Econometrica, Vol. 75, No. 1 (January, 2007), 259–276

EFFICIENT SEMIPARAMETRIC ESTIMATION OF QUANTILE TREATMENT EFFECTS

BY SERGIO FIRPO[1]

This paper develops estimators for quantile treatment effects under the identifying restriction that selection to treatment is based on observable characteristics. Identification is achieved without requiring computation of the conditional quantiles of the potential outcomes. Instead, the identification results for the marginal quantiles lead to an estimation procedure for the quantile treatment effect parameters that has two steps: nonparametric estimation of the propensity score and computation of the difference between the solutions of two separate minimization problems. Root-N consistency, asymptotic normality, and achievement of the semiparametric efficiency bound are shown for that estimator. A consistent estimation procedure for the variance is also presented. Finally, the method developed here is applied to evaluation of a job training program and to a Monte Carlo exercise. Results from the empirical application indicate that the method works relatively well even for a data set with limited overlap between treated and controls in the support of covariates. The Monte Carlo study shows that, for a relatively small sample size, the method produces estimates with good precision and low bias, especially for middle quantiles.

KEYWORDS: Quantile treatment effects, propensity score, semiparametric efficiency bounds, efficient estimation, semiparametric estimation.

1. INTRODUCTION

IN PROGRAM EVALUATION STUDIES, it is often important to learn about distributional impacts beyond the average effects of the program. For example, a policy-maker might be interested in the effect of a treatment on the dispersion of an outcome or its effect on the lower tail of the outcome distribution.

One way to capture this effect in a setting with binary treatment and scalar outcomes is to compute the quantiles of the distribution of treated and control outcomes. Using quantiles, discretized versions of the distribution functions of treated and control can be calculated. Quantiles can also be used to obtain many inequality measurements, for instance, quantile ratios, interquantile ranges, concentration functions, and the Gini coefficient. Finally, differences

[1]This paper is based on the second chapter of my Ph.D. dissertation at UC—Berkeley, written under the supervision of Guido Imbens and Jim Powell. I am indebted to them for their advice, support, and many suggestions. I am grateful for helpful comments from two anonymous referees and the co-editor on earlier drafts. I have also benefited from comments from Marcelo Fernandes, Fernando Ferreira, Carlos Flores, Gustavo Gonzaga, Jinyong Hahn, Michael Jansson, Shakeeb Khan, David Lee, Thomas Lemieux, Thierry Magnac, Rosa Matzkin, Dan McFadden, João Pinho de Mello, Deb Nolan, David Reinstein, James Robins, Paul Ruud, Jeffrey Wooldridge, and participants of seminars at EPGE-FGV, Georgetown, MSU, PUC-Rio, UBC, UC—Berkeley, UC—Davis, University of Miami, University of Toronto, University of Washington, CESG-2004, and USP-RP. Special thanks go to Geert Ridder for many detailed suggestions. Financial support from CAPES—Brazil is acknowledged. All errors are mine.

in quantiles are important because the effects of a treatment may be heterogeneous, varying along the outcome distribution.

A parameter of interest in the presence of heterogeneous treatment effects is the *quantile treatment effect* (QTE). As originally defined by Doksum (1974) and Lehmann (1974), the QTE corresponds, for any fixed percentile, to the horizontal distance between two cumulative distribution functions. In defining QTE as the treatment effect at the individual level, both Doksum and Lehmann implicitly argued that an observed individual would maintain his rank in the distribution regardless of his treatment status. This paper will refer to this type of assumption as a *rank preservation* assumption.

Rank preservation is a strong assumption because it requires the relative value (rank) of the potential outcome for a given individual to be the same regardless of whether that individual is in the treatment or in the control groups. There are two ways to deal with cases in which rank preservation is an unreasonable assumption. First, as suggested by Heckman, Smith, and Clements (1997), one could compute bounds for the distribution of treatment effects, allowing for several possibilities of reordering of the ranks. Second, even when rank preservation is violated, a meaningful parameter for policy purposes might be the simple difference in parameters of two marginal distributions: the distribution of outcome under treatment and the distribution of outcome under nontreatment.

Consider the latter case, in which all the policy-maker is interested in learning about the marginal distributions of the potential outcomes. A convenient way to summarize interesting aspects of these distributions is by computing their quantiles. In this case, quantile treatment effects are simple differences between quantiles of the marginal distributions of potential outcomes.[2] However, if rank preservation holds, then the simple differences in quantiles turn out to be the quantiles of the treatment effect.[3]

This definition of quantile treatment effects, together with the selection on observables assumption, allows identification of various QTE parameters that differ by the subpopulation to which they refer. Following the approach of Heckman and Robb (1986), Hahn (1998), Hirano, Imbens, and Ridder (2003; henceforth HIR), and Imbens (2004), two QTE parameters will be the primary object of study in this paper. They are labeled the *overall quantile treatment effect* (QTE) and the *quantile treatment effect on the treated* (QTT); the former is the QTE parameter for the whole population under consideration and the latter is the parameter for those individuals subject to treatment. Before we formally define those parameters, let us introduce some useful notation.

Define T as the indicator variable of treatment. For an individual i, if $T_i = 1$, we observe $Y_i(1)$; otherwise, if $T_i = 0$, we observe $Y_i(0)$. Here $Y_i(1)$ and $Y_i(0)$

[2]Note that conditional quantiles, which are the objects of interest in quantile regression methods, are not approached in this paper.

[3]There is no similar problem in estimation of the average treatment effect, because differences in means always coincide with means of differences.

are, respectively, the potential outcomes of receiving and not receiving the treatment. Whereas a given individual i is either treated or not, we define the observed outcome as $Y_i = Y_i(1) \cdot T_i + Y_i(0) \cdot (1 - T_i)$. Assume that we also observe a random vector X_i of covariates with support $\mathcal{X} \subset \mathbb{R}^r$.

Let τ be a real in $(0, 1)$. The QTE and QTT parameters can then be expressed as follows:

- The **QTE** is written as $\Delta_\tau = q_{1,\tau} - q_{0,\tau}$, where $q_{j,\tau} \equiv \inf_q \Pr[Y(j) \leq q] \geq \tau$, $j = 0, 1$.

- The **QTT** is written as $\Delta_{\tau|T=1} = q_{1,\tau|T=1} - q_{0,\tau|T=1}$, where $q_{j,\tau|T=1} \equiv \inf_q \Pr[Y(j) \leq q|T = 1] \geq \tau$, $j = 0, 1$.

