To motivate, consider Y_i as the observed earnings of individual i in a model of the impact of a job training program on worker earnings. In this example, T_i is the indicator for the receipt of training.

Potential outcomes depend on both observed and unobserved individual characteristics. For each individual i , let ε_i be a vector of unobservable attributes. In a job training program model, for example, earnings of each individual are a function of their pre-program observable characteristics, such as past earnings, employment status, education, age, job experience, gender, and union status; they are also a function of unobservable attributes, such as ability, motivation and some possible idiosyncratic shock.

Suppose the impact of X and ε on the potential outcomes is given by:

$$Y_i(1) = G_1(X_i, \varepsilon_i) \quad (2)$$

$$Y_i(0) = G_0(X_i, \varepsilon_i) \quad (3)$$

We assume self-selection into treatment: individuals can decide whether or not to be treated. When an individual i faces the decision whether or not to join the job training program, he will weigh the gains and costs of both situations. Assume that an individual i predicts his (or her) expected earnings (given his vector X_i) and his (or her) costs for each of the alternatives. In other words, the individual i chooses the state that yields the largest expected utility:

$$\max \{ \mathbb{E}[u(Y_i(1)) | X_i, \eta_i] - C_1(X_i, \eta_i); \mathbb{E}[u(Y_i(0)) | X_i, \eta_i] - C_0(X_i, \eta_i) \} \quad (4)$$

where $u(\cdot)$ is utility function, $C_1(\cdot, \cdot)$ and $C_0(\cdot, \cdot)$ are some costs associated respectively with joining the training program and not joining it, and η_i is a vector of variables that is unobserved to the econometrician but not to the individual. Also, η_i is assumed to be independent of ε_i . The effect of η_i on the individual's utility will depend on whether or not he enters the job program. For example, η_i might be a reservation wage that enters as an argument to a foregone earnings function. Individual i will then choose to take part in the program if $\mathbb{E}[u(Y(1)) | X_i, \eta_i] - C_1(X_i, \eta_i) \geq \mathbb{E}[u(Y(0)) | X_i, \eta_i] - C_0(X_i, \eta_i)$. That is:⁷

$$T_i = \mathbb{I}\{ \mathbb{E}[u(Y(1)) - u(Y(0)) | X_i, \eta_i] - (C_1(X_i, \eta_i) - C_0(X_i, \eta_i)) \geq 0 \} \quad (5)$$

Note how this model fits into the Roy model (1951) of income distribution.⁸ In the Roy

⁷The indicator function $\mathbb{I}\{A\}$ is equal to one if A is true and zero otherwise.

⁸See also Heckman and Honoré (1990).

model, an individual chooses the largest of the potential earnings given by two different occupations. Here, the choice is based on the individual's expected earnings and on some individual cost. Thus, after controlling for X_i , the choice of getting treatment will be independent of the individual potential earnings, which depends only on X_i and ε_i . That will hold as long as η_i and ε_i are independent and the functional form of potential earnings is the one described in Equations (2) and (3). The independence result can be written as:

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp T_i \mid X_i \quad \forall i \quad (6)$$

Equation (6) is the unconfoundedness (or ignorability) assumption labeled by Rubin (1977). Unless there is a gain in insight to writing the model with the structure presented in Equations (2)-(5), Equation (6) could actually have been our starting point.

Assume now that there is a social welfare function, W , that depends on a vector ν of parameters of the earnings distribution. We can write ν as a real function of the earnings distribution, that is, letting \mathcal{F}_ν be a class of distribution functions such that $F \in \mathcal{F}_\nu$ if $\nu(F) < +\infty$, we define $\nu : \mathcal{F}_\nu \rightarrow \mathbb{R}$. An example of W and ν is the case where ν is an inequality measure and W its inverse: The smaller the inequality the larger the welfare.

In order to simplify the argument, imagine that there are two possible scenarios: we either treat everyone or treat no one.⁹ Under the first scenario, the distribution of earnings is then equal to distribution of $Y(1)$, which has the distribution function $F_{Y(1)}$; while in the second scenario, the earnings distribution equals that of $Y(0)$, whose distribution function is $F_{Y(0)}$. We define the **Inequality Treatment Effect**, Δ_ν , as:

$$\Delta_\nu = \nu(F_{Y(1)}) - \nu(F_{Y(0)})$$

which is the difference in outcome (earnings, in this example) inequality between providing everyone the training and not providing it at all. Other parameters could be defined to sub-populations. In particular, consider the **Inequality Treatment Effect on the Treated**, $\Delta_{\nu|T=1}$:

$$\Delta_{\nu|T=1} = \nu(F_{Y(1)|T=1}) - \nu(F_{Y(0)|T=1})$$

where $F_{Y(1)|T=1}$ and $F_{Y(0)|T=1}$ are respectively the conditional distribution functions of the potential outcome of being treated and of not being treated given the treatment.

⁹Alternatives, as discussed in Manski (1997), include allowing individuals to choose their treatment status or assigning them to treatment based on observed characteristics.

3 Identification of Inequality Treatment Effects

We now set up assumptions for identification of Δ_ν . Remember that because $Y(1)$ and $Y(0)$ are never fully observable, we need to impose some identifying assumptions for parameters based on their marginal distributions, such as Δ_ν is. Before doing so and giving sequence to the definitions, let us define the “propensity-score”, $p(x)$, as the probability that, given a value x of X , an individual will be in the treatment group, that is, $p(x) = \Pr[T = 1|X = x]$. The unconditional probability is p , which will be assumed to be positive.

Now we are able to state and discuss the following assumptions.

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

This assumption has become popular in empirical research after a series of papers by Rubin and coauthors and by Heckman and coauthors¹⁰. Assumption 1 should be analyzed in a case-by-case situation, as for many exercises it is plausible not to hold.

We also make an overlap assumption:

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

Under Assumption 2 there is overlap in observable characteristics across groups, in the sense that it does not exist a value of x in \mathcal{X} such that it is only observed among individuals of one of the groups.

Another necessary assumption for identification is uniqueness. This can be written as:

ASSUMPTION 3 [*Uniqueness*] *Let F_A and $F_B \in \mathcal{F}_\nu$ where \mathcal{F}_ν is a class of distribution functions such that $F \in \mathcal{F}_\nu$ if $\nu(F) < +\infty$. If $F_A = F_B$ then, $\nu(F_A) = \nu(F_B)$.*

Finally, the main identification result will follow as a corollary of the next lemma.

LEMMA 1 *Let $F_{Y(1)}(y)$, $F_{Y(0)}(y)$, $F_{Y(1)|T=1}(y)$, and $F_{Y(0)|T=1}(y)$ be the respective cumulative distribution functions (c.d.f.) of $Y(1)$, $Y(0)$, $Y(1)|T = 1$, and $Y(0)|T = 1$ evaluated at a real*

¹⁰See, for instance, Rosenbaum and Rubin (1983 and 1984), Heckman, Ichimura, and Todd (1997) and Heckman, Ichimura, Smith, and Todd, (1998).

number y . Under Assumptions 1 and 2 these c.d.f.s can be expressed as:

$$\begin{aligned} F_{Y(1)}(y) &= \mathbb{E} \left[\frac{T}{p(X)} \cdot \mathbb{I}\{Y_i \leq y\} \right], \\ F_{Y(0)}(y) &= \mathbb{E} \left[\left(\frac{1-T}{1-p(X)} \right) \cdot \mathbb{I}\{Y_i \leq y\} \right], \\ F_{Y(1)|T=1}(y) &= \mathbb{E} \left[\frac{T}{p} \cdot \mathbb{I}\{Y_i \leq y\} \right], \\ F_{Y(0)|T=1}(y) &= \mathbb{E} \left[\left(\frac{p(X)}{p} \right) \cdot \left(\frac{1-T}{1-p(X)} \right) \cdot \mathbb{I}\{Y_i \leq y\} \right]. \end{aligned}$$

COROLLARY 1 Under Assumptions 1, 2 and 3, Δ_ν and $\Delta_{\nu|T=1}$ are identifiable.

Once we know that those inequality treatment effects are identifiable, we can now turn our attention to estimation and inference.

4 Estimation and Asymptotically Valid Inference

We now focus our attention to estimation of $\nu(F_{Y(1)})$ as estimation and inference of $\nu(F_{Y(0)})$, $\nu(F_{Y(1)|T=1})$, and $\nu(F_{Y(0)|T=1})$ follow by analogy. As the main objects of this paper are $\Delta_\nu = \nu(F_{Y(1)}) - \nu(F_{Y(0)})$ and $\Delta_{\nu|T=1} = \nu(F_{Y(1)|T=1}) - \nu(F_{Y(0)|T=1})$, after we show how to estimate and derive the asymptotic distribution of the estimator of $\nu(F_{Y(1)})$, we will also show how those results can be extended for estimation and inference regarding Δ_ν . An extension to $\Delta_{\nu|T=1}$ is not presented here although it would easily follow by analogy.

4.1 Estimation

Estimation of $\nu(F_{Y(1)})$ follows from the sample analogy principle, as the empirical distribution function of $Y(1)$ can be written as $\widehat{F}_{Y(1)}(y) = \frac{1}{N} \sum_{i=1}^N \frac{T_i}{\widehat{p}(X_i)} \cdot \mathbb{I}\{Y_i \leq y\}$. The key step that can be noticed is estimation of the propensity-score by $\widehat{p}(\cdot)$. As argued before, this is done in a first step, which we explain now.

Start by defining $H_K(x) = [H_{K,j}(x)]$ ($j = 1, \dots, K$), a vector of length K of polynomial functions of $x \in \mathcal{X}$ satisfying the following properties: (i) $H_K : \mathcal{X} \rightarrow \mathbb{R}^K$; and (ii) $H_{K,1}(x) = 1$. If we want $H_K(x)$ to include polynomials of x up to the order n , then it is sufficient to choose K such that $K \geq (n+1)^R$. In what follows, we will assume that K is a function of the sample size N , $K = K(N) \rightarrow +\infty$ as $N \rightarrow +\infty$.

Next, the $p(x)$ is estimated. Let $\hat{p}(x) = L(H_K(x)' \hat{\pi})$, where $L : \mathbb{R} \rightarrow \mathbb{R}$, $L(z) = (1 + \exp(-z))^{-1}$; and

$$\hat{\pi} = \arg \max_{\pi} \frac{1}{N} \sum_{i=1}^N (T_i \log(L(H_K(X_i)' \pi)) + (1 - T_i) \log(1 - L(H_K(X_i)' \pi))).$$

Also, when we need to estimate p , the unconditional probability of being treated, we consider the simple estimator $\hat{p} = \sum_{i=1}^N T_i / N$.

The second step follows a reweighting argument. The estimator of $\nu_{Y(1)} = \nu(F_{Y(1)})$ is $\hat{\nu}_{Y(1)} = \nu(\hat{F}_{Y(1)})$. It can be computed as one would compute any inequality measure using data from the empirical distribution of Y , the only difference is the necessary usage of the weights $\frac{T_i}{\hat{p}(X_i)}$. As examples, consider the following two inequality measures, the variance and the 75-25 inter-quartile range.

Example 1: Variance. The variance of $Y(1)$, $\nu_{Y(1)}^V$, can be written as:

$$\nu_{Y(1)}^V = \mathbb{E} \left[(Y(1) - \mathbb{E}[Y(1)])^2 \right].$$

However, as the distribution of $Y(1)$ is not observable, under assumptions 1-3, the reweighting scheme will be used to write $\nu_{Y(1)}^V$ as

$$\nu_{Y(1)}^V = \mathbb{E} \left[\frac{T}{p(X)} \cdot \left(Y - \mathbb{E} \left[\frac{T}{p(X)} \cdot Y \right] \right)^2 \right], \quad (7)$$

and its estimator as

$$\hat{\nu}_{Y(1)}^V = \frac{1}{N} \sum_{i=1}^N \frac{T_i}{\hat{p}(X_i)} \cdot \left(Y_i - \frac{1}{N} \sum_{j=1}^N \frac{T_j}{\hat{p}(X_j)} \cdot Y_j \right)^2. \quad (8)$$

Example 2: The 75-25 Interquartile Range. Even when the inequality measure is not an expectation, as the variance is, we can apply the reweighting method. The interquartile range is an example of such case. Fortunately, we can use the results derived in Firpo (2004) to express the 75-25 inter-quartile range of the distribution of $Y(1)$, $\nu_{Y(1)}^{75-25}$, as:

$$\begin{aligned} \nu_{Y(1)}^{75-25} &= q_{Y(1),.75} - q_{Y(1),.25} \\ &= \arg \min_q \mathbb{E} \left[\frac{T}{p(X)} \cdot (Y - q) (.75 - \mathbb{I}\{Y - q \leq 0\}) \right] \\ &\quad - \arg \min_q \mathbb{E} \left[\frac{T}{p(X)} \cdot (Y - q) (.25 - \mathbb{I}\{Y - q \leq 0\}) \right], \end{aligned} \quad (9)$$

and its estimator is defined as:

$$\begin{aligned}
\hat{\nu}_{Y(1)}^{75-25} &= \hat{q}_{Y(1),.75} - \hat{q}_{Y(1),.25} \\
&= \arg \min_q \frac{1}{N} \sum_{i=1}^N \frac{T_i}{\hat{p}(X_i)} \cdot (Y_i - q) (.75 - \mathbb{I}\{Y_i - q \leq 0\}) \\
&\quad - \arg \min_q \frac{1}{N} \sum_{j=1}^N \frac{T_j}{\hat{p}(X_j)} \cdot (Y_j - q) (.25 - \mathbb{I}\{Y_j - q \leq 0\}).
\end{aligned} \tag{10}$$

4.2 Inference

We now devote our attention to the asymptotic behavior of our estimators. This subsection is divided in the following parts. We first show how to estimate the first step, that is, how the propensity-score function is estimated. Then, we invoke the hypothetical case of full observability of potential outcome to establish conditions for asymptotic normality and efficiency of our estimators, which are shown in part 3. Finally we show how to consistently estimate the standard errors of our estimators.

4.2.1 The First Step

An important assumption for the final ITE estimators to be asymptotically normal are that the first step estimator of the propensity score be uniformly consistent. In order to guarantee uniform consistency of this first step procedure, we assume that:

ASSUMPTION 4 [*Distribution of X and Smoothness*] (i) \mathcal{X} is a compact subset of \mathbb{R}^r ; (ii) the density of X , $f(x)$, satisfies $0 < \inf_{x \in \mathcal{X}} f(x) \leq \sup_{x \in \mathcal{X}} f(x) < \infty$ (iii) $p(x)$ is s -times continuously differentiable, where $s \geq 7r$ and r is the dimension of X ; (iv) the order of $H_K(x)$, K , is of the form $K = C \cdot N^\alpha$ where C is a constant and $\alpha \in \left(\frac{1}{4 \cdot (\frac{s}{r} - 1)}, \frac{1}{9}\right)$.

