City Research Online

City St George's, University of London

Citation: Montes-Rojas, G. (2010). Nonparametric estimation of ATE and QTE: an application of Fractile Graphical Analysis (10/06). London, UK: Department of Economics, City University London.

This is the unspecified version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/1486/

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

School of Social Sciences
CITY UNIVERSITY LONDON

Department of Economics

Nonparametric Estimation of ATE and QTE:An Application of Fractile Graphical Analysis

Gabriel Montes-Rojas*
City University London

Department of Economics
Discussion Paper Series
No. 10/06

* Department of Economics, City University London, Social Sciences Bldg, Northampton Square, London EC1V 0HB, UK.
Email: Gabriel.Montes-Rojas.1@city.ac.uk

Nonparametric Estimation of ATE and QTE : An Application of Fractile Graphical Analysis

Gabriel V. Montes-Rojas*
City University London

Abstract

Nonparametric estimators for average and quantile treatment effects are constructed using Fractile Graphical Analysis, under the identifying assumption that selection to treatment is based on observable characteristics. The proposed method has two-steps: first, the propensity score is estimated, and second, a blocking estimation procedure using this estimate is used to compute treatment effects. In both cases, the estimators are proved to be consistent. Monte Carlo results show a better performance than other procedures based on the propensity score. Finally, these estimators are applied to a job training dataset.

JEL: C14, C31

Keywords: Treatment Effects, Fractile Graphical Analysis

*Corresponding Author: Department of Economics, City University London, D306 Social Sciences Bldg, Northampton Square, London EC1V 0HB, UK. Email: Gabriel.Montes-Rojas.1@city.ac.uk

1 Introduction

Econometric methods for estimating the effects of certain programs (such as job search assistance or classroom teaching programs) has been widely developed since the pioneering work of Ashenfelter (1978), LaLonde (1986) and others. In this case, a treatment refers to a certain program whose benefits are potentially obtainable by those selected for participation (treated) and it has no effect on a control group (non-treated).

Estimating average treatment effects (ATE), which refers to the mean effect of the program on a given outcome variable in parametric and nonparametric environments (see Angrist and Krueger, 1999, Imbens, 2004) has been a central issue in the literature. Lehmann (1974) and Docksum (1974) introduced the concept of quantile treatment effects (QTE) as the difference of the quantiles of the treated and control outcome distributions. In this case, it is implicitly assumed that individuals have an intrinsic heterogeneity which cannot be controlled for using observables. Bitler, Gelbach, and Hoynes (2006) discuss the costs of focusing on average treatment estimation instead of other statistics.

Provided that in nonexperimental settings selection into treatment is not random, ordinary least squares (OLS) and quantile regression techniques are inconsistent. As stated by Heckman and Navarro-Lozano (2004), three different approaches were used to overcome this problem. First, the control function approach explicitly models the selection mechanism and its relation to the outcome equation; second, instrumental variables; and third, local estimation and aggregation. In the latter, under the unconfoundness assumption, which states that conditional on a given set of exogenous covariates

(observables) treatment occurrence is statistically independent of the potential outcomes, local unbiased estimates can be obtained by conditioning on this set of covariates. The identification strategy we follow rely on this assumption. Rosenbaum and Rubin (1983, 1984) show that adjusting solely for differences between treated and control units in a scalar function of the pre-treatment covariates, the propensity score, also removes the entire bias associated with differences in pre-treatment variables.

Several estimation methods have been proposed for estimating ATE by conditioning on the propensity score. *Matching* estimators are widely used in empirical settings, and in particular *propensity score matching*. In this case, each treated (non-treated) individual is *matched* to a non-treated (treated) individual (or aggregate of individuals) by means of their proximity in terms of the propensity score. Only in a few cases matching on more than one dimension has been used (see for instance Abadie and Imbens, 2002), because of the computational burden that multivariate matching requires. Moreover, Hirano, Imbens, and Ridder (2003) method uses a series estimator of the propensity score to obtain efficient (in the sense of Hahn, 1998) ATE estimators.

Estimation of QTE have been developed using the minimization of convex check functions as in Koenker and Bassett (1978). Abadie, Angrist, and Imbens (2002) and Chernozhukov and Hansen (2004, 2005) develop this methodology using instrumental variables. On the other hand, Firpo (2007) does not require instrumental variables, and his methodology follows a two-step procedure: in the first stage, he estimates the propensity score using a series estimator, while in the second, he uses a weighted quantile regression method.

Bitler, Gelbach, and Hoynes (2006) compute QTE using the empirical distribution function and derives an equivalent estimator. Diamond (2005) uses matching to construct comparable treated and non-treated groups, and then computes the difference between the matched sample quantiles.

An alternative source of heterogeneity comes from the consideration of observables only. Treatment effects may vary depending on the amount of human capital or on the income and job status of their families. Differences in terms of these covariates determines that one may be interested in the conditional treatment effect, that is conditional on some value of the observables. For instance, in terms of the propensity score, individuals which are more likely to receive a treatment may have a different effect than those which are less likely to receive it. As we show in this paper, how observables are treated determines differences in the parameter of interest for QTE but not for ATE. We define as the *average conditional quantile treatment effect* as our parameter of interest, which can be described as the average of local QTEs. This parameter is equivalent to the standard *unconditional* QTE only in the case that the quantile treatment effect is constant.

In many cases, one would be more interested in the dependence of the outcome variable on the fractiles (i.e. quantiles) of the covariates rather than the covariates themselves. Mahalanobis (1960) fractile graphical analysis (FGA) methodology was developed to account for this heterogeneity in observables. This method has awaken recent interest in the literature as a nonparametric regression technique (Bera and Gosh, 2006, Sen, 2005).

For our purposes, this methodology can be used as an alternative to matching, and it allows not only for estimating average but quantile treat-

ment effects. The idea is simple: divide the covariates space into fractiles and obtain the conditional regression (or quantile) by a step function. Provided that the number of fractile groups increases with the number of observations, we obtain consistent estimates of these functions, as the local estimators would satisfy the unconfoundedness assumption[1].

FGA can be viewed as a histogram-type smoother and it shares the convergence rate of histograms as opposed to kernel based methods that have a better performance. In the classification of Imbens (2004) it can be associated with the "blocking on the propensity score" methods. An advantage of this procedure is that only the number of fractile groups needs to be chosen as a smoothing parameter.

In spirit, this method is very similar to matching. The latter matches every treated individual to a control (non-treated) individual whose characteristics are similar. Then, using the unconfoundness assumption, it integrates over the covariates as the matched sample is similar to the treated. FGA decomposes the covariates distribution into fractiles. Then within each fractile treated and non-treated individuals are compared. Finally, it integrates over the covariates (in this case over the fractile groups) as matching does. However, this nonparametric technique allows us to recover the complete graph for the conditional expectation or quantiles. In the latter, we show

[1]Quoting Koenker and Hallock (2001, p.147): "(...) segmenting the sample into subsets defined according to the conditioning covariates is always a valid option. Indeed, such local fitting underlies all nonparametric quantile regression approaches. In the most extreme cases, we have p distinct cells corresponding to different settings of the covariate vector, x, and quantile regression reduces simply to computing univariate quantiles for each of these cells".

that the graph contains more information than the comparison of treated and non-treated separately.

The propensity score FGA estimators are compared to other estimators based on the propensity score. In particular we compare it to propensity score matching estimators and Hirano, Imbens, and Ridder (2003) estimator for ATE, and to Firpo (2007) for QTE.

The paper is organized as follows. Section 2 describes the general framework and defines the parameters of interest. Section 3 reviews the literature on FGA. Section 4 derives ATE estimators and Section 5 does it for QTE. Section 6 presents Monte Carlo evidence on the performance of these estimators while Section 7 applies them to a well-known job training dataset. Conclusions appear in Section 8.

2 A general set-up for non-random experiments and main estimands

2.1 Unconditional treatment effects

To more formally characterize the model we follow the potential-outcome notation used in Imbens (2004) , which dates back to Fisher (1935), Neyman (1990) and Rubin (1974, 1977, 1978) and it is standard in the literature.

Consider N individuals indexed by $i = 1, 2, ..., N$ who may receive a certain "treatment" (e.g. receiving job training), indicated by the binary variable $W_i = 0, 1$. Each individual has a pair of potential outcomes (Y_{1i}, Y_{0i}) that corresponds to the outcome with and without treatment respectively.

The fundamental problem, of course, is the inability to observe at the same time the same individual both with and without the treatment effect. That is, we only observe $Y_i = W_i \times Y_{1i} + (1 - W_i) \times Y_{0i}$ and a set of exogenous variables X_i. We are interested in measuring the "effect" of the W-treatment (e.g. whether job training increase salaries, or the chances of being employed).

A parameter of interest is the average treatment effect, ATE,

$$\delta = E[Y_1 - Y_0] \tag{2.1}$$

which tells us whether on average, the W-treatment has an effect on the population.