As is the case for any treatment effect parameter, identification restrictions are necessary for consistent estimation. In this paper, the relevant restriction is the assumption that selection to treatment is based on observable variables (exogeneity assumption). In other words, it is assumed that given a set of observed covariates, individuals are randomly assigned either to the treatment group or to the control group. That assumption was termed by Rubin (1977) the *unconfoundedness assumption* and it characterizes the *selection on observables* branch of the program evaluation literature. Barnow, Cain, and Goldberger (1980), Heckman, Ichimura, Smith, and Todd (1998), Dehejia and Wahba (1999), and HIR are important examples. Further discussion of these identifying assumptions will be provided in a later section.

Estimation of average treatment effects (ATE) under this exogeneity assumption is often performed by first computing a conditional average treatment effect and then integrating over the distribution of covariates to recover the unconditional average treatment effect. However, because the mean of the quantiles is not equal to the quantile of the mean, integrating a first-stage computation of the conditional quantiles (of the treated and the control outcomes) will not yield the marginal quantiles. Instead, this paper demonstrates how to use the selection on observables assumption to calculate the marginal quantiles for the treated and for the control outcomes without computing the corresponding conditional quantiles. The role that the observable covariates play in identifying both ATE and QTE is made clearer in the QTE case, because for the latter, the covariates serve only to remove the selection bias.

Despite the relevance of QTE, the program evaluation literature on this topic is not as vast as that of its main competitor, ATE. Traditionally, expectations have received more attention in the literature than quantiles. Pioneering papers on quantile estimation, such as those by Koenker and Bassett (1978) and, in an instrumental variables setting, by Amemiya (1982) and Powell (1983), have helped to bridge this gap. In the treatment effects literature, some recent contributions have also been made to the study of the distributional effects of the treatment. Among them, Abadie, Angrist, and Imbens (2002) and Chernozhukov and Hansen (2005) explicitly deal with the fact that treatment effects may be nonmonotonic along the outcome distribution and they propose methods to estimate QTE parameters.

There are two important differences between the approaches of Abadie, Angrist, and Imbens (2002) and Chernozhukov and Hansen (2005), and our approach. The first difference is that they consider a conditional on covariates version of QTE. As stated before, conditional quantiles are not directly useful if we are interested in quantiles of the marginal distribution; therefore, it is not clear how to recover the unconditional QTE from their settings. The second difference involves the choice of the identifying set of assumptions. In this paper we show how to identify QTE parameters under unconfoundedness, whereas Abadie, Angrist, and Imbens (2002) and Chernozhukov and Hansen (2005) show how instrumental variables can be used to identify QTE when selection to treatment is based on unobservable variables. However, there is one important similarity between Abadie, Angrist, and Imbens (2002) and our method. Both methods use Koenker and Bassett's (1978) representation for quantiles as minimizers of expectations of check functions and both show how to identify QTE by introducing proper weighting functions to those expectations. Because the QTE parameters identified in our work are essentially different from those in Abadie, Angrist, and Imbens (2002), the weighting functions will reflect this difference.

Recently, an extension to identification and estimation of a class of parameters more general than QTE, termed structural quantile functions, was proposed by Imbens and Newey (2003) and applied to the case of continuous treatments. Unlike Abadie, Angrist, and Imbens (2002) and Chernozhukov and Hansen (2005), who use instrumental variables, Imbens and Newey's (2003) identification strategy is based on control functions. For a general control variable, the key identifying requirement is that observable and unobservable factors that explain the response variable are independent, given that control variable. With an additional common support assumption, they show how to identify the conditional cumulative distribution function (c.d.f.) of the response variable, given covariates, and to obtain quantiles of its marginal distribution, which results from integrating the conditional c.d.f. and inverting it. Unlike Imbens and Newey (2003), our identification results are applied directly to the quantiles of the marginal distribution and, therefore, we do not need to work with c.d.f.'s as a first step.

Imbens and Rubin (1997) and Abadie (2002) proposed methods to estimate some distributional features for a subset of the treated units. Their proposal also was used in an instrumental variables setting and looked at c.d.f.'s, not quantiles. Athey and Imbens (2006) focused on the effects at quantiles in the situation in which repeated cross sections or longitudinal data are available. Distributional effects have also been studied empirically in Card (1996), DiNardo, Fortin, and Lemieux (1996), and Bitler, Gelbach, and Hoynes (2006). Particularly in the paper by DiNardo, Fortin, and Lemieux, quantile treatment effects have been indirectly computed as an estimation byproduct of nonparametric potential outcomes densities for the treated subpopulation.

Herein, a semiparametric method of estimating each QTE parameter is presented. This estimation technique requires a nonparametric first step in which the propensity score is estimated. The final estimators will be equal to the differences between two quantiles, which can be expressed as solutions of minimization problems, where the minimands are sums of check functions, which are convex empirical processes. Using the empirical process literature, consistency and asymptotic normality results are derived. The semiparametric efficiency bound is computed using the techniques suggested in Newey (1990) and Bickel, Klaassen, Ritov, and Wellner (1993), and it is shown that the asymptotic variance of the QTE estimator equals that bound.

2. IDENTIFICATION OF QTE PARAMETERS

Let the propensity score, $\Pr[T = 1|X = x]$, be written as $p(x)$ and let the marginal probability of being treated, $\Pr[T = 1] = E[p(X)]$, be written as p. The following identifying assumption is used here.

ASSUMPTION 1—Strong Ignorability (Rosenbaum and Rubin (1983)): Let $(Y(1), Y(0), T, X)$ have a joint distribution. Then, for all x in $\mathcal{X}$, the support of X, the following conditions hold:

(i) *Unconfoundedness*: Given X, $(Y(1), Y(0))$ is jointly independent from T.

(ii) *Common support*: For some $c > 0$, $c < p(x) < 1 - c$.

Although part (i) of Assumption 1 is a strong assumption, it has been used in several studies on the effect of treatments or programs. Prominent examples are Heckman, Ichimura, Smith, and Todd (1998) and Dehejia and Wahba (1999). Part (ii) states that for almost all values of X, both treatment assignment levels have a positive probability of occurrence.

Now consider each one of the four types of quantiles defined previously: $q_{1,\tau}$, $q_{0,\tau}$, $q_{1,\tau|T=1}$, and $q_{0,\tau|T=1}$. We will assume that for some values of $\tau \in (0, 1)$, these quantiles are well defined and that the respective quantiles are unique. We do so by assuming that the distribution functions of the potential outcomes are continuous and not flat at the τ-percentile. These conditions are stated in the following assumption:

ASSUMPTION 2—Uniqueness of Quantiles: For $j = 0, 1$, $Y(j)$ is a continuous random variable with support in $\mathbb{R}$ and where the following statements apply:

(i) There are nonempty sets Υ_1 and Υ_0, such that $\Upsilon_j = \{\tau \in (0, 1); \Pr[Y(j) \leq q_{j,\tau} - c] < \Pr[Y(j) \leq q_{j,\tau} + c], \forall\, c \in \mathbb{R}, c > 0\}$.