Newey (1995, 1997) established that for orthogonal polynomials $H_K(x)$ and compact \mathcal{X} , $\zeta(K) = \sup_{x \in \mathcal{X}} \|H_K(x)\| \leq C \cdot K$, where C is a generic constant. Note then that because of part (iv) of Assumption 4, ζ will be a function of N since K is assumed to be a function of N . Following Hirano, Imbens and Ridder (2003), we can invoke state the following result:¹¹

LEMMA 2 *Under Assumptions 1 and 4 the following results hold: (i) $\sup_{x \in \mathcal{X}} |p(x) - p_K(x)| \leq C \cdot \zeta(K) K^{-s/2r} \leq C \cdot \zeta^{1-s/2r} \leq C \cdot N^{(1-s/2r)\alpha} = o(1)$; where $p_K(x) = L(H_K(x)^\top \pi_K)$ and*

¹¹Henceforth sometimes we will refer to their paper as HIR.

$\pi_K = \arg \max_{\pi} \mathbb{E} \left\{ p(X) \log(L(H_K(X)^\top \pi)) + (1 - p(X)) \log(1 - L(H_K(X)^\top \pi)) \right\}$; (ii) $\mathbb{E} \|\hat{\pi} - \pi_K\|^2 \leq C \cdot \frac{\zeta(K)}{N} \leq C \cdot N^{\alpha-1} = o(1)$; (iii) there is $\delta > 0$: $\lim_{N \rightarrow \infty} \Pr[\delta < \inf_{X \in \mathcal{X}} \hat{p}(X) \leq \sup_{X \in \mathcal{X}} \hat{p}(X) < 1 - \delta] = 1$.

An important point to make here is that Assumption 4 is stronger than the necessary for Lemma 2 to hold. This can be seen more clearly in the proof by HIR. In fact, the condition that $s \geq 7r$ is too strong for uniform consistency of the first-step: $4r$ times differentiable would be enough. Assumption 4 is maintained though as it is necessary for the asymptotic normality of the final estimators of the inequality treatment effects.

4.2.2 A Hypothetical Benchmark Case: Full Observability of Potential Outcomes

We will need to establish some extra conditions to be able to derive the asymptotic normality of the inequality estimators just proposed. Among those conditions, two are presented now: That their influence functions under the hypothetical case of full observability of potential outcomes have (i) finite variances and (ii) continuously differentiable conditional expectations given $X = x$.

Let us now define $\psi_\nu(y; \theta_{\nu, Y(1)})$ as:

$$\psi_\nu(y; \theta_{\nu, Y(1)}) = \frac{d \nu((1-s) \cdot F_{Y(1)} + s \cdot \delta_y)}{d s} \Big|_{s=0}, \quad (11)$$

where δ_y denotes the c.d.f. of the degenerated distribution that places probability one at the point y . The function $\psi_\nu(y; \theta_{\nu, Y(1)})$ is then the derivative with respect to s of $\nu((1-s) \cdot F_{Y(1)} + s \cdot \delta_y)$ evaluated at $s = 0$, or said in another way, it is the Gâteaux derivative of $\nu(\cdot)$ at y . It is also an influence function associated with the parameter $\nu_{Y(1)}$ under the assumption of *full observability* of $Y(1)$.¹² The vector $\theta_{\nu, Y(1)}$ is a parameter vector that takes values from the distribution $F_{Y(1)}$ to \mathbb{R}^L and includes $\nu_{Y(1)}$. By analogy, $\psi_\nu(\cdot; \theta_{\nu, Y(0)})$ is an influence function of the parameter $\nu_{Y(0)}$ under the assumption of full observability of $Y(0)$.

Now, we state some assumptions regarding the variance and the conditional expectation given X of $\psi_\nu(Y(1); \theta_{\nu, Y(1)})$ and $\psi_\nu(Y(0); \theta_{\nu, Y(0)})$.

¹²The influence function of an estimator places a key role in the robust statistics literature, so the natural reference for the definitions used in this section is Huber (1981).

ASSUMPTION 5 [*The Influence Function*] Assume that:

- (i) $\mathbb{E} \left[(\psi_\nu(Y(1); \theta_{\nu, Y(1)}))^2 \right] < \infty$ and $\mathbb{E} \left[(\psi_\nu(Y(0); \theta_{\nu, Y(0)}))^2 \right] < \infty$;
- (ii) $\mathbb{E} [\psi_\nu(Y(1); \theta_{\nu, Y(1)}) \mid X = x]$ and $\mathbb{E} [\psi_\nu(Y(0); \theta_{\nu, Y(0)}) \mid X = x]$ are continuously differentiable for all x in \mathcal{X} .

The above assumption is crucial for the derivation of our limiting distribution theory. Part (i) is the usual condition of square integrability of the influence function, whereas part (ii) is a regularity condition used in bounding a remainder term. As we will see in the proof of the main theorem of this paper, Assumption 5 plays the same role that HIR's Assumption 3 plays in the proof of their Theorem 1.

Given the generality of the parameters in this paper, for the analysis of the asymptotic behavior of their estimators we will restrict ourselves to a specific class of parameters. Under full observability of potential outcomes, we will assume that each parameter in that class will be estimated by an asymptotically linear estimator whose influence function lies on the tangent space of the statistical model (under full observability of potential outcomes). Furthermore, ν will be assumed to be pathwise differentiable, that is, ψ will be the efficient influence function of ν . Before we state more formally these points as assumptions, let us first define S_1^{FULL} and S_0^{FULL} the tangent spaces of the models under full observability of potential outcomes, as $S_1^{FULL} = \{S_1 : \mathbb{R} \times \{0, 1\} \times \mathcal{X} \rightarrow \mathbb{R} \mid S_1(y, t, x) = s_1(y, |x) + a(x) \cdot (t - p(x)) + s_X(x); \text{ such that } \mathbb{E}[s_1(Y(1)|X)] = \mathbb{E}[s_X(x)] = 0 \text{ and } a(x) \text{ is square-integrable function of } x\}$ and

$S_0^{FULL} = \{S_0 : \mathbb{R} \times \{0, 1\} \times \mathcal{X} \rightarrow \mathbb{R} \mid S_0(y, t, x) = s_0(y, |x) + a(x) \cdot (t - p(x)) + s_X(x); \text{ such that } \mathbb{E}[s_0(Y(0)|X)] = \mathbb{E}[s_X(x)] = 0 \text{ and } a(x) \text{ is square-integrable function of } x\}$.

Also define $W_1 = [Y(1), T, X^\top]^\top$ and $W_0 = [Y(0), T, X^\top]^\top$. Under this terminology, we define Δ_ν as:

$$\begin{aligned} \Delta_\nu = & \nu \left(p \cdot \int dF_{W_1}(\cdot, 1, x) dx + (1 - p) \cdot \int dF_{W_1}(\cdot, 0, x) dx \right) \\ & - \nu \left(p \cdot \int dF_{W_0}(\cdot, 1, x) dx + (1 - p) \cdot \int dF_{W_0}(\cdot, 0, x) dx \right) \end{aligned}$$

Also, we define the estimators under the hypothetical case of full observability as functionals of the following empirical distributions, which have the superscript U , as in “unfeasible”, to remind us that the potential outcomes are in reality never observed. They are: $\hat{F}_{Y(1)}^U(y)$

$$= \frac{1}{N} \sum_{i=1}^N \mathbb{I}\{Y_i(1) \leq y\} \text{ and } \widehat{F}_{Y(0)}^U(y) = \frac{1}{N} \sum_{i=1}^N \mathbb{I}\{Y_i(0) \leq y\}. \text{ Thus, } \widehat{\Delta}_\nu^U = \nu\left(\widehat{F}_{Y(1)}^U\right) - \nu\left(\widehat{F}_{Y(0)}^U\right).$$

We are now ready to state two key assumptions for our asymptotic results:

ASSUMPTION 6 [*Asymptotic Linearity of Unfeasible Estimator*] *We assume that if random samples of $\{Y_i(1), T_i, X_i\}$ and $\{Y_i(0), T_i, X_i\}$ were available then, $\widehat{\Delta}_\nu^U$ would be a regular estimator of Δ_ν such that:*

$$\widehat{\Delta}_\nu^U - \Delta_\nu - \frac{1}{N} \sum_{i=1}^N (\psi_\nu(Y_i(1); \theta_{\nu, Y(1)}) - \psi_\nu(Y_i(0); \theta_{\nu, Y(0)})) = o_p(1/\sqrt{N}).$$

Also, we consider that the following assumption holds:

ASSUMPTION 7 [*Efficiency of Unfeasible Estimator*] *Let $F_{W_1}(w; \theta)$ and $F_{W_0}(w; \theta) \in \mathcal{F}_\nu$ be model such that their tangent spaces are respectively characterized by \mathcal{S}_1^{FULL} and \mathcal{S}_0^{FULL} . We assume that: $\psi_\nu(\cdot; \theta_{Y(1)}) \in \mathcal{S}_1^{FULL}$ and $\psi_\nu(\cdot; \theta_{Y(0)}) \in \mathcal{S}_0^{FULL}$.*

Although apparently too restrictive, assumptions 6 and 7 are not. Consider the examples already given, the variance and the 75-25 interquartile range.

Example 1: Variance (continued). The unfeasible estimator of the variance is defined as:

$$\widehat{\nu}_{Y(1)}^{V,U} = \frac{1}{N} \sum_{i=1}^N \left(Y_i(1) - \frac{1}{N} \sum_{j=1}^N Y_j(1) \right)^2,$$

and we can verify that

$$\begin{aligned} \widehat{\nu}_{Y(1)}^{V,U} - \nu_{Y(1)}^V - \frac{1}{N} \sum_{i=1}^N \left((Y_i(1) - \mathbb{E}[Y(1)])^2 - \nu_{Y(1)}^V \right) \\ = \widehat{\nu}_{Y(1)}^{V,U} - \nu_{Y(1)}^V - \frac{1}{N} \sum_{i=1}^N \psi_\nu^V(Y_i(1); \theta_{Y(1)}) = o_p(1/\sqrt{N}), \end{aligned}$$

which will hold by the fact that under full observability of $Y(1)$, $\frac{1}{N} \sum_{i=1}^N Y_i(1) - \mathbb{E}[Y(1)] = o_p(1)$. Therefore, applying the same results to $Y(0)$, we have that:

$$\widehat{\Delta}_\nu^{V,U} - \Delta_\nu^V - \frac{1}{N} \sum_{i=1}^N (\psi_\nu^V(Y_i(1); \theta_{Y(1)}) - \psi_\nu^V(Y_i(0); \theta_{Y(0)})) = o_p(1/\sqrt{N}).$$

Also, $\psi_\nu^V(\cdot; \theta_{\nu, Y(1)}) \in \mathcal{S}_1^{FULL}$ and $\psi_\nu^V(\cdot; \theta_{\nu, Y(0)}) \in \mathcal{S}_0^{FULL}$ as such examples fit into Example 3.3.2 and Proposition A.5.2 of Bickel, Klassen, Ritov and Wellner (1993).¹³

¹³Henceforth simply BKRW.

Example 2: *The 75-25 Interquartile Range* (continued). The same is true for interquartile range. To see that, note that the unfeasible estimator for $Y(1)$ is defined as:

$$\begin{aligned}\widehat{\nu}_{Y(1)}^{75-25,U} &= \widehat{q}_{Y(1),.75}^U - \widehat{q}_{Y(1),.25}^U \\ &= \arg \min_q \frac{1}{N} \sum_{i=1}^N (Y_i(1) - q) (.75 - \mathbb{I}\{Y_i(1) - q \leq 0\}) \\ &\quad - \arg \min_q \frac{1}{N} \sum_{j=1}^N (Y_j(1) - q) (.25 - \mathbb{I}\{Y_j(1) - q \leq 0\}).\end{aligned}$$

And then,

$$\begin{aligned}\widehat{\nu}_{Y(1)}^{75-25,U} - \nu_{Y(1)}^{75-25} &- \frac{1}{N} \sum_{i=1}^N \left(\frac{(.75 - \mathbb{I}\{Y_i(1) - (\nu_{Y(1)}^{75-25} + q_{Y(1),.25}) \leq 0\})}{f_{Y(1)}(\nu_{Y(1)}^{75-25} + q_{Y(1),.25})} - \frac{(.25 - \mathbb{I}\{Y_i(1) - q_{Y(1),.25} \leq 0\})}{f_{Y(1)}(q_{Y(1),.25})} \right) \\ &= \widehat{\nu}_{Y(1)}^{75-25,U} - \nu_{Y(1)}^{75-25} - \frac{1}{N} \sum_{i=1}^N \psi_{\nu}^{75-25}(Y_i(1); \theta_{Y(1)}) = o_p(1/\sqrt{N})\end{aligned}$$

where $\theta_{\nu, Y(1)} = [\nu_{Y(1)}^{75-25}, q_{Y(1),.25}]^\top$. We assume that $f_{Y(1)}(q_{Y(1),.75}) \neq 0$ and $f_{Y(1)}(q_{Y(1),.25}) \neq 0$. The above equation will hold as a result from an application of the ordinary delta-method to the difference $\widehat{q}_{Y(1),.75}^U - \widehat{q}_{Y(1),.25}^U$, where for $\tau = \{.25, .75\}$, as showed, for example, in Corollary 21.5 of van der Vaart (1998),

$$\begin{aligned}\widehat{q}_{Y(1),\tau}^U &= q_{Y(1),\tau} + \frac{1}{N} \sum_{i=1}^N \frac{(\tau - \mathbb{I}\{Y_i(1) - q_{Y(1),\tau} \leq 0\})}{f_{Y(1)}(q_{Y(1),\tau})} + o_p(1/\sqrt{N}) \\ &= q_{Y(1),\tau} + \frac{1}{N} \sum_{i=1}^N \psi_{\nu^q}(Y_i(1); q_{Y(1),\tau}) + o_p(1/\sqrt{N}).\end{aligned}$$

Also, one can check that $\psi_{\nu^q}(\cdot; q_{Y(1),.25})$ and $\psi_{\nu^q}(\cdot; q_{Y(1),.75})$ lie on the tangent space \mathcal{S}_1^{FULL} and furthermore, that $q_{Y(1),.25}$ and $q_{Y(1),.75}$ are pathwise differentiable (as in BKRW's Definition 3.3.1). Thus, as an application of Proposition 3.3.1 and Corollary 3.3.2 of BKRW, $\widehat{\nu}_{Y(1)}^{75-25,U}$ will be asymptotically efficient under full observability of $Y(1)$.

