Nonparametric quantile treatment effects under endogeneity

Markus Frölich, Blaise Melly

Department of Economics, University of St.Gallen

Last changes: 18 May 2007

Abstract:

In this paper we develop nonparametric estimators of the *unconditional* quantile treatment effects when treatment selection is *endogenous*.

The identification and estimation of quantile treatment effects (QTE) has received recently a lot of interest, particularly with applied policy evaluation studies. In this we propose and analyze new nonparametric estimators of the *unconditional* QTE when treatment selection is *endogenous*. In contrast to conditional QTE, i.e. the effects conditional on a large number of covariates X, the unconditional QTE summarizes the effects of a treatment for the entire population and is usually of most interest in policy evaluations. The results can easily be conveyed and summarized. In addition, and no less important, is that unconditional QTE can be estimated fully nonparametrically at $\sqrt{n}$ rate, which is obviously impossible for conditional QTE (unless all X are discrete or one imposes parametric assumptions). Hence, unconditional QTE will usually be much less noisy than estimates of conditional QTE.

In this paper we extend the identification of quantile treatment effects to endogenous treatments. We propose new nonparametric IV estimators, which are root-N consistent. We derive the semiparametric efficiency bound and give conditions under which our estimator is efficient. In a Monte Carlo study, we compare the various estimators. The estimator is then applied to returns to college using distance to college as an instrument.

Keywords: Quantile treatment effects, nonparametric regression, instrumental variables
JEL classification: C13, C14

..... We thank David Card for providing us with the data. Address for correspondence: University of St. Gallen, Bodanstrasse 8, SIAW, 9000 St. Gallen, Switzerland; markus.froelich@unisg.ch, blaise.melly@unisg.ch

1 Introduction

The identification and estimation of quantile treatment effects (QTE) has received recently a lot of interest, particularly with applied policy evaluation studies. In this we propose and analyze new nonparametric estimators of the *unconditional* QTE when treatment selection is *endogenous*. In contrast to conditional QTE, i.e. the effects conditional on a large number of covariates X, the unconditional QTE summarizes the effects of a treatment for the entire population and is usually of most interest in policy evaluations. The results can easily be conveyed and summarized. In addition, and no less important, is that unconditional QTE can be estimated fully nonparametrically at $\sqrt{n}$ rate, which is obviously impossible for conditional QTE (unless all X are discrete or one imposes parametric assumptions). Hence, unconditional QTE will usually be much less noisy than estimates of conditional QTE.[1]

For the case where treatment selection is exogenous, conditional on X,[2]Koenker and Bassett (1978) analyzed estimation of quantile effects in a parametric framework, whereas Chaudhuri (1991) analyzed nonparametric estimation of conditional QTE. Nonparametric estimation of *unconditional* QTE was recently examined in Firpo (2007), Frölich (2007) and Melly (2006).

If treatment selection is *endogenous*, different instrumental variable type approaches have been considered. Abadie, Angrist, and Imbens (2002), Chernozhukov and Hansen (2005) and Chernozhukov and Hansen (2006) pursue a parametric approach,[3] whereas Chernozhukov, Imbens, and Newey (2007) and Horowitz and Lee (2007) examine nonparametric estimation of conditional QTE. In this paper, we analyze nonparametric identification and estimation of *unconditional* QTE.

A further distinction to Chernozhukov, Imbens, and Newey (2007) and Chernozhukov and Hansen (2005) is that our identification approach relies on a monotonicity assumption in the treatment *choice equation*, whereas they require monotonicity in the *outcome equation*. Which of these two assumptions is more appropriate clearly depends on the particular empirical ap-

[1] By conditional QTE we refer to conditioning on a higher dimensional vector X, which is often needed to make certain instrumental variables assumptions plausible. If one were interested in estimating the QTE separately for men and for women, we would subsume this as unconditional QTE since distinction by gender does not introduce any dimensionality problems. This will become more clear below.

[2] Also callsed "selection on observables".

[3] Ist das Chernoz Econometrica Papier wirklich parametrisch, oder nur nichtparametrisch aber für conditional QTE ???

plication. In some applications, the former assumption may be more plausible, in other applications the latter assumption may be more appropriate.[4] Apart from that, our estimators also appear to be easier to implement.

In a sense, our approach is thus more in the spirit of Abadie, Angrist, and Imbens (2002) in that we estimate effects for compliers. We show, however, that their approach is not applicable for estimating unconditional QTE.

We also derive results for the entire process of quantile effects. This will allow us to do tests for treatment effect heterogeneity and the like.

In this paper, we first propose several new nonparametric instrumental variable type estimators of the quantile treatment effect (QTE) when treatment choice is endogenous. We compare then the finite sample properties of these estimators in a Monte Carlo study and subsequently derive the asymptotic properties, showing $\sqrt{n}$-consistency, asymptotic normality and semiparametric efficiency.

Finally, in the last section, we apply these estimators to estimate the returns to college in the USA, an important issue in the debate on inequality. (See e.g. Smith (??) for returns to college in the USA.) We find .....

2 Nonparametric Quantile Treatment Effect

2.1 QTE matching estimator

In this section we consider the effect of a *binary* treatment variable D on a continuous outcome variable Y. (Extensions are considered in the next subsection). Let Y_i^1 and Y_i^0 be the potential outcomes of individual i. Hence, Y_i^1 would be realized if individual i were to receive treatment 1 and Y_i^0 would be realized otherwise. Most interest has focussed on the estimation of average treatment effects

$$E[Y^1 - Y^0]$$

or average treatment effects on the treated

$$E[Y^1 - Y^0|D = 1].$$

[4] In future work, we are going to examine the combination of both assumptions.

Instead of considering only average effects, it is often considerable interest to compare the distributional effects of the treatment as well. A standard example may be the impact of some treatment on income inequality. Another example which has received considerable public interest is *educational equality*, where many societies would prefer to provide every child with a fair chance into adult live. Here, Y is a measure of cognitive ability (e.g. obtained from Math and language tests) and D may be the introduction of computers in classroom (teaching).

In this paper, we will identify and estimate the entire distribution functions of Y^1 and Y^0, but will focus our attention on *quantile treatment effects* (QTE):

$$\Delta^\tau = Q^\tau_{Y^1} - Q^\tau_{Y^0},$$

where $Q^\tau_{Y^1}$ is the τ quantile of Y^1. In the earnings example, $\Delta^{0.9}$ would be the impact of D on the high income part of the distribution. In fact, we will identify the entire processes $Q^\tau_{Y^1}$ and $Q^\tau_{Y^0}$ for $\tau \in (0,1)$, which will also to derive estimates and inference e.g. for the treatment effect on *inequality measures* such as the interquantile spread. For example, a typical inequality measure would be the ratio of earnings at the upper decile and the lower decile and the treatment effect could be defined as

$$\frac{Q^{0.9}_{Y^1}}{Q^{0.1}_{Y^1}} - \frac{Q^{0.9}_{Y^0}}{Q^{0.1}_{Y^0}}$$

or as

$$\frac{Q^{0.9}_{Y^1}}{Q^{0.1}_{Y^1}} \frac{Q^{0.1}_{Y^0}}{Q^{0.9}_{Y^0}}.$$

Our main focus is on *unconditional* treatment effects, i.e. the effects of D in the population at large. We might also be interested in the effects in subpopulations defined by some, usually broadly defined, set A, e.g. women below the age of 25, which we define as

$$\Delta^\tau_A = Q^\tau_{Y^1|A} - Q^\tau_{Y^0|A}$$

where $Q^\tau_{Y^1|A}$ is the quantile in the subpopulation A. Notice that this focus differs from AAI, who focus on *conditional* treatment effects, i.e. conditional on a set of variables X. We call our effects unconditional in the sense that A usually contains a very broadly defined set, while X usually consists of a large set of covariates often including continuous variables as well.[5]

[5] In our analysis, we will also need in a first step to condition on a large set of regressors X to make the instrumental variables conditions hold, but then average over the support of X to obtain unconditional effects.

Conditional and unconditional effects are interesting in their own rights. Whereas conditional effects may be more interesting in economic analysis of effects heterogeneity, for public policy unconditional effects will usually be more relevant. The reason for this is not only that policy and the public need more aggregated results for decision making, but also that unconditional effects can be estimated (nonparametrically) much more precisely than conditional effects. We can achieve $\sqrt{n}$-consistency for *unconditional* QTE, whereas nonparametric estimation of *conditional* QTE will always be estimated at a lower rate (unless all X are discrete).

The usual concern with estimating treatment effects is endogeneity of D and we will rely on exclusion restrictions for some variables Z. Our setup is related to the recent literature on nonparametric identification of nonseparable models. Consider a nonparametric nonseparable model:

$$\begin{aligned} Y_i &= \varphi(D_i, X_i, U_i) \\ D_i &= \zeta(Z_i, X_i, V_i), \end{aligned} \tag{1}$$

where U and V are, possibly, related unobservables and X are additional covariates. We will assume that, after having included X in the model, that Z is excluded from the function φ. The corresponding potential outcomes are

$$\begin{aligned} Y_i^d &= \varphi(d, X_i, U_i) \\ D_i^z &= \zeta(z, X_i, V_i). \end{aligned}$$

In contrast to Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2007) and Chesher (2007), we impose triangularity, i.e. assume that Y does not enter in ζ. On the other hand, we do not need to assume any kind of monotonicity or rank invariance about φ.[6]

We do impose, on the other hand, that the function ζ is (weakly) monotonous in its first argument, i.e. assume that an exogenous increase in Z_i can never decrease the value of D_i. This is the monotonicity assumption of Imbens and Angrist (1994). This assumption may be more plausible than monotonicity in φ in some applications, whereas in other applications it may be less appealing.