(ii) There are nonempty sets $\Upsilon_{1|T=1}$ and $\Upsilon_{0|T=1}$, such that $\Upsilon_{j|T=1} = \{\tau \in (0, 1); \Pr[Y(j) \leq q_{j,\tau|T=1} - c|T = 1] < \Pr[Y(j) \leq q_{j,\tau|T=1} + c|T = 1], \forall\, c \in \mathbb{R}, c > 0\}$.

Under Assumptions 1 and 2, both QTE and QTT become estimable from the data on (Y, T, X). To show this, we first prove that the quantiles of the potential outcome distributions can be written as implicit functions of the observed data.[4]

LEMMA 1—Identification of Quantiles: *Under Assumptions 1 and 2,* $q_{1,\tau}$, $q_{0,\tau}$, $q_{1,\tau|T=1}$, *and* $q_{0,\tau|T=1}$ *can be written as implicit functions of observed data*: (i) $\tau = E[\frac{T}{p(X)} \cdot \mathbb{1}\{Y \leq q_{1,\tau}\}]$, $\forall\, \tau \in Y_1$; (ii) $\tau = E[\frac{1-T}{1-p(X)} \cdot \mathbb{1}\{Y \leq q_{0,\tau}\}]$, $\forall\, \tau \in Y_0$; (iii) $\tau = E[\frac{T}{p} \cdot \mathbb{1}\{Y \leq q_{1,\tau|T=1}\}]$, $\forall\, \tau \in Y_{1|T=1}$; *and* (iv) $\tau = E[\frac{1-T}{p} \cdot \frac{p(X)}{(1-p(X))} \cdot \mathbb{1}\{Y \leq q_{0,\tau|T=1}\}]$, $\forall\, \tau \in Y_{0|T=1}$.

Note that Assumption 1 plays no role in the identification of $q_{1,\tau|T=1}$. Heckman, Ichimura, and Todd (1997) have already shown an analogous result in the search for identification conditions for the average treatment effects on the treated. Finally, identification of quantile treatment effects is a straightforward consequence of Lemma 1, as stated in the next corollary.

COROLLARY 1—Identification of Quantile Treatment Effect Parameters: *Under Assumptions 1 and 2, the following parameters are identified from data on* (Y, T, X): (i) *the overall quantile treatment effect* Δ_τ *for* $\tau \in Y_1 \cap Y_0$; (ii) *the quantile treatment effect on the treated* $\Delta_{\tau|T=1}$ *for* $\tau \in Y_{1|T=1} \cap Y_{0|T=1}$.[5]

Estimation of the QTT parameter based on the results of Lemma 1 has been used in the applied literature. DiNardo, Fortin, and Lemieux (1996) proposed estimating the counterfactual density of $Y(0)|T = 1$ using a reweighting method that has a population counterpart that is similar to the result for the identification of $q_{0,\tau|T=1}$. They argued that once the counterfactual density is estimated, it is possible to recover the counterfactual quantiles and, therefore, the difference between the quantiles of the treated group and the counterfactual quantiles of the control group. However, as is made clear by Lemma 1 there is no need to compute densities first if the ultimate goal is the estimation of quantiles.

3. ESTIMATION

In this section, we use the sample analogy principle (Manski (1988)) to motivate estimators for Δ_τ and $\Delta_{\tau|T=1}$ that are differences in solutions of minimization problems. We present their large sample properties and show how to estimate their asymptotic variances consistently.

[4]All proofs of the results can be found in the supplement to this article (Firpo (2007)). The supplement also includes details about a Monte Carlo study, which we briefly report in Section 5 of this article.

[5]Throughout the rest of the paper, we implicitly assume that when we consider Δ_τ, the possible values of τ will be such that $\tau \in Y_1 \cap Y_0$; analogously, for $\Delta_{\tau|T=1}$, $\tau \in Y_{1|T=1} \cap Y_{0|T=1}$.

The estimation technique proposed here is semiparametric in the sense that it does not impose any restriction on the joint distribution of (Y, T, X). It extends to quantile treatment effects the characteristics previously proposed for average treatment effects. Examples of semiparametric estimation for ATE can be found in Hahn (1998), Heckman, Ichimura, Smith, and Todd (1998), and HIR.

3.1. *Two-Step Approach*

Using the identification expression from Lemma 1, we present here an estimation method that is a reweighed version of the procedure proposed by Koenker and Bassett (1978) for the quantile estimation problem. Let the estimators of functionals of (Y, T, X) be denoted by a "hat." For example, the nonparametric estimator of the propensity score is $\hat{p}(x)$. The estimator for the QTE parameter Δ_τ is $\hat{\Delta}_\tau \equiv \hat{q}_{1,\tau} - \hat{q}_{0,\tau}$, where, for $j = 0, 1$,

$$\hat{q}_{j,\tau} \equiv \arg\min_q \sum_{i=1}^{N} \hat{\omega}_{j,i} \cdot \rho_\tau(Y_i - q), \tag{1}$$

where the check function $\rho_\tau(\cdot)$ evaluated at a real number a is $\rho_\tau(a) = a \cdot (\tau - \mathbb{1}\{a \le 0\})$ and, finally, where the weights $\hat{\omega}_{1,i}$ and $\hat{\omega}_{0,i}$ are

$$\hat{\omega}_{1,i} = \frac{T_i}{N \cdot \hat{p}(X_i)} \quad \text{and} \quad \hat{\omega}_{0,i} = \frac{1 - T_i}{N \cdot (1 - \hat{p}(X_i))}. \tag{2}$$

The definition of the estimator in Equation (1) relies on the fact that sample quantiles can be found by minimizing a sum of check functions, as pointed out by Koenker and Bassett (1978). In our particular case, we have a weighted sum of check functions, which reflects the fact that the distribution of the covariates differs in the two groups.

We focus on the sample quantile of the $Y(1)$ distribution $\hat{q}_{1,\tau}$. This object is defined as the minimizer of a weighted sum, where the weight of each unit is given by $\hat{\omega}_{1,i}$. To get some intuition on why $\hat{q}_{1,\tau}$ is consistent for $q_{1,\tau}$, note that an approximate first derivative of Equation (1) using the weight defined in the first part of Equation (2) and evaluated at $\hat{q}_{1,\tau}$ is equal to $(1/N)\sum_{i=1}^{N}(T_i/(\hat{p}(X_i))) \cdot (\mathbb{1}\{Y_i \le \hat{q}_{1,\tau}\} - \tau)$. Whereas $\hat{q}_{1,\tau}$ is the minimizer of the convex function expressed in Equation (1) using the weight $\hat{\omega}_1$, $(1/N)\sum_{i=1}^{N}(T_i/(\hat{p}(X_i))) \cdot (\mathbb{1}\{Y_i \le \hat{q}_{1,\tau}\} - \tau)$ will converge in probability to zero as N increases.