4.2.3 Asymptotic Normality and Efficiency

Let us now define some objects that we will see are important in getting an expression for the asymptotic distribution of the estimators $\widehat{\Delta}_\nu$.

$$\psi_{\Delta}(Y, X, T; \theta_{\nu}) = \varphi_{\Delta}(Y, X, T; \theta_{\nu}) + \alpha_{\Delta}(T, X; \theta_{\nu})$$

where

$$\varphi_{\Delta}(Y, X, T; \theta_{\nu}) = \frac{T}{p(X)} \cdot \psi_{\nu}(Y; \theta_{\nu, Y(1)}) - \left(\frac{1-T}{1-p(X)} \right) \cdot \psi_{\nu}(Y; \theta_{\nu, Y(0)}),$$

and

$$\begin{aligned} \alpha_{\Delta}(T, X; \theta_{\nu}) = & \\ & - \left(\frac{\mathbb{E}[\psi_{\nu}(Y; \theta_{\nu, Y(1)}) \mid X, T=1]}{p(X)} + \frac{\mathbb{E}[\psi_{\nu}(Y; \theta_{\nu, Y(0)}) \mid X, T=0]}{1-p(X)} \right) \cdot (T - p(X)). \end{aligned}$$

where $\theta_{\nu} = [\theta_{\nu, Y(1)}, \theta_{\nu, Y(0)}]^{\top}$.

Some words about the previous functions. As will be stated in the next theorem, the function ψ_{Δ} and is the influence function of the estimator $\widehat{\Delta}_{\nu}$. It is a sum of two functions: φ_{Δ} and α_{Δ} . When the parameter ν is an expectation, then the first term can be interpreted as the function generating the moment condition: $\Delta_{\nu} - \mathbb{E}[\varphi_{\Delta}(Y, X, T; \theta_{\nu})] = 0$.

The α_{Δ} function represents the effect on the influence function of estimating $p(\cdot)$. Its first multiplicative factor, $-\left(\frac{\mathbb{E}[\psi_{\nu}(Y; \theta_{\nu, Y(1)}) \mid X, T=1]}{p(X)} + \frac{\mathbb{E}[\psi_{\nu}(Y; \theta_{\nu, Y(0)}) \mid X, T=0]}{1-p(X)}\right)$, is the conditional expectation of the derivative of with respect to $p(x)$. Hence, the function ψ_{Δ} linearizes the effect of the propensity-score, $p(x)$.

Finally, define:

$$\begin{aligned} \sigma_{\Delta_{\nu}}^2 &= \mathbb{E}[(\psi_{\Delta}(Y, X, T; \theta))^2] = \mathbb{E}[(\varphi_{\Delta}(Y, X, T; \theta_{\nu}) + \alpha_{\Delta}(T, X; \theta_{\nu}))^2] \\ &= \mathbb{E}\left[\frac{\mathbb{V}[\psi_{\nu}(Y; \theta_{\nu, Y(1)}) \mid X, T=1]}{p(X)} + \frac{\mathbb{V}[\psi_{\nu}(Y; \theta_{\nu, Y(0)}) \mid X, T=0]}{1-p(X)} \right. \\ &\quad \left. + (\mathbb{E}[\psi_{\nu}(Y; \theta_{\nu, Y(1)}) \mid X, T=1] - \mathbb{E}[\psi_{\nu}(Y; \theta_{\nu, Y(0)}) \mid X, T=0])^2\right]. \end{aligned} \quad (12)$$

We are now ready to establish the key inference result of the paper.

THEOREM 1 *Under Assumptions 1-6:*

$$(i) \sqrt{N}(\widehat{\Delta}_{\nu} - \Delta_{\nu}) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \psi_{\Delta}(Y_i, X_i, T_i; \theta_{\nu}) + o_p(1) \xrightarrow{D} N(0, \sigma_{\Delta_{\nu}}^2)$$

(ii) and adding an extra condition, Assumption 7, $\sigma_{\Delta_{\nu}}^2$ is the semiparametric efficiency bound of Δ_{ν} .

4.2.4 Variance Estimation

The variance term $\sigma_{\Delta_\nu}^2$ can be estimated by replacing expectations by sample averages. There are two important points in the variance estimation procedure. One has been addressed by HIR and is the fact that the α function depends on conditional expectations that need to be estimated, and following their approach, we use the same type of estimation that we used in the estimation of the propensity-score. The other point, which, unlike HIR is relevant for our case, is that both the φ function and part of the θ_ν parameter might need to be estimated as well. Thus, the estimator for the bound $\sigma_{\Delta_\nu}^2$ will be:

$$\hat{\sigma}_{\Delta_\nu}^2 = \frac{1}{N} \sum_{i=1}^N \left(\hat{\varphi}_\Delta \left(Y_i, X_i, T_i; \hat{\theta}_\nu \right) + \hat{\alpha}_\Delta \left(T_i, X_i; \hat{\theta}_\nu \right) \right)^2$$

where $\hat{\theta}_\nu$ is an estimates of θ_ν of the type:

$$\hat{\theta}_\nu = \left[\hat{\theta}_{\nu, Y(1)}, \hat{\theta}_{\nu, Y(0)} \right]^\top = \left[\theta_{\nu, Y(1)} \left(\hat{F}_{Y(1)} \right), \theta_{\nu, Y(0)} \left(\hat{F}_{Y(0)} \right) \right]^\top$$

and where $\hat{\varphi}_\Delta$ is the following estimate of φ_Δ :

$$\hat{\varphi}_\Delta \left(Y_i, X_i, T_i; \hat{\theta}_\nu \right) = \frac{T_i}{\hat{p}(X_i)} \cdot \hat{\psi}_\nu \left(Y_i; \hat{\theta}_{\nu, Y(1)} \right) - \left(\frac{1 - T_i}{1 - \hat{p}(X_i)} \right) \cdot \hat{\psi}_\nu \left(Y_i; \hat{\theta}_{\nu, Y(0)} \right).$$

The function $\hat{\alpha}_\Delta \left(T, X; \hat{\theta}_\nu \right)$ is:

$$\begin{aligned} \hat{\alpha}_\Delta \left(t, x; \hat{\theta}_\nu \right) = & - \sum_{i=1}^N \left(\frac{T_i \cdot \hat{\psi}_\nu \left(Y_i; \hat{\theta}_{\nu, Y(1)} \right)}{(\hat{p}(X_i))^2} + \frac{(1 - T_i) \cdot \hat{\psi}_\nu \left(Y_i; \hat{\theta}_{\nu, Y(0)} \right)}{(1 - \hat{p}(X_i))^2} \right) \cdot H_K(X_i)^\top \\ & \cdot \left(\sum_{j=1}^N H_K(X_j) \cdot H_K(X_j)^\top \right)^{-1} \cdot H_K(x) \cdot (t - \hat{p}(x)) \end{aligned}$$

The $\hat{\psi}_\nu$ function considered here converges pointwise in probability to ψ_ν . We state that as the following assumption:

ASSUMPTION 8 [Consistency of the “Proxy” for the Influence Function] Assume that (i) $\hat{\psi}_\nu(y, \theta_{Y(1)}) - \psi_\nu(y, \theta_{Y(1)}) = o_p(1)$ for any y in the support of $Y(1)$; (ii) $\hat{\psi}_\nu(y, \hat{\theta}_{Y(0)}) - \psi_\nu(y, \theta_{Y(0)}) = o_p(1)$ for any y in the support of $Y(0)$; and $\psi_\nu(\cdot, \cdot)$ is continuous in its second argument at $\theta_{Y(1)}$ and $\theta_{Y(0)}$.

Consider now the examples already discussed, the variance and the interquartile range. We can see in both cases, computation of $\widehat{\theta}_\nu$, $\widehat{\varphi}_\Delta$ and $\widehat{\alpha}_\Delta$ can be performed by the following way:

Example 1: Variance (continued):

$$\widehat{\theta}_{\nu^V} = \left[\frac{1}{N} \sum_{i=1}^N \frac{T_i}{\widehat{p}(X_i)} \cdot Y_i, \widehat{\nu}_{Y(1)}^V, \frac{1}{N} \sum_{i=1}^N \left(\frac{1-T_i}{1-\widehat{p}(X_i)} \right) \cdot Y_i, \widehat{\nu}_{Y(0)}^V \right]^\top,$$

$$\begin{aligned} \widehat{\varphi}_{\Delta,V}(y, x, t; \widehat{\theta}_{\nu^V}) &= \frac{t}{\widehat{p}(x)} \cdot \left(\left(y - \frac{1}{N} \sum_{j=1}^N \frac{T_j}{\widehat{p}(X_j)} \cdot Y_j \right)^2 - \widehat{\nu}_{Y(1)}^V \right) \\ &\quad - \left(\frac{1-t}{1-\widehat{p}(x)} \right) \cdot \left(\left(y - \frac{1}{N} \sum_{l=1}^N \left(\frac{1-T_l}{1-\widehat{p}(X_l)} \right) \cdot Y_l \right)^2 - \widehat{\nu}_{Y(0)}^V \right), \end{aligned}$$

and

$$\begin{aligned} \widehat{\alpha}_{\Delta,V}(t, x; \widehat{\theta}_{\nu^V}) &= - \sum_{i=1}^N \left(\frac{T_i}{(\widehat{p}(X_i))^2} \cdot \left(\left(Y_i - \frac{1}{N} \sum_{j=1}^N \frac{T_j}{\widehat{p}(X_j)} \cdot Y_j \right)^2 - \widehat{\nu}_{Y(1)}^V \right) \right. \\ &\quad \left. + \frac{(1-T_i)}{(1-\widehat{p}(X_i))^2} \cdot \left(\left(Y_i - \frac{1}{N} \sum_{l=1}^N \left(\frac{1-T_l}{1-\widehat{p}(X_l)} \right) \cdot Y_l \right)^2 - \widehat{\nu}_{Y(0)}^V \right) \right) \\ &\quad \cdot H_K(X_i)^\top \cdot \left(\sum_{m=1}^N H_K(X_m) \cdot H_K(X_m)^\top \right)^{-1} \cdot H_K(x) \cdot (t - \widehat{p}(x)) \end{aligned}$$

Notice that in this case, the function $\widehat{\psi}_{\nu^V}(\cdot, \cdot)$ is exactly equal to $\psi_{\nu^V}(\cdot, \cdot)$ thus, in order to check whether Assumption 8 holds, we just have to be concerned about (a) $\psi_{\nu^V}(\cdot, \cdot)$ being continuous in its second argument at the true value of θ and (b) $\widehat{\theta}_\nu - \theta_\nu = o_p(1)$. But (a) and (b) can be easily verified to hold in this case.

Example 2: The 75-25 Interquartile Range (continued):

$$\widehat{\theta}_{\nu^{75-25}} = \left[\widehat{q}_{Y(1),.1}, \widehat{\nu}_{Y(1)}^{75-25}, \widehat{q}_{Y(0),.1}, \widehat{\nu}_{Y(0)}^{75-25} \right]^\top,$$

also

$$\begin{aligned}\widehat{\varphi}_{\Delta,75-25}(y, x, t; \widehat{\theta}_{\nu^{75-25}}) &= \frac{t}{\widehat{p}(x)} \cdot \left(\frac{(.75 - \mathbb{I}\{y - (\widehat{\nu}_{Y(1)}^{75-25} + \widehat{q}_{Y(1),.25}) \leq 0\})}{\widehat{f}_{Y(1)}(\widehat{\nu}_{Y(1)}^{75-25} + \widehat{q}_{Y(1),.25})} \right. \\ &\quad \left. - \frac{(.25 - \mathbb{I}\{y - \widehat{q}_{Y(1),.25} \leq 0\})}{\widehat{f}_{Y(1)}(\widehat{q}_{Y(1),.25})} \right) \\ &\quad - \left(\frac{1-t}{1-\widehat{p}(x)} \right) \cdot \left(\frac{(.75 - \mathbb{I}\{y - (\widehat{\nu}_{Y(0)}^{75-25} + \widehat{q}_{Y(0),.25}) \leq 0\})}{\widehat{f}_{Y(0)}(\widehat{\nu}_{Y(0)}^{75-25} + \widehat{q}_{Y(0),.25})} \right. \\ &\quad \left. - \frac{(.25 - \mathbb{I}\{y - \widehat{q}_{Y(0),.25} \leq 0\})}{\widehat{f}_{Y(0)}(\widehat{q}_{Y(0),.25})} \right)\end{aligned}$$

and finally:

$$\begin{aligned}\widehat{\alpha}_{\Delta,75-25}(t, x; \widehat{\theta}_{\nu^{75-25}}) &= - \sum_{i=1}^N \left(\frac{T_i}{(\widehat{p}(X_i))^2} \cdot \left(\frac{(.75 - \mathbb{I}\{Y_i - (\widehat{\nu}_{Y(1)}^{75-25} + \widehat{q}_{Y(1),.25}) \leq 0\})}{\widehat{f}_{Y(1)}(\widehat{\nu}_{Y(1)}^{75-25} + \widehat{q}_{Y(1),.25})} \right. \right. \\ &\quad \left. \left. - \frac{(.25 - \mathbb{I}\{Y_i - \widehat{q}_{Y(1),.25} \leq 0\})}{\widehat{f}_{Y(1)}(\widehat{q}_{Y(1),.25})} \right) \right. \\ &\quad \left. + \frac{(1-T_i)}{(1-\widehat{p}(X_i))^2} \cdot \left(\frac{(.75 - \mathbb{I}\{Y_i - (\widehat{\nu}_{Y(0)}^{75-25} + \widehat{q}_{Y(0),.25}) \leq 0\})}{\widehat{f}_{Y(0)}(\widehat{\nu}_{Y(0)}^{75-25} + \widehat{q}_{Y(0),.25})} \right. \right. \\ &\quad \left. \left. - \frac{(.25 - \mathbb{I}\{Y_i - \widehat{q}_{Y(0),.25} \leq 0\})}{\widehat{f}_{Y(0)}(\widehat{q}_{Y(0),.25})} \right) \right) \cdot H_K(X_i)^\top \cdot \left(\sum_{j=1}^N H_K(X_j) \cdot H_K(X_j)^\top \right)^{-1} \cdot H_K(x) \cdot (t - \widehat{p}(x))\end{aligned}$$

where $\widehat{\theta}_{\nu^{75-25}}$ was previously defined, and $\widehat{f}_{Y(1)}(y)$ and $\widehat{f}_{Y(0)}(y)$ are estimated similarly as in DiNardo, Fortin and Lemieux (1996) as:

$$\begin{aligned}\widehat{f}_{Y(1)}(y) &= \frac{1}{N} \sum_{i=1}^N \frac{T_i}{\widehat{p}(X_i)} \cdot \frac{1}{h} \cdot K\left(\frac{Y_i - y}{h}\right) \\ \widehat{f}_{Y(0)}(y) &= \frac{1}{N} \sum_{i=1}^N \left(\frac{1-T_i}{1-\widehat{p}(X_i)} \right) \cdot \frac{1}{h} \cdot K\left(\frac{Y_i - y}{h}\right)\end{aligned}$$

where $K(\cdot)$ is a kernel function and h is a bandwidth. Under typical conditions for (a) kernel density estimation to be consistent; (b) the estimators $\widehat{\nu}_{Y(1)}^{75-25}$ and $\widehat{q}_{Y(1),.25}$ to be consistent for $\nu_{Y(1)}^{75-25}$ and $q_{Y(1),.25}$, Assumption 8 will hold for this case.