[6] Chernozhukov and Hansen (2005), Chernozhukov, Imbens, and Newey (2007) and Chesher (2007) assume that φ is monotonous in its third argument.

Our method is well suited for D binary and, with some qualifications, for discrete D with few mass points. Imbens and Newey (2003) and Chesher (2003) analyzed identification for continuous D and Chesher (2005) examined interval identification with discrete D. Heckman and Vytlacil (2005) analyzed (marginal) average treatment effects for continuous Z and focussed on treatment effects conditional on X, whereas we aim for unconditional effects.

There are only relatively few contributions that examine explicitly distributional impacts of treatment. Firpo (2007), Frölich (2007) and Melly (2006) considered estimation of treatment effects, when D is exogenous conditional on X.

We will focus our attention on the subgroup of *compliers*, which we define as all individuals who are responsive to a change in Z within the support of Z. For individuals for whom $D_i^z = \zeta(z, X_i, V_i)$ does not vary with z in the support of Z,[7] we cannot identify their reaction of D on Y. If the instruments Z are sufficiently powerful to move everyone from $D_i = 0$ to $D_i = 1$, this will lead to the treatment effects in the population. But, in very many applications, however, the instruments available are not so powerful such that it is interesting to consider effects in the *largest subpopulation* for which the effects are identified. But, in very many applications, the instruments available are not powerful enough to achieve $\Pr(D = 1|X, Z) = 1$ and $\Pr(D = 0|X, Z) = 0$ for every value of X and some values of Z.[8] Hence, in almost all applications we can identify an effect only for a subpopulation which reacts to the instrument. In addition, if Y is bounded, we can derive bounds on the overall treatment effects because the size of the subpopulation of compliers is identified as well. We will focus on the QTE for the compliers, which is the *largest* subpopulation for which an effect can be identified:

$$\Delta_c^\tau = Q_{Y^1|c}^\tau - Q_{Y^0|c}^\tau$$

where $Q_{Y^1|c}^\tau = \inf_q \Pr\left(Y^1 \leq q|T = c\right) \geq \tau$, where $T_i = c$ means that individual i is a complier, as defined below.

{Blaise: Bitte prüfen, ob nachfolgender Paragraph verständlich geschrieben ist.} If Z consisted only of a single *binary* variable and if this has a (weakly) monotonous impact in the relationship determining D_i, the largest subpopulation affected by moving the instrument consists of those individuals who would change D if Z was to be increased from 0 to 1. More generally,

[7] These are the always-participants or never-participants in the language of Imbens and Angrist (1994).

[8] This is related to the identification at infinity argument.

the largest subpopulation affected would be obtained by moving Z from the smallest point of its support to its largest point, as long as Z has a weakly monotonous impact on D. If there is only a single instrument Z with support $\mathcal{Z} = [z_{\min}, z_{\max}]$, this corresponds to hypothetically moving Z_i from $z_{\min}$ to $z_{\max}$ for every individual. If Z contains several instrumental variables, the largest subpopulation affected would be obtained by moving the instruments from z_1^* to z_2^* where

$$(z_1^*, z_2^*) = \arg\max_{z_1, z_2 \in \mathcal{Z}} \left| \int \left(E\left[D|X, Z = z_2\right] - E\left[D|X, Z = z_1\right] \right) dF_X \right|,$$

where the integral expression measures the size of this subpopulation, as further discussed below. With monotonicity z_1^* and z_2^* will be at the boundary of the support of $\mathcal{Z}$.[9]

In the following we will assume throughout that z_1^* and z_2^* are known (and not estimated) and that $\Pr(Z = z_1^*) > 0$ and $\Pr(Z = z_2^*) > 0$.[10] To ease notation we will use the values 0 and 1 subsequently instead of $z_{\min}$ to $z_{\max}$ or z_1^* and z_2^*, respectively. Furthermore, we will in the following only refer to the effectively used sample $\{i : Z_i \in \{0, 1\}\}$ or in other words assume that $\Pr(Z = z_1^*) + \Pr(Z = z_2^*) = 1$. This is appropriate for our application where the single instrument Z is truly binary. In other applications, where $\Pr(Z = z_1^*) + \Pr(Z = z_2^*) < 1$, our results apply with reference to the subsample $\{i : Z_i \in \{0, 1\}\}$.[11]

By considering only the endpoints of the support of Z, and recoding Z as 0 and 1, and with D being a binary treatment variable, we can partition the population into 4 groups defined as $\mathcal{T}_i = a$ if $D_i^1 = D_i^0 = 1$ (always treated), $\mathcal{T}_i = n$ if $D_i^1 = D_i^0 = 0$ (never treated), $\mathcal{T}_i = c$ if $D_i^1 > D_i^0$ (compliers), $\mathcal{T}_i = d$ if $D_i^1 < D_i^0$ (defiers). We assume that

Assumption 1:

i) Existence of compliers:	$\Pr(\mathcal{T} = c) > 0$
ii) Monotonicity:	$\Pr(\mathcal{T} = d) = 0$
iii) Independent instrument:	$(Y^d, \mathcal{T}) \perp\!\!\!\perp Z \mid X$
iv) Common support:	$0 < p(X) < 1 \quad a.s.$

[9] This may not be the case, if the impact of Z is monotonous only given X, such that the relationship determining D may be decreasing in z for some x and increasing for other x.

[10] The latter condition requires purely continuous variables to be discretized at their endpoints.

[11] Consider $\Pr(Z = z_1^*) + \Pr(Z = z_2^*) = r < 1$ with $plim \frac{n}{N} = r$ where N is the total sample size and $n = \sum_{i=1}^{N} 1(Z_i \in \{0, 1\}$ the number of observations at the endpoints of the support of Z. When calculating the variance approximation for a particular application, the sample size n should be used. (If r is much smaller than 1, there could be finite-sample precision gains by smoothing over Z. We leave this for future research.)

where $p(x) = \Pr(Z = 1|X = x)$.

The first two assumptions are often referred to as monotonicity. They require that D_i^z either weakly increases with z for all individuals (or decreases for all individuals). They also require that at least some individuals react to movements in the instrument. The third assumption is the main instrumental variable assumption. It implicitly requires an exclusion restriction (=triangularity) and an unconfounded instrument restriction. In other words, Z_i should not affect the potential outcomes of individual i directly and those individuals for whom $Z = z$ is observed should not differ in their relevant unobserved characteristics from individuals with $Z \neq z$. Unless the instrument has been randomly assigned, this last restriction is often very unlikely to hold. However, *conditional* on a large set of covariates X, this assumption can often be plausible.[12] Note further that we do *not* need X to be exogenous. X can be related to U and V in (1) in any way. This may be important in many applications where X often contains lagged (dependent) variables that may well be related to unobserved ability U. See also Frölich (2006b).[13]

The fourth assumption requires that the support of X is identical in the $Z = 0$ and the $Z = 1$ subpopulation. This assumption is needed since we first condition on X to make the instrumental variables assumption valid but then integrate out to obtain the unconditional treatment effects.

Theorem 1 (Identification: Matching on X) *Under Assumption 1, the potential outcome distributions for the compliers are nonparametrically identified as*

$$\begin{aligned} F_{Y^1|c}(u) &= \frac{\int \left(E\left[1\left(Y \leq u\right) \cdot D|X, Z = 1\right] - E\left[1\left(Y \leq u\right) \cdot D|X, Z = 0\right]\right) dF_X}{\int \left(E\left[D|X, Z = 1\right] - E\left[D|X, Z = 0\right]\right) dF_X} \\ F_{Y^0|c}(u) &= \frac{\int \left(E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z = 1\right] - E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z = 0\right]\right) dF_X}{\int \left(E\left[D|X, Z = 1\right] - E\left[D|X, Z = 0\right]\right) dF_X} \end{aligned} \tag{2}$$

[12] In our application .... Clearly the instrument is not randomly assigned and individuals with $Z = 1$ are certainly different from those with $Z = 0$. After conditioning on a number of X variables that capture the **endogenous location choice** the assumption becomes more plausible. **Text anpassen sobald Anwendung fertig.**

[13] If X were exogenous, we could strengthen the above assumption and identify the effect for a larger subpopulation in Section 2.@@.

which gives the QTE as the difference between the quantiles:

$$Q^{\tau}_{Y^1|c} = F^{-1}_{Y^1|c}(\tau) \qquad\qquad Q^{\tau}_{Y^0|c} = F^{-1}_{Y^0|c}(\tau).$$

We can estimate the potential outcome distributions and then the QTE, by plugging in nonparametric estimators for the conditional expectation functions and the empirical distribution function for F_X.

The following Lemma shows that

Lemma 2 (Propensity score matching) *Let $P = p(X)$ and dF_P be the distribution of P. Under Assumption 1 it follows that:*

$$\begin{aligned} F_{Y^1|c}(u) &= \frac{\int \left(E\left[1\left(Y \leq u\right) \cdot D|P, Z=1\right] - E\left[1\left(Y \leq u\right) \cdot D|P, Z=0\right]\right) dF_P}{\int \left(E\left[D|P, Z=1\right] - E\left[D|P, Z=0\right]\right) dF_P} \\ F_{Y^0|c}(u) &= \frac{\int \left(E\left[1\left(Y \leq u\right) \cdot (D-1)|P, Z=1\right] - E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z=0\right]\right) dF_P}{\int \left(E\left[D|P, Z=1\right] - E\left[D|P, Z=0\right]\right) dF_P} \end{aligned}$$

(Proof see appendix.)