The same line of reasoning can be applied to estimation of $\Delta_{\tau|T=1}$. The proposed estimator is defined as the difference between the solutions of two minimizations of sums of weighted check functions: $\hat{\Delta}_{\tau|T=1} \equiv \hat{q}_{1,\tau|T=1} - \hat{q}_{0,\tau|T=1}$, where $\hat{q}_{1,\tau|T=1} \equiv \arg\min_q \sum_{i=1}^{N} \hat{\omega}_{1,i|T=1} \cdot \rho_\tau(Y_i - q)$ and $\hat{q}_{0,\tau|T=1} \equiv$

$\arg\min_q \sum_{i=1}^N \hat{\omega}_{0,i|T=1} \cdot \rho_\tau(Y_i - q)$. The weights used in those definitions are given by

$$\hat{\omega}_{1,i|T=1} = \frac{T_i}{\sum_{l=1}^N T_l} \quad \text{and} \tag{3}$$

$$\hat{\omega}_{0,i|T=1} = \frac{\hat{p}(X_i)}{1-\hat{p}(X_i)} \cdot \frac{1-T_i}{\sum_{l=1}^N T_l}.$$

For the remainder of the paper, we restrict the discussion to the estimator $\hat{q}_{1,\tau}$ of $q_{1,\tau}$ because extensions for $\hat{q}_{0,\tau}$, $\hat{q}_{1,\tau|T=1}$, and $\hat{q}_{0,\tau|T=1}$ follow immediately. Also note that for the derivation of the asymptotic properties, it is enough to concentrate on $\hat{q}_{1,\tau}$, because the differences $\hat{\Delta}_\tau$ and $\hat{\Delta}_{\tau|T=1}$ involve independent terms and, therefore, the covariance term will be zero.[6] Focusing thus on the estimator $\hat{q}_{1,\tau}$, we see that it is a two-step estimator. In the first step, we estimate the propensity score nonparametrically. In the second stage, we minimize

$$G_{\tau,N}(q;\hat{p}) = \frac{1}{N}\sum_{i=1}^N \frac{T_i}{\hat{p}(X_i)} \cdot (Y_i - q) \cdot (\tau - \mathbb{1}\{Y_i \le q\}), \tag{4}$$

which is a weighted sum of check functions whose weights are introduced here to correct for the selection (on observables) problem.

Let us now turn our attention to computation of the weights used in this subsection. In particular, let us concentrate on the calculation of $\hat{\omega}_{1,i}$.

3.2. *First Step*

Following the propensity score estimation strategy employed by HIR, we use a logistic power series approximation, i.e., a series of functions of X is used to approximate the log-odds ratio of the propensity score. The log-odds ratio of $p(x)$ is equal to $\log(p(x)/(1-p(x)))$. These functions are chosen to be polynomials of x and the coefficients that correspond to those functions are estimated by a pseudo-maximum likelihood method.

Start by defining $H_K(x) = [H_{K,j}(x)]$ $(j = 1, \ldots, K)$, a vector of length K of polynomial functions of $x \in \mathcal{X}$ that satisfy the following properties: (i) H_K: $\mathcal{X} \to \mathbb{R}^K$, (ii) $H_{K,1}(x) = 1$, and (iii) if $K > (n+1)^r$, then $H_K(x)$ includes all polynomials up to order n.[7] In what follows, we assume that K is a function of the sample size N such that $K(N) \to \infty$ as $N \to \infty$.

[6]Because the weights $\hat{\omega}_1$ and $\hat{\omega}_0$ are independent, the same situation occurs with $\hat{\omega}_{1|T=1}$ and $\hat{\omega}_{0|T=1}$.

[7]Further details regarding the choice of $H_K(x)$ and its asymptotic properties can be found in the supplement and in HIR.

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

$$\hat{\pi}_K = \arg\max_{\pi\in\mathbb{R}^K} \frac{1}{N}\sum_{i=1}^{N}\big(T_i\cdot\log\big(L(H_K(X_i)'\pi)\big) + (1-T_i)\cdot\log\big(1-L(H_K(X_i)'\pi)\big)\big). \tag{5}$$

After estimating the propensity score, we minimize $G_{\tau,N}(q;\hat{p})$ with respect to q, obtaining $\hat{q}_{1,\tau}$.

3.3. *Large Sample Properties*

In this subsection we show the main asymptotic result of the paper: $\hat{\Delta}_\tau$ is (i) root-N consistent for Δ_τ and (ii) asymptotically normal. This result is shown by concentrating on the case of $\hat{q}_{1,\tau}$ and by arguing that the same type of result holds, by analogy, for $\hat{q}_{0,\tau}$. Before stating the result as a theorem, let us define some important quantities. For $i=1,\dots,N$,

$$\psi_{1,\tau}(y,t,x) = \frac{t}{p(x)}\cdot g_{1,\tau}(y) - \frac{t-p(x)}{p(x)}\cdot E[g_{1,\tau}(Y)|x,T=1], \tag{6}$$

$$\psi_{0,\tau}(y,t,x) = \frac{1-t}{1-p(x)}\cdot g_{0,\tau}(y) + \frac{t-p(x)}{1-p(x)}\cdot E[g_{0,\tau}(Y)|x,T=0], \tag{7}$$

$$g_{j,\tau}(y) = -\frac{\mathbb{1}\{y\le q_{j,\tau}\}-\tau}{f_j(q_{j,\tau})} \qquad (j=0,1), \tag{8}$$

where functions $g_{j,\tau}(\cdot)$ would be the influence functions of the sample quantiles of the potential outcomes $Y(j)$, $j=0,1$, if they were fully observable.