We now state the result regarding the consistency of the variance estimator.

THEOREM 2 *Under Assumptions 1-8, $\widehat{\sigma}_{\Delta_\nu}^2 - \sigma_{\Delta_\nu}^2 = o_p(1)$.*

5 A Monte Carlo Exercise

In this section we report the results of Monte Carlo exercises. The interest is in learning how the estimators for the overall inequality treatment effect (OITE) and estimators of their asymptotic variances behave in finite samples. The generated data follows a very simple specification: $X = [X_1, X_2]^T \sim \text{Bivariate } N([\mu_{X_1}, \mu_{X_2}]^T, \Omega_X)$, $T = \mathbb{I}\{\delta_0 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_1^2 + \eta > 0\}$ where η has a standard logistic c.d.f. $F_\eta(n) = (1 + \exp(-\pi n/\sqrt{3}))^{-1}$, $Y(0) = \theta_1 X_1 + \theta_2 X_2 + \gamma_0 \epsilon$ and $Y(1) = \beta + \theta_1 X_1 + \theta_2 X_2 + \gamma_1 \epsilon$, where ϵ is distributed as $N(\mu_\epsilon, \sigma_\epsilon^2)$. The variables X , η , ϵ_0 and ϵ_1 are mutually independent. The parameters were chosen to be: $\mu_{X_1} = 1$, $\mu_{X_2} = 5$, $\Omega_{X,11} = \sigma_{X_1}^2 = 1$, $\Omega_{X,22} = \sigma_{X_2}^2 = 1$, $\Omega_{X,12} = \sigma_{X_1, X_2} = 0$, $\delta_0 = -1$, $\delta_1 = 5$, $\delta_2 = -5$, $\delta_3 = -.05$, $\theta_1 = -5$, $\theta_2 = 1$, $\beta = 5$, $\mu_\epsilon = 0$, $\sigma_\epsilon^2 = 1$, $\gamma_0 = 5$ and $\gamma_1 = .5$. Under this specification, $Y(1) \sim N(5, 26.25)$ and $Y(0) \sim N(0, 51)$. This specification leads to selection on observables, and estimation of treatment effects not taking the selection into account will inevitably produce inconsistent estimates.

One hundred replications of this experiment with three different sample sizes were considered: 100, 1,000 and 10,000 observations. As $Y(1)$ and $Y(0)$ are known for each observation i , we can also compute “unfeasible” estimators of parameters of the marginal distributions of $Y(1)$ and $Y(0)$.

Results can be found in Tables 1-4. Each one of Tables 1-3 reports estimates of different treatment effect parameters; they differ in the sample size considered. In each of those tables, we report average, variance, quantile (lower quartile, median, upper quartile) and inter-quartile range treatment effects. Analytical standard errors are also computed and they appear on the right side of each of those tables.¹⁴

The results indicate that the reweighting estimator performs well according to the MSE criteria. Also, looking separately at bias and variance terms, it is clear that the bias vanishes relatively fast as the sample increases for all of the parameters being estimated by the reweighting method, the same not occurring with the “wrong” estimator, which is basically an estimator of $\nu(F_{Y(1)|T=1}) - \nu(F_{Y(0)|T=0})$. Analytical standard errors tend to be (either looking at the average or at the median) close to the bootstrapped standard errors for all sample sizes and all

¹⁴The polynomial order for both the propensity-score estimation and the asymptotic variance estimation can be determined by cross-validation. In order to simplify our computations we fixed the propensity-score to have the true polynomial structure and also used it for the variance estimation.

parameters. This indicates that bootstrapping may be a good alternative to analytical standard errors estimation.

There is one important point that can be seen clearly through the analytical standard errors formulae and which reveals itself to be relevant in the Monte Carlo exercise. In samples with “weak” common support, that is, when the estimated propensity-score may assume values “close” to either 0 or 1, the standard errors will reflect that situation. From Equation 12 we can see for the treated, a low propensity score will increase standard errors, the same happening with a high propensity-score for the untreated. In Table 4 we report the minimum and maximum values among all samples and replications for a given sample size respectively for the treated and the control units. As expected, higher biases and larger standard errors tend to occur when there is weak common support problems.

This last fact allows us to draw an analogy with the famous discussion of multicollinearity in Goldberger’s textbook, a discussion also known by introducing the “micronumerosity” problem in the literature.¹⁵ Samples with poor overlapping in terms of covariates among treated and controls will produce imprecise and uninformative estimates of treatment effects, the same happening with multicollinear regressors and regression coefficient estimates. Trimming with respect to distribution of the propensity-score has the same flavour of dropping a regressor that is highly correlated with some other one in order to “solve” the multicollinearity problem: It leads to an estimate of another parameter, not the one we were initially interested in. The assertion that “more data will be of no help if it is more of the same” is also true here: More data will be of no help if we still have no or little intersection in the supports of the covariates empirical distributions given treatment assignment. Finally, putting into brackets our comments so we can draw the analogy more precisely, we cite Goldberger’s first remark about multicollinearity on page 251: “Researchers should not be concerned with whether or not “there really is collinearity” [no-overlapping in the support in our case]. They may well be concerned with whether the variances of the coefficient estimates are too large - for whatever reason - to provide useful estimates of the regression coefficients [treatment effects in our case].” We may conclude then that the claim that reweighting method is of little importance because it does not solve the common support problem sounds similar to say that OLS is of little importance because it is not robust to multicollinearity.

¹⁵See Goldberger (1991), chapter 23.

6 Conclusion

In this paper we proposed estimators for the effects of a treatment on some inequality measures. This was achieved by first estimating, through a reweighing method, the inequality measures of the potential outcomes, and then taking the difference between those estimates. This estimation strategy is useful for policy-making purposes when the individual decision to participate into the social program (the treatment) depends on observable characteristics. If the identification restrictions hold, then the reweighing method allows identify the distribution of potential outcomes and, therefore, many of their inequality parameters.

We showed that the inequality treatment effect estimators are root- N consistent and asymptotically normal. We also calculated the semiparametric efficiency bounds and proved that the proposed estimators achieve them.

Finally, we performed a series of Monte Carlo exercises. The reweighing estimator performed well in terms of the MSE criteria. It also seemed that finite sample bias vanishes fast as the sample size increases. Analytical standard errors behaved similarly to bootstrapping ones, revealing that bootstrap may be a valid alternative. However, the interpretation gain from the analytical expression is important: if the covariates in the treated just “weakly” share the same support of distribution with covariates in the control group, then the standard errors of the reweighing estimator will reflect the fact that the comparisons across groups will be poor and that treatment effects estimates will not be very useful.

REFERENCES

- ABADIE, A., J. ANGRIST, AND G. IMBENS, (2002), "Instrumental Variables Estimation of Quantile Treatment Effects," *Econometrica*, 70, 91-117.
- BICKEL, P., C. KLASSEN, Y. RITOV, AND J. WELLNER, (1993), *Efficient and Adaptive Estimation for Semiparametric Models*. New York, Springer-Verlag.
- BITLER, M., J. GELBACH, AND H. HOYNES, (2004), "What Mean Impacts MissL Distributional Effects of Welfare Reform Experiments," preprint.
- CHEN, X., H. HONG, AND A. TAROZZI, (2004), "Semiparametric Efficiency in GMM Models of Nonclassical Measurement Errors, Missing Data and Treatment Effects," preprint.
- COWELL, F., (2000), "Measurement of Inequality," in *Handbook of Income Distribution*, ed. by A. Atkinson, and F. Bourguignon, North Holland, Amsterdam.
- COWELL, F., (2003), "Theil, Inequality and the Structure of Income Distribution," *Distributional Analysis Discussion Paper, STICERD, LSE*, No. 67.
- DiNARDO, J., N. FORTIN, AND T. LEMIEUX, (1996), "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach," *Econometrica*, 64, 1001-1044.
- FIRPO, S. (2004), "Efficient Semiparametric Estimation of Quantile Treatment Effects," *UBC Department of Economics Discussion Paper*, No. 04-01.
- GOLDBERGER, A. (1991), *A Course in Econometrics*. Cambridge, Harvard University Press.
- HAHN, J., (1998), "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66, 315-331.
- HECKMAN, J., AND B. HONORÉ (1990), "The Empirical Content of the Roy Model," *Econometrica*, 58, 1121-1149.
- HECKMAN, J., H. ICHIMURA, AND P. TODD, (1997), "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65(2), 261-294.

- HECKMAN, J., H. ICHIMURA, J. SMITH, AND P. TODD, (1998), "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017-1098.
- HECKMAN, J., J. SMITH, AND N. CLEMENTS, (1997), "Making the Most out of Programme Evaluations and Social Experiments Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64(4), 487-535.
- HIRANO, K., G. IMBENS, AND G. RIDDER, (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161-1189.
- HUBER, P. (1981), *Robust Statistics*. New York, John Wiley & Sons.
- IMBENS, G., AND D. RUBIN, (1997), "Estimating Outcome Distributions for Compliers in Instrumental Variable Models," *Review of Economic Studies*, October, 555-574.
- MANSKI, C. (1997), "The Mixing Problem in Programme Evaluation," *Review of Economic Studies*, October, 537-554.
- NEWHEY, W., (1990), "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99-135.
- NEWHEY, W., (1994), "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349-1382.
- NEWHEY, W., (1995), "Convergence Rates for Series Estimators," in *Advances in Econometrics and Qualitative Economics: Essays in Honor of C.R. Rao*, G. Maddala, P.C. Phillips, and T.N. Srinivasan, eds., Cambridge US, Basil-Blackwell.
- NEWHEY, W., (1997), "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147-168.
- ROY, A., (1951), "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135-146.
- ROSENBAUM, P., AND D. RUBIN, (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.

- ROSENBAUM, P., AND D. RUBIN, (1984), "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score," *Journal of the American Statistical Association*, 79, 516-574.
- RUBIN, D., (1977), "Assignment to Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 2(1), 1-26.
- VAN DER VAART, A. (1998), *Asymptotic Statistics*. Cambridge, Cambridge University Press.

APPENDIX: Proofs

Proof of Lemma 1:

Fix y , where $y \in \text{Supp}(Y(1))$, the support of $Y(1)$. Let us work with $F_{Y(1)}(y)$ as the other c.d.f.'s follow by simple analogy. $F_{Y(1)}(y) = \Pr[Y(1) \leq y] = \mathbb{E}[\Pr[Y(1) \leq y | X]] = \mathbb{E}[\Pr[Y(1) \leq y | X, T = 1]] = \mathbb{E}[\Pr[Y \leq y | X, T = 1]] = \mathbb{E}[\mathbb{E}[T \mathbb{I}\{Y \leq y\} | X, T = 1]] = \mathbb{E}\left[\frac{1}{p(X)} \mathbb{E}[T \mathbb{I}\{Y \leq y\} | X]\right] = \mathbb{E}\left[\frac{T}{p(X)} \mathbb{I}\{Y \leq y\}\right].$

The first equality follows from the definition of the c.d.f.. The second is an application of the law of iterated expectations. The third equality follows from the ignorability assumption (Assumption 1). The fourth results from the definition of $Y = TY(1) + (1 - T)Y(0)$. The fifth equality comes from $\mathbb{E}[\mathbb{I}\{A\}] = \Pr[A]$ (where A is some event) and from the fact that the expectation is conditional on $T = 1$. The sixth is a consequence from $\mathbb{E}[Z | X] = p(X)\mathbb{E}[Z | X, T = 1] + (1 - p(X))\mathbb{E}[Z | X, T = 0]$, where Z is some random variable with finite variance and from the common support assumption (Assumption 2). Finally, the last equality is a backward application of the law of iterated expectations. Analogous results for $F_{Y(0)}(y)$ and $F_{Y(0)|T=1}(y)$ could have been derived following essentially the same steps as above. The result for $F_{Y(1)|T=1}(y)$ could have been easily proved but with not using the ignorability assumption as that c.d.f. is identifiable from data on Y for $T = 1$. ■

Proof of Corollary 1:

Let us define $F_Y^{\omega: T/p(X)}$ and $F_Y^{\omega: (1-T)/(1-p(X))}$ the reweighted distribution functions of Y using $T/p(X)$ and $(1 - T)/(1 - p(X))$ as weights. From Lemma 1, $F_Y^{\omega: T/p(X)} = F_{Y(1)}$ and $F_Y^{\omega: (1-T)/(1-p(X))} = F_{Y(0)}$. From Assumption 3, as $F_{Y(1)} = F_Y^{\omega: T/p(X)}$, we have that $\nu(F_{Y(1)}) = \nu(F_Y^{\omega: T/p(X)})$; and as $F_{Y(0)} = F_Y^{\omega: (1-T)/(1-p(X))}$, we have that $\nu(F_{Y(0)}) = \nu(F_Y^{\omega: (1-T)/(1-p(X))})$. Therefore, Δ_ν is identified from data on (Y, T, X) . The same holds by analogy to $\Delta_{\nu|T=1}$. ■

Proof of Lemma 2:

See Hirano, Imbens and Ridder (2003), Lemmas 1 and 2. ■

Proof of Theorem 1: [Part (i)]

Let us again concentrate on $\widehat{\nu}_{Y(1)}$, since extensions to other functionals estimates follow again by simple analogy.