2.2 QTE weighting estimator

Starting from (2) one can also derive an expression for the unconditional distribution functions by some kind of weighting through the propensity score. In particular, the expressions in (2) can be shown to be equivalent to

Lemma 3 (QTE weighting estimator) *The potential outcome distributions are identified as*

$$\begin{aligned} F_{Y^1|c}(u) &= \frac{E\left[1\left(Y < u\right) \cdot D \cdot \frac{Z-p(X)}{p(X)(1-p(X))}\right]}{P_c} \\ F_{Y^0|c}(u) &= \frac{E\left[1\left(Y < u\right) \cdot (D-1) \cdot \frac{Z-p(X)}{p(X)(1-p(X))}\right]}{P_c} \end{aligned} \tag{3}$$

where P_c is the denominator in (2), which can also be written as $P_c = E\left[D \cdot \frac{Z-p(X)}{p(X)(1-p(X))}\right]$.

Hence, one could estimate the QTE by the difference

$$q_1 - q_0$$

of the solutions of the two moment conditions

$$\begin{aligned} E\left[1\left(Y < q_1\right) \cdot D \frac{Z - p(X)}{p(X)\left(1 - p(X)\right)}\right] &= \tau P_c \\ E\left[1\left(Y < q_0\right) \cdot (D-1) \frac{Z - p(X)}{p(X)\left(1 - p(X)\right)}\right] &= \tau P_c \end{aligned} \tag{4}$$

or equivalently

$$\begin{aligned} E\left[\left\{1\left(Y < q_1\right) - \tau\right\} W D\right] &= 0 \\ E\left[\left\{1\left(Y < q_0\right) - \tau\right\} W \left(1 - D\right)\right] &= 0 \end{aligned} \tag{5}$$

where the weights W are defined as followed.

$$W = \frac{Z - p\left(X\right)}{p(X)\left(1 - p(X)\right)}\left(2D - 1\right) \tag{6}$$

We could thus estimate q_0 and q_1 by univariate quantile regressions in the $D = 0$ and $D = 1$ populations. Alternatively, we could estimate the treatment effect directly by a weighted quantile regression:

Lemma 4 *The solution of the following optimization problem*

$$\left(\alpha, \beta\right) = \arg\min_{a,b} E\left[\rho_\tau(Y - a - bD) \cdot W\right], \tag{7}$$

where $\rho_\tau(u) = u \cdot \{\tau - 1\left(u < 0\right)\}$*, is equivalent to the solutions to the moment conditions (5) in that the solution for* a *corresponds to* q_0 *and the solution for* b *corresponds to* $q_1 - q_0$*. (The proof is straightforward, because (5) are the first order conditions of the minimization problem, see appendix.)*

Note that W_i is negative for $Z_i \neq D_i$. Therefore, if we use these weights the sample objective function to (7), i.e. where the expectation operator is replaced by the sample average, will typically be non-convex. This complicates the optimization problem because local optima can exist. AAI notice a similar problem in their approach. The problem is not so serious here since we need to estimate only a scalar in the $D = 1$ population and in the $D = 0$ population.

This *one-dimensional* estimation problem can be solved by grid-search supported with visual inspection of the objective function for local minima.

This problem can altogether be avoided by estimating the cdf via (3) instead of the quantiles via (7), particularly if one is interested in the entire distribution anyhow instead of only the effect at one single quantile, e.g. the median.

Alternatively, we could apply an iterated expectations argument to obtain

$$\begin{aligned}(\alpha,\beta) &= \arg\min_{a,b} E\left[\rho_\tau(Y-a-bD)\cdot W\right] = \arg\min_{a,b} E\left[\rho_\tau(Y-a-bD)E\left[W|Y,D\right]\right] \quad (8)\\ &= \arg\min_{a,b} E\left[\rho_\tau(Y-a-bD)\cdot W_+\right] \quad (9)\end{aligned}$$

where

$$W_+ = E\left[W\,|Y,D\right] = E\left[\frac{Z-p\left(X\right)}{p(X)\left(1-p(X)\right)}\,|Y,D\right](2D-1)\,. \quad (10)$$

These new weights W_+ are always nonnegative as shown below. Hence, the sample objective function of the rightmost term in (8) is globally convex as it would be the sum of convex functions, and fast linear programming (LP) algorithms can be used. However, we would need to estimate (10) first. Note that AAI suggest a similar projection approach, but their weights are conditional on Y, D and X. Hence, nonparametric estimation of their weights could be more difficult and computationally demanding, whereas estimation of (10) requires only *univariate* nonparametric regression separately for the $D=0$ and $D=1$ population.

We show now that these weights W_+ are always non-negative. The following relationship is helpful

$$E\left[D|X,Z=0\right] \leq E\left[D|X\right] \leq E\left[D|X,Z=1\right],$$

which follows from the proof of Theorem (1). Via Bayes' theorem one can show that this implies also

$$E\left[Z|X,D=0\right] \leq p(X) \leq E\left[Z|X,D=1\right].$$

Now, if $D=1$ the weights would be negative if $E\left[Z-p\left(X\right)|Y,D\right]<0$. However,

$$\begin{aligned}E\left[Z-p\left(X\right)|Y,D=1\right] &= E\left[E\left[Z-p\left(X\right)\big|X,Y^1,D=1\right]\big|Y^1,D=1\right]\\ &= E\left[E\left[Z|X,D=1\right]-p\left(X\right)|Y^1,D=1\right] \qquad \text{(by Assumption 1 iii)}\\ &\geq E\left[p(X)-p\left(X\right)|Y^1,D=1\right]=0.\end{aligned}$$

Hence, W_+ is always non-negative. On the other hand, if $D = 0$ the weights would be negative if $E\left[Z - p\left(X\right)|Y, D\right] > 0$. However,

$$\begin{aligned} E\left[Z - p\left(X\right)|Y, D = 0\right] &= E\left[E\left[Z - p\left(X\right)\middle|X, Y^1, D = 0\right]\middle|Y^1, D = 0\right] \\ &= E\left[E\left[Z|X, D = 0\right] - p\left(X\right)|Y^1, D = 0\right] \qquad \text{(by Assumption 1 iii)} \\ &\leq E\left[p(X) - p\left(X\right)|Y^1, D = 0\right] = 0. \end{aligned}$$

Therefore, the weights W_+ are always non-negative.

2.3 Relationship to AAI

These results bear some resemblance with Abadie, Angrist, and Imbens (2002), who suggested to estimate a weighted linear quantile regression

$$\arg\min_{\alpha,\beta} E\left[\rho_\tau(Y - \alpha D - \beta' X) \cdot \left(1 - \frac{D(1-Z)}{1-p(X)} - \frac{(1-D)Z}{p(X)}\right)\right]. \tag{11}$$

However, the interpretation is very different. Abadie, Angrist, and Imbens (2002) impose a linear parametric specification, whereas our approach is entirely nonparametric. Furthermore, they are interested in the conditional treatment effects, i.e. conditional on X, whereas we are interested in the *unconditional* treatment effects. As argued above, the unconditional treatment effects have substantial advantages when the number of covariates is large: First, they avoid the curse of dimensionality without any linearity assumption. Second, policy makers and users of evaluation studies are often interested in summary measures that summarize the effects in the population or large subpopulations, instead of examining heterogeneity in every of the many dimensions of X.

Note that the approach of Abadie, Angrist, and Imbens (2002) *cannot* be used for estimating unconditional treatment effects if one does not want to impose a parametric specification since the weights in (11) are not appropriate for that case. In other words, one might be thinking to run a weighted quantile regression of Y on a constant and D by using equation (11) and replacing X by a constant in the first term in (11). For that purpose, however, the weights of Abadie, Angrist, and Imbens (2002) are *not correct*. Suppose one used the approach of (11) with X replaced by a constant, i.e.

$$\arg\min_{\alpha,\beta} E\left[\rho_\tau(Y - \alpha - \beta D) \cdot \left(1 - \frac{D(1-Z)}{1-p(X)} - \frac{(1-D)Z}{p(X)}\right)\right]. \tag{12}$$

Proposition 5 *As shown in the appendix, the solution of (12) for β gives the difference between the τ quantiles of the* treated compliers *and* non-treated compliers*, respectively:*

$$\beta = F^{-1}_{Y^1|c,D=1}(\tau) - F^{-1}_{Y^0|c,D=0}(\tau)$$

where

$$\begin{aligned} F_{Y^1|c,D=1}(u) &= \Pr(Y^1 \leq u | D=1, T=c) \\ F_{Y^0|c,D=0}(u) &= \Pr(Y^0 \leq u | D=0, T=c). \end{aligned}$$

This difference is not very meaningful as one compares the Y^1 outcomes among the treated with the Y^0 outcomes among the non-treated. (Only if the instrumental variables assumptions are valid *without* conditioning on X would this correspond to the QTE of interest.) Therefore the weights of Abadie, Angrist, and Imbens (2002) are only useful in combination with a parametric assumption for conditional quantiles.

Hence, if one is interested in nonparametric estimation of the unconditional QTE, one should use the weights in (7) but not those in (12). On the other hand, if one were interested in estimating conditional QTE using a parametric specification, the weights W we propose in (6) could also be used:

Proposition 6 *If one assumes a linear model for the conditional quantile for the compliers*

$$F^{-1}_Y(\tau | X, D, T=c) = X'\beta^\tau_0 + \alpha^\tau_0 D,$$

a weighted quantile regression with weights W would identify α^τ_0 and β^τ_0.[14]

This follows from showing that the weighted objective function

$$(\alpha, \beta) = \arg\min_{a,b} E\left[\rho_\tau(Y - X'b - aD) \cdot W\right] \tag{13}$$

is equivalent to

$$\begin{aligned} &= \arg\min_{a,b} E\left[\rho_\tau(Y - X'b - aD) \cdot W \,|\mathcal{T} = c\right] \\ &= \arg\min_{a,b} E\left[\left(\frac{D}{p(X)} + \frac{1-D}{1-p(X)}\right) \cdot \rho_\tau(Y - X'\beta - \alpha D) \;\; |\mathcal{T} = c\right], \end{aligned} \tag{14}$$

[14] Instead of W one could also use $E[W|Y, X, D]$, which are always nonnegative, but usually not the weights $W_+ = E[W|Y, D]$ as one would also need to condition on X here.

which is the objective function of a weighted linear quantile regression for compliers. (See appendix.) Note that all weights are strictly positive and finite because we assume that $0 < p(X) < 1$. Therefore, standard quantile regression results (see for instance Koenker (2005) theorem 4.1 and 5.1) imply that this function is minimized at α_0^τ and β_0^τ as long as $E\left[\left(\frac{D}{p(X)} + \frac{1-D}{1-p(X)}\right)(X', D)'(X', D)\right]$ is positive definite, what is assumed.