The key identification assumption is the *unconfoundedness assumption*[2] (Rubin, 1978, Rosenbaum and Rubin, 1983), which states that conditional on the exogenous variables, the treatment indicator is independent of the potential outcomes. More formally,

Assumption 1 *(Unconfoundedness)*

$$W(Y_1, Y_0)|X$$

where denotes statistical independence. Under this assumption we can identify the ATE (see Imbens, 2004) if both treated and non-treated have a common support, that is, comparable X-values:

[2]Rosenbaum and Rubin (1983) called this strongly ignorable treatment assignment assumption, Heckman, Ichimura, and Todd (1997) and Lechner (1999, 2000) conditional independence assumption.

$$
\begin{aligned}
\delta &= E[Y_1 - Y_0] = E_X[E[Y_1 - Y_0|X]] \\
&= E_X[E[Y_1|X, W = 1]] - E_X[E[Y_0|X, W = 0]] \\
&= E_X[E[Y|X, W = 1]] - E_X[E[Y|X, W = 0]]
\end{aligned}
$$

In some cases, we are interested not only on the average effect but on the effect on a sub-group of the population. Average treatment effects do not fully describe all the distributional features of the W-treatment. For instance, high ability individuals may benefit differently from program participation than low ability ones, even if they have the same value of covariates. This determines that the effect of a certain treatment would vary according to unobservable characteristics. A parameter of interest in the presence of heterogeneous treatment effects is the quantile treatment effect (QTE). As originally defined in Docksum (1974) and Lehmann (1974), the QTE corresponds, for any fixed percentile, to the horizontal distance between two cumulative distribution functions. Let F_0 and F_1 be the control and treated distribution of a certain outcome and let $\Delta(y)$ denote the horizontal distance at y between F_0 and F_1, i.e. $F_0(y) = F_1(y + \Delta(y))$ or $\Delta(y) = F_1^{-1}(F_0(y)) - y$. We can express this effect not in terms of y, but on the quantiles of the same variable, and the QTE is then

$$
\delta_\tau = \Delta(F_0^{-1}(\tau)) = F_1^{-1}(\tau) - F_0^{-1}(\tau) \equiv Q_{\tau j} - Q_{\tau j} \tag{2.2}
$$

where $Q_{\tau j}, j = 0, 1$ are the quantiles of the treated and non-treated outcome distributions.

The key identification assumption here is the *rank invariance* assumption

(which is implied by the unconfoundedness assumption): in both treatment statuses, all individuals would mantain their *rank* in the distribution (see Heckman, Smith, and Clemments, 1997, for a general discussion about this assumption). Therefore, using a similar argument as in the ATE case, Firpo (2007) shows that this assumption provides a way of identifying the QTE:

$$\begin{aligned}\tau = &\quad E\left[P[Y_1 \leq Q_{\tau 1}|X]\right] = E\left[P[Y_0 \leq Q_{\tau 0}|X]\right] \\ = &\quad E\left[P[Y \leq Q_{\tau 1}|X, W=1]\right] = E\left[P[Y \leq Q_{\tau 0}|X, W=0]\right]\end{aligned}$$

where the last two expectations can be estimated from the observable data.

In both cases, Assumption 1 suggests that by constructing cells of homogenous values of X, we would be able to get an unbiased estimate of the treatment effect. However this becomes increasingly difficult and computationally imposible as the dimension of X increases. Rosenbaum and Rubin (1983) argue that the unconfoundedness assumption can be re-stated in terms of the propensity score, $p(X) \equiv P[W = 1|X = x]$, under the following assumption

Assumption 2 *(Common support)*

For all $x \in domain(X)$ we have that

$$0 < \underline{p} \leq p(x) \leq \bar{p} < 1$$

In this case, we have:

Lemma 1 *Assumptions 1 and 2 imply that*

$$W(Y_1, Y_0)|p(X)$$

Proof: See Rosenbaum and Rubin (1983).

Therefore the problem can be reduced to the dimension of $p(X)$. Through this paper we consider estimators based only on the propensity score.

2.2 Conditional treatment effects

Let $Y_j(X) = E[Y_i|X]$ and $F_j(.|X)$,$j = 0, 1$ be the outcome distribution functions conditional on X, and let $H(.)$ be the distribution function of X. Then the ATE can be defined as:

$$\begin{aligned}\int \left[\int Y_1(X)dH(X)\right] dF_1(Y_1) - \int \left[\int Y_0(X)dH(X)\right] dF_0(Y_0) \\ = \int \left[\int Y_1(X)dF_1(Y_1|X) - \int Y_0(X)dF_0(Y_0|X)\right] dH(X)\end{aligned}$$

Therefore, ATE can be obtained by comparing the unconditional mean outcome for the treated and non-treated, or by obtaining first the conditional ATE and then integrating over the covariates space.

Now define

$$Q_{\tau j}(x) = F_j^{-1}(\tau|X = x) \equiv \inf\{Y_j : F_j(Y_j|X = x) \geq \tau\}, j = 0, 1 \qquad (2.3)$$

as the conditional τ-th quantile. In general

$$E_X[Q_{\tau j}(X)] \neq Q_{\tau j}, j = 0, 1 \qquad (2.4)$$

In other words, the above equivalence cannot be applied to QTE: comparing the unconditional quantiles of the outcome distributions is not equivalent

to computing the conditional quantiles and then aggregating. Chernozhukov and Hansen (2004, 2005) define the conditional quantile treatment effect (CQTE) as

$$\delta_\tau(x) = Q_{\tau 1}(x) - Q_{\tau 0}(x) \tag{2.5}$$

Define the *average conditional quantile treatment effect* (ACQTE) as

$$\bar{\delta}_\tau = E_X[Q_{\tau 1}(X) - Q_{\tau 0}(X)] \tag{2.6}$$

Strictly speaking, differences in $Q_{\tau 1}(X) - Q_{\tau 0}(X)$ can either be attributed to differences in the treatment effect or differences in the effect of the X's on the treated and non-treated. For instance, in a linear regression set-up, we may have $Q_{\tau j}(X) = \alpha(\tau, X)j + \beta(\tau, j)X, j = 0, 1$. In the job training example, we may have that training increases salaries and returns to schooling, where years of schooling is X. However, in general, both parameters cannot be identified separately, and the literature often attributes to the treatment the whole conditional difference, that is $\beta(\tau, j) = \beta(\tau), j = 0, 1$.

In order to see these differences consider the following simple example with one outcome variable. Let X be a uniform random variable on $(0, 1)$ and let

$$Y(X) = \begin{cases} 0 & \text{with prob. } 0.5 \\ 0.5 & \text{with prob. } 0.5 \end{cases} \quad \text{if } X \leq 0.5$$
$$\begin{cases} 0.5 & \text{with prob. } 0.5 \\ 1 & \text{with prob. } 0.5 \end{cases} \quad \text{if } X > 0.5$$

Here note that $E[Y] = E\left[E_X[Y]\right]$ by the Law of Iterated Expectations.

Let Q_τ be the quantile of the Y distribution and let $Q_\tau(X)$ be the conditional quantile of Y conditional on X. In this case,

$$Q_\tau = \begin{cases} 0 & \text{if } \tau < 0.25 \\ 0.5 & \text{if } 0.25 \leq \tau < 0.75 \\ 1 & \text{if } \tau \geq 0.75 \end{cases}$$

But,

$$E_X[Q_\tau(X)] = \begin{cases} 0.25 \text{ if } \tau < 0.5 \\ 0.75 \text{ if } \tau \geq 0.5 \end{cases}$$

This determines that recovering the complete graph $\{X, Q_{\tau j}(X)\}, j = 0, 1$ provides additional information that cannot be recovered by computing unconditional quantiles. Firpo (2007), Bitler, Gelbach, and Hoynes (2006) and Diamond (2005) estimators obtain unconditional quantiles because their estimators compute the difference between the treated and non-treated quantiles.

If we add X to the model and the treatment effect is constant across X, we have the following expression:

$$Q_\tau(X) = \alpha(\tau) + \beta(\tau)\mathbf{1}[X > 0.5] = 0.5 + 0.5 \times \mathbf{1}[X > 0.5], \forall \tau \qquad (2.7)$$

However, in this case we would be attributing no difference across quantiles. If we consider differences in the treatment effect across X:

$$Q_\tau(X) = \alpha(\tau, X) = \begin{cases} 0 & \text{if } \tau < 0.5 \\ 0.5 & \text{if } \tau \geq 0.5 \end{cases} \text{ if } X \leq 0.5$$
$$\begin{cases} 0.5 & \text{if } \tau < 0.5 \\ 1 & \text{if } \tau \geq 0.5 \end{cases} \text{ if } X > 0.5$$

We assume that $Y_j(.), Q_{\tau j}(.), j = 0, 1$ can be expressed as a function of p. In particular, for QTE we assume that the CQTE is of the form $Q_{\tau 1}(p) - Q_{\tau 0}(p) = \alpha(\tau, p)$, and therefore the ACQTE becomes

$$\bar{\delta}_\tau = E_p[Q_{\tau 1}(p) - Q_{\tau 0}(p)] = E_p[\delta_\tau(p)] \tag{2.8}$$

which is our parameter of interest.

3 Fractile graphical analysis

Fractile graphical analysis (FGA) is a nonparametric estimation method developed first by Mahalanobis (1960) based on conditioning on the fractiles of the X's. It was specifically designed to compare two populations where the X variable was influenced by inflation and therefore not directly comparable. It has the same properties as other histogram-type estimators (Bhattacharya and Müller, 1993). Moreover, Bhattacharya (1963) developed a conditional quantile estimation method based on FGA. Our proposal is to use FGA to develop estimators for both ATE and QTE. FGA produces a histogram-type smoother by blocking on the fractiles (i.e. quantiles) of the propensity score.

FGA was originally developed for one covariate (i.e. $dim(X) = 1$), but Bhattacharya (1963) and others showed that it can be extended to more covariates. However, we will only consider FGA based on a single covariate, the propensity score. One dimensional FGA allows us to recover the graphs $\{p, \gamma(p)\}$ where γ is any function of the propensity score.