There are two reasons for $g_{1,\tau}$ and $g_{0,\tau}$ to be, respectively, different from $\psi_{1,\tau}$ and $\psi_{0,\tau}$. The first reason is related to the partial unobservability of the potential outcomes. Because we cannot observe two potential outcomes for the same unit, the moment conditions associated with $q_{1,\tau}$ and $q_{0,\tau}$, as seen from Lemma 1, are, respectively,

$$\varphi_{1,\tau}(y,t,x) = \frac{t}{p(x)}\cdot g_{1,\tau}(y) \quad\text{and}\quad \varphi_{0,\tau}(y,t,x) = \frac{1-t}{1-p(x)}\cdot g_{0,\tau}(y) \tag{9}$$

because $E[\varphi_{1,\tau}(Y,T,X)] = E[\varphi_{0,\tau}(Y,T,X)] = 0$.[8] The second difference is related to the fact that, in addition to not fully observing the potential out-

[8]If we could observe the potential outcomes, the moment conditions associated with the quantiles of the potential outcomes would simply be $E[\mathbb{1}\{Y(1)\le q_{1,\tau}\}-\tau] = E[\mathbb{1}\{Y(0)\le q_{0,\tau}\}-\tau] = 0$.

comes, we also do not know the true propensity score. The effect of estimating $p(\cdot)$ on the influence functions of the sample quantiles is given by the conditional expectation of the derivatives of $\varphi_{1,\tau}$ and $\varphi_{0,\tau}$ with respect to $p(\cdot)$, which correspond to $\alpha_{1,\tau}(t,x) = -E[g_{1,\tau}(Y)|x, T=1] \cdot (t - p(x))/p(x)$ and $\alpha_{0,\tau}(t,x) = E[g_{0,\tau}(Y)|x, T=0] \cdot (t - p(x))/(1-p(x))$. Define now $\psi_\tau(y,t,x) = \psi_{1,\tau}(y,t,x) - \psi_{0,\tau}(y,t,x)$, $\varphi_\tau(y,t,x) = \varphi_{1,\tau}(y,t,x) - \varphi_{0,\tau}(y,t,x)$, and $\alpha_\tau(t,x) = \alpha_{1,\tau}(t,x) - \alpha_{0,\tau}(t,x)$. Finally, define [9]

$$
\begin{aligned}
V_\tau &= E\big[(\psi_\tau(Y,T,X))^2\big] = E\big[(\varphi_\tau(Y,T,X) + \alpha_\tau(T,X))^2\big] \\
&= E\Bigg[\frac{V[g_{1,\tau}(Y)|X,T=1]}{p(X)} + \frac{V[g_{0,\tau}(Y)|X,T=0]}{1-p(X)} \\
&\qquad + \big(E[g_{1,\tau}(Y)|X,T=1] - E[g_{0,\tau}(Y)|X,T=0]\big)^2\Bigg].
\end{aligned}
$$

We now state the main asymptotic result:

THEOREM 1—Asymptotic Properties of $\hat{\Delta}_\tau$: *Under Assumptions* 1 *and* 2 *herein and Assumptions* A.1 *and* A.2 *in supplements*,

$$
\sqrt{N}(\hat{\Delta}_\tau - \Delta_\tau) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \psi_\tau(Y_i, T_i, X_i) + o_p(1) \xrightarrow{D} N(0, V_\tau).
$$

We show a sketch of the proof of Theorem 1. For the complete proof, see the supplement.

We start with the results derived in HIR, in turn based on Newey (1995), for the asymptotic properties of the nonparametric estimation of the propensity score in the first step by means of a power series approximation. Their approach to estimating the propensity score guarantees, under certain regularity conditions, that $\hat{p}(x)$, the estimator of the propensity score, is uniformly consistent for the true $p(x)$. That set of regularity conditions is listed in the supplement as Assumption A.1 and the uniform convergence result is presented there as Lemma A.1.

We then focus on the case of $\hat{q}_{1,\tau}$. The next step in the proof is to define $Q_{\tau,N}(u;\hat{p}) \equiv N \cdot (G_{\tau,N}(q;\hat{p}) - G_{\tau,N}(q_{1,\tau};\hat{p}))$, where $q = q_{1,\tau} + u/\sqrt{N}$ and u is some real number. Therefore, at $\hat{u}_\tau = \sqrt{N}(\hat{q}_{1,\tau} - q_{1,\tau})$, $Q_{\tau,N}(\cdot;\hat{p})$ reaches its minimum value. We thus show that $Q_{\tau,N}(u;\hat{p}) - \tilde{Q}_{\tau,N}(u)$ is $o_p(1)$ for any fixed u, where $\tilde{Q}_{\tau,N}(u)$, which does not depend on $\hat{p}(x)$, is a quadratic random function. For that result to be true, we impose the technical requirement that $\Pr[Y(1) \leq q_{1,\tau}|X=x]$ is continuously differentiable in x (Assumption A.2).

[9] In what follows, $V[\cdot]$ is the variance operator.

This is analogous to imposing continuity of the first derivative of the conditional expectation of $Y(1)$ given X, as HIR do for ATE. The result that shows that $Q_{\tau,N}(u;\hat{p}) - \tilde{Q}_{\tau,N}(u) = o_p(1)$ appears in Lemma A.2.

Let $\tilde{u}_\tau$ be the argument that minimizes $\tilde{Q}_{\tau,N}(\cdot)$. We show in Lemma A.3 that $\tilde{u}_\tau = O_p(1)$ and $\tilde{u}_\tau \xrightarrow{D} N(0, V_{1,\tau})$, and we derive an expression for $V_{1,\tau}$. To get results about $\hat{u}_\tau$ and, consequently, about $\hat{q}_{1,\tau}$ we use a result in Hjört and Pollard (1993) on the nearness of minimizers of convex random functions. In particular, we apply Hjört and Pollard's Lemma 2 directly to our case, showing in Lemma A.4 that $\hat{u}_\tau - \tilde{u}_\tau = o_p(1)$. This concludes the proof that $\hat{q}_{1,\tau}$ is (i) root-N consistent for $q_{1,\tau}$ and (ii) asymptotically normal, which are shown in Lemma A.5. Extending this result to $\hat{q}_{0,\tau}$, Theorem 1 is easily shown to hold.

Finally, note that estimation of the QTT parameter, $\Delta_{\tau|T=1}$, will yield a similar result, which could have been obtained using steps analogous to those used for the QTE parameter Δ_τ to get results similar to Theorem 1. For the sake of completeness, we present now the normalized asymptotic variance of the QTT estimator, $\hat{\Delta}_{\tau|T=1}$:

$$
\begin{aligned}
V_{\tau|T=1} = E\Bigg[\frac{p(X)}{p} \cdot \Bigg(& \frac{V[g_{1,\tau|T=1}(Y)|X, T=1]}{p} \\
& + \frac{p(X)}{p} \cdot \frac{V[g_{0,\tau|T=1}(Y)|X, T=0]}{1-p(X)} \\
& + \frac{(E[g_{1,\tau|T=1}(Y)|X, T=1] - E[g_{0,\tau|T=1}(Y)|X, T=0])^2}{p} \Bigg) \Bigg],
\end{aligned} \tag{10}
$$

$$
g_{j,\tau|T=1}(Y) = -\frac{\mathbb{1}\{Y \le q_{j,\tau|T=1}\} - \tau}{f_{j|T=1}(q_{j,\tau|T=1})} \qquad (j = 0, 1),
$$

and $f_{1|T=1}(\cdot)$ and $f_{0|T=1}(\cdot)$ are the densities of $Y(1)|T=1$ and $Y(0)|T=1$, respectively.