Hence, both types of weights, i.e. those of AAI and those in (6), would identify the conditional quantile treatment effects, but it is not clear which would be more efficient. The weights in (14) vary with x whereas the weights in AAI are identical to one for every complier. In any case, both types of weights would be generally inefficient since they do not incorporate the density function of U at the τ quantile. Hence, if one were mainly interested in estimating *conditional* QTE with a parametric specification, more efficient estimators could be developed.

2.4 Comparison of weighting versus matching

2.5 Extensions to non-binary treatments and non-binary instruments

3 Asymptotic properties

In the previous section a number of different estimators for the unconditional quantile treatment effects have been developed. In this section, the large sample properties are examined. We will first derive the semiparametric efficiency bound. Therefore we impose assumptions that guarantee that the quantile is well defined:

Assumption 2:

The random variables $Y^1|c$ and $Y^0|c$ are continuous with positive density in a neighbourhood of $Q^\tau_{Y^1|c}$ and $Q^\tau_{Y^0|c}$, respectively.

Theorem 7 (Efficiency bound) *a) The efficiency bound for* $\Delta_c^\tau = Q^\tau_{Y^1|c} - Q^\tau_{Y^0|c}$ *is*

$$\begin{aligned}
\mathcal{V} = & \frac{1}{P_c^2 f^2_{Y^1|c}(Q^\tau_{Y^1|c})} E\left[\frac{\pi(X,1)}{p(X)} F_{Y|D=1,Z=1,X}(Q^\tau_{Y^1|c})\left(1 - F_{Y|D=1,Z=1,X}(Q^\tau_{Y^1|c})\right)\right] \\
& + \frac{1}{P_c^2 f^2_{Y^1|c}(Q^\tau_{Y^1|c})} E\left[\frac{\pi(X,0)}{1-p(x)} F_{Y|D=1,Z=0,X}(Q^\tau_{Y^1|c})\left(1 - F_{Y|D=1,Z=0,X}(Q^\tau_{Y^1|c})\right)\right] \\
& + \frac{1}{P_c^2 f^2_{Y^0|c}(Q^\tau_{Y^0|c})} E\left[\frac{1-\pi(X,1)}{p(X)} F_{Y|D=0,Z=1,X}(Q^\tau_{Y^0|c})\left(1 - F_{Y|D=0,Z=1,X}(Q^\tau_{Y^0|c})\right)\right] \\
& + \frac{1}{P_c^2 f^2_{Y^0|c}(Q^\tau_{Y^0|c})} E\left[\frac{1-\pi(X,0)}{1-p(X)} F_{Y|D=0,Z=0,X}(Q^\tau_{Y^0|c})\left(1 - F_{Y|D=0,Z=0,X}(Q^\tau_{Y^0|c})\right)\right] \\
& + \frac{1}{P_c^2} Var\left(\frac{F_{Y|D=1,Z=0,X}(Q^\tau_{Y^1|c}) - F_{Y|D=1,Z=1,X}(Q^\tau_{Y^1|c})}{f_{Y^1|c}(Q^\tau_{Y^1|c})} + \frac{F_{Y|D=0,Z=0,X}(Q^\tau_{Y^0|c}) - F_{Y|D=0,Z=1,X}(Q^\tau_{Y^0|c})}{f_{Y^0|c}(Q^\tau_{Y^0|c})}\right)
\end{aligned}$$

where

$$\begin{aligned}
f_{Y^1|c}(u) &= \left\{\int \left(f_{Y|X,D=1,Z=1}(u)\pi(x,1) - f_{Y|X,D=1,Z=0}(u)\pi(x,0)\right) dF_X\right\}/P_c \\
f_{Y^0|c}(u) &= -\left\{\int \left(f_{Y|X,D=0,Z=1}(u)(1-\pi(x,1)) - f_{Y|X,D=0,Z=0}(u)(1-\pi(x,0))\right) dF_X\right\}/P_c
\end{aligned}$$

where $\pi(x,z) = \Pr(D=1|X=x,Z=z)$ *and* $P_c = \int (\pi(x,1) - \pi(x,0))\, dF_X$ *is the fraction of compliers.*

We suppose to estimate everything by llr or kernel regression. We need the following assumptions:

Theorem 8 (Asymptotic distribution) *Under the previous conditions, the estimator is* $\sqrt{n}$ *consistent and asymptotically normal (and efficient)*

$$\sqrt{n}\left(\hat{\Delta}_c^\tau - \Delta_c^\tau\right) \xrightarrow{d} N(\cdot,\cdot)$$

To be continued: The proof needs to be typed in.

4 Monte Carlo

In this section we examine the finite sample performance of the proposed estimator and compare it to AAI and some other versions. The basic model setup is

$$\begin{aligned} Y_i &= \varphi(D_i, X_i, U_i) \\ D_i &= 1\,(\,\zeta(Z_i, X_i, V_i) \geq 0) \qquad \text{weakly monotonous in } Z \\ Z_i &= 1\,(\,\xi(X_i, \eta_i) \geq 0) \end{aligned}$$

where D and Z are both binary, U, V, η are error terms (where η is independent of U, V) and X are covariates that affect Y, D and Z. We examine a variety of different designs characterized by the following parameters:

1) Distribution of the X regressors: In a first design, X consists of 3 $\chi^2_{(1)}$ variables. In the second design, X consists of 3 Cauchy variables.

2) Distribution of the random variable $p(X) = E[Z|X]$: We measure the density mass of the events $0.05 \leq p(X) \leq 0.95$ and of $0.01 \leq p(X) \leq 0.99$. In those designs, where more mass is in the tails, the estimation problem is much harder.

3) The degree of differentiability of $p(X)$. (Differentiability may be particularly relevant in the tail areas.)

4) The fraction of compliers: I.e. the percentage of observations that would change D if Z were increased.

5) The degree of endogeneity measured as the correlation between U and V.

6) The degree of non-linearity of φ as a function of X, measured as the average of the square of the second derivative (so wie in der spline smoothing Literatur).

Obviously, it would be impossible to run a MonteCarlo for all combinations of these different parameters. Therefore, we choose a few designs below and later regress the estimated MSE on these 7 parameters defined before to obtain an indication of which of these characteristics of the designs are most beneficial or harmful for the MSE.

Funktionen zu wählen, die relativ symmetrisch in den 3 X sind, so dass wir den Verlauf der Kurve graphisch leichter darstellen könnten z.B.

$$Z = 1\,(\,\{1(X_1 > 1) + 1(X_2 > 1) + 1(X_3 > 1)\}\,\eta + \eta \;\geq 0)\,.$$

$$Z = 1\,(\ \{1(X_1 > 1)X_1 + 1(X_2 > 1)X_2 + 1(X_3 > 1)X_3\}\,\eta + \eta\ \geq 0)$$

und die Parameter so zu wählen, dass $\pi(X)$ auch viel Masse in den Aussenbereichen hat. Danach eine Funktion für D und erst zum Schluss eine Funktion für Y^1 und Y^0.

5 Returns to college

The returns to education have received a lot of attention with recent research interests aiming also particularly at higher education, see e.g. @@@, @Smith, @@ about the returns to college. In this section we apply the new quantile estimators to estimate the *returns to college* using *distance to college* as an instrument. The data is taken from Card (1995), who found that the 2SLS estimates of the returns to schooling were about 13% and thus twice as large as the corresponding OLS estimates. Here, we focus particularly on the treatment effect of having attended college. The data stems from the National Longitudinal Survey of Young Men (NLSYM), which began in 1966 with 5525 men between 14 to 24 years old. The sampling frame of the NLSYM oversampled neighbourhoods with a large fraction of non-white residents.

We follow Card (1995) in that we examine wages in the year 1976 to mitigate the influence of attrition. About 20% of the sample attrited in the first three years of the survey and the total attrition rate was about 29% in 1976. Total attrition increased further to 35% until the 1981 wave. In 1976 the respondents were between 24 to 34 years old such that most of them should have completed college at that time. As pointed out in Card (1995), the sample interviewed in 1976 is very similar in most characterstics to the original 1966 sample apart from a smaller fraction of blacks. Descriptive statistics for some variables are given in Table 1. The variable of interest Y is hourly wage in 1976, measured in *cents per hour*.

The binary indicator D having *attended college* is also taken from the 1976 wave. About 50% have attended college, while the other 50% did not. Most of the other variables are taken from the baseline survey in 1966. This includes an indicator for the presence of an accredited 4-year college in the local labour market. Almost 70% of the observations were living in such neighbourhoods in 1966. In the baseline survey, individuals were also asked about their family situation when they were 14 years old. 77% lived with both their father and mother, 12% lived only with their mother, and the other lived with relatives of at least one step-parent. The data

also contains information on mother's and father's education and race and region of residence: rural versus urban (Standard Metropolitan Statistical Area, SMSA) and geographic location.