Assume first that the propensity score is known and it has a distribution function $H(p)$. Further, assume that $H(.)$ is continuous and strictly increas-

ing, and p satisfies Assumption 2. Construct R fractile groups (indexed by r) on the propensity score:

$$\Im_p^r = \left\{ p \in [\underline{p}, \bar{p}] : \xi_{\frac{r-1}{R}} < p \leq \xi_{\frac{r}{R}}, \xi_{\frac{r-1}{R}} = H^{-1}\left(\frac{r-1}{R}\right), \xi_{\frac{r}{R}} = H^{-1}\left(\frac{r}{R}\right) \right\} \tag{3.1}$$

$$r = 1, 2, ...R$$

where $H^{-1}(\tau) = \inf\{p : H(p) \geq \tau\}$.

Each fractile group contains a similar number of observations (i.e. about N/R) and it has an associated interval on the domain of p defined by the order statistics $(\xi_{\frac{r-1}{R}}, \xi_{\frac{r}{R}}]$, such that $P\left[p \in (\xi_{\frac{r-1}{R}}, \xi_{\frac{r}{R}}]\right] \simeq 1/R$. As the number of fractiles increases, the divergence in terms of p for all observations within the same fractile group becomes smaller, and therefore we would be gradually constructing groups with the same p-characteristics. In that case, estimates within each fractile group asymptotically satisfy the unconfoundedness assumption, provided that the conditioning set converges to a single propensity score value.

The following lines provide a short review of the asymptotic properties of FGA, which can be found in Bhattacharya and Müller (1993) and Bera and Gosh (2006). Let $g(p) = E[Y|P = p]$ and $\sigma^2(p) = VAR[Y|P = p]$ be the conditional expectation and variance in terms of the propensity score, and consider the following notation: $h(t) = g \circ H^{-1}(t)$ and $k(t) = \sigma^2 \circ H^{-1}(t)$ for $t = (r - 1 + \alpha)/R$ with $0 \leq \alpha \leq 1$. Suppose that $h(.)$ has bounded second derivative and $k(.)$ has bounded first derivative. Then as $N \to \infty$ and $R \to \infty$ so that $R/N \to 0$ for fixed t, the bias and the variance of an

FGA estimator of $h(t)$, $\hat{h}(t)$ become:

$$BIAS: \ E[\hat{h}(t) - h(t)] = -(2R)^{-1}h'(t)[1 + o(1)] = O(1/R)$$

$$VARIANCE: \ VAR[\hat{h}(t)] = (R/N)\left[(1-\alpha)^2 + \alpha^2\right]k(t)[1+o(1)] = O(R/N)$$

so that the mean-squared error of $\hat{h}$ is

$$MSE: \ MSE[\hat{h}(t)] = \left[(4R^2)^2(h'(t)) + (R/N)\{(1-\alpha)^2 + \alpha^2\}k(t)\right][1 + o(1)]$$

where $0 \leq \alpha = Rt - [Rt] < 1$. Therefore the best rate of convergence of fractile graphs is obtained by letting $R = O(N^{1/3})$, which yields a rate of $O(N^{-2/3})$ for the Integrated MSE.

If p is not known it has to be estimated. In practice any estimate $\hat{p} = p + o_p(1)$ remove the bias. However, they will differ in the variance of the estimator, provided that the first stage (i.e. the estimation of the propensity score) need to be taken into account. Hahn (1998) shows that by using the estimated propensity score, instead of the true propensity score, efficiency is achieved. Hirano, Imbens, and Ridder (2003) and Firpo (2007) use a semiparametric series estimator of the propensity score which produces this result.

We impose the following assumption regarding the use of the estimated propensity score:

Assumption 3 *(Convergence of propensity score fractile groups)*

Let $\hat{p}$ be an estimator of the propensity score. Then for fixed R and for all r,

$$\lim_{N\to\infty} P\left[\Im_p^r = \Im_{\hat{p}}^r\right] = 1$$

4 ATE estimators

FGA ATE estimators are based on imputing the unobserved outcome in each fractile group. Let

$$\hat{Y}_{1i} = \left\{ \begin{array}{c} Y_i \text{ if } W_i = 1 \\ \check{Y}_{1i} \text{ if } W_i = 0 \end{array} \right\}$$

where $\check{Y}_{1i} = \frac{\sum_{k=1}^{N} W_k Y_k \mathbf{1}[\hat{p}_k \in \Im_{\hat{p}}^{r_i}]}{\sum_{k=1}^{N} W_k \mathbf{1}[\hat{p}_k \in \Im_{\hat{p}}^{r_i}]}$

and

$$\hat{Y}_{0i} = \left\{ \begin{array}{c} Y_i \text{ if } W_i = 0 \\ \check{Y}_{0i} \text{ if } W_i = 1 \end{array} \right\}$$

where $\check{Y}_{0i} = \frac{\sum_{k=1}^{N} (1-W_k) Y_k \mathbf{1}[\hat{p}_k \in \Im_{\hat{p}}^{r_i}]}{\sum_{k=1}^{N} (1-W_k) \mathbf{1}[\hat{p}_k \in \Im_{\hat{p}}^{r_i}]}$

Therefore the FGA ATE estimator is

$$\hat{\delta} = \frac{1}{N} \sum_{i=1}^{N} \hat{Y}_{1i} - \hat{Y}_{0i} = \frac{1}{N} \sum_{i=1}^{N} \check{Y}_{1i} - \check{Y}_{0i} \tag{4.1}$$

Similarly, it can be expressed as

$$\hat{\delta} = \frac{1}{R} \sum_{r=1}^{R} \hat{\delta}^{(r)} \tag{4.2}$$

where

$$\hat{\delta}^{(r)} = \frac{\sum_{i=1}^{N} W_i Y_i \mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^{r}]}{\sum_{i=1}^{N} W_i \mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^{r}]} - \frac{\sum_{i=1}^{N} (1 - W_i) Y_i \mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^{r}]}{\sum_{i=1}^{N} (1 - W_i) \mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^{r}]}$$

The logic of this estimator is based on that of Hahn (1998) "nonparametric imputation". In this case, witihin each fractile group, $E[WY|\Im_p^r]$,

$E[(1-W)Y|\Im_p^r]$ and $E[W|\Im_p^r]$ are estimated nonparametrically using the previously estimated propensity score ($\hat{p}$).

Alternatively we construct a similar estimator using the weighting technique described in Hirano, Imbens, and Ridder (2003). Let

$$\tilde{Y}_{1i} = \left\{ \begin{array}{c} Y_i \text{ if } W_i = 1 \\ \breve{Y}_{1i} \text{ if } W_i = 0 \end{array} \right\}$$

where $\breve{Y}_{1i} = \sum_{k=1}^{N} \frac{W_k Y_k}{\hat{p}_k} \mathbf{1}[\hat{p}_k \in \Im_{\hat{p}}^{r_i}]$

and

$$\tilde{Y}_{0i} = \left\{ \begin{array}{c} Y_i \text{ if } W_i = 0 \\ \breve{Y}_{0i} \text{ if } W_i = 1 \end{array} \right\}$$

where $\breve{Y}_{0i} = \sum_{k=1}^{N} \frac{(1-W_k)Y_k}{1-\hat{p}_k} \mathbf{1}[\hat{p}_k \in \Im_{\hat{p}}^{r_i}]$.

Then

$$\tilde{\delta} = \frac{1}{R} \sum_{r=1}^{R} \tilde{\delta}^{(r)} \tag{4.3}$$

where

$$\tilde{\delta}^{(r)} = \frac{R}{N} \sum_{i=1}^{N} \frac{W_i}{\hat{p}_i} Y_i \mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^r] - \frac{R}{N} \sum_{i=1}^{N} \frac{1-W_i}{1-\hat{p}_i} Y_i \mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^r] \tag{4.4}$$

This estimator suffers from the same problems of Hirano, Imbens, and Ridder (2003) estimator, that is, the presence of occassional high/low values of the propensity score produce a very bad empirical performance.

The following theorem shows that the FGA ATE estimators are consistent. The intuition behind the proof is that as N increases, and R does it but at a smaller rate, each fractile group will have individuals with similar

propensity score values. In the limit, the differences among them is negligible, and therefore the unconfoundedness assumption can be applied. In this case, the local (i.e. for a given propensity score value) ATE can be obtained by constructing the difference of the average treated and control individuals with that propensity score value.

Theorem 1 *(Consistency of ATE estimator)*

Consider Assumptions 1,2 and 3 and assume that

1. *The distribution functions of* p *and* $(Y_1, Y_0)|p$ *are continuous and strictly increasing.*

2. $E[Y_1^2] < \infty, E[Y_0^2] < \infty$

Then, $\hat{\delta} \xrightarrow{P} \delta$ *and* $\tilde{\delta} \xrightarrow{P} \delta$ *as* $N, R \to \infty$,$R/N \to 0$.

Proof: See Appendix.

5 QTE estimators

Define the within fractile conditional quantiles:

$$\hat{Q}_{\tau 1}^{(r)} = argmin_q \frac{\sum_{i=1}^{N} \mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^r] W_i (Y_i - q)(\tau - \mathbf{1}[Y_i \leq q])}{\sum_{i=1}^{N} \mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^r] W_i}$$

$$\hat{Q}_{\tau 0}^{(r)} = argmin_q \frac{\sum_{i=1}^{N} \mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^r] (1 - W_i)(Y_i - q)(\tau - \mathbf{1}[Y_i \leq q])}{\sum_{i=1}^{N} \mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^r] (1 - W_i)}$$

Therefore the QTE estimator is

$$\hat{\delta}_\tau = \frac{1}{R}\sum_{r=1}^{R}\hat{\delta}_\tau^{(r)} = \frac{1}{R}\sum_{r=1}^{R}\hat{Q}_{\tau 1}^{(r)} - \hat{Q}_{\tau 0}^{(r)} \tag{5.1}$$

Similarly we define

$$\tilde{Q}_{\tau 1}^{(r)} = argmin_q \sum_{i=1}^{N} \frac{1}{\hat{p}_i}\mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^r]W_i(Y_i - q)(\tau - \mathbf{1}[Y_i \leq q])$$

$$\tilde{Q}_{\tau 0}^{(r)} = argmin_q \sum_{i=1}^{N} \frac{1}{1-\hat{p}_i}\mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^r](1 - W_i)(Y_i - q)(\tau - \mathbf{1}[Y_i \leq q])$$

and

$$\tilde{\delta}_\tau = \frac{1}{R}\sum_{r=1}^{R}\tilde{\delta}_\tau^{(r)} = \frac{1}{R}\sum_{r=1}^{R}\tilde{Q}_{\tau 1}^{(r)} - \tilde{Q}_{\tau 0}^{(r)} \tag{5.2}$$

The following theorem proves the consistency of both QTE estimators.