3.4. *Variance Estimation*

In this section we present an estimator for V_τ, the normalized asymptotic variance $\hat{\Delta}_\tau$. We then state some sufficient conditions for that estimator to be consistent, which is proved in the supplement. Note that the same argument we subsequently show could be easily extended to the estimation of $V_{\tau|T=1}$, the normalized asymptotic variance $\hat{\Delta}_{\tau|T=1}$.

The normalized asymptotic variance of $\hat{\Delta}_\tau$ is $V_\tau = E[(\varphi_\tau(Y,T,X) + \alpha_\tau(Y,T,X))^2]$. A natural procedure to estimate that variance term is by using $\hat{V}_\tau = \frac{1}{N}\sum_{i=1}^N (\hat{\varphi}_{\tau,i} + \hat{\alpha}_{\tau,i})^2$, where

$$
\hat{\varphi}_{\tau,i} = \frac{T_i}{\hat{p}(X_i)} \cdot \hat{g}_{1,\tau}(Y_i) - \frac{1-T_i}{1-\hat{p}(X_i)} \cdot \hat{g}_{0,\tau}(Y_i),
$$

$$\hat{g}_{j,\tau}(y) = -\frac{\mathbb{1}\{y \le \hat{q}_{j,\tau}\} - \tau}{\hat{f}_j(\hat{q}_{j,\tau})} \qquad (j = 0, 1),$$

and $\hat{f}_j(\cdot)$ is an estimator of the density of the potential outcome $Y(j)$. Note that a simple application of iterated expectations to the definition of $\alpha_\tau(t, x)$ yields

$$\alpha_\tau(t, x) = -E\left[\frac{T \cdot g_{1,\tau}(Y)}{(p(X))^2} + \frac{(1-T) \cdot g_{0,\tau}(Y)}{(1-p(X))^2}\Big| X = x\right] \cdot (t - p(x)),$$

which leads to its estimation through

$$\hat{\alpha}_{\tau,i} = -\hat{E}\left[\frac{T \cdot \hat{g}_{1,\tau}(Y)}{(\hat{p}(X))^2} + \frac{(1-T) \cdot \hat{g}_{0,\tau}(Y)}{(1-\hat{p}(X))^2}\Big| X = X_i\right] \cdot (T_i - \hat{p}(X_i)).$$

Further details of the estimation procedure of such conditional expectation can be found in the supplement.

Finally, we show that under the same assumptions used to prove the main asymptotic result plus a set of regularity conditions presented in the supplement as Assumption A.3, $\hat{V}_\tau$ will be consistent for V_τ:

THEOREM 2—Consistent Estimation of the Asymptotic Variance of $\hat{\Delta}_\tau$: *Under Assumptions* 1, 2, A.1, A.2, *and* A.3, $\hat{V}_\tau - V_\tau = o_p(1)$.

4. SEMIPARAMETRIC EFFICIENCY BOUNDS

We now show that the respective estimators $\hat{\Delta}_\tau$ and $\hat{\Delta}_{\tau|T=1}$ of QTE and QTT are indeed efficient in the class of semiparametric estimators. To show this, we calculate the semiparametric efficiency bounds for QTE and QTT parameters under unconfoundedness of the treatment and unknown propensity score.

Efficiency bounds for semiparametric models were first introduced by Stein (1956), but have become more popular in the econometric literature only recently after the systematic presentations by Bickel, Klaassen, Ritov, and Wellner (1993), and Newey (1990, 1994). For semiparametric models with missing data under the "missing at random" assumption, Robins, Rotnitzky, and Zhao (1994), Robins and Rotnitzky (1995), and Rotnitzky and Robins (1995) have shown how to compute the efficiency bounds by deriving the efficient score of that model from the efficient score under random sampling. Under unconfoundedness, Hahn (1998) and HIR have computed the bounds for the average treatment effect (ATE) and for average treatment effects for given subpopulations, with particular emphasis on the bounds for the average treatment effect on the treated ATT.

For the quantile treatment effects setting, we computed bounds for two parameters, namely, Δ_τ and $\Delta_{\tau|T=1}$. With Assumptions 1 and 2, the semiparametric efficiency bounds for Δ_τ and $\Delta_{\tau|T=1}$ can be calculated:

THEOREM 3—Bounds for Δ_τ and $\Delta_{\tau|T=1}$: *Under Assumptions 1 and 2, the semiparametric efficiency bounds for* Δ_τ *and* $\Delta_{\tau|T=1}$ *are, respectively, equal to* V_τ *and* $V_{\tau|T=1}$.

Note that the bounds V_τ and $V_{\tau|T=1}$ are similar to the bounds computed by Hahn (1998) for the mean case. There are two reasons for the similarities. First, both QTE and ATE are parameters from the same statistical model and, therefore, can be expressed as functionals of the same distribution of the data. Second, both can be written as differences in solutions to moment conditions (implicitly for the QTE case) over the same density. The latter can be seen from the definitions of $g_{j,\tau}(\cdot)$ and $g_{j,\tau|T=1}(\cdot)$ $(j = 0, 1)$ as the influence functions of the sample quantiles of $Y(1)$ and $Y(0)$ under full observability of potential outcomes. Because, for example, $E[g_{j,\tau}(Y(j))] = 0$, we know that $\Delta_\tau = \arg\text{zero}_q E[\mathbb{1}\{Y(1) \leq q\} - \tau] - \arg\text{zero}_q E[\mathbb{1}\{Y(0) \leq q\} - \tau]$. Also, by defining $\beta_1 = E[Y(1)]$, $\beta_0 = E[Y(0)]$, and $\beta = \beta_1 - \beta_0$, then $\beta = \arg\text{zero}_b E[Y(1) - b] - \arg\text{zero}_b E[Y(0) - b]$.[10]

5. SUMMARY OF RESULTS

We have considered two applications for the method developed here: the first is the evaluation of a well studied job training program; the second is a Monte Carlo study. Further details of both applications, as well as all figures and tables can be found in the supplement[11] (Firpo (2007)).

5.1. *Empirical Application*

The empirical exercise used the job training program data set first analyzed by LaLonde (1986) and later by many others, including Heckman and Hotz (1989), Dehejia and Wahba (1999), Smith and Todd (2001, 2005), and Imbens (2003). The interest is to identify the causal effects of job training on future earnings. For our application, we have used one of the many data subsets LaLonde constructed based on the original sample. We have both the experimental and the observational data sets, the latter using information from the Panel Study of Income Dynamics (PSID) and the former using information from the National Supported Work Program (NSW). Following Dehejia and Wahba (1999), we restricted the analysis to a subset of 185 treated units, 260 control units, and 2,490 comparison units from the PSID.[12] Some summary statistics are presented in Table I in the supplement.