Under Assumption 6, the unfeasible estimator of $\nu_{Y(1)}$ is asymptotically linear: $\nu(\widehat{F}_{Y(1)}^U) - \nu(F_{Y(1)}) - \frac{1}{N} \sum_{i=1}^N \psi_\nu(Y_i(1); \theta_{\nu, Y(1)}) = o_p(1/\sqrt{N})$. Now consider that we have a new sampling scheme, based on ignorability. Thus, we just replace $\widehat{F}_{Y(1)}^U$, the empirical distribution under random sampling or full observability of $Y(1)$, by $\widehat{F}_{Y(1)}$, the empirical distribution under ignorability. We have that:

$$\nu(\widehat{F}_{Y(1)}) - \nu(F_{Y(1)}) - \frac{1}{N} \sum_{i=1}^N \frac{T_i}{\widehat{p}(X_i)} \cdot \psi_\nu(Y_i; \theta_{\nu, Y(1)}) = o_p(1/\sqrt{N}). \quad (\text{A-1})$$

Now, define

$$\begin{aligned} \xi_N = & \frac{1}{N} \sum_{i=1}^N \left(\frac{T_i}{\widehat{p}(X_i)} - \frac{T_i}{p(X_i)} \right) \cdot \psi_\nu(Y_i; \theta_{\nu, Y(1)}) \\ & + \frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) \mid X_i, T=1]}{p(X_i)} \cdot (T_i - p(X_i)) \end{aligned}$$

Thus Equation A-1 can be rewritten as:

$$\begin{aligned} \nu(\widehat{F}_{Y(1)}) - \nu(F_{Y(1)}) - \frac{1}{N} \sum_{i=1}^N \frac{T_i}{p(X_i)} \cdot \psi_\nu(Y_i; \theta_{\nu, Y(1)}) \\ + \frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) \mid X_i, T=1]}{p(X_i)} \cdot (T_i - p(X_i)) = \xi_N + o_p(1/\sqrt{N}). \end{aligned}$$

And we need to show that $\xi_N = o_p(1/\sqrt{N})$. In order to show that, rewrite ξ_N as:

$$\xi_N = \frac{1}{N} \sum_{i=1}^N \left(\frac{T_i \cdot \psi_\nu(Y_i; \theta_{\nu, Y(1)})}{\widehat{p}(X_i)} - \frac{T_i \cdot \psi_\nu(Y_i; \theta_{\nu, Y(1)})}{p(X_i)} + \frac{T_i \cdot \psi_\nu(Y_i; \theta_{\nu, Y(1)})}{p^2(X_i)} (\widehat{p}(X_i) - p(X_i)) \right) \quad (\text{A-2})$$

$$- \frac{1}{N} \sum_{i=1}^N \left(\frac{T_i \cdot \psi_\nu(Y_i; \theta_{\nu, Y(1)})}{p^2(X_i)} (\widehat{p}(X_i) - p(X_i)) \right) + \mathbb{E} \left[\frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) \mid X, T=1]}{p(X)} (\widehat{p}(X) - p(X)) \right] \quad (\text{A-3})$$

$$- \mathbb{E} \left[\frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) \mid X, T=1]}{p(X)} (\widehat{p}(X) - p(X)) \right] - \frac{1}{N} \sum_{i=1}^N \tilde{\delta}(X_i, u) \frac{T_i - p_K(X_i)}{\sqrt{p_K(X_i)(1 - p_K(X_i))}} \quad (\text{A-4})$$

$$+ \frac{1}{N} \sum_{i=1}^N (\tilde{\delta}(X_i, u) - \delta_K(X_i, u)) \frac{T_i - p_K(X_i)}{\sqrt{p_K(X_i)(1 - p_K(X_i))}} \quad (\text{A-5})$$

$$+ \frac{1}{N} \sum_{i=1}^N \delta_K(X_i, u) \frac{T_i - p_K(X_i)}{\sqrt{p_K(X_i)(1 - p_K(X_i))}} - \frac{1}{N} \sum_{i=1}^N \delta(X_i, u) \frac{T_i - p(X_i)}{\sqrt{p(X_i)(1 - p(X_i))}} \quad (\text{A-6})$$

where:

$$\begin{aligned}\tilde{\delta}(X_i, u) &= -\mathbb{E} \left[\frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) \mid X, T = 1]}{p(X)} L'(H_K(X)' \tilde{\pi}) H_K(X)' \right] \tilde{\Sigma}^{-1} \sqrt{L'(H_K(X_i)' \pi_K) H_K(X_i)} \\ \delta_K(X_i, u) &= -\mathbb{E} \left[\frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) \mid X, T = 1]}{p(X)} L'(H_K(X)' \pi_K) H_K(X)' \right] \Sigma_K^{-1} \sqrt{L'(H_K(X_i)' \pi_K) H_K(X_i)} \\ \delta(X_i, u) &= -\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) \mid X_i, T = 1] \frac{\sqrt{p(X_i)(1 - p(X_i))}}{p(X_i)} \\ \tilde{\Sigma} &= \frac{1}{N} \sum_{i=1}^N H_K(X_i) H_K(X_i)' L'(H_K(X_i)' \tilde{\pi})\end{aligned}$$

and

$$\Sigma = \mathbb{E}[H_K(X) H_K(X)' L'(H_K(X)' \pi_K)].$$

Hirano, Imbens and Ridder (2003) have computed their first step in the exact same way we do. Also, in their Theorem 1 they have a remainder term to bound very similar to ξ_N . The main difference is that instead of $\psi_\nu(Y; \theta_{\nu, Y(1)})$ they have Y , where $\mathbb{E}[Y^2]$ is assumed to be finite and have continuously differentiable conditional expectation given X and $T = 1$. We have made a similar assumption, but to the influence function $\psi_\nu(Y; \theta_{\nu, Y(1)})$, by imposing Assumption 5. It is then possible to bound ξ_N using exactly the same arguments they used. Firpo (2004) has shown how to draw the analogy between HIR's proof for the average treatment effects and the quantile treatment effects. In order to avoid repetition of the same argument, we refer the reader to their proofs (Lemma 3 in Firpo (2004), Theorem 1 in Hirano, Imbens and Ridder (2003)). A complete proof is, however, available upon request. Therefore, replacing $Y - \beta$ by $\psi_\nu(Y; \theta_{\nu, Y(1)})$ in HIR, we can conclude that $\xi_N = o_p(1/\sqrt{N})$.

Extending the result to $\hat{\nu}_{Y(0)}$ is trivial. Thus, making use of the fact that Δ_ν is just the difference between $\nu_{Y(1)}$ and $\nu_{Y(0)}$ we can conclude that: $\sqrt{N}(\hat{\Delta}_\nu - \Delta_\nu) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \psi_\Delta(Y_i, X_i, T_i; \theta_\nu) + o_p(1) \xrightarrow{D} N(0, \sigma_{\Delta_\nu}^2)$. ■

Proof of Theorem 1: [Part (ii)]

We have now to show that $\sigma_{\Delta_\nu}^2$ is indeed the semiparametric efficiency bound of Δ_ν . This proof is an extension to this general parameter case of the proofs by Hahn (1998) and Hirano, Imbens and Ridder (2003) for the mean case and by Firpo (2004) for the quantile case. These references use the machinery presented by Bickel, Klassen, Ritov, and Wellner (1993), Newey (1990) and Newey (1994).

By the same reason as before, we concentrate attention to $\nu(F_{Y(1)})$ and its semiparametric efficiency bound. An extension to $\nu(F_{Y(0)})$ and therefore to the treatment effect Δ_ν will follow by analogy.

Let us divide the problem into two parts. In the first part we discuss the model in which there is full observability of $Y(1)$; in the second part we discuss what happens when we have ignorability instead.

Under full observability the tangent space is characterized by \mathcal{S}_1^{FULL} , which has been defined previously. Any efficient estimator of parameters of this model will have to have its asymptotic influence function belonging to \mathcal{S}_1^{FULL} . By Assumption 7 $\psi_\nu(y; \theta_{Y(1)}) = \phi_\nu(y, t, x; \theta_{Y(1)}) \in \mathcal{S}_1^{FULL}$. As $\psi_\nu(y; \theta_{Y(1)})$ does not depend on (t, x) we conclude that it has to satisfy $0 = \mathbb{E}[\psi_\nu(Y(1); \theta_{Y(1)})]$. Also, because $\nu(\widehat{F}_{Y(1)}^U)$ is asymptotically linear and regular by Assumption 6, then by BKRW's Proposition 3.3.1 ν is pathwise differentiable, that is, for any regular parametric sub-model indexed by a finite dimensional vector ϑ , we have:

$$\begin{aligned} & \nu(F_{Y(1)}(\vartheta)) - \nu(F_{Y(1)}) \\ &= \int \psi_\nu(y; \theta_{Y(1)}) \cdot dF_{Y(1)}(y; \vartheta) - \int \psi_\nu(y; \theta_{Y(1)}) \cdot dF_{Y(1)}(y) + o(\|\vartheta - \vartheta_0\|) \\ &= \int \psi_\nu(y; \theta_{Y(1)}) \cdot dF_{Y(1)}(y; \vartheta) + o(\|\vartheta - \vartheta_0\|) \end{aligned}$$

where we have used the following notational normalization: $F_{Y(1)}(\vartheta_0) = F_{Y(1)}$, and where ϑ_0 is the true population parameter. Thus, the derivative of $\nu(F_{Y(1)}(\vartheta))$ with respect to ϑ evaluated at ϑ_0 has to be equal to:

$$\frac{\partial \nu(F_{Y(1)}(\vartheta_0))}{\partial \vartheta} = \int \psi_\nu(y; \theta_{Y(1)}) \cdot s_1(y) \cdot dF_{Y(1)}(y)$$

where $s_1(y)$ is the score function of the distribution $F_{Y(1)}$ evaluated at the true parameter model.

Under ignorability, start defining the densities, with respect to some σ -finite measure, of $(Y(1), T, X)$ and of the observed data (Y, T, X) . Under Assumption 1, both densities represent the same statistical model and are, therefore, equivalent. These densities can be written as $f_{Y(1), T, X}(y, t, x) = (f_{Y(1)|X}(y|x) \cdot p(x))^t \cdot f_X(x)$ and $f_{Y, T, X}(y, t, x) = (f_{Y|X, T=1}(y|x) \cdot p(x))^t \cdot f_X(x)$. Working with the density of observed data, consider the regular parametric sub-model indexed by ϑ :

$$f_{Y, T, X}(y, t, x | \vartheta) = (f_{Y|X, T=1}(y|x; \vartheta) \cdot p(x | \vartheta))^t \cdot f_X(x | \vartheta),$$

By a normalization argument, let $f_{Y,T,X}(y, t, x) = f_{Y,T,X}(y, t, x | \vartheta_0)$. The score of a parametric sub-model indexed by ϑ is given by:

$$s(y, t, x | \vartheta) = t \cdot s_1(y | x; \vartheta) + \frac{t}{p(x | \vartheta)} \cdot p'(x | \vartheta) + s_X(x | \vartheta)$$

where, for $s_1(y | x; \vartheta) = \frac{\partial}{\partial \theta} \log f_{Y|X,T=1}(y | x; \vartheta)$; $p'(x | \vartheta) = \frac{\partial}{\partial \theta} p(x | \vartheta)$; and $s_X(x | \vartheta) = \frac{\partial}{\partial \theta} \log f(x | \vartheta)$. Again we normalize: $s(y, t, x) = s(y, t, x | \vartheta_0)$. In order to find the efficient influence functions of the parameters of interest, $\nu(F_{Y(1)})$ we need first to define the tangent space of this statistical model. This will be the set \mathcal{S} of all possible score functions, and it is defined as: $\mathcal{S} = \{S : \mathbb{R} \times \{0, 1\} \times \mathcal{X} \rightarrow \mathbb{R} \mid S(y, t, x) = t \cdot s_1(y | x) + a(x) \cdot t + s_X(x); \text{ and } \mathbb{E}[s_1(Y | X) | X = x, T = 1] = \mathbb{E}[s_X(X)] = 0, \forall x \text{ and where } a(x) \text{ is square-integrable function of } x\}$. Next note that ν is still pathwise differentiable under the ignorability model, as $\nu(\widehat{F}_{Y(1)})$ is a regular asymptotically linear estimator of $\nu(F_{Y(1)})$ as consequence of part (i) of Theorem 1. Thus $\nu(F_{Y(1)}(\vartheta))$ can be written as:

$$\begin{aligned} & \nu(F_{Y(1)}(\vartheta)) - \nu(F_{Y(1)}) \\ &= \int \phi_\nu^*(y; \theta_{Y(1)}) \cdot dF_{Y(1)}(y; \vartheta) - \int \phi_\nu^*(y; \theta_{Y(1)}) \cdot dF_{Y(1)}(y) + o(\|\vartheta - \vartheta_0\|) \\ &= \int \int \psi_\nu^*(y, t, x; \theta_{Y(1)}) \cdot f_{Y|X,T=1}(y | x; \vartheta) \cdot f(x | \vartheta) \cdot dy \cdot dx + o(\|\vartheta - \vartheta_0\|) \end{aligned}$$

where $\phi_\nu^*(y; \theta_{Y(1)})$ is a bounded linear functional of $F_{Y(1)}$ such that

$$0 = \mathbb{E}[\phi_\nu^*(Y; \theta_{Y(1)})] = \mathbb{E}[\mathbb{E}[\psi_\nu^*(Y, T, X; \theta_{Y(1)}) | X, T = 1]] .$$

Thus, the derivative of $\nu(F_{Y(1)}(\vartheta))$ with respect to ϑ evaluated at ϑ_0 has to be equal to:

$$\begin{aligned} \frac{\partial \nu(F_{Y(1)}(\vartheta_0))}{\partial \vartheta} &= \int \int \psi_\nu^*(y, t, x; \theta_{Y(1)}) \cdot s_1(y | x) \cdot f_{Y|X,T=1}(y | x) \cdot f(x) \cdot dy \cdot dx \\ &\quad + \int \left(\int \psi_\nu^*(y, t, x; \theta_{Y(1)}) \cdot f_{Y|X,T=1}(y | x) \cdot dy \right) \cdot s_X(x) \cdot f(x) \cdot dx. \end{aligned}$$

We know that $\psi_\nu^*(y, t, x; \theta_{Y(1)})$ will be the efficient influence function of $\nu(F_{Y(1)})$ if $\psi_\nu^*(y, t, x; \theta_{Y(1)}) \in \mathcal{S}$. By the uniqueness of the efficient influence function, $\phi_\nu^*(y; \theta_{Y(1)}) = \psi_\nu(y; \theta_{Y(1)})$, and therefore, a natural candidate for $\psi_\nu^*(y, t, x; \theta_{Y(1)})$ is:

$$\psi_\nu^*(y, t, x; \theta_{Y(1)}) = \frac{t}{p(x)} \cdot (\psi_\nu(y; \theta_{\nu, Y(1)}) - \mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) | x, T = 1]) + \mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) | x, T = 1] \quad (\text{A-7})$$

which belongs to the set of scores and is, indeed, the influence function of $\nu(\widehat{F}_{Y(1)})$. Thus, $\nu(\widehat{F}_{Y(1)})$ is the efficient estimator of $\nu(F_{Y(1)})$ under the class of semiparametric estimators for the statistical model under ignorability sampling.