Theorem 2 *(Consistency of QTE estimator)*

Consider Assumptions 1,2 and 3 and assume that:

The distribution function of p is continuous and strictly increasing. The distribution function of $(Y_1, Y_0)|p$ *is continuous, strictly increasing and continuously differentiable.*

Then, for $\tau \in (0,1)$, $\hat{\delta}_\tau \xrightarrow{P} \delta_\tau$ *and* $\tilde{\delta}_\tau \xrightarrow{P} \delta_\tau$ *as* $N, R \to \infty$,$R/N \to 0$.

Proof: See Appendix.

6 Monte Carlo experiments

We evaluate the performance of the proposed estimators with respect to other estimators based on the propensity score. We compute propensity score matching estimators using nearest neighbor procedures (with 1, 2 and 4 matches per observation), kernel and spline estimates. These estimators were designed by Barbara Sianesi for `STATA 9.1` and they are available in the `psmatch2` package. Additionally we compute Hirano, Imbens, and Ridder (2003) semiparametric efficient estimator. In the case of QTE we compute Firpo (2007) and Bitler, Gelbach, and Hoynes (2006) estimators. We also compute QTE matching estimators following Diamond (2005). In this case, for each observation, the matching procedure constructs the corresponding matched pair (i.e. imputes the "closest" observation with the opposite treatment status). Then, we compute the unconditional quantiles of the imputed treated and non-treated distributions. A succinct description of some estimators appears in the Appendix.

Our baseline model is

$$X_1, X_2, X_3, u, e \sim N(0,1)$$

$$W = \mathbf{1}[X_1 - X_2 + X_3 + e > 0]$$

$$Y_1 = \delta + X_1 + X_2 + u$$

$$Y_0 = X_1 + X_2 + X_3 + u$$

In this simple model QTEs are equal to ATE for all quantiles. We set $\delta = 2$. We generate 1000 replications of the baseline models for sample sizes in $\{100, 200, 500, 1000, 2000\}$, and we compute mean square error (MSE) and

mean absolute error (MAE). Table 1 reports ATE estimators while Table 2 shows QTE estimators for τ in $\{.10, .25, .50, .75, .90\}$. For FGA the number of fractile groups are $R = [N^{1/3}]$ which minimizes the Integrated MSE (see Bhattacharya and Müller, 1993), and we also consider doubling the number of fractile groups (i.e. $R \times 2$). We consider the two FGA estimators discussed above, that is, $\hat{\delta}$ and $\tilde{\delta}$.

The FGA ATE estimator has reasonable good performance in terms of both MSE and MAE. In almost every case, doubling the number of fractile groups results in a better performance of the $\hat{\delta}$ estimator. However, the contrary occurs to the $\tilde{\delta}$ estimator. FGA ATE $\hat{\delta}$ $(R \times 2)$ achieves the same values of the best matching estimators (using 4 neighbors and splines). Increasing the sample size reduces both MSE and MAE at similar rates in all estimators. Overall the Hirano, Imbens, and Ridder (2003) and FGA ATE $\tilde{\delta}$ estimators show extremely high values, mainly because a random draw may contain occasional values of the propensity score very close to the boundary (i.e. 0 or 1).

FGA QTE $\hat{\delta}$ estimators outperform that of Firpo (2007) for all sample sizes and quantiles. All the estimators show consistency, although FGA QTE reduce both MSE and MAE at higher rates than the Firpo's estimator. As in the last paragraph, doubling the number of fractile groups improves the estimator performance and FGA QTE $\hat{\delta}$ outperform $\tilde{\delta}$. As expected, better estimates are found in the median case than in the extreme quantiles. Matching estimators show a relatively good performance. However, only in a few cases they outperform the FGA QTE estimator. In particular the spline matching estimator shows an outstanding performance for $\tau = 0.9$.

Overall nonparametric FGA estimators where the propensity score is re-estimated nonparametrically (i.e. $\hat{\delta}$) show the best performance.

7 Empirical application

We apply the estimators proposed in the paper to a widely used job training dataset first analyzed by LaLonde (1986), the "National Supported Work Program" (NSW). The same database was used in other applications such as Heckman and Hotz (1989), Dehejia and Wahba (1999, 2002), Abadie and Imbens (2002) and Firpo (2007) among others.

The program was designated as a random experiment for applicants who if selected would had received work experience (treatment) in a wide range of possible activities, like learning to operate a restaurant, a child care or a construction work, for a period not exceeding twelve months. Eligible participants were targeted from recipients of AFDC, former addicts, former offenders and young school dropouts. Candidates eligible for the NSW were randomized into the program between March 1975 and July 1977. The NSW data set consists of information on earnings and employment in 1978 (outcome variables); whether treated or not; information on earnings and employment in 1974 and 1975; and background characteristics such as education, ethnicity, marital status and age. We use the database provided by Guido Imbens (http://emlab.berkley.edu/users/imbens), which consist on 455 individuals, 185 treated and 260 control observations. This particular subset is the one constructed by Dehejia and Wahba (1999) and described there in more detail.

We will focus on the possible effect on participants' earnings in 1978 (if any), that is, we answer the question: what is the effect of this particular training program on future earnings? Provided that earnings is a continuous variable we would be able to apply quantile analysis. A main drawback of this variable is that those unemployed in 1978 report earnings of zero. In 1978, 92 control and 45 treated individuals were unemployed. The average (standard deviation) of earnings in 1978 is \$5300 (\$6631), which breaks into \$6349 (\$578) for treated and \$4554 (\$340) for control individuals. Without considering covariates, the difference between treated and non-treated is \$1794 (\$671), which in a two-sample t test rejects the null hypothesis of equal values (t-stat 2.67, p-value 0.0079). We also observe differences in terms of the percentiles in the earnings distribution. The 10th percentile for the treated (control) is \$0 (\$0); the 25th percentile \$485 (\$0); the median is \$4232 (\$3139); the 75th percentile 9643(7292); and the 90th percentile is \$14582 (\$11551). Therefore, assuming the rank invariance property discussed above, higher quantiles of the earnings distribution seems to be associated with larger treatment effects.

The propensity score is estimated by a probit model where the dependent variable is participation and the covariates used are the individual characteristics and employment and earnings in 1974 and 1975. Note that the propensity score is of no particular interest by itself, provided that participants were randomly selected in the experiment. In this case, no particular covariate is individually significant and a likelihood ratio test of joint significance gets chi-squared(8)=8.30, p-value=0.4050.

Figure 1 presents the kernel density estimate of the estimated propen-

sity score together with a smoothed version of earnings in 1978, separately for both treated and non-treated. It is clearly observed a dissimilar behavior. As the propensity score increases, the earnings of the treated increases as well, while those of non-treated show no clear pattern. In other words, those individuals who were more likely to be selected benefit more from the treatment.

As we mention above, a common support in the propensity score domain is necessary to make meaningful comparisons among treated and non-treated individuals. The empirical relevance of this assumption was pointed out by Heckman, Ichimura, and Todd (1998) and it was identified as one of the major sources of bias. In our case, this has special importance since consistent estimates of treatment effect requires that both the number of treated and control is eventually large enough to apply large sample theory. Moreover, if there are no treated (controls) in a given fractile group, no within fractile estimate can be obtained. We use two different trimming procedures. First, provided that we may assume that $F_1(p) \leq F_0(p)$, we only consider propensity score values in the range:

$$\underline{p}^* = \min_p(p_i, W_i = 1) \leq p \leq \max_p(p_i, W_i = 0) = \bar{p}^*$$

By doing this we drop 8 observations, and we refer to this sample as Trim 1. We also trim 2.5% in each tail of the propensity score distribution (Trim 2) droppping 23 observations.

Table 3 reports the propensity score estimates used in the Monte Carlo simulation, applied to the LaLonde's data set. The first column contains the ATE estimate, while the second and third contain the average and standard

deviation of a bootstrapping experiment with 1000 random samples with replacement of the original database. The last column calculates the ATE estimator for the two different trimming procedures discussed above. Table 4 estimates the QTE for the same quantiles analyzed in Table 2. The results confirm a positive average impact of training on earnings. FGA ATE estimators get $1572 and $1537, which are of the same magnitude as the kernel and spline propensity score matching estimates and the Hirano, Imbens, and Ridder (2003) estimates. However, nearest neighbor estimates are below these estimates by $100. Figure 2 plots the estimated local ATE in terms of the propensity score the FGA ATE estimators. All estimators are very responsive to the trimming procedure, which as in Dehejia and Wahba (1999) article, casts doubts about their robustness.