As a first step to estimating the QTE of college attendance on wages, we examine the relationship between the instrumental variable Z and other background characteristics X that are likely to have a strong influence on earnings in 1976. Table 2 shows a probit regression of Z on X and the following Figure 1 shows the distribution of $p(X) = \Pr(Z = 1|X)$ in the $Z = 1$ and $Z = 0$ subpopulations. This figures shows that those individuals living near to a college ($Z = 1$) and those with $Z = 0$ do indeed seem to differ with respect to their family characteristics X. On the other hand, there does not seem to be a problem with respect to common support since the support of $p(x)$ is rather similar in these two subpopulations.

The following figures show the estimators of the quantile treatment effects. Figure 2 gives the results for the matching estimator, the propensity score weighting estimator and the propensity score weighting estimator with positive weights.[15] Bandwidths are chosen by cross-validation.[16] All three estimators produce relatively similar results. The treatment effects are positive throughout and increasing with the quantiles. At the lower end the QTE is about 1.5$ per hour and increases up to 4$ to 8$ per hour for higher quantiles.

Figure 3 compares the estimated QTE to the estimates of the QTE when college attendance is *not* instrumented for, i.e. when it is considered as exogenous, conditional on X. The nonparametric estimator of Frölich (2006a) is used for estimating the QTE when D is exogenous, conditional on X. The results clearly show that the QTE when college choice is handled as endogenous are much larger and the absolute gap increases for larger quantiles. Figure 4 shows the relative comparison of the results of Figure 3. It shows that the instrumentation doubles the QTE for lower quantiles, whereas for higher quantiles the QTE with endogenous college choice are about 2.5 times as large than when college choice is treated as exogenous.

Finally, Figure 5 shows bootstrap confidence intervals for the propensity score weighting estimator with positive weights of Figure 2. Except for the very low quantiles, the estimates are signifikantly different from zero. (However, only 50 bootstrap replications so far.)

[15] The propensity score matching estimator will follow soon.

[16] Markus: Explain the smoothing over discrete and continuous regressors with two different bandwidths.

6 Conclusions

...

A Appendix (Proof of theorems)

A.1 Proof of Theorem (1):

We consider the derivation of

$$F_{Y^0|c}(u) = \frac{\int \left(E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z=1\right] - E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z=0\right]\right) dF_X}{\int \left(E\left[D|X, Z=1\right] - E\left[D|X, Z=0\right]\right) dF_X}. \quad (15)$$

(The results for $F_{Y^1|c}$ are analogous and are omitted.)

Consider first the expression

$$E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z=1\right] - E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z=0\right]$$

which by the law of total probability can be partitioned into the four subpopulations:

$$\begin{aligned} = \quad & E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z=1, \mathcal{T}=a\right] \Pr\left(\mathcal{T}=a|X, Z=1\right) - E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z=0, \mathcal{T}=a\right] \\ & + E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z=1, \mathcal{T}=n\right] \Pr\left(\mathcal{T}=n|X, Z=1\right) - E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z=0, \mathcal{T}=n\right] \\ & + E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z=1, \mathcal{T}=c\right] \Pr\left(\mathcal{T}=c|X, Z=1\right) - E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z=0, \mathcal{T}=c\right] \end{aligned}$$

Noting that the value of Z and $\mathcal{T}$ together determine the value of D and using that $\mathcal{T} \perp\!\!\!\perp Z|X$ from assumption 1, we obtain

$$\begin{aligned} = \quad & 0 \cdot \Pr\left(\mathcal{T}=a|X\right) \\ & - \left\{E\left[1\left(Y^0 \leq u\right)|X, Z=1, \mathcal{T}=n\right] - E\left[1\left(Y^0 \leq u\right)|X, Z=0, \mathcal{T}=n\right]\right\} \Pr\left(\mathcal{T}=n|X\right) \\ & + \left\{0 + E\left[1\left(Y^0 \leq u\right)|X, Z=0, \mathcal{T}=c\right]\right\} \Pr\left(\mathcal{T}=c|X\right). \end{aligned}$$

Now we use $Y^d \perp\!\!\!\perp Z|X, \mathcal{T}$ by assumption 1 to obtain

$$\begin{aligned} = \quad & - \left\{E\left[1\left(Y^0 \leq u\right)|X, \mathcal{T}=n\right] - E\left[1\left(Y^0 \leq u\right)|X, \mathcal{T}=n\right]\right\} \Pr\left(\mathcal{T}=n|X\right) \\ & + \left\{E\left[1\left(Y^0 \leq u\right)|X, \mathcal{T}=c\right]\right\} \Pr\left(\mathcal{T}=c|X\right) \end{aligned}$$

$$= E\left[1\left(Y^0 \leq u\right)|X, \mathcal{T}=c\right] \Pr\left(\mathcal{T}=c|X\right).$$

Now we insert this result into the numerator of (15) to obtain

$$\begin{aligned}
& \int \left(E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z=1\right] - E\left[1\left(Y \leq u\right) \cdot (D-1)|X, Z=0\right]\right) dF_X \\
= & \int E\left[1\left(Y^0 \leq u\right)|X, \mathcal{T}=c\right] \Pr\left(\mathcal{T}=c|X\right) dF_X \\
= & \int E\left[1\left(Y^0 \leq u\right)|X, \mathcal{T}=c\right] dF_{X|c} \cdot P_c \\
= & E\left[1\left(Y^0 \leq u\right)|\mathcal{T}=c\right] \cdot P_c
\end{aligned}$$

where the second last equality made use of Bayes' theorem:

$$dF_{X|c} \cdot P_c = \Pr\left(\mathcal{T}=c|X\right) \cdot dF_X.$$

Now consider the denominator of (15) and proceed as before. First notice that conditional on X

$$E\left[D|X, Z=1\right] - E\left[D|X, Z=0\right]$$

$$\begin{aligned}
= \quad & E\left[D|X, Z=1, \mathcal{T}=a\right] \Pr\left(\mathcal{T}=a|X, Z=1\right) - E\left[D|X, Z=0, \mathcal{T}=a\right] \Pr\left(\mathcal{T}=a|X, Z=0\right) \\
& + E\left[D|X, Z=1, \mathcal{T}=c\right] \Pr\left(\mathcal{T}=c|X, Z=1\right) - E\left[D|X, Z=0, \mathcal{T}=c\right] \Pr\left(\mathcal{T}=c|X, Z=0\right)
\end{aligned}$$

using that $\mathcal{T} \perp\!\!\!\perp Z|X$ from assumption 1, we obtain

$$\begin{aligned}
= \quad & \Pr\left(\mathcal{T}=a|X\right) - \Pr\left(\mathcal{T}=a|X\right) + \Pr\left(\mathcal{T}=c|X\right) \\
= \quad & \Pr\left(\mathcal{T}=c|X\right).
\end{aligned}$$

Inserting this into the denominator of (15) and again making use of Bayes' theorem

$$\int \Pr\left(\mathcal{T}=c|X\right) dF_X = \int dF_{X|c} \cdot P_c = P_c.$$

Putting these results together we obtain for the right hand side of (15)

$$\frac{E\left[1\left(Y^0 \leq u\right)|\mathcal{T}=c\right] \cdot P_c}{P_c} = \Pr\left(Y^0 \leq u|\mathcal{T}=c\right) = F_{Y^0|c}(u).$$

A.2 Proof of Lemma (2):

The following two variants of using iterated expectations show the equality for a typical component of the estimator

$$E\left[\frac{1\left(Y\leq u\right)DZ}{p(X)}\right]=\int E\left[\frac{1\left(Y\leq u\right)DZ}{p(X)}|X\right]dF_X=\int E\left[1\left(Y\leq u\right)D|X,Z=1\right]dF_X$$

$$\begin{aligned}E\left[\frac{1\left(Y\leq u\right)DZ}{p(X)}\right] &= E\left[E\left[\frac{1\left(Y\leq u\right)DZ}{p(X)}|p(X)=\rho\right]\right]=E\left[E\left[\rho\frac{1\left(Y\leq u\right)D}{p(X)}|p(X)=\rho,Z=1\right]\right]\\ &= E\left[E\left[1\left(Y\leq u\right)D|p(X)=\rho,Z=1\right]\right].\end{aligned}$$

For the corresponding components in the $Z=0$ population, Z is replaced by $1-Z$ and $p(X)$ is replaced by $1-p(X)$ in the previous derivations.

A.3 Proof of Lemma (3):

Note that by iterated expectations

$$E\left[\frac{1\left(Y\leq u\right)DZ}{p(X)}\right]=\int E\left[\frac{1\left(Y\leq u\right)DZ}{p(X)}|X\right]dF_X=\int E\left[1\left(Y\leq u\right)D|X,Z=1\right]dF_X$$

and

$$E\left[\frac{1\left(Y\leq u\right)D(1-Z)}{1-p(X)}\right]=\int E\left[\frac{1\left(Y\leq u\right)D(1-Z)}{1-p(X)}|X\right]dF_X=\int E\left[1\left(Y\leq u\right)D|X,Z=0\right]dF_X$$

Hence, the equation (2) can be written as

$$F_{Y^1|c}(u)=\frac{\int\left(E\left[1\left(Y\leq u\right)D\left(\frac{Z-p(X)}{p(X)(1-p(X))}\right)\right]\right)dF_X}{P_c}$$

and analogously for $F_{Y^0|c}(u)$.