Applying the same reasoning to $\nu(\widehat{F}_{Y(0)})$ we could show that $\nu(\widehat{F}_{Y(0)})$ is efficient. We omit that part of the proof as it follows once more by analogy. We are now able to conclude that:

$$\begin{aligned}
& AVar \left[\sqrt{N} (\widehat{\Delta}_\nu - \Delta_\nu) \right] \\
&= AVar \left[\sqrt{N} \left(\nu(\widehat{F}_{Y(1)}) - \nu(\widehat{F}_{Y(0)}) - \Delta_\nu \right) \right] \\
&= AVar \left[\frac{1}{\sqrt{N}} \sum_{i=1}^N \psi_\nu^*(Y_i, X_i, T_i; \boldsymbol{\theta}_{Y(1)}) - \psi_\nu^*(Y_i, X_i, T_i; \boldsymbol{\theta}_{Y(0)}) \right] \\
&= AVar \left[\frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{T_i}{p(X_i)} \cdot \psi_\nu(Y_i; \boldsymbol{\theta}_{Y(1)}) - \frac{\mathbb{E}[\psi_\nu(Y; \boldsymbol{\theta}_{Y(1)}) \mid X_i, T=1]}{p(X_i)} \cdot (T_i - p(X_i)) \right. \\
&\quad \left. - \left(\frac{1 - T_i}{1 - p(X_i)} \right) \cdot \psi_\nu(Y_i; \boldsymbol{\theta}_{Y(0)}) + \frac{\mathbb{E}[\psi_\nu(Y; \boldsymbol{\theta}_{Y(0)}) \mid X_i, T=0]}{1 - p(X_i)} \cdot (T_i - p(X_i)) \right] \\
&= \mathbb{E} \left[(\psi_\Delta(Y, X, T; \boldsymbol{\theta}))^2 \right] \\
&= \sigma_{\Delta_\nu}^2.
\end{aligned}$$

That is, the asymptotic variance of $\widehat{\Delta}_\nu$ reaches the efficiency bound. ■

Proof of Theorem 2:

The goal here is to show that $\widehat{\sigma}_{\Delta_\nu}^2 = \frac{1}{N} \sum_{i=1}^N \left(\widehat{\varphi}_\Delta(Y_i, X_i, T_i; \widehat{\boldsymbol{\theta}}_\nu) + \widehat{\alpha}_\Delta(T_i, X_i; \widehat{\boldsymbol{\theta}}_\nu) \right)^2$ is $o_p(1)$ away from $\sigma_{\Delta_\nu}^2$. Thus, let us write $|\widehat{\sigma}_{\Delta_\nu}^2 - \sigma_{\Delta_\nu}^2|$ as:

$$\begin{aligned}
& |\hat{\sigma}_{\Delta_\nu}^2 - \sigma_{\Delta_\nu}^2| \\
= & \left| \frac{1}{N} \sum_{i=1}^N \left(\hat{\varphi}_\Delta \left(Y_i, X_i, T_i; \hat{\boldsymbol{\theta}}_\nu \right) + \hat{\alpha}_\Delta \left(T_i, X_i; \hat{\boldsymbol{\theta}}_\nu \right) \right)^2 \right. \\
& \left. - \frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[\left(\varphi_\Delta \left(Y, X, T; \boldsymbol{\theta}_\nu \right) + \alpha_\Delta \left(T, X; \boldsymbol{\theta}_\nu \right) \right)^2 \right] \right| \\
\leq & \left| \frac{1}{N} \sum_{i=1}^N \left(\hat{\varphi}_\Delta \left(Y_i, X_i, T_i; \hat{\boldsymbol{\theta}}_\nu \right) + \hat{\alpha}_\Delta \left(T_i, X_i; \hat{\boldsymbol{\theta}}_\nu \right) + \mathbb{E} [\varphi_\Delta \left(Y, X, T; \boldsymbol{\theta}_\nu \right)] + \mathbb{E} [\alpha_\Delta \left(T, X; \boldsymbol{\theta}_\nu \right)] \right) \right| \\
& \cdot \left| \frac{1}{N} \sum_{i=1}^N \left(\hat{\varphi}_\Delta \left(Y_i, X_i, T_i; \hat{\boldsymbol{\theta}}_\nu \right) - \mathbb{E} [\varphi_\Delta \left(Y, X, T; \boldsymbol{\theta}_\nu \right)] + \hat{\alpha}_\Delta \left(T_i, X_i; \hat{\boldsymbol{\theta}}_\nu \right) - \mathbb{E} [\alpha_\Delta \left(T, X; \boldsymbol{\theta}_\nu \right)] \right) \right| \\
\leq & \left(\frac{1}{N} \sum_{i=1}^N |\hat{\varphi}_\Delta \left(Y_i, X_i, T_i; \hat{\boldsymbol{\theta}}_\nu \right) - \varphi_\Delta \left(Y_i, X_i, T_i; \hat{\boldsymbol{\theta}}_\nu \right)| + \frac{1}{N} \sum_{i=1}^N |\hat{\alpha}_\Delta \left(T_i, X_i; \hat{\boldsymbol{\theta}}_\nu \right) - \alpha_\Delta \left(T_i, X_i; \hat{\boldsymbol{\theta}}_\nu \right)| \right. \\
& + \frac{1}{N} \sum_{i=1}^N |\varphi_\Delta \left(Y_i, X_i, T_i; \hat{\boldsymbol{\theta}}_\nu \right) + \mathbb{E} [\varphi_\Delta \left(Y, X, T; \boldsymbol{\theta}_\nu \right)]| \\
& + \frac{1}{N} \sum_{i=1}^N |\alpha_\Delta \left(T_i, X_i; \hat{\boldsymbol{\theta}}_\nu \right) + \mathbb{E} [\alpha_\Delta \left(T, X; \boldsymbol{\theta}_\nu \right)]| \Bigg) \\
& \cdot \left(\frac{1}{N} \sum_{i=1}^N |\hat{\varphi}_\Delta \left(Y_i, X_i, T_i; \hat{\boldsymbol{\theta}}_\nu \right) - \varphi_\Delta \left(Y_i, X_i, T_i; \hat{\boldsymbol{\theta}}_\nu \right)| \right. \\
& + \frac{1}{N} \sum_{i=1}^N |\varphi_\Delta \left(Y_i, X_i, T_i; \hat{\boldsymbol{\theta}}_\nu \right) - \mathbb{E} [\varphi_\Delta \left(Y, X, T; \boldsymbol{\theta}_\nu \right)]| \\
& + \frac{1}{N} \left(\sum_{i=1}^N |\hat{\alpha}_\Delta \left(T_i, X_i; \hat{\boldsymbol{\theta}}_\nu \right) - \alpha_\Delta \left(T_i, X_i; \hat{\boldsymbol{\theta}}_\nu \right)| + \sum_{i=1}^N |\alpha_\Delta \left(T_i, X_i; \hat{\boldsymbol{\theta}}_\nu \right) - \mathbb{E} [\alpha_\Delta \left(T, X; \boldsymbol{\theta}_\nu \right)]| \right) \Bigg)
\end{aligned}$$

Now, doing the last three lines term by term:

$$\begin{aligned}
& \frac{1}{N} \sum_{i=1}^N |\widehat{\varphi}_{\Delta}(Y_i, X_i, T_i; \widehat{\boldsymbol{\theta}}_{\nu}) - \varphi_{\Delta}(Y_i, X_i, T_i; \widehat{\boldsymbol{\theta}}_{\nu})| \\
& \leq \frac{1}{N} \sum_{i=1}^N \left| \frac{T_i}{\widehat{p}(X_i)} \cdot \widehat{\psi}_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) - \frac{T_i}{p(X_i)} \cdot \psi_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) \right| \\
& \quad + \frac{1}{N} \sum_{i=1}^N \left| \left(\frac{1-T_i}{1-\widehat{p}(X_i)} \right) \cdot \widehat{\psi}_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(0)}) - \left(\frac{1-T_i}{1-p(X_i)} \right) \cdot \widehat{\psi}_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(0)}) \right| \\
& \leq \frac{1}{N} \sum_{i=1}^N \left| \frac{T_i}{\widehat{p}(X_i) \cdot p(X_i)} \cdot \left(\widehat{\psi}_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) \cdot p(X_i) - \psi_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) \cdot \widehat{p}(X_i) \right) \right| \\
& \quad + \frac{1}{N} \sum_{i=1}^N \left| \left(\frac{1-T_i}{(1-\widehat{p}(X_i)) \cdot (1-p(X_i))} \right) \right. \\
& \quad \cdot \left(\widehat{\psi}_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(0)}) \cdot (1-\widehat{p}(X_i)) - \widehat{\psi}_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(0)}) \cdot (1-p(X_i)) \right) \left. \right| \\
& \leq \left(\inf_{x \in \mathcal{X}} (\widehat{p}(x)) \cdot \inf_{x \in \mathcal{X}} (p(x)) \right)^{-1} \cdot \frac{1}{N} \cdot \sum_{i=1}^N \left| \left(\widehat{\psi}_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) - \psi_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) \right) \cdot p(X_i) \right| \\
& \quad + \left(\inf_{x \in \mathcal{X}} (\widehat{p}(x)) \cdot \inf_{x \in \mathcal{X}} (p(x)) \right)^{-1} \cdot \frac{1}{N} \cdot \sum_{i=1}^N \left| (\widehat{p}(X_i) - p(X_i)) \cdot \psi_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) \right| \\
& \quad + \left(\sup_{x \in \mathcal{X}} (1-\widehat{p}(x)) \cdot \sup_{x \in \mathcal{X}} (1-p(x)) \right)^{-1} \cdot \frac{1}{N} \cdot \sum_{i=1}^N \left| \left(\widehat{\psi}_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) - \psi_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) \right) \cdot (1-p(X_i)) \right| \\
& \quad + \left(\sup_{x \in \mathcal{X}} (1-\widehat{p}(x)) \cdot \sup_{x \in \mathcal{X}} (1-p(x)) \right)^{-1} \cdot \frac{1}{N} \cdot \sum_{i=1}^N \left| (\widehat{p}(X_i) - p(X_i)) \cdot \psi_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) \right|
\end{aligned}$$

and term by term:¹⁶

$$\begin{aligned}
& \left(\inf_{x \in \mathcal{X}} (\widehat{p}(x)) \cdot \inf_{x \in \mathcal{X}} (p(x)) \right)^{-1} \cdot \frac{1}{N} \cdot \sum_{i=1}^N \left| \left(\widehat{\psi}_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) - \psi_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) \right) \cdot p(X_i) \right| \\
& \leq C_0 \cdot \sup_{x \in \mathcal{X}} (p(x)) \cdot \frac{1}{N} \cdot \sum_{i=1}^N \left| \left(\widehat{\psi}_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) - \psi_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) \right) \right| \\
& \leq C_1 \cdot \frac{1}{N} \cdot \sum_{i=1}^N \left| \left(\widehat{\psi}_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) - \psi_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) \right) \right| \\
& = o_p(1)
\end{aligned}$$

where the last lines follow from Lemma 2 and Assumptions 2 and 8;

¹⁶In what follows, the C 's are generic constants.

$$\begin{aligned}
& \left(\inf_{x \in \mathcal{X}} (\widehat{p}(x)) \cdot \inf_{x \in \mathcal{X}} (p(x)) \right)^{-1} \cdot \frac{1}{N} \cdot \sum_{i=1}^N |(\widehat{p}(X_i) - p(X_i)) \cdot \psi_\nu(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)})| \\
& \leq C_0 \cdot \frac{1}{N} \cdot \sum_{i=1}^N |\psi_\nu(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)})| \cdot \sup_{x \in \mathcal{X}} (|\widehat{p}(x) - p(x)|) \leq C_1 \cdot \sup_{x \in \mathcal{X}} (|\widehat{p}(x) - p(x)|) \\
& = o_p(1)
\end{aligned}$$

where the last lines follow from Lemma 2 and Assumptions 2 and 5. Finally, note that:

$$\begin{aligned}
& \left(\sup_{x \in \mathcal{X}} (1 - \widehat{p}(x)) \cdot \sup_{x \in \mathcal{X}} (1 - p(x)) \right)^{-1} \cdot \frac{1}{N} \cdot \sum_{i=1}^N |(\widehat{\psi}_\nu(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)}) - \psi_\nu(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)})) \cdot (1 - p(X_i))| \\
& = o_p(1)
\end{aligned}$$

and

$$\left(\sup_{x \in \mathcal{X}} (1 - \widehat{p}(x)) \cdot \sup_{x \in \mathcal{X}} (1 - p(x)) \right)^{-1} \cdot \frac{1}{N} \cdot \sum_{i=1}^N |(\widehat{p}(X_i) - p(X_i)) \cdot \psi_\nu(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)})| = o_p(1)$$

by the respectively the same type of arguments. Now consider:

$$\begin{aligned}
& \frac{1}{N} \sum_{i=1}^N |\varphi_\Delta(Y_i, X_i, T_i; \widehat{\boldsymbol{\theta}}_\nu) - \mathbb{E}[\varphi_\Delta(Y, X, T; \boldsymbol{\theta}_\nu)]| \\
& \leq \frac{1}{N} \sum_{i=1}^N |\varphi_\Delta(Y_i, X_i, T_i; \widehat{\boldsymbol{\theta}}_\nu) - \varphi_\Delta(Y_i, X_i, T_i; \boldsymbol{\theta}_\nu)| \\
& \quad + \frac{1}{N} \sum_{i=1}^N |\varphi_\Delta(Y_i, X_i, T_i; \boldsymbol{\theta}_\nu) - \mathbb{E}[\varphi_\Delta(Y, X, T; \boldsymbol{\theta}_\nu)]| = o_p(1)
\end{aligned}$$

by Assumptions 5 and 8 and Lemma 1.