QTE estimates show considerable variability across quantiles (see Table 4). For the 10th quantile estimates are not statistically different from zero. The median quantile is almost two thirds of the ATE estimates, reflecting the presence of outliers in the sample or different distributional properties. Finally for the 90th quantile, the estimates produce up to a $3000 impact, twice the ATE. In other words, those who benefit more are those with a high level of unobservables. Unfortunately, all the estimators show high bootstrap standard errors.

8 Conclusions and suggestions for future research

FGA provides a simple methodology for constructing nonparametric estimators of average and quantile treatment effects, under the assumption of selection on observables. In this paper we develop estimators using the estimated propensity score and we prove its consistency. Moreover, FGA QTE estimators show a better performance than that of Firpo (2007) QTE estimator, which constitutes the most relevant estimator in the literature using the propensity score.

Similar estimators can be derived for FGA in more than one dimension (see for instance the discussion in Bhattacharya, 1963), although its computational burden is unknown. Moreover, more efficient estimators may be obtained by applying smoothing techniques within or between fractiles (Sen, 2005).

Figure 1: Empirical density of propensity score and earnings

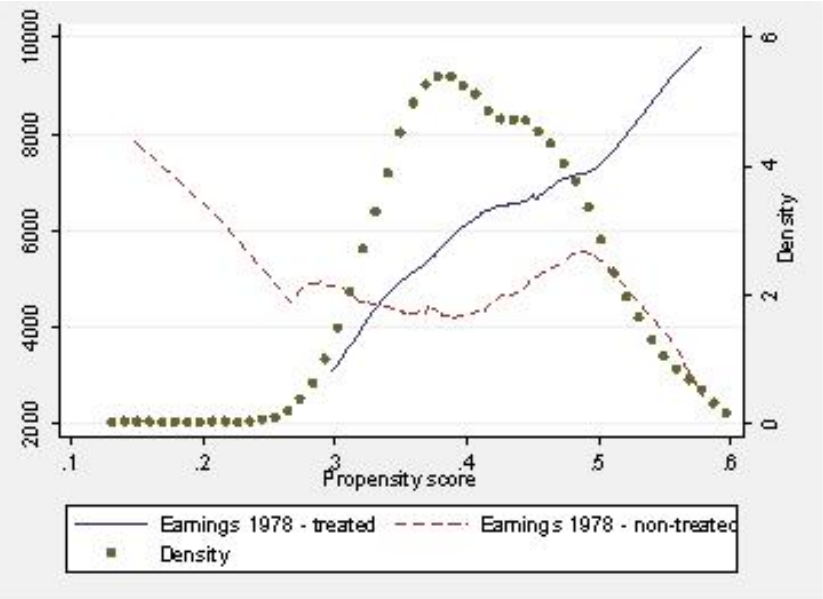

Figure 2: Empirical density of propensity score and FGA ATE

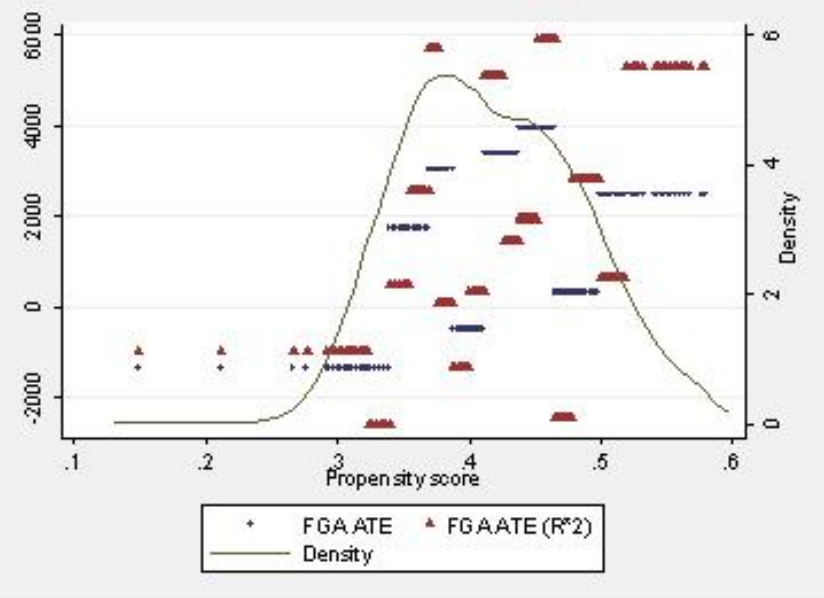

Table 1: ATE Monte Carlo simulations

	MSE					MAE				
Estimator	100	200	500	1000	2000	100	200	500	1000	2000
FGA ATE (a)	0.221	0.157	0.079	0.057	0.036	0.372	0.308	0.221	0.187	0.147
FGA ATE ($R \times 2$) (a)	0.190	0.109	0.053	0.037	0.023	0.346	0.261	0.185	0.155	0.121
FGA ATE (b)	0.686	0.657	0.630	0.607	1.699	0.464	0.456	0.369	0.343	0.336
FGA ATE ($R \times 2$) (b)	0.646	0.615	0.612	0.597	1.683	0.462	0.450	0.368	0.341	0.335
Hirano, Imbens, and Ridder (2003)	0.732	0.679	0.644	0.618	1.710	0.472	0.460	0.371	0.344	0.337
PS Matching										
Nearest neighbor (1)	0.467	0.331	0.202	0.145	0.099	0.540	0.443	0.358	0.298	0.251
Nearest neighbor (2)	0.290	0.192	0.126	0.091	0.063	0.429	0.347	0.285	0.238	0.198
Nearest neighbor (4)	0.170	0.124	0.074	0.057	0.039	0.329	0.281	0.217	0.188	0.157
Kernel	0.285	0.226	0.127	0.074	0.033	0.418	0.369	0.283	0.217	0.146
Spline	0.233	0.139	0.058	0.034	0.021	0.384	0.297	0.192	0.149	0.117

Notes: (a) $\hat{\delta}_\tau$, (b) $\tilde{\delta}_\tau$. MSE: Mean Squared Error. MAE: Mean Absolute Error. Monte Carlo simulations based on 1000 replications of the baseline model.

Table 2: QTE Monte Carlo simulations

	MSE					MAE				
Estimator	100	200	500	1000	2000	100	200	500	1000	2000
	$\tau = 0.10$									
FGA QTE (a)	0.964	0.585	0.328	0.247	0.155	0.791	0.611	0.461	0.404	0.317
FGA QTE ($R \times 2$) (a)	0.825	0.433	0.224	0.178	0.117	0.713	0.525	0.379	0.339	0.279
FGA QTE (b)	0.863	0.491	0.243	0.146	0.090	0.802	0.593	0.409	0.312	0.241
FGA QTE ($R \times 2$) (b)	1.334	0.849	0.391	0.241	0.146	1.074	0.847	0.561	0.433	0.339
Firpo (2007)	0.879	0.633	0.438	0.332	0.332	0.733	0.639	0.525	0.467	0.453
Bitler, Gelbach, and Hoynes (2006)	0.835	0.628	0.432	0.332	0.332	0.733	0.638	0.524	0.467	0.453
PS Matching										
Nearest neighbor (1)	0.775	0.649	0.455	0.341	0.383	0.697	0.636	0.532	0.486	0.478
Nearest neighbor (2)	0.551	0.442	0.315	0.269	0.236	0.603	0.555	0.484	0.462	0.444
Nearest neighbor (4)	0.562	0.429	0.309	0.285	0.259	0.591	0.557	0.492	0.494	0.484
Kernel	0.748	0.566	0.467	0.401	0.361	0.691	0.617	0.580	0.587	0.581
Spline	0.626	0.519	0.400	0.389	0.367	0.646	0.622	0.575	0.593	0.591
	$\tau = 0.25$									
FGA QTE (a)	0.523	0.333	0.184	0.133	0.083	0.585	0.459	0.336	0.292	0.230
FGA QTE ($R \times 2$) (a)	0.409	0.246	0.131	0.088	0.058	0.507	0.397	0.291	0.236	0.194
FGA QTE (b)	0.866	0.519	0.270	0.169	0.116	0.823	0.631	0.450	0.352	0.290
FGA QTE ($R \times 2$) (b)	1.400	0.867	0.462	0.292	0.195	1.126	0.874	0.634	0.503	0.410
Firpo (2007)	0.721	0.527	0.360	0.268	0.226	0.641	0.543	0.436	0.377	0.338
Bitler, Gelbach, and Hoynes (2006)	0.687	0.527	0.360	0.268	0.226	0.635	0.543	0.436	0.377	0.338
PS Matching										
Nearest neighbor (1)	0.952	0.757	0.471	0.309	0.209	0.731	0.621	0.517	0.420	0.364
Nearest neighbor (2)	0.563	0.381	0.287	0.195	0.147	0.579	0.472	0.403	0.339	0.306
Nearest neighbor (4)	0.312	0.204	0.153	0.110	0.089	0.438	0.352	0.296	0.255	0.231
Kernel	0.519	0.390	0.299	0.175	0.101	0.553	0.469	0.400	0.313	0.246
Spline	0.491	0.274	0.154	0.104	0.095	0.533	0.406	0.312	0.276	0.278
	$\tau = 0.50$									