[10]Note that this argument could very well be applied to the comparison between the "on the treated" parameters $\Delta_{\tau|T=1}$ and $E[Y(1)|T = 1] - E[Y(0)|T = 1]$.

[11]The Matlab program used in the empirical application and in the Monte Carlo study is available on the *Econometrica Supplementary Materials* web site (Firpo (2007)).

[12]This corresponds to the control sample labeled by LaLonde (1986) and Dehejia and Wahba (1999) as PSID-1.

Using that data set, we generated estimates of the QTT. We also performed an experimental QTE estimation, by simply taking the difference between the quantiles of the treated and the experimental controls, without any reweighing. Our results are presented in Table II and Figures 1–5 in the supplement.

Figures 1–3 (in the supplement) show that when experimental controls are used, treatment effects tend to be more homogeneous than in the observational setting. With a nonexperimental comparison group, treatment effects seem to increase along the distribution, starting around the median until almost the upper end of the distribution. Also, although nonexperimental ATT is not statistically significantly different from zero (point estimate of \$1,163 with a standard error of \$1,736), as shown in Table II, QTT estimates are positive and significant from the median to the 85th percentile.

Despite the fact that observational treatment effects seem to be more heterogeneous along the distribution than the experimental effects, confidence bands the size of 2 standard errors around the difference between the QTT estimates include the zero for almost all quantiles, as shown by Figure 4. Table II reveals the same pattern for the mean. The difference between the experimental ATT (\$1,794) and nonexperimental ATT (\$1,163) estimates is \$631, which is relatively small given the magnitude its standard errors (\$1,860).

We also checked the common support assumption. As Table I reveals, comparison and treated groups are largely different for most of the covariates. The estimated propensity score is, however, bounded away from 0 and 1, which allows us to proceed with the analysis and to compute weights that are well defined for all observations in the sample.[13] A closer inspection of the data reveals that the distributions of the estimated propensity scores for treated and comparison groups are importantly different. Figure 5 shows the histograms of these two distributions. Most of the data in the comparison group present values for the estimated propensity score below 0.1. In fact, there are only 132 out of 2,490 comparison observations that have an estimated propensity score above 0.1. This means that although we are using all 2,490 comparison units, only a few of them will effectively be important for comparisons with the treated group, because the weights $\hat{\omega}_{0|T=1}$ used to estimate the quantiles of that counterfactual distribution are proportional to $\hat{p}(x)/(1-\hat{p}(x))$ and, therefore, we will weigh down observations in the comparison group with estimated propensity scores close to zero. We should expect that because we are effectively using just few observations from PSID, the standard errors of QTT estimates should be affected. Our findings do indicate that this is exactly what happens: there are some QTT estimates in the middle of the distribution that have larger standard errors than those QTT estimates at the upper end of the distribution. As we show in more detail in the supplement, this is due to the

[13] As we are considering the quantile treatment effect on the treated, a sufficient common support assumption is that $p(x) < 1$.

presence of some comparison units that have large values of the estimated propensity score in the middle of the empirical distribution of earnings.

The main findings of the empirical application are that observational QTT point estimates are relatively close to the experimental ones, but that they are imprecisely estimated at some parts of the earnings distribution. The large standard errors we report for the observational QTT can be attributed to the limited overlap in the covariates distribution between treated and comparison groups. Limited overlap in the support of the empirical distribution of covariates does not mean, however, that the proposed method is inadequate for evaluation of this program. In fact, as discussed in Imbens (2004), the reweighting method applied to ATT estimation is well designed to cope with limited overlap because point estimates are not substantially affected by the inclusion of controls that have propensity score close to 0, but standard errors would increase with the inclusion of controls that have propensity score close to 1. The same is true for the QTT estimation method presented in this paper.

5.2. *Monte Carlo*

In the Monte Carlo study, 1,000 replications with sample sizes of 500 and 5,000 were considered. The data design implied that assignment to the treatment was not completely random, but satisfied the ignorability assumption. Hence, estimation of treatment effects that do not take the selection into account would inevitably produce inconsistent estimates. We have actually performed such calculations without any correction for the selection problem: they are called the "naive" estimators. Because $Y(1)$ 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 $Y(0)$. Finally, we compared the naive and the unfeasible estimators with the one proposed in this article in terms of bias, root mean squared error (RMSE), median absolute error (MAE), and coverage rate (CR) of 90% confidence intervals. As we expected, the unfeasible estimator had the best performance, but it was closely followed by the reweighting estimator in all those measures. The asymptotic variance was estimated using the analytical expression provided in this paper and it was compared to the variance estimate generated through the 1,000 experiment replications.

The results indicate that the reweighting estimator performs well according to RMSE and MAE criteria. Coverage rates of the reweighting estimator were close to those from the unfeasible estimator and to 90%. The naive estimator had, as expected, the worst performance in all those measures. 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 quantiles being estimated by the reweighting method; the same situation does not occur with the naive estimator. Analytical standard errors tend to be (either looking at the average or at the median) close to the Monte Carlo standard errors for all sample sizes

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

6. CONCLUSION

For applications in which the exogeneity assumption is likely to hold, this paper has shown how to estimate the quantile treatment effects using a two-step procedure. The estimator is shown to be root-N consistent and asymptotically normal. We also calculated the semiparametric efficiency bound and proved that this quantile treatment effects estimator achieves it.

We investigated identification, estimation, and inference for QTE and QTT parameters, which are differences in solutions of minimization problems. Under the assumptions used in this article, the method developed here could be extended to quantities that are solutions to the general problem

$$
\begin{aligned}
(11) \qquad \zeta_1 - \zeta_0 &= \arg\min_{\zeta} E\big[m(Y(1);\zeta)\big] - \arg\min_{\zeta} E\big[m(Y(0);\zeta)\big] \\
&= \arg\min_{\zeta} E\left[\frac{T}{p(X)} \cdot m(Y;\zeta)\right] \\
&\quad - \arg\min_{\zeta} E\left[\frac{1-T}{1-p(X)} \cdot m(Y;\zeta)\right]
\end{aligned}
$$

for some known real-valued function m. There are several examples of functions m, but two of the most important ones are $m(Y;\zeta) = (Y-\zeta)^2$, which allows us to use the identification result of Equation (11) to compute average treatment effects, and $m(Y;\zeta) = \rho_\tau(Y-\zeta)$, the check function, which was used in this paper to compute quantile treatment effects. Finally, note that if a random sample of (Y, T, X) were available, estimation of ζ_1 and ζ_0 would follow by replacing population moments by their sums and replacing the weights by $\hat{\omega}_1$ and $\hat{\omega}_0$, respectively.