Now we look at into:

$$\begin{aligned}
& \frac{1}{N} \cdot \sum_{i=1}^N |\widehat{\alpha}_{\Delta}(T_i, X_i; \widehat{\boldsymbol{\theta}}_{\nu}) - \alpha_{\Delta}(T_i, X_i; \widehat{\boldsymbol{\theta}}_{\nu})| + \frac{1}{N} \cdot \sum_{i=1}^N |\alpha_{\Delta}(T_i, X_i; \widehat{\boldsymbol{\theta}}_{\nu}) - \mathbb{E}[\alpha_{\Delta}(T, X; \boldsymbol{\theta}_{\nu})]| \\
\leq & \frac{1}{N} \cdot \sum_{j=1}^N \left| \sum_{i=1}^N \left(\frac{T_i \cdot \widehat{\psi}_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(1)})}{(\widehat{p}(X_i))^2} + \frac{(1 - T_i) \cdot \widehat{\psi}_{\nu}(Y_i; \widehat{\boldsymbol{\theta}}_{\nu, Y(0)})}{(1 - \widehat{p}(X_i))^2} \right) \right. \\
& \cdot H_K(X_i)^{\top} \cdot \left(\sum_{l=1}^N H_K(X_l) \cdot H_K(X_l)^{\top} \right)^{-1} \cdot H_K(X_j) \cdot (T_j - \widehat{p}(X_j)) \\
& \left. - \left(\frac{\mathbb{E}[\psi_{\nu}(Y; \boldsymbol{\theta}_{\nu, Y(1)}) \mid X_j, T = 1]}{p(X_j)} + \frac{\mathbb{E}[\psi_{\nu}(Y; \boldsymbol{\theta}_{\nu, Y(0)}) \mid X_j, T = 0]}{1 - p(X_j)} \right) \cdot (T_j - p(X_j)) \right| \\
& + \frac{1}{N} \cdot \sum_{i=1}^N |\alpha_{\Delta}(T_i, X_i; \widehat{\boldsymbol{\theta}}_{\nu}) - \alpha_{\Delta}(T_i, X_i; \boldsymbol{\theta}_{\nu})| + \frac{1}{N} \cdot \sum_{i=1}^N |\alpha_{\Delta}(T_i, X_i; \boldsymbol{\theta}_{\nu}) - \mathbb{E}[\alpha_{\Delta}(T, X; \boldsymbol{\theta}_{\nu})]|
\end{aligned}$$

The last line is $o_p(1)$ by Assumptions 2, 5, 8 and Lemma 1. We now have to find bounds for:

$$\begin{aligned}
& \frac{1}{N} \cdot \sum_{j=1}^N \left| \sum_{i=1}^N \left(\frac{T_i \cdot \widehat{\psi}_\nu(Y_i; \widehat{\theta}_{\nu, Y(1)})}{(\widehat{p}(X_i))^2} + \frac{(1 - T_i) \cdot \widehat{\psi}_\nu(Y_i; \widehat{\theta}_{\nu, Y(0)})}{(1 - \widehat{p}(X_i))^2} \right) \right. \\
& \cdot H_K(X_i)^\top \cdot \left(\sum_{l=1}^N H_K(X_l) \cdot H_K(X_l)^\top \right)^{-1} \cdot H_K(X_j) \cdot (T_j - \widehat{p}(X_j)) \\
& \left. - \left(\frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) \mid X_j, T = 1]}{p(X_j)} + \frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(0)}) \mid X_j, T = 0]}{1 - p(X_j)} \right) \cdot (T_j - p(X_j)) \right| \\
& \leq \frac{1}{N} \cdot \sum_{j=1}^N \left| \sum_{i=1}^N \left(\frac{T_i \cdot (\widehat{\psi}_\nu(Y_i; \widehat{\theta}_{\nu, Y(1)}) - \psi_\nu(Y_i; \widehat{\theta}_{\nu, Y(1)}))}{(\widehat{p}(X_i))^2} \right. \right. \\
& \left. \left. + \frac{(1 - T_i) \cdot (\widehat{\psi}_\nu(Y_i; \widehat{\theta}_{\nu, Y(0)}) - \psi_\nu(Y_i; \widehat{\theta}_{\nu, Y(0)}))}{(1 - \widehat{p}(X_i))^2} \right) \right. \\
& \cdot H_K(X_i)^\top \cdot \left(\sum_{l=1}^N H_K(X_l) \cdot H_K(X_l)^\top \right)^{-1} \cdot H_K(X_j) \cdot (T_j - \widehat{p}(X_j)) \left| \right. \\
& + \frac{1}{N} \cdot \sum_{j=1}^N \left| \sum_{i=1}^N \left(\left(\frac{T_i \cdot \psi_\nu(Y_i; \widehat{\theta}_{\nu, Y(1)})}{(\widehat{p}(X_i))^2} + \frac{(1 - T_i) \cdot \psi_\nu(Y_i; \widehat{\theta}_{\nu, Y(0)})}{(1 - \widehat{p}(X_i))^2} \right) \right. \right. \\
& \cdot H_K(X_i)^\top \cdot \left(\sum_{l=1}^N H_K(X_l) \cdot H_K(X_l)^\top \right)^{-1} \cdot H_K(X_j) \\
& \left. \left. - \left(\frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) \mid X_j, T = 1]}{p(X_j)} + \frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(0)}) \mid X_j, T = 0]}{1 - p(X_j)} \right) \right) \cdot (T_j - \widehat{p}(X_j)) \right| \\
& + \left(\frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) \mid X_j, T = 1]}{p(X_j)} + \frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(0)}) \mid X_j, T = 0]}{1 - p(X_j)} \right) \cdot (p(X_j) - \widehat{p}(X_j)) \left| \right.
\end{aligned}$$

and those expressions will be bounded because:

$$\begin{aligned}
& \frac{1}{N} \cdot \sum_{i=1}^N \left| \frac{T_i \cdot (\widehat{\psi}_\nu(Y_i; \widehat{\theta}_{\nu, Y(1)}) - \psi_\nu(Y_i; \widehat{\theta}_{\nu, Y(1)}))}{(\widehat{p}(X_i))^2} + \frac{(1 - T_i) \cdot (\widehat{\psi}_\nu(Y_i; \widehat{\theta}_{\nu, Y(0)}) - \psi_\nu(Y_i; \widehat{\theta}_{\nu, Y(0)}))}{(1 - \widehat{p}(X_i))^2} \right| \\
& \leq \left(\inf_{x \in \mathcal{X}} (\widehat{p}(x)) \right)^{-2} \cdot \frac{1}{N} \cdot \sum_{i=1}^N |\widehat{\psi}_\nu(Y_i; \widehat{\theta}_{\nu, Y(1)}) - \psi_\nu(Y_i; \widehat{\theta}_{\nu, Y(1)})| \\
& + \left(\inf_{x \in \mathcal{X}} (1 - \widehat{p}(x)) \right)^{-2} \cdot \frac{1}{N} \cdot \sum_{i=1}^N |\widehat{\psi}_\nu(Y_i; \widehat{\theta}_{\nu, Y(0)}) - \psi_\nu(Y_i; \widehat{\theta}_{\nu, Y(0)})| \\
& = o_p(1)
\end{aligned}$$

by Lemma 2 and Assumption 8. And for $x \in \mathcal{X}$:

$$\begin{aligned}
& + \frac{1}{N} \cdot \sum_{i=1}^N \left| \left(\frac{T_i \cdot \psi_\nu(Y_i; \hat{\theta}_{\nu, Y(1)})}{(\hat{p}(X_i))^2} + \frac{(1 - T_i) \cdot \psi_\nu(Y_i; \hat{\theta}_{\nu, Y(0)})}{(1 - \hat{p}(X_i))^2} \right) \right. \\
& \quad \cdot H_K(X_i)^\top \cdot \left(\sum_{l=1}^N H_K(X_l) \cdot H_K(X_l)^\top \right)^{-1} \cdot H_K(x) - \\
& \quad \left. \left(\frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) \mid x, T = 1]}{p(x)} + \frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(0)}) \mid x, T = 0]}{1 - p(x)} \right) \right| \\
& = o_p(1)
\end{aligned}$$

by the same type of polynomial approximation arguments used to show consistency of the estimated propensity score. See Hirano, Imbens and Ridder's (2003) Theorem 2 for a detailed proof of this part. The important thing to notice here is that because of ignorability and common support,

$$\mathbb{E} \left[\frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) \mid X, T = 1]}{p(X)} \right] = \mathbb{E} \left[\frac{T \cdot \psi_\nu(Y; \theta_{\nu, Y(1)})}{(p(X))^2} \right]$$

and

$$\mathbb{E} \left[\frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(0)}) \mid X, T = 0]}{1 - p(X)} \right] = \mathbb{E} \left[\frac{(1 - T) \cdot \psi_\nu(Y; \theta_{\nu, Y(0)})}{(1 - p(X))^2} \right]$$

which led us to divide by the squared of $\hat{p}(\cdot)$ and $1 - \hat{p}(\cdot)$ in the estimation of those conditional expectations. Finally:

$$\begin{aligned}
& \frac{1}{N} \cdot \sum_{j=1}^N \left| \left(\frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) \mid X_j, T = 1]}{p(X_j)} + \frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(0)}) \mid X_j, T = 0]}{1 - p(X_j)} \right) \cdot (p(X_j) - \hat{p}(X_j)) \right| \\
& \leq \sup_{x \in \mathcal{X}} \left| \frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(1)}) \mid x, T = 1]}{p(x)} + \frac{\mathbb{E}[\psi_\nu(Y; \theta_{\nu, Y(0)}) \mid x, T = 0]}{1 - p(x)} \right| \cdot o_p(1) \\
& = o_p(1)
\end{aligned}$$

by Lemma 2 and Assumption 5. The only part that is missing to complete the proof is to show that the following expression is bounded in probability:

$$\begin{aligned}
& \frac{1}{N} \sum_{i=1}^N |\hat{\varphi}_\Delta(Y_i, X_i, T_i; \hat{\theta}_\nu) - \varphi_\Delta(Y_i, X_i, T_i; \hat{\theta}_\nu)| + \frac{1}{N} \sum_{i=1}^N |\hat{\alpha}_\Delta(T_i, X_i; \hat{\theta}_\nu) - \alpha_\Delta(T_i, X_i; \hat{\theta}_\nu)| \\
& + \frac{1}{N} \sum_{i=1}^N |\varphi_\Delta(Y_i, X_i, T_i; \hat{\theta}_\nu) + \mathbb{E}[\varphi_\Delta(Y, X, T; \theta_\nu)]| + \frac{1}{N} \sum_{i=1}^N |\alpha_\Delta(T_i, X_i; \hat{\theta}_\nu) + \mathbb{E}[\alpha_\Delta(T, X; \theta_\nu)]|
\end{aligned}$$

We have already shown that the first line is $o_p(1)$. The last line is $O_p(1)$ and should be bounded in probability by $2 \cdot (|\mathbb{E}[\varphi_\Delta(Y, X, T; \boldsymbol{\theta}_\nu)] + \mathbb{E}[\alpha_\Delta(T, X; \boldsymbol{\theta}_\nu)]|)$. This concludes the proof. Thus, we can conclude that $\widehat{\sigma}_{\Delta_\nu}^2 - \sigma_{\Delta_\nu}^2 = o_p(1)$. ■

Table 1.a. Sample Size=100, Replications=100, Mean Treatment Effects

| | MEAN | | | Analytical S.E. | |
|----------|------------|-------|-------------|-----------------|-------------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | 5.00 | 5.00 | 5.00 | Average | 0.83 |
| Average | 4.93 | 3.47 | 4.76 | Median | 0.73 |
| Median | 4.96 | 3.33 | 4.87 | Minimum | 0.50 |
| Minimum | 3.17 | -0.11 | 2.13 | Maximum | 3.00 |
| Maximum | 5.90 | 6.76 | 6.85 | Variance | 0.11 |
| Variance | 0.25 | 1.92 | 0.66 | S.E. | 0.34 |
| S.E. | 0.50 | 1.39 | 0.81 | | |
| bias | -0.07 | -1.53 | -0.24 | | |
| RMSE | 0.50 | 2.06 | 0.85 | | |

Table 1.b. Sample Size=100, Replications=100, Variance Treatment Effects

| | VARIANCE | | | Analytical S.E. | |
|----------|------------|--------|--------------|-----------------|-------------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | -24.75 | -24.75 | -24.75 | Average | 8.59 |
| Average | -25.37 | -25.63 | -26.31 | Median | 7.96 |
| Median | -25.07 | -25.49 | -25.04 | Minimum | 5.47 |
| Minimum | -41.41 | -62.05 | -53.09 | Maximum | 28.19 |
| Maximum | -5.08 | 7.40 | 4.00 | Variance | 8.66 |
| Variance | 53.18 | 171.70 | 102.92 | S.E. | 2.94 |
| S.E. | 7.29 | 13.10 | 10.15 | | |
| bias | -0.62 | -0.88 | -1.56 | | |
| RMSE | 7.32 | 13.13 | 10.26 | | |

Table 1.c. Sample Size=100, Replications=100, Quantile(.25) Treatment Effects

| | QTE(.25) | | | Analytical S.E. | |
|----------|------------|-------|-------------|-----------------|-------------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | 6.36 | 6.36 | 6.36 | Average | 1.46 |
| Average | 6.33 | 4.94 | 6.30 | Median | 1.40 |
| Median | 6.42 | 4.94 | 6.32 | Minimum | 1.08 |
| Minimum | 3.91 | 0.48 | 2.84 | Maximum | 3.00 |
| Maximum | 8.11 | 9.89 | 9.19 | Variance | 0.08 |
| Variance | 0.80 | 3.63 | 1.73 | S.E. | 0.29 |
| S.E. | 0.89 | 1.90 | 1.32 | | |
| bias | -0.03 | -1.42 | -0.06 | | |
| RMSE | 0.89 | 2.38 | 1.32 | | |

Table 1.d. Sample Size=100, Replications=100, Quantile(.50) Treatment Effects

| | QTE(.50) | | | Analytical S.E. | |
|----------|------------|-------|-------------|-----------------|-------------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | 5.00 | 5.00 | 5.00 | Average | 1.45 |
| Average | 4.96 | 3.65 | 4.98 | Median | 1.37 |
| Median | 5.01 | 3.38 | 4.84 | Minimum | 1.04 |
| Minimum | 3.13 | -0.66 | 1.63 | Maximum | 2.98 |
| Maximum | 6.72 | 9.11 | 7.82 | Variance | 0.10 |
| Variance | 0.54 | 3.62 | 1.69 | S.E. | 0.31 |
| S.E. | 0.74 | 1.90 | 1.30 | | |
| bias | -0.04 | -1.35 | -0.02 | | |
| RMSE | 0.74 | 2.33 | 1.30 | | |