Continued on next page

Table 2 – continued from previous page

	MSE					MAE				
Estimator	100	200	500	1000	2000	100	200	500	1000	2000
FGA QTE (a)	0.305	0.191	0.105	0.072	0.046	0.441	0.344	0.259	0.211	0.168
FGA QTE ($R \times 2$) (a)	0.252	0.143	0.070	0.049	0.029	0.394	0.301	0.212	0.175	0.138
FGA QTE (b)	0.941	0.616	0.349	0.242	0.169	0.871	0.703	0.529	0.435	0.367
FGA QTE ($R \times 2$) (b)	1.541	0.977	0.567	0.370	0.270	1.187	0.940	0.716	0.576	0.496
Firpo (2007)	0.658	0.540	0.358	0.249	0.206	0.606	0.521	0.401	0.332	0.285
Bitler, Gelbach, and Hoynes (2006)	0.629	0.540	0.358	0.249	0.206	0.603	0.521	0.401	0.332	0.285
PS Matching										
Nearest neighbor (1)	1.288	0.748	0.443	0.271	0.184	0.921	0.697	0.534	0.423	0.350
Nearest neighbor (2)	0.993	0.608	0.359	0.229	0.152	0.809	0.624	0.489	0.390	0.317
Nearest neighbor (4)	0.656	0.465	0.292	0.194	0.137	0.654	0.547	0.442	0.360	0.298
Kernel	0.367	0.250	0.130	0.084	0.060	0.476	0.396	0.292	0.234	0.200
Spline	0.648	0.349	0.181	0.121	0.093	0.642	0.476	0.343	0.285	0.235
	$\tau = 0.75$									
FGA QTE (a)	0.391	0.256	0.138	0.092	0.069	0.497	0.411	0.295	0.242	0.205
FGA QTE ($R \times 2$) (a)	0.352	0.218	0.112	0.071	0.051	0.466	0.376	0.267	0.215	0.180
FGA QTE (b)	1.043	0.721	0.452	0.320	0.250	0.919	0.768	0.610	0.515	0.461
FGA QTE ($R \times 2$) (b)	1.635	1.094	0.675	0.486	0.368	1.229	1.001	0.787	0.670	0.585
Firpo (2007)	0.652	0.561	0.356	0.346	0.250	0.615	0.548	0.420	0.390	0.322
Bitler, Gelbach, and Hoynes (2006)	0.652	0.561	0.356	0.346	0.250	0.615	0.548	0.420	0.390	0.322
PS Matching										
Nearest neighbor (1)	1.439	1.138	0.710	0.492	0.317	0.888	0.801	0.624	0.532	0.440
Nearest neighbor (2)	0.768	0.647	0.509	0.396	0.257	0.668	0.599	0.527	0.464	0.387
Nearest neighbor (4)	0.429	0.374	0.326	0.293	0.230	0.507	0.466	0.416	0.389	0.349
Kernel	0.486	0.576	0.407	0.328	0.255	0.534	0.556	0.484	0.445	0.400
Spline	0.713	0.596	0.342	0.284	0.258	0.631	0.578	0.438	0.412	0.408
	$\tau = 0.90$									
FGA QTE (a)	0.711	0.398	0.229	0.167	0.125	0.666	0.505	0.385	0.328	0.287
FGA QTE ($R \times 2$) (a)	0.707	0.333	0.185	0.142	0.100	0.671	0.457	0.345	0.304	0.257

Continued on next page

Table 2 – continued from previous page

	MSE					MAE				
Estimator	100	200	500	1000	2000	100	200	500	1000	2000
FGA QTE (b)	1.218	0.837	0.560	0.417	0.345	0.979	0.824	0.678	0.588	0.544
FGA QTE ($R \times 2$) (b)	1.771	1.176	0.809	0.589	0.473	1.269	1.029	0.860	0.736	0.664
Firpo (2007)	0.754	0.717	0.444	0.462	0.340	0.651	0.625	0.479	0.461	0.401
Bitler, Gelbach, and Hoynes (2006)	0.754	0.717	0.444	0.462	0.340	0.651	0.625	0.479	0.461	0.401
PS Matching										
Nearest neighbor (1)	0.898	0.930	0.674	0.789	0.631	0.699	0.690	0.566	0.562	0.510
Nearest neighbor (2)	0.424	0.370	0.231	0.287	0.241	0.506	0.449	0.352	0.338	0.310
Nearest neighbor (4)	0.337	0.204	0.112	0.122	0.113	0.456	0.356	0.249	0.228	0.217
Kernel	0.697	0.708	0.581	0.394	0.131	0.613	0.591	0.504	0.395	0.267
Spline	0.432	0.287	0.105	0.078	0.058	0.501	0.411	0.256	0.228	0.209

Notes: (a) $\hat{\delta}_\tau$, (b) $\tilde{\delta}_\tau$. MSE: Mean Squared Error. MAE: Mean Absolute Error. Monte Carlo simulations based on 1000 replications of the baseline model.

Table 3: ATE estimators of the effect of training on earnings

	Non-trimmed			Trim 1			Trim 2		
Estimator	Coef.	Average	Std.Error	Coef.	Average	Std.Error	Coef.	Average	Std.Error
FGA ATE	1572.2	1589.0	634.0	1608.6	1620.2	645.3	1480.9	1565.2	665.8
FGA ATE ($R \times 2$)	1537.1	1606.4	659.0	1520.2	1634.4	659.8	1208.6	1572.4	682.3
FGA ATE (b)	1584.1	1561.9	616.7	1672.5	1625.7	637.0	1676.1	1536.1	655.5
FGA ATE ($R \times 2$) (b)	1563.2	1511.4	606.6	1604.5	1576.1	627.3	1530.2	1483.4	644.5
Hirano, Imbens, and Ridder (2003)	1598.2	1612.2	631.3	1731.3	1691.9	661.0	1862.0	1589.1	682.6
PS Matching									
Nearest neighbor (1)	997.2	1393.8	736.4	1101.6	1389.2	721.3	869.8	1334.4	744.8
Nearest neighbor (2)	1156.2	1471.6	710.6	1262.0	1477.8	702.8	984.2	1424.2	729.9
Nearest neighbor (4)	1471.8	1559.0	670.0	1552.7	1571.7	671.6	1346.5	1525.1	700.4
Kernel	1629.0	1639.6	631.1	1638.9	1636.3	634.3	1482.4	1590.8	657.3
Spline	1587.0	1616.6	641.4	1614.4	1613.8	638.8	1380.3	1565.6	660.9

Notes: (a) $\hat{\delta}_\tau$, (b) $\tilde{\delta}_\tau$. Averages and Standard Errors are obtained using 1000 bootstrapping random samples with replacement of the original database.

Table 4: QTE estimators of the effect of training on earnings

	Non-trimmed			Trim 1			Trim 2		
Estimator	Coef.	Average	Std.Error	Coef.	Average	Std.Error	Coef.	Average	Std.Error
	$\tau = 0.10$								
FGA QTE (a)	0.0	149.4	188.2	0.0	146.8	190.9	0.0	145.5	183.0
FGA QTE ($R \times 2$) (a)	95.7	382.5	325.2	78.0	371.3	311.2	101.9	392.4	340.2
FGA QTE (b)	0.0	150.9	190.5	0.0	148.6	190.9	0.0	145.0	185.2
FGA QTE ($R \times 2$) (b)	95.7	369.7	317.7	78.0	364.6	312.0	240.1	385.1	335.9
Firpo (2007)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Bitler, Gelbach, and Hoynes (2006)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PS Matching									
Nearest neighbor (1)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nearest neighbor (2)	0.0	0.7	15.2	0.0	0.7	15.2	0.0	1.1	21.0
Nearest neighbor (4)	0.0	73.8	184.5	0.0	69.7	179.3	0.0	74.5	188.7
Kernel	0.0	292.5	356.3	0.0	281.0	351.0	0.0	298.9	365.8
Spline	0.0	295.3	355.7	0.0	278.5	349.5	0.0	291.2	361.0
	$\tau = 0.25$								
FGA QTE	409.6	653.5	480.2	376.4	686.1	475.9	623.2	701.7	502.8
FGA QTE ($R \times 2$)	628.8	853.5	546.4	487.4	864.3	540.9	836.9	880.2	547.8
FGA QTE (b)	414.2	649.8	488.0	361.9	685.8	484.2	500.9	705.8	511.8
FGA QTE ($R \times 2$) (b)	694.5	843.3	546.3	363.1	853.3	554.6	856.6	872.1	561.9
Firpo (2007)	0.0	295.8	341.3	0.0	291.3	340.4	289.8	276.8	343.2
Bitler, Gelbach, and Hoynes (2006)	0.0	295.8	341.3	0.0	291.3	340.4	289.8	276.8	343.2
PS Matching									
Nearest neighbor (1)	0.0	159.1	290.8	0.0	145.5	278.6	0.0	134.9	273.8
Nearest neighbor (2)	1067.5	804.7	523.1	1254.6	783.1	527.8	1152.1	763.5	763.5
Nearest neighbor (4)	1568.0	1455.8	738.6	1682.1	1454.2	733.0	803.4	1440.0	748.0
Kernel	3068.7	2808.3	986.1	3180.7	2864.5	992.4	3074.4	2892.7	1036.1
Spline	2628.7	2363.8	1019.7	2708.2	2486.6	1023.5	2715.1	2470.7	1053.1
	$\tau = 0.50$								