Whereas there are many inequality measures that are functions of quantiles, an obvious extension of the method presented here would involve estimation of inequality measures for the potential outcomes. Several relevant inequality measures are of interest in the applied literature. The framework developed here could be extended to estimate and infer the response of such inequality measures to a treatment.

Another extension would deal with the fact that under the current assumptions, the quantiles of the treatment effects distribution are not point identified and, therefore, a strategy to compute bounds, in the spirit of Manski (2003), could be an alternative. Thus, a natural extension is to provide conditions for interval identification for those quantiles and to provide estimates of those intervals. A comparison between the two methods would be useful in the sense of establishing how close differences in quantiles of potential outcomes are to the

quantiles of the treatment effects even when no rank preservation assumptions are invoked.

Dept. de Economia, Pontifícia Universidade Católica do Rio de Janeiro, Rua Marquês de São Vicente, 225-Gávea, 22453 Rio de Janeiro, RJ, Brazil; firpo@econ.puc-rio.br, http://www.econ.puc-rio.br/firpo/homepage_puc.htm.

Manuscript received August, 2004; final revision received June, 2006.

REFERENCES

ABADIE, A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284–292. [262]

ABADIE, A., J. ANGRIST, AND G. IMBENS (2002): "Instrumental Variables Estimation of Quantile Treatment Effects," *Econometrica*, 70, 91–117. [261,262]

AMEMIYA, T. (1982): "Two-Stage Least Absolute Deviations Estimators," *Econometrica*, 50, 689–711. [261]

ATHEY, S., AND G. IMBENS (2006): "Identification and Inference in Nonlinear Difference-in-Differences Models," *Econometrica*, 74, 431–497. [262]

BARNOW, B. S., G. G. CAIN, AND A. S. GOLDBERGER (1980): "Issues in the Analysis of the Selectivity Bias," in *Evaluation Studies*, Vol. 5, ed. by E. Stromsdorfer and G. Farkas. San Francisco: Sage, 43–59. [261]

BICKEL, P., C. A. J. KLAASSEN, Y. RITOV, AND J. A. WELLNER (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press. [263,270]

BITLER, M., J. GELBACH, AND H. HOYNES (2006): "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," *American Economic Review*, 96, 988–1012. [262]

CARD, D. (1996): "The Effect of Unions on the Structure of Wages: A Longitudinal Analysis," *Econometrica*, 64, 957–979. [262]

CHERNOZHUKOV, V., AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261. [261,262]

DEHEJIA, R., AND S. WAHBA (1999): "Causal Effects in Nonexperimental Studies: Revaluating the Evaluation of Training Programs," *Journal of the American Statistical Association*, 94, 1053–1062. [261,263,271]

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. [262,264]

DOKSUM, K. (1974): "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *The Annals of Statistics*, 2, 267–277. [260]

FIRPO, S. (2007): "Supplement to 'Efficient Semiparametric Estimation of Quantile Treatment Effects'," *Econometrica Supplementary Materials*, 75, http://www.econometricsociety.org/ecta/supmat/proofs.pdf. [264,271]

HAHN, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects," *Econometrica*, 66, 315–331. [260,265,270,271]

HECKMAN, J., AND J. HOTZ (1989): "Choosing among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," *Journal of the American Statistical Association*, 84, 862–874. [271]

HECKMAN, J., H. ICHIMURA, J. SMITH, AND P. TODD (1998): "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 66, 1017–1098. [261,263,265]

HECKMAN, J., H. ICHIMURA, AND P. TODD (1997): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261–294. [264]

HECKMAN, J., AND R. ROBB (1986): "Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes," in *Drawing Inferences from Self-Selected Samples*, ed. by H. Wainer. Berlin: Springer-Verlag, 63–107. [260]

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, 487–535. [260]

HIRANO, K., G. IMBENS, AND G. RIDDER (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189. [260]

HJÖRT, N., AND D. POLLARD (1993): "Asymptotics for Minimizers of Convex Processes," typescript, www.stat.yale.edu/∼pollard/Papers/convex.pdf. [269]

IMBENS, G. (2003): "Sensitivity to Exogeneity Assumptions in Program Evaluation," *The American Economic Review*, 93, 126–132. [271]

——— (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economics and Statistics*, 86, 4–29. [260,273]

IMBENS, G., AND W. NEWEY (2003): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Manuscript, Department of Economics, UC Berkeley. [262]

IMBENS, G., AND D. RUBIN (1997): "Estimating Outcome Distributions for Compliers in Instrumental Variable Models," *Review of Economic Studies*, 64, 555–574. [262]

KOENKER, R., AND G. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33–50. [261, 262,265]

LALONDE, R. J. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *The American Economic Review*, 76, 604–620. [271]

LEHMANN, E. (1974): *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day. [260]

MANSKI, C. (1988): *Analog Estimation Methods in Econometrics*. New York: Chapman & Hall. [264]

——— (2003): *Partial Identification of Probability Distributions*. New York: Springer-Verlag. [274]

NEWEY, W. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99–135. [263,270]

——— (1994): "Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382. [270]

——— (1995): "Convergence Rates for Series Estimators," in *Statistical Methods of Economics and Quantitative Economics: Essays in Honor of C. R. Rao*, ed. by G. S. Maddalla, P. C. B. Phillips, and T. N. Srinavasan. Cambridge, MA: Blackwell, 254–275. [268]

POWELL, J. (1983): "The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators," *Econometrica*, 51, 1569–1576. [261]

ROBINS, J., AND A. ROTNITZKY (1995): "Semiparametric Efficiency in Multivariate Regression Models with Missing Data," *Journal of the American Statistical Association*, 90, 122–129. [270]

ROBINS, J., A. ROTNITZKY, AND L. P. ZHAO (1994): "Estimation of Regression Coefficients when Some Regressions Are not Always Observed," *Journal of the American Statistical Association*, 89, 846–866. [270]

ROSENBAUM, P., AND D. RUBIN (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55. [263]

ROTNITZKY, A., AND J. ROBINS (1995): "Semiparametric Regression Estimation in the Presence of Dependent Censoring," *Biometrika*, 82, 805–820. [270]

RUBIN, D. (1977): "Assignment to Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, 2, 1–26. [261]

SMITH, J. A., AND P. E. TODD (2001): "Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods," *The American Economic Review*, 91, 112–118. [271]

——— (2005): "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*, 125, 305–353. [271]

STEIN, C. (1956): "Efficient Nonparametric Testing and Estimation," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1. Berkeley, CA: University of California Press, 187–195. [270]