A.4 Proof of Lemma (4):

If the objective function has a unique interior solution, it follows that

$$\arg\min_{a,b}E\left[W\cdot\rho_\tau(Y-a-bD)\right] \tag{16}$$

$$= \arg zeroE\left[W\cdot\left\{\tau-1\left(Y<a+bD\right)\right\}\cdot\begin{pmatrix}1\\D\end{pmatrix}\right]. \tag{17}$$

This implies the moment conditions:

$$\begin{aligned} E\left[\frac{Z-p(X)}{p(X)\left(1-p(X)\right)}\left(2D-1\right)\left\{\tau-1\left(Y<a+bD\right)\right\}\right] &= 0 \\ E\left[\frac{Z-p(X)}{p(X)\left(1-p(X)\right)}\left(2D-1\right)\left\{\tau-1\left(Y<a+bD\right)\right\}\cdot D\right] &= 0. \end{aligned}$$

Multiplying the first moment condition with $(D+(1-D))$ inside the expectation operator and inserting the second moment condition gives:

$$\begin{aligned} E\left[\frac{Z-p(X)}{p(X)\left(1-p(X)\right)}\left(2D-1\right)\left\{\tau-1\left(Y<a+bD\right)\right\}\cdot(1-D)\right] &= 0 \\ E\left[\frac{Z-p(X)}{p(X)\left(1-p(X)\right)}\left(2D-1\right)\left\{\tau-1\left(Y<a+bD\right)\right\}\cdot D\right] &= 0 \end{aligned}$$

which is equivalent to

$$\begin{aligned} E\left[\frac{Z-p(X)}{p(X)\left(1-p(X)\right)}\left\{\tau-1\left(Y<a\right)\right\}\cdot(1-D)\right] &= 0 \\ E\left[\frac{Z-p(X)}{p(X)\left(1-p(X)\right)}\left\{\tau-1\left(Y<a+b\right)\right\}\cdot D\right] &= 0. \end{aligned}$$

Renaming a with q_0 and $a+b$ with q_1 and subtracting the term $E\left[\tau\frac{Z-p(X)}{p(X)(1-p(X))}\right]$, which is zero, from the first moment condition gives

$$\begin{aligned} E\left[\frac{Z-p(X)}{p(X)\left(1-p(X)\right)}\left\{1\left(Y<q_0\right)\left(D-1\right)-\tau D)\right\}\right] &= 0 \\ E\left[\frac{Z-p(X)}{p(X)\left(1-p(X)\right)}\left\{\tau-1\left(Y<q_1\right)\right\}\cdot D\right] &= 0, \end{aligned}$$

which are identical to (4).

A.5 Proof of Proposition (5)

Replicating the previous proofs in reverse order, one can first show that the first order conditions to

$$\arg\min_{a,b} E\left[\left(1-\frac{D(1-Z)}{1-p(X)}-\frac{(1-D)Z}{p(X)}\right)\rho_\tau(Y-\alpha-\beta D)\right]$$

are:

$$\begin{aligned} E\left[D\left(\frac{Z-p(X)}{1-p(X)}\right)1\left(Y<\alpha+\beta\right)\right] &= \tau\cdot P(T=c,D=1) \\ E\left[(D-1)\left(\frac{Z-p(X)}{p(X)}\right)1\left(Y<\alpha\right)\right] &= \tau\cdot P(T=c,D=0). \end{aligned} \tag{18}$$

where $\Pr(T=c,D=1)=\Pr(D=1|T=c)\Pr(T=c)=E\left[D\frac{Z-p(X)}{1-p(X)}\right]$ is the fraction of 'treated compliers' and $\Pr(T=c,D=0)=\Pr(D=0|T=c)\Pr(T=c)=E\left[(D-1)\frac{Z-p(X)}{p(X)}\right]$ is the fraction of 'non-treated compliers'. (Since the proof is very similar to the previous one it is ommitted.)

Define the distributions of the potential outcomes for *treated compliers* and *non-treated compliers* as

$$\begin{aligned} F_{Y^1|c,D=1}(u) &= \Pr(Y^1\leq u|D=1,T=c) \\ F_{Y^0|c,D=0}(u) &= \Pr(Y^0\leq u|D=0,T=c). \end{aligned}$$

Analogously to the previous proofs one can show that these distributions are identified as

$$\begin{aligned} F_{Y^1|c,D=1}(u) &= \frac{E\left[1\left(Y<u\right)\cdot D\cdot\frac{Z-p(X)}{1-p(X)}\right]}{\Pr(T=c,D=1)} \\ F_{Y^0|c,D=0}(u) &= \frac{E\left[1\left(Y<u\right)\cdot (D-1)\cdot\frac{Z-p(X)}{p(X)}\right]}{\Pr(T=c,D=0)}. \end{aligned}$$

Hence, $\alpha+\beta$ and α in (18) define the quantiles in the sense that $F_{Y^1|c,D=1}(\alpha_0+\beta_0)=\tau=F_{Y^0|c,D=0}(\alpha_0)$. This implies then that $F^{-1}_{Y^1|c,D=1}(\tau)=\alpha_0+\beta_0$ and $F^{-1}_{Y^0|c,D=0}(\tau)=\alpha_0$ and that

$$\beta_0=F^{-1}_{Y^1|c,D=1}(\tau)-F^{-1}_{Y^0|c,D=0}(\tau).$$

A.6 Proof of Proposition (6):

We show that $E\left[\rho_\tau(Y-X'b-aD)\cdot W\right]$ has expectation zero in the subpopulation of always- and never-participants, for every value of a and b. Hence, $E\left[\rho_\tau(Y-X'b-aD)\cdot W\,|\mathcal{T}=a\right]\cdot\Pr\left(\mathcal{T}=a\right)$ such that the relationship below Proposition (6) follows.

Note first that

$$\begin{aligned} E\left[\rho_\tau(Y-X'\beta-\alpha D)W\,|\mathcal{T}=a\right] &= E\left[\rho_\tau(Y-X'\beta-\alpha D)\frac{Z-p\left(X\right)}{p(X)\left(1-p(X)\right)}\left(2D-1\right)|\mathcal{T}=a\right] \\ &= E\left[E\left[\rho_\tau(Y^1-X'\beta-\alpha)\frac{Z-p\left(X\right)}{p(X)\left(1-p(X)\right)}|X,Z,\mathcal{T}=a\right]|\mathcal{T}=a\right] \\ &= E\left[E\left[\rho_\tau(Y^1-X'\beta-\alpha)|X,\mathcal{T}=a\right]\frac{E\left[Z|X,T=a\right]-p\left(X\right)}{p(X)\left(1-p(X)\right)}|\mathcal{T}=a\right] \\ &= E\left[E\left[\rho_\tau(Y^1-X'\beta-\alpha)|X,\mathcal{T}=a\right]\frac{E\left[Z|X\right]-p\left(X\right)}{p(X)\left(1-p(X)\right)}|\mathcal{T}=a\right] \\ &= 0. \end{aligned}$$

where $Y^d \perp\!\!\!\perp Z|X, \mathcal{T}$ and $\mathcal{T} \perp\!\!\!\perp Z|X$ has been used which follows from Assumption 1. The same result holds for the never-taker. Therefore the expression (13) is equivalent to

$$\begin{aligned}(\alpha, \beta) &= \arg\min_{a,b} E\left[\rho_\tau(Y - X'b - aD) \cdot W \,|\mathcal{T} = c\right] \\ &= \arg\min_{a,b} E\left[\left(\frac{D}{p(X)} + \frac{1-D}{1-p(X)}\right) \cdot \rho_\tau(Y - X'\beta - \alpha D) \;\; |\mathcal{T} = c\right],\end{aligned}$$

which is the objective function of a weighted linear quantile regression for compliers. Note that all weights are strictly positive and finite because we assume that $0 < p(X) < 1$. Therefore, standard quantile regression results (see for instance Koenker (2005) theorem 4.1 and 5.1) imply that this function is minimized at α_0^τ and β_0^τ as long as $E\left[\left(\frac{D}{p(X)} + \frac{1-D}{1-p(X)}\right)(X', D)'(X', D)\right]$ is positive definite, what is assumed.

A.7 Proof of Theorem (7):

Semiparametric efficiency bounds were introduced by Stein (1956) and developed by Koshevnik and Levit (1976), Pfanzagl and Wefelmeyer (1982), Begun, Hall, Huang, and Wellner (1983) and Bickel, Klaassen, Ritov, and Wellner (1993). See also the survey of Newey (1990) or Newey (1994).

We need to derive the efficiency bound for

$$\Delta_c^\tau = Q^\tau_{Y^1|c} - Q^\tau_{Y^0|c}.$$

For this it will be helpful to have an expression for $f_{Y^1|c}$ and $f_{Y^0|c}$. From Theorem (1) it follows that

$$\begin{aligned}f_{Y^1|c}(u) &= \left\{\int \left(f_{Y|X,D=1,Z=1}(u)\pi(x,1) - f_{Y|X,D=1,Z=0}(u)\pi(x,0)\right) dF_X\right\}/P_c \\ f_{Y^0|c}(u) &= -\left\{\int \left(f_{Y|X,D=0,Z=1}(u)(1-\pi(x,1)) - f_{Y|X,D=0,Z=0}(u)(1-\pi(x,0))\right) dF_X\right\}/P_c\end{aligned}$$

where $\pi(x,z) = \Pr(D=1|X=x, Z=z)$ and $P_c = \int (\pi(x,1) - \pi(x,0))\, dF_X$ is the fraction of compliers.

By **Assumption 2** the quantiles $Q^\tau_{Y^1|c}$ and $Q^\tau_{Y^0|c}$ are unique and defined as

$$\begin{aligned}0 &= E\left[1(Y^1 \le Q^\tau_{Y^1|c}) - \tau | T = c\right] = \int \left(1(u \le Q^\tau_{Y^1|c}) - \tau\right) \cdot f_{Y^1|c}(u) du \qquad (19) \\ 0 &= E\left[1(Y^0 \le Q^\tau_{Y^0|c}) - \tau | T = c\right] = \int \left(1(u \le Q^\tau_{Y^0|c}) - \tau\right) \cdot f_{Y^0|c}(u) du\end{aligned}$$

where $f_{Y^d|c}$ are given above. We thus have expressed the quantiles in terms of the densities of the observed variables.