Continued on next page

Table 4 – continued from previous page

	Non-trimmed			Trim 1			Trim 2		
Estimator	Coef.	Average	Std.Error	Coef.	Average	Std.Error	Coef.	Average	Std.Error
FGA QTE	1131.0	1379.9	851.0	1321.2	1404.8	870.5	1026.0	1376.2	869.2
FGA QTE ($R \times 2$)	914.2	1466.4	887.3	965.5	1501.7	866.6	385.5	1488.7	873.5
FGA QTE (b)	1193.9	1403.5	871.6	1306.0	1432.7	891.8	1066.3	1384.7	883.4
FGA QTE ($R \times 2$) (b)	1078.8	1476.8	939.7	981.7	1509.1	919.9	262.5	1492.6	916.5
Firpo (2007)	1063.0	1178.5	901.6	1254.9	1260.9	938.1	763.8	1257.8	973.9
Bitler, Gelbach, and Hoynes (2006)	1063.0	1178.5	901.6	1254.9	1260.9	938.1	763.8	1257.8	973.9
PS Matching									
Nearest neighbor (1)	5.2	910.5	1034.6	368.5	984.4	1042.5	77.5	990.4	1066.2
Nearest neighbor (2)	388.1	1050.9	859.9	563.4	1093.0	877.5	284.4	1088.4	898.4
Nearest neighbor (4)	695.5	1171.6	810.5	801.8	1210.6	814.5	616.7	1201.0	842.8
Kernel	1567.1	1351.2	752.2	1643.1	1365.3	759.6	1587.8	1357.7	783.7
Spline	846.6	1261.7	776.9	1843.3	1259.1	783.4	1578.1	1272.3	811.9
	$\tau = 0.75$								
FGA QTE	2110.9	1948.9	1180.7	1960.5	1986.9	1190.8	1710.1	1950.1	1233.8
FGA QTE ($R \times 2$)	1614.6	2162.8	1376.1	1640.0	2215.2	1369.5	742.6	2131.1	1424.7
FGA QTE (b)	2240.9	1968.0	1191.9	1815.5	2010.4	1221.5	1706.4	1954.1	1264.7
FGA QTE ($R \times 2$) (b)	1351.3	2222.8	1447.2	1549.0	2253.0	1440.5	789.2	2159.8	1449.7
Firpo (2007)	2274.1	2060.9	919.6	2258.0	2071.3	942.1	2214.5	2034.2	979.1
Bitler, Gelbach, and Hoynes (2006)	2274.1	2060.9	919.6	2258.0	2071.3	942.1	2214.5	2034.2	979.1
PS Matching									
Nearest neighbor (1)	1537.0	1663.9	1111.7	1921.1	1699.6	1081.8	1263.0	1637.5	1116.7
Nearest neighbor (2)	585.5	1598.8	1106.4	877.6	1640.8	1097.4	393.7	1589.5	1115.2
Nearest neighbor (4)	1280.1	1634.1	1046.5	1243.8	1653.2	1047.7	1239.6	1595.4	1084.2
Kernel	1399.3	1735.3	972.3	1399.5	1754.3	971.5	1180.6	1710.6	984.8
Spline	1381.6	1963.1	1088.7	1320.6	1954.4	1087.3	1108.5	1870.9	1101.3
	$\tau = 0.90$								
FGA QTE	3093.8	3345.7	2315.2	3426.5	3401.9	2364.0	2789.1	3233.7	2459.3
FGA QTE ($R \times 2$)	3170.8	3301.2	2157.9	4126.5	3317.0	2174.2	4161.9	3118.0	2096.5

Continued on next page

Table 4 – continued from previous page

	Non-trimmed			Trim 1			Trim 2		
Estimator	Coef.	Average	Std.Error	Coef.	Average	Std.Error	Coef.	Average	Std.Error
FGA QTE (b)	3093.8	3493.2	2421.2	3518.0	3512.3	2454.9	3189.5	3189.9	2442.7
FGA QTE ($R \times 2$) (b)	3142.1	3343.0	2221.1	4819.4	3351.2	2213.4	3998.6	3154.9	2133.0
Firpo (2007)	2713.4	2854.3	1890.4	2150.6	2797.8	1889.2	2126.2	2715.5	1895.0
Bitler, Gelbach, and Hoynes (2006)	2713.4	2854.3	1890.4	2150.6	2797.8	1889.2	2126.2	2715.5	1895.0
PS Matching									
Nearest neighbor (1)	1861.2	2425.7	2150.1	1392.3	2372.7	2111.8	444.7	2249.4	2130.0
Nearest neighbor (2)	1279.4	2388.4	1915.6	1278.0	2336.6	1900.4	45.4	2172.0	1892.5
Nearest neighbor (4)	2047.9	2180.9	1716.8	2153.7	2181.3	1743.9	2153.7	2030.9	1758.1
Kernel	175.0	691.5	1389.5	208.0	770.6	1431.7	327.6	566.2	1415.4
Spline	571.5	881.2	1520.4	261.8	899.4	1477.9	327.6	755.2	1458.8

Notes: (a) $\hat{\delta}_\tau$, (b) $\tilde{\delta}_\tau$. Averages and Standard Errors are obtained using 1000 bootstrapping random samples with replacement of the original database.

Appendix

Proofs of Theorems

Proof of Theorem 1

Proof: Let $N \to \infty$, R and r be fixed. Then,

$$
\begin{aligned}
plim_{N\to\infty}\hat{\delta}^{(r)} &= plim_{N\to\infty}\frac{\sum_{i=1}^{N} W_i Y_i \mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^r]}{\sum_{i=1}^{N} W_i \mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^r]} - \frac{\sum_{i=1}^{N} (1-W_i) Y_i \mathbf{1}[\hat{p}_j \in \Im_{\hat{p}}^r]}{\sum_{i=1}^{N} (1-W_i) \mathbf{1}[\hat{p}_i \in \Im_{\hat{p}}^r]} \\
&\text{(by Law of Large Numbers and Assumptions 2 and 3)} \\
&= \frac{E\big[W\times Y|\Im_p^r\big]}{P\big[W|\Im_p^r\big]} - \frac{E\big[(1-W)\times Y|\Im_p^r\big]}{P\big[(1-W)|\Im_p^r\big]} \\
&= \frac{E\Big[E[W\times Y|p]|\Im_p^r\Big]}{P\big[W|\Im_p^r\big]} - \frac{E\Big[E[(1-W)\times Y|p]|\Im_p^r\Big]}{P\big[(1-W)|\Im_p^r\big]} \\
&\text{(by Law of Iterated Expectations)} \\
&= \frac{E\Big[E[W|p]\times E[Y_1|p]|\Im_p^r\Big]}{P\big[W|\Im_p^r\big]} - \frac{E\Big[E[(1-W)|p]\times E[Y_0|p]|\Im_p^r\Big]}{P\big[(1-W)|\Im_p^r\big]} \\
&\text{(by Assumption 1)}
\end{aligned}
$$

Let $E[W|p] = p, P[W|\Im_p^r] \equiv \bar{p}^{(r)}, E[Y_1|p] = g_1(p), E[Y_0|p] = g_0(p), \delta^{(r)} = E[Y_1 - Y_0|\Im_p^r]$.

Then

$$
\begin{aligned}
\Big|plim_{N\to\infty}\hat{\delta}^{(r)} - \delta^{(r)}\Big| &= \left|\frac{E\big[(p-\bar{p}^{(r)})g_1(p)|\Im_p^r\big]}{\bar{p}^{(r)}} + \frac{E\big[(p-\bar{p}^{(r)})g_0(p)|\Im_p^r\big]}{1-\bar{p}^{(r)}}\right| \\
&= \left|\frac{COV\big[p;g_1(p)|\Im_p^r\big]}{\bar{p}^{(r)}} + \frac{COV\big[p;g_0(p)|\Im_p^r\big]}{1-\bar{p}^{(r)}}\right| \\
&\leq \sqrt{VAR\left[p|\Im_p^r\right]} \times C^{(r)}
\end{aligned}
$$

where

$$C^{(r)} = \left(\sqrt{VAR[g_1(p)|\Im_p^r]}/\bar{p}^{(r)} + \sqrt{VAR[g_0(p)|\Im_p^r]}/(1-\bar{p}^{(r)})\right)$$

Now let $R, N \to \infty$,$R/N \to 0$.

$$\begin{aligned}\left|plim_{N\to\infty}\hat{\delta} - \delta\right| &= \left|plim_{N\to\infty}\tfrac{1}{R(N)}\textstyle\sum_{r=1}^{R(N)}\left(\hat{\delta}^{(r)} - \delta^{(r)}\right)\right| \\ &\leq \lim_{N\to\infty}\tfrac{1}{R(N)}\textstyle\sum_{r=1}^{R(N)}\sqrt{VAR\left[p|\Im_p^r\right]}C^{(r)} \\ &\leq \lim_{N\to\infty}\max_r\sqrt{VAR\left[p|\Im_p^r\right]}C^{(r)}\end{aligned}$$

By assumptions $C^{(r)}$ is bounded and $\sqrt{VAR\left[p|\Im_p^r\right]} \leq \sup_{p\in\Im_p^r}(p) - \inf_{p\in\Im_p^r}(p) \leq 1/R, \forall r$. Then $\hat{\delta} - \delta = O_p(1/R)$.

The consistency of $\tilde{\delta}$ can be easily proved by noting that within each fractile group, the estimator is equivalent to that of Hirano, Imbens, and Ridder (2003). *Q.E.D.*

Proof of Theorem 2

Note that as $N \to \infty$, R and r fixed, by convergence of sample quantiles

$$\hat{Q}_{\tau 1}^{(r)} \xrightarrow{p} F_{r,W=1}^{-1}(\tau)$$

where

$$F_{r,W=1}(q) \equiv P[Y \leq q|\Im_p^r, W=1] = E[\mathbf{1}[Y \leq q]|\Im_p^r, W=1] = \frac{E\left[W\mathbf{1}[Y \leq q]|\Im_p^r\right]}{\bar{p}^{(r)}}$$

and $\bar{p}^{(r)} = P[W = 1|\Im_p^r]$ is defined in the proof of Theorem 1.