Table 1.e. Sample Size=100, Replications=100, Quantile(.75) Treatment Effects

| | QTE(.75) | | | Analytical S.E. | |
|----------|------------|-------|-------------|-----------------|-------------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | 3.64 | 3.64 | 3.64 | Average | 1.68 |
| Average | 3.38 | 1.93 | 3.38 | Median | 1.51 |
| Median | 3.43 | 1.95 | 3.34 | Minimum | 1.15 |
| Minimum | 1.61 | -2.95 | 0.30 | Maximum | 4.18 |
| Maximum | 5.43 | 5.79 | 6.28 | Variance | 0.32 |
| Variance | 0.73 | 3.23 | 1.68 | S.E. | 0.56 |
| S.E. | 0.85 | 1.80 | 1.30 | | |
| bias | -0.25 | -1.71 | -0.26 | | |
| RMSE | 0.89 | 2.48 | 1.30 | | |

Table 1.f. Sample Size=100, Replications=100, InterQuartile Treatment Effects

| | INTER-QUARTILE RANGE | | | Analytical S.E. | |
|----------|----------------------|-------|-------------|-----------------|-------------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | -2.72 | -2.72 | -2.72 | Average | 2.06 |
| Average | -2.94 | -3.01 | -2.92 | Median | 1.98 |
| Median | -2.95 | -3.32 | -2.88 | Minimum | 1.48 |
| Minimum | -5.81 | -8.47 | -6.75 | Maximum | 3.52 |
| Maximum | 0.73 | 2.58 | 1.42 | Variance | 0.12 |
| Variance | 1.43 | 5.15 | 3.58 | S.E. | 0.35 |
| S.E. | 1.20 | 2.27 | 1.89 | | |
| bias | -0.22 | -0.28 | -0.20 | | |
| RMSE | 1.22 | 2.29 | 1.90 | | |

Table 2.a. Sample Size=1000, Replications=100, Mean Treatment Effects

| | MEAN | | | Analytical S.E. | |
|----------|------------|-------|-------------|-----------------|-------------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | 5.00 | 5.00 | 5.00 | Average | 0.23 |
| Average | 4.96 | 3.40 | 4.96 | Median | 0.20 |
| Median | 4.97 | 3.40 | 4.96 | Minimum | 0.18 |
| Minimum | 4.68 | 2.59 | 4.47 | Maximum | 1.43 |
| Maximum | 5.39 | 4.46 | 5.39 | Variance | 0.02 |
| Variance | 0.02 | 0.15 | 0.04 | S.E. | 0.13 |
| S.E. | 0.15 | 0.39 | 0.20 | | |
| bias | -0.04 | -1.60 | -0.04 | | |
| RMSE | 0.15 | 1.65 | 0.20 | | |

Table 2.b. Sample Size=1000, Replications=100, Variance Treatment Effects

| | VARIANCE | | | Analytical S.E. | |
|----------|------------|--------|-------------|-----------------|-------------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | -24.75 | -24.75 | -24.75 | Average | 2.90 |
| Average | -24.64 | -25.00 | -24.60 | Median | 2.61 |
| Median | -24.93 | -25.00 | -24.31 | Minimum | 2.24 |
| Minimum | -30.08 | -34.06 | -32.09 | Maximum | 21.62 |
| Maximum | -18.47 | -16.51 | -14.65 | Variance | 3.87 |
| Variance | 5.30 | 12.59 | 10.22 | S.E. | 1.97 |
| S.E. | 2.30 | 3.55 | 3.20 | | |
| bias | 0.11 | -0.25 | 0.15 | | |
| RMSE | 2.30 | 3.56 | 3.20 | | |

Table 2.c. Sample Size=1000, Replications=100, Quantile(.25) Treatment Effects

| | QTE(.25) | | | Analytical S.E. | |
|----------|------------|-------|-------------|-----------------|-------------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | 6.36 | 6.36 | 6.36 | Average | 0.41 |
| Average | 6.33 | 4.82 | 6.35 | Median | 0.41 |
| Median | 6.32 | 4.79 | 6.33 | Minimum | 0.38 |
| Minimum | 5.72 | 3.31 | 5.59 | Maximum | 0.53 |
| Maximum | 6.91 | 6.22 | 7.06 | Variance | 0.00 |
| Variance | 0.06 | 0.27 | 0.13 | S.E. | 0.02 |
| S.E. | 0.25 | 0.52 | 0.35 | | |
| bias | -0.03 | -1.54 | -0.01 | | |
| RMSE | 0.26 | 1.63 | 0.35 | | |

Table 2.d. Sample Size=1000, Replications=100, Quantile(.50) Treatment Effects

| | QTE(.50) | | | Analytical S.E. | |
|----------|------------|-------|-------------|-----------------|-------------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | 5.00 | 5.00 | 5.00 | Average | 0.39 |
| Average | 4.98 | 3.45 | 4.99 | Median | 0.39 |
| Median | 4.99 | 3.45 | 5.06 | Minimum | 0.36 |
| Minimum | 4.30 | 2.33 | 4.09 | Maximum | 0.64 |
| Maximum | 5.56 | 4.90 | 5.81 | Variance | 0.00 |
| Variance | 0.06 | 0.26 | 0.14 | S.E. | 0.03 |
| S.E. | 0.25 | 0.51 | 0.37 | | |
| bias | -0.02 | -1.55 | -0.01 | | |
| RMSE | 0.25 | 1.64 | 0.37 | | |

Table 2.e. Sample Size=1000, Replications=100, Quantile(.75) Treatment Effects

| | QTE(.75) | | | Analytical S.E. | |
|----------|------------|-------|-------------|-----------------|-------------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | 3.64 | 3.64 | 3.64 | Average | 0.45 |
| Average | 3.59 | 1.97 | 3.59 | Median | 0.44 |
| Median | 3.59 | 2.05 | 3.55 | Minimum | 0.39 |
| Minimum | 2.63 | 0.57 | 2.30 | Maximum | 1.30 |
| Maximum | 4.28 | 3.20 | 4.64 | Variance | 0.01 |
| Variance | 0.08 | 0.31 | 0.18 | S.E. | 0.09 |
| S.E. | 0.29 | 0.56 | 0.42 | | |
| bias | -0.04 | -1.67 | -0.05 | | |
| RMSE | 0.29 | 1.76 | 0.42 | | |

Table 2.f. Sample Size=1000, Replications=100, InterQuartile Treatment Effects

| | INTER-QUARTILE RANGE | | | Analytical S.E. | |
|----------|----------------------|-------|-------------|-----------------|-------------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | -2.72 | -2.72 | -2.72 | Average | 0.61 |
| Average | -2.74 | -2.85 | -2.77 | Median | 0.60 |
| Median | -2.74 | -2.79 | -2.75 | Minimum | 0.56 |
| Minimum | -3.77 | -4.43 | -4.17 | Maximum | 1.12 |
| Maximum | -1.85 | -1.16 | -1.46 | Variance | 0.00 |
| Variance | 0.15 | 0.36 | 0.36 | S.E. | 0.06 |
| S.E. | 0.39 | 0.60 | 0.60 | | |
| bias | -0.02 | -0.13 | -0.04 | | |
| RMSE | 0.39 | 0.61 | 0.60 | | |

Table 3.a. Sample Size=10,000, Replications=100, Mean Treatment Effects

| | MEAN | | | Analytical S.E. | |
|----------|------------|-------|-------------|-----------------|-------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | 5.00 | 5.00 | 5.00 | Average | 0.064 |
| Average | 5.00 | 3.43 | 4.99 | Median | 0.064 |
| Median | 4.99 | 3.44 | 4.99 | Minimum | 0.061 |
| Minimum | 4.87 | 3.12 | 4.81 | Maximum | 0.087 |
| Maximum | 5.12 | 3.80 | 5.10 | Variance | 0.000 |
| Variance | 0.00 | 0.02 | 0.00 | S.E. | 0.003 |
| S.E. | 0.05 | 0.14 | 0.06 | | |
| bias | 0.00 | -1.57 | -0.01 | | |
| RMSE | 0.05 | 1.57 | 0.06 | | |

Table 3.b. Sample Size=10,000, Replications=100, Variance Treatment Effects

| | VARIANCE | | | Analytical S.E. | |
|----------|------------|--------|-------------|-----------------|-------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | -24.75 | -24.75 | -24.75 | Average | 0.848 |
| Average | -24.62 | -24.80 | -24.65 | Median | 0.828 |
| Median | -24.61 | -24.82 | -24.66 | Minimum | 0.769 |
| Minimum | -26.11 | -27.62 | -26.96 | Maximum | 1.605 |
| Maximum | -23.25 | -21.51 | -22.38 | Variance | 0.008 |
| Variance | 0.46 | 1.53 | 0.84 | S.E. | 0.091 |
| S.E. | 0.68 | 1.24 | 0.91 | | |
| bias | 0.13 | -0.05 | 0.10 | | |
| RMSE | 0.69 | 1.24 | 0.92 | | |

Table 3.c. Sample Size=10,000, Replications=100, Quantile(.25) Treatment Effects

| | QTE(.25) | | | Analytical S.E. | |
|----------|------------|-------|-------------|-----------------|-------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | 6.36 | 6.36 | 6.36 | Average | 0.129 |
| Average | 6.36 | 4.80 | 6.36 | Median | 0.128 |
| Median | 6.37 | 4.79 | 6.36 | Minimum | 0.123 |
| Minimum | 6.13 | 4.22 | 5.95 | Maximum | 0.135 |
| Maximum | 6.64 | 5.29 | 6.74 | Variance | 0.000 |
| Variance | 0.01 | 0.04 | 0.02 | S.E. | 0.003 |
| S.E. | 0.10 | 0.20 | 0.13 | | |
| bias | 0.00 | -1.56 | 0.00 | | |
| RMSE | 0.10 | 1.57 | 0.13 | | |

Table 3.d. Sample Size=10,000, Replications=100, Quantile(.50) Treatment Effects

| | QTE(.50) | | | Analytical S.E. | |
|----------|------------|-------|-------------|-----------------|-------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | 5.00 | 5.00 | 5.00 | Average | 0.120 |
| Average | 5.00 | 3.45 | 5.00 | Median | 0.120 |
| Median | 5.00 | 3.43 | 4.98 | Minimum | 0.116 |
| Minimum | 4.78 | 3.08 | 4.70 | Maximum | 0.126 |
| Maximum | 5.23 | 3.95 | 5.34 | Variance | 0.000 |
| Variance | 0.01 | 0.03 | 0.02 | S.E. | 0.002 |
| S.E. | 0.08 | 0.17 | 0.13 | | |
| bias | 0.00 | -1.55 | 0.00 | | |
| RMSE | 0.08 | 1.56 | 0.13 | | |

Table 3.e. Sample Size=10,000, Replications=100, Quantile(.75) Treatment Effects

| | QTE(.75) | | | Analytical S.E. | |
|----------|------------|-------|-------------|-----------------|-------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | 3.64 | 3.64 | 3.64 | Average | 0.136 |
| Average | 3.65 | 2.06 | 3.65 | Median | 0.136 |
| Median | 3.65 | 2.05 | 3.64 | Minimum | 0.128 |
| Minimum | 3.45 | 1.61 | 3.34 | Maximum | 0.142 |
| Maximum | 3.84 | 2.59 | 3.97 | Variance | 0.000 |
| Variance | 0.01 | 0.03 | 0.02 | S.E. | 0.003 |
| S.E. | 0.08 | 0.18 | 0.13 | | |
| bias | 0.01 | -1.58 | 0.01 | | |
| RMSE | 0.08 | 1.59 | 0.13 | | |

Table 3.f. Sample Size=10,000, Replications=100, InterQuartile Treatment Effects

| | INTER-QUARTILE RANGE | | | Analytical S.E. | |
|----------|----------------------|-------|-------------|-----------------|-------|
| | Unfeasible | Wrong | Reweighting | | |
| TRUE | -2.72 | -2.72 | -2.72 | Average | 0.188 |
| Average | -2.71 | -2.75 | -2.71 | Median | 0.188 |
| Median | -2.71 | -2.76 | -2.71 | Minimum | 0.180 |
| Minimum | -3.10 | -3.28 | -3.33 | Maximum | 0.197 |
| Maximum | -2.45 | -2.17 | -2.29 | Variance | 0.000 |
| Variance | 0.01 | 0.05 | 0.04 | S.E. | 0.003 |
| S.E. | 0.12 | 0.23 | 0.19 | | |
| bias | 0.01 | -0.03 | 0.01 | | |
| RMSE | 0.12 | 0.23 | 0.19 | | |

N.Cham. P/EPGE SPE F527i

Autor: Firpo, Sergio

Título: Inequality treatment effects.



000364941

97473

Nº Pat.:364941

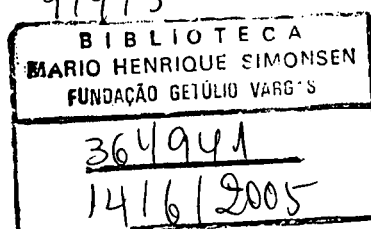


Table 4: Outlier Observations in Terms of Estimated Propensity-Score

| | N=100 | | N=1,000 | | N=10,000 | |
|----------|------------------|------------------|-------------------|------------------|------------------|------------------|
| | Min (for T=1) | Max (for T=0) | Min (for T=1) | Max (for T=0) | Min (for T=1) | Max (for T=0) |
| P-SCORE | 0.04 | 0.82 | 0.03 | 0.67 | 0.04 | 0.68 |
| MEAN | 5.04 (1.31) | 5.05 (0.66) | 5.39 (1.43) | 4.98 (0.22) | 4.89 (0.09) | 5.01 (0.07) |
| VLOG | -28.65 (9.48) | -22.75 (8.00) | -17.37 (21.62) | -22.65 (2.75) | -24.13 (1.60) | -24.52 (0.84) |
| QTE(.25) | 7.70 (2.21) | 6.13 (2.07) | 6.63 (0.53) | 6.08 (0.40) | 6.13 (0.13) | 6.43 (0.13) |
| QTE(.50) | 4.79 (2.92) | 4.86 (1.31) | 4.64 (0.64) | 4.96 (0.41) | 4.90 (0.12) | 4.89 (0.12) |
| QTE(.75) | 3.15 (3.38) | 4.53 (1.34) | 3.55 (1.30) | 4.05 (0.46) | 3.45 (0.14) | 3.81 (0.14) |
| QRANGE | -4.55 (2.58) | -1.59 (2.32) | -3.08 (1.12) | -2.03 (0.59) | -2.68 (0.19) | -2.62 (0.19) |

Note: Entries correspond to estimates in a given replication that produced extreme values of propensity-score.
Analytical standard errors in parentheses.