The joint density of the observed variables (Y, D, Z, X) with D and Z binary can be written as

$$f(y,d,z,x) = f(y|d,z,x)\, f(d|z,x)\, f(z|x)\, f(x)$$
$$= f(y|d,z,x)\left\{\pi(x,z)^d \cdot (1-\pi(x,z))^{1-d}\right\}\left\{p(x)^z \cdot (1-p(x))^{1-z}\right\} f(x).$$

Consider a regular parametric submodel indexed by θ with θ_0 corresponding to the true model: $f(y,d,z,x;\theta_0) = f(y,d,z,x)$. The density $f(y,d,z,x;\theta)$ can be written as

$$\begin{aligned} f(y,d,z,x;\theta) &= f^{11}(y|x;\theta)^{dz} \cdot f^{10z}(y|x;\theta)^{d(1-z)} \cdot f^{01}(y|x;\theta)^{(1-d)z} \cdot f^{00}(y|x;\theta)^{(1-d)(1-z)} \\ &\quad \left\{\pi(x,z;\theta)^d \cdot (1-\pi(x,z;\theta))^{1-d}\right\}\left\{p(x;\theta)^z \cdot (1-p(x;\theta))^{1-z}\right\} f(x;\theta), \end{aligned}$$

where $f^{dz}(y|x;\theta) = f(y|d,z,x;\theta)$.

We will assume throughout that all terms of the previous equation admit an interchange of the order of integration and differentiation, such that

$$\int \frac{\partial f(y,d,z,x;\theta)}{\partial\theta} dydddzdx = \frac{\partial}{\partial\theta}\int f(y,d,z,x;\theta)\, dydddzdx = 0.$$

Sufficient conditions for permitting interchanging differentiation and integration are, for example, given by Theorem 1.3.2 of Amemiya (1985).

The corresponding score of $f(y,d,z,x;\theta)$ is

$$\begin{aligned} s(y,d,z,x;\theta) &= \frac{\partial \ln f(y,d,z,x;\theta)}{\partial\theta} \\ &= dz\breve{f}^{11}(y|x;\theta) + d(1-z)\breve{f}^{10}(y|x;\theta) + (1-d)z\breve{f}^{01}(y|x;\theta) + (1-d)(1-z)\breve{f}^{00}(y|x;\theta) \\ &\quad + \frac{d-\pi(x,z;\theta)}{1-\pi(x,z;\theta)}\breve{\pi}(x,z,\theta) + \frac{z-p(x;\theta)}{1-p(x;\theta)}\breve{p}(x,\theta) + \breve{f}(x;\theta), \end{aligned}$$

where the subscript $\breve{f}$ defines a derivative of the log, i.e. $\breve{f}(x;\theta) = \partial \ln f(x;\theta)/\partial\theta$.

At the true value θ_0 the expectation of the score is zero. The tangent space of the model is the set of functions that are mean zero and satisfy the additive structure of the score:

$$\Im = \left\{ \begin{array}{c} dzs^{11}(y|x) + d(1-z)s^{10}(y|x) + (1-d)zs^{01}(y|x) + (1-d)(1-z)s^{00}(y|x) \\ + (d-\pi(x,z)) \cdot s_\pi(x,z) + (z-p(x)) \cdot s_p(x) + s_x(x) \end{array} \right\} \quad (20)$$

for any functions $s^{11}, s^{10}, s^{01}, s^{00}, s_x$ satisfying the mean-zero property: $E\left[s^{dz}|D=d, Z=z, X\right] = 0 = E\left[s_x(x)\right]$ and $s_\pi(x,z)$ and $s_p(x)$ being square-integrable measurable functions.

The *semiparametric variance bound* of Δ_c^τ is the variance of the projection on $\Im$ of a function $\psi(Y,D,Z,X)$ (with $E\left[\psi\right]=0$ and $E[\|\psi(\cdot)\|^2] < \infty$) that satisfies for all regular parametric submodels

$$\frac{\partial \Delta_c^\tau(F_\theta)}{\partial \theta}_{|\theta=\theta_0} = E\left[\psi(Y,D,Z,X)\cdot s(Y,D,Z,X)\right]_{|\theta=\theta_0} \tag{21}$$

If ψ itself already lies in the tangent space, the variance bound is given by $E\left[\psi^2\right]$.

As a first step to calculating the variance bound, we need to derive

$$\frac{\partial \Delta_c^\tau(\theta)}{\partial \theta} = \frac{\partial Q^\tau_{Y^1|c}(\theta)}{\partial \theta} - \frac{\partial Q^\tau_{Y^0|c}(\theta)}{\partial \theta}.$$

The identify (19) holds for all submodels θ such that we obtain

$$\frac{\partial}{\partial \theta}\int \left(1(u \le Q^\tau_{Y^1|c}(\theta)) - \tau\right)\cdot f_{Y^1|c}(u;\theta)du = 0$$

$$= (1-\tau)\frac{\partial}{\partial\theta}\int_{-\infty}^{Q^\tau_{Y^1|c}(\theta)} f_{Y^1|c}(u;\theta)du - \tau\frac{\partial}{\partial\theta}\int_{Q^\tau_{Y^1|c}(\theta)}^{\infty} f_{Y^1|c}(u;\theta)du$$

$$= f_{Y^1|c}(Q^\tau_{Y^1|c}(\theta);\theta)\cdot\frac{\partial Q^\tau_{Y^1|c}(\theta)}{\partial\theta} + \int\left(1(u\le Q^\tau_{Y^1|c}(\theta)) - \tau\right)\frac{\partial}{\partial\theta}f_{Y^1|c}(u;\theta)du = 0.$$

by Leibniz's rule of differentiation. We thus obtain that the derivative evaluated at the true θ_0 is

$$\frac{\partial \Delta_c^\tau(\theta)}{\partial\theta}_{|\theta=\theta_0} = \frac{\int\left(\tau - 1(u\le Q^\tau_{Y^1|c})\right)\frac{\partial}{\partial\theta}f_{Y^1|c}(u;\theta_0)du}{f_{Y^1|c}(Q^\tau_{Y^1|c})} - \frac{\int\left(\tau - 1(u\le Q^\tau_{Y^0|c})\right)\frac{\partial}{\partial\theta}f_{Y^0|c}(u;\theta_0)du}{f_{Y^0|c}(Q^\tau_{Y^0|c})}$$

where

$$\frac{\partial}{\partial\theta}f_{Y^1|c}(u;\theta_0) = \frac{1}{P_c}\frac{\partial}{\partial\theta}\left\{\int\left(f_{Y|X,D=1,Z=1}(u)\pi(x,1) - f_{Y|X,D=1,Z=0}(u)\pi(x,0)\right)f(x)dx\right\}$$
$$- f_{Y^1|c}(u;\theta_0)\frac{\partial \ln P_c(\theta_0)}{\partial\theta}$$

such that

$$\frac{\partial \Delta_c^\tau(\theta)}{\partial \theta}_{|\theta=\theta_0}$$

$$= \frac{\int \left(\tau - 1(u \leq Q^\tau_{Y^1|c})\right) \frac{\partial}{\partial \theta} \left\{\int \left(f_{Y|X,D=1,Z=1}(u)\pi(x,1) - f_{Y|X,D=1,Z=0}(u)\pi(x,0)\right) f(x)dx\right\} du}{P_c \cdot f_{Y^1|c}(Q^\tau_{Y^1|c})}$$

$$+ \frac{\int \left(\tau - 1(u \leq Q^\tau_{Y^0|c})\right) \frac{\partial}{\partial \theta} \left\{\int \left(f_{Y|X,D=0,Z=1}(u)\left(1-\pi(x,1)\right) - f_{Y|X,D=0,Z=0}(u)\left(1-\pi(x,0)\right)\right) f(x)dx\right\} du}{P_c \cdot f_{Y^0|c}(Q^\tau_{Y^0|c})}$$

Define

$$\chi_{dz}(x) = \frac{1}{P_c \cdot f_{Y^d|c}(Q^\tau_{Y^d|c})} \left\{\left(\tau - 1(y \leq Q^\tau_{Y^d|c})\right) - E\left[\tau - 1(y \leq Q^\tau_{Y^d|c}) | D=d, Z=z, X=x\right]\right\}$$

and

$$\vartheta_{dz}(x) = \frac{1}{P_c \cdot f_{Y^d|c}(Q^\tau_{Y^d|c})} E\left[\tau - 1(y \leq Q^\tau_{Y^d|c}) | D=d, Z=z, X=x\right]$$

and choose $\psi(Y, D, Z, X)$ as

$$\begin{aligned}\psi(y,d,z,x) &= \frac{zd}{p(x)}\chi_{11}(x) - \frac{(1-z)d}{1-p(x)}\chi_{10}(x) + \frac{z(1-d)}{p(x)}\chi_{01}(x) - \frac{(1-z)(1-d)}{1-p(x)}\chi_{00}(x) \\ &\quad + \vartheta_{11}(x) - \vartheta_{10}(x) + \vartheta_{01}(x) - \vartheta_{00}(x) - E\left[\vartheta_{11}(x) - \vartheta_{10}(x) + \vartheta_{01}(x) - \vartheta_{00}(x)\right],\end{aligned}$$

which, after some tedious calculations, can be shown to satisfy (21).