Therefore,

$$\tau = E\left[\frac{W}{\bar{p}^{(r)}}\mathbf{1}[Y \leq \hat{Q}_{\tau 1}^{(r)}]|\Im_p^r\right] = E\left[\frac{p}{\bar{p}^{(r)}}E[\mathbf{1}[Y_1 \leq \hat{Q}_{\tau 1}^{(r)}]|p]|\Im_p^r\right]$$

However, in general

$$\tau \neq E\left[E[\mathbf{1}[Y_1 \leq \hat{Q}_{\tau 1}^{(r)}]|p]|\Im_p^r\right] = E\left[\mathbf{1}[Y_1 \leq \hat{Q}_{\tau 1}^{(r)}]|\Im_p^r\right]$$

This divergence can be expressed as

$$\left|\tau - E[\mathbf{1}[Y_1 \leq \hat{Q}_{\tau 1}^{(r)}]|\Im_p^r]\right| = \left|\tau - F_{r,1}(\hat{Q}_{\tau 1}^{(r)})\right| = \frac{1}{\bar{p}^{(r)}}\left|COV\left[p, E[\mathbf{1}[Y_1 \leq \hat{Q}_{\tau 1}^{(r)}]|p]|\Im_p^r\right]\right| \leq \frac{1}{R}\times K_1^{(r)}$$

where $F_{r,1}(q) = P[Y_1 \leq q|\Im_p^r]$ and $K_1^{(r)} = \frac{\sqrt{VAR[\mathbf{1}[Y_0 \leq \hat{Q}_{\tau 1}^{(r)}]|\Im_p^r]}}{\bar{p}^{(r)}}$ is bounded by assumptions (see Theorem 1).

How does this translate into the divergence of $\hat{Q}_{\tau 1}^{(r)}$ and $Q_{\tau 1}^{(r)}$? By Taylor's theorem,

$$\hat{Q}_{\tau 1}^{(r)} - Q_{\tau 1}^{(r)} = \frac{F_{r,1}(\hat{Q}_{\tau 1}^{(r)}) - \tau}{f_{r,1}(Q_{\tau 1}^{(r)})} + o_p(\frac{1}{R}K_1^{(r)}) = O_p(1/R)$$

Consider now the case that $N, R \to \infty$, $R/N \to 0$,

$$\left|\frac{1}{R(N)}\sum_{r=1}^{R(N)}\hat{Q}_{\tau 1}^{(r)} - E[Q_{\tau 1}(p)]\right| = O_p(1/R)$$

where $E[\mathbf{1}[Y_1 \leq Q_{\tau 1}(p)]|p] = \tau$ for all $p \in [\underline{p}, \bar{p}]$. The same argument can be applied to show the consistency of $\hat{Q}_{\tau 0}$.

Therefore,

$$\hat{\delta}_\tau = \frac{1}{R(N)} \sum_{r=1}^{R(N)} (\hat{Q}_{\tau 1}^{(r)} - \hat{Q}_{\tau 0}^{(r)}) = \delta_\tau + o_p(1)$$

The consistency of $\tilde{\delta}_\tau$ can be easily proved by noting that within each fractile group, the estimator is equivalent to that of Firpo (2007). *Q.E.D.*

Other ATE and QTE estimators

Hirano, Imbens, and Ridder (2003) semi-parametric efficient ATE estimator is

$$\sum_{i=1}^{N} \left(\frac{W_i Y_i}{\hat{p}_i} - \frac{(1-W_i)Y_i}{1-\hat{p}_i} \right)$$

where $\hat{p}$ is a semi-parametric series estimator of the propensity score.

Bitler, Gelbach, and Hoynes (2006) QTE estimator is obtained by finding the empirical quantiles of the weighted empirical distributions:

$$\hat{F}_0(q) = \frac{\sum_{i=1}^{N} \frac{(1-W_i)\mathbf{1}[Y_i \leq q]}{1-\hat{p}_i}}{\sum_{i=1}^{N} \frac{(1-W_i)}{1-\hat{p}_i}}$$

and

$$\hat{F}_1(q) = \frac{\sum_{i=1}^{N} \frac{W_i \mathbf{1}[Y_i \leq q]}{\hat{p}_i}}{\sum_{i=1}^{N} \frac{W_i}{\hat{p}_i}}$$

that is, $\hat{F}_0^{-1}(\tau)$ and $\hat{F}^{-1}(\tau)$.

Firpo (2007) obtains the same results by minimizing weighted convex check functions:

$$\hat{F}_0^{-1}(\tau) = argmin_q \sum_{i=1}^{N} \frac{(1-W_i)}{1-\hat{p}_i} (Y_i - q)(\tau - \mathbf{1}[Y_i \leq q])$$

$$\hat{F}_1^{-1}(\tau) = argmin_q \sum_{i=1}^{N} \frac{W_i}{\hat{p}_i}(Y_i - q)(\tau - \mathbf{1}[Y_i \leq q])$$

References

Abadie, A., J. D. Angrist, and G. W. Imbens (2002): "Instrumental Variables Estimated of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70(1), 91–117.

Abadie, A., and G. W. Imbens (2002): "Simple and Bias-Corrected Matching Estimators for Average Treatment Effects," NBER Technical Working Paper.

Angrist, J. D., and A. Krueger (1999): "Empirical Strategies in Labor Economics," in *Handbook of Labor Economics, Volume 1*, ed. by O. Ashenfelter, and D. Card, pp. 1277–1366. Elsevier, first edn.

Ashenfelter, O. (1978): "Estimating the Effect of Training Programs on Earnings," *Review of Economics and Statistics*, 60(1), 47–57.

Bera, A. K., and A. Gosh (2006): "Fractile regression and its applications," University of Illinois at Urbana-Champaign.

Bhattacharya, P. (1963): "On an analog of regression analysis," *The Annals of Mathematical Statistics*, 34(4), 1459–1473.

Bhattacharya, P., and H.-G. Müller (1993): "Asymptotics for nonparametric regression," *Sankhya: The Indian Journal of Statistics*, 55, Series A(Part 3), 420–441.

Bitler, M. P., J. B. Gelbach, and H. W. Hoynes (2006): "What mean impact miss: Distributional effects of welfare reform experiment," *American Economic Review*, 96(4), 988–1012.

Chernozhukov, V., and C. Hansen (2004): "The effects of 401(k) participation on the wealth distribution: An instrumental quantile regression analysis," *Review of Economic and Statistics*, 86(3), 735–751.

——— (2005): "An IV model of quantile treatment effects," *Econometrica*, 73(1), 245–261.

Dehejia, R. H., and S. Wahba (1999): "Causal effects in nonexperimental studies: reevaluating the evaluation of training programs," *Journal of the American Statistical Association*, 94(448), 1053–1062.

——— (2002): "Propensity score-matching methods for nonexperimental causal studies," *Review of Economics and Statistics*, 84(1), 151–191.

Diamond, A. (2005): "Reliable estimation of average and quantile causal effects in non-experimental settings," Harvard University.

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

Firpo, S. (2007): "Efficient semiparametric estimation of quantile treatment effects," *Econometrica*, 75(1), 259–276.

Fisher, R. (1935): *The design of experiments*. Boyd, London.

Hahn, J. (1998): "On the role of the propensity score in efficient semiparametric estimation of average treatment effects," *Econometrica*, 66(2), 315–331.

Heckman, J. J., and V. 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(408), 862–874.

Heckman, J. J., H. Ichimura, and P. Todd (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, 64(4), 605–654.

——— (1998): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65(2), 261–294.

Heckman, J. J., and S. Navarro-Lozano (2004): "Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models," *Review of Economics and Statistics*, 86(1), 30–57.

Heckman, J. J., J. Smith, and N. Clemments (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. W. Imbens, and G. Ridder (2003): "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71(4), 1161–1190.

Imbens, G. W. (2004): "Nonparametric estimation of average treatment effects under exogeneity: a review," *Review of Economics and Statistics*, 86(1), 4–29.

Koenker, R., and G. W. Bassett (1978): "Regression quantiles," *Econometrica*, 46(1), 33–50.

Koenker, R., and K. Hallock (2001): "Quantile regression," *Journal of Economic Perspectives*, 15(1), 143–156.

LaLonde, R. (1986): "Evaluating the econometric evaluations of training programs with experimental data," *American Economic Review*, 76(4), 604–620.

Lechner, M. (1999): "Earnings and employment effects of continuous off-the-job training in East Germany after unification," *Journal of Business and Economic Statistics*, 17(1), 74–90.

——— (2000): "An evaluation of public sector sponsored continuous vocational training programs in East Germany," *Journal of Human Resources*, 35(2), 347–375.

Lehmann, E. L. (1974): *Nonparametrics: Statistical methods based on ranks*. Holde-Bay, San Francisco.

Mahalanobis, P. C. (1960): "A method of fractile graphical analysis," *Econometrica*, 28(2), 325–351.

Neyman, J. (1990): "On the Application of Probability Theory to Agricul-

tural Experiments. Essays on Principles. Section 9 (translated)," *Statistical Science*, 5(4), 465–480.

Rosenbaum, P. R., and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70(1), 41–55.

Rosenbaum, P. R., and D. B. Rubin (1984): "Reducing bias in observational studies using subclassification on the propensity score," *Journal of the American Statistical Society*, 79(387), 516–524.

Rubin, D. B. (1974): "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, 66(5), 688–701.

——— (1977): "Assignment to a treatment group on the basis of a covariate," *Journal of Educational Statistics*, 2(1), 1–26.

——— (1978): "Bayesian inference for causal effects: The role of randomization," *The Annals of Statistics*, 6(1), 34–58.

Sen, B. (2005): "Estimation and comparison of fractile graphs using kernel smoothing techniques," *Sankhya: The Indian Journal of Statistics*, 67(Part 2), 305–334.