Since ψ is mean zero and lies in the tangent space, the variance bound is

$$E\left[\psi(y,d,z,x)^2\right]$$

$$\begin{aligned}&= E\left[\left(\frac{zd}{p(x)}\chi_{11}(x) - \frac{(1-z)d}{1-p(x)}\chi_{10}(x) + \frac{z(1-d)}{p(x)}\chi_{01}(x) - \frac{(1-z)(1-d)}{1-p(x)}\chi_{00}(x)\right)^2\right] \\ &\quad + \frac{1}{P_c^2} Var\left(\frac{F_{Y|D=1,Z=0,X}(Q^\tau_{Y^1|c}) - F_{Y|D=1,Z=1,X}(Q^\tau_{Y^1|c})}{f_{Y^1|c}(Q^\tau_{Y^1|c})} + \frac{F_{Y|D=0,Z=0,X}(Q^\tau_{Y^0|c}) - F_{Y|D=0,Z=1,X}(Q^\tau_{Y^0|c}}{f_{Y^0|c}(Q^\tau_{Y^0|c})}\right.\end{aligned}$$

$$\begin{aligned}&= \frac{1}{P_c^2 f^2_{Y^1|c}(Q^\tau_{Y^1|c})} E\left[\left(\frac{zd}{p(x)}\chi_{11}(x)\right)^2\right] + \frac{1}{P_c^2 f^2_{Y^1|c}(Q^\tau_{Y^1|c})} E\left[\left(\frac{(1-z)d}{1-p(x)}\chi_{10}(x)\right)^2\right] \\ &\quad + \frac{1}{P_c^2 f^2_{Y^0|c}(Q^\tau_{Y^0|c})} E\left[\left(\frac{z(1-d)}{p(x)}\chi_{01}(x)\right)^2\right] + \frac{1}{P_c^2 f^2_{Y^0|c}(Q^\tau_{Y^0|c})} E\left[\left(\frac{(1-z)(1-d)}{1-p(x)}\chi_{00}(x)\right)^2\right] \\ &\quad + \frac{1}{P_c^2} Var\left(\frac{F_{Y|D=1,Z=0,X}(Q^\tau_{Y^1|c}) - F_{Y|D=1,Z=1,X}(Q^\tau_{Y^1|c})}{f_{Y^1|c}(Q^\tau_{Y^1|c})} + \frac{F_{Y|D=0,Z=0,X}(Q^\tau_{Y^0|c}) - F_{Y|D=0,Z=1,X}(Q^\tau_{Y^0|c}}{f_{Y^0|c}(Q^\tau_{Y^0|c})}\right.\end{aligned}$$

$$
\begin{aligned}
= \quad & \frac{1}{P_c^2 f^2_{Y^1|c}(Q^\tau_{Y^1|c})} E\left[\frac{\pi(X,1)}{p(X)} F_{Y|D=1,Z=1,X}(Q^\tau_{Y^1|c})\left(1 - F_{Y|D=1,Z=1,X}(Q^\tau_{Y^1|c})\right)\right] \\
& + \frac{1}{P_c^2 f^2_{Y^1|c}(Q^\tau_{Y^1|c})} E\left[\frac{\pi(X,0)}{1-p(x)} F_{Y|D=1,Z=0,X}(Q^\tau_{Y^1|c})\left(1 - F_{Y|D=1,Z=0,X}(Q^\tau_{Y^1|c})\right)\right] \\
& + \frac{1}{P_c^2 f^2_{Y^0|c}(Q^\tau_{Y^0|c})} E\left[\frac{1-\pi(X,1)}{p(X)} F_{Y|D=0,Z=1,X}(Q^\tau_{Y^0|c})\left(1 - F_{Y|D=0,Z=1,X}(Q^\tau_{Y^0|c})\right)\right] \\
& + \frac{1}{P_c^2 f^2_{Y^0|c}(Q^\tau_{Y^0|c})} E\left[\frac{1-\pi(X,0)}{1-p(X)} F_{Y|D=0,Z=0,X}(Q^\tau_{Y^0|c})\left(1 - F_{Y|D=0,Z=0,X}(Q^\tau_{Y^0|c})\right)\right] \\
& + \frac{1}{P_c^2} Var\left(\frac{F_{Y|D=1,Z=0,X}(Q^\tau_{Y^1|c}) - F_{Y|D=1,Z=1,X}(Q^\tau_{Y^1|c})}{f_{Y^1|c}(Q^\tau_{Y^1|c})} + \frac{F_{Y|D=0,Z=0,X}(Q^\tau_{Y^0|c}) - F_{Y|D=0,Z=1,X}(Q^\tau_{Y^0|c}}{f_{Y^0|c}(Q^\tau_{Y^0|c})}\right.
\end{aligned}
$$

because

$$
E\left[\left(\frac{DZ}{p(X)}\chi_{11}(X)\right)^2\right] = E\left[E\left[\frac{DZ}{p^2(X)}\chi^2_{11}(X)|X\right]\right] = E\left[E\left[\frac{\pi(X,1)p(X)}{p^2(X)} E\left[\chi^2_{11}(X)|D=Z=1,X\right]|X\right]\right]
$$

and

$$
E\left[\chi^2_{11}(X)|D=Z=1,X\right] = F_{Y|D=1,Z=1,X}(Q^\tau_{Y^1|c})\left(1 - F_{Y|D=1,Z=1,X}(Q^\tau_{Y^1|c})\right)
$$

and analogously for the other terms.

References

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

Amemiya, T. (1985): *Advanced Econometrics*. Harvard University Press, Cambridge, Mass.

Begun, J., W. Hall, W. Huang, and J. Wellner (1983): "Information and Asymptotic Efficiency in Parametric-Nonparametric Models," *Annals of Statistics*, 11, 432–452.

Bickel, P., C. Klaassen, Y. Ritov, and J. Wellner (1993): *Efficient and Adaptive Estimation for Semiparametric Models*. John Hopkins University Press, Baltimore.

Card, D. (1995): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*, ed. by L. Christofides, E. Grant, and R. Swidinsky, pp. 201–222. University of Toronto Press, Toronto.

Chaudhuri, P. (1991): "Global nonparametric estimation of conditional quantile functions and their derivatives," *Journal of Multivariate Analysis*, 39, 246–269.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

——— (2006): "Instrumental quantile regression inference for structural and treatment effect models," *Journal of Econometrics*, 132, 491–525.

Chernozhukov, V., G. Imbens, and W. Newey (2007): "Instrumental variable estimation of nonseparable models," *forthcoming in Journal of Econometrics*, 0, X–X.

Chesher, A. (2003): "Identification in nonseparable models," *Econometrica*, 71, 1405–1441.

——— (2005): "Nonparametric identification under discrete variation," *Econometrica*, 73, 1525–1550.

——— (2007): "Endogeneity and discrete outcomes," mimeo.

Firpo, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *forthcoming in Econometrica*, 1, —.

Frölich, M. (2006a): "Nonparametric IV Estimation of Local Average Treatment Effects with Covariates," *forthcoming in Journal of Econometrics*.

——— (2006b): "A Note on Parametric and Nonparametric Regression in the Presence of Endogenous Control Variables," *IZA Discussion Paper*, 2126.

——— (2007): "Propensity score matching without conditional independence assumption - with an application to the gender wage gap in the UK," *forthcoming in Econometrics Journal*.

Heckman, J., and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669–738.

Horowitz, J., and S. Lee (2007): "Nonparametric Instrumental Variables Estimation of a Quantile Regression Model," *forthcoming in Econometrica*, -, –.

Imbens, G., and J. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Imbens, G., and W. Newey (2003): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," presented at the EC2 conference London December 2003.

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

Koshevnik, Y., and B. Levit (1976): "On a Non-parametric Analogue of the Information Matrix," *Theory of Probability and Applications*, 21, 738–753.

Melly, B. (2006): "Estimation of counterfactual distribution using quantile regression," mimeo.

Newey, W. (1990): "Semiparametric Efficiency Bounds," *Journal of Applied Econometrics*, 5, 99–135.

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

Pfanzagl, J., and W. Wefelmeyer (1982): *Contributions to a General Asymptotic Statistical Theory*. Springer Verlag, Heidelberg.

Stein, C. (1956): "Efficient Nonparametric Testing and Estimation," in *Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. University of California Press, Berkeley.

Table 1: Descriptive statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
ID	3613	2609.78	1498.51	2.0000000	5225.00
NEARC2	3613	0.4317741	0.4953919	0	1.0000000
NEARC4	**3613**	**0.6781068**	**0.4672669**	**0**	**1.0000000**
NEARC4A	3613	0.4921118	0.5000070	0	1.0000000
NEARC4B	3613	0.1859950	0.3891565	0	1.0000000
ED76	3613	13.2252975	2.7497411	0	18.0000000
ED66	3613	10.7428730	2.4590854	0	18.0000000
AGE76	3613	28.1752007	3.1718104	24.0000000	34.0000000
DADED	3613	10.0028785	3.2960212	0	18.0000000
NODADED	3613	0.2241904	0.4171058	0	1.0000000
MOMED	3613	10.3421672	3.0293785	0	18.0000000
NOMOMED	3613	0.1143094	0.3182308	0	1.0000000
WEIGHT	3613	320318.35	168006.76	75607.00	1752340.00
MOMDAD14	3613	0.7921395	0.4058326	0	1.0000000
SINMOM14	3613	0.1001937	0.3002997	0	1.0000000
STEP14	3613	0.0384722	0.1923599	0	1.0000000
REG661	3613	0.0445613	0.2063671	0	1.0000000
REG662	3613	0.1549958	0.3619508	0	1.0000000
REG663	3613	0.1940216	0.3955003	0	1.0000000
REG664	3613	0.0691946	0.2538199	0	1.0000000
REG665	3613	0.2095212	0.4070232	0	1.0000000
REG666	3613	0.0929975	0.2904691	0	1.0000000
REG667	3613	0.1101578	0.3131296	0	1.0000000
REG668	3613	0.0309992	0.1733394	0	1.0000000
REG669	3613	0.0935511	0.2912434	0	1.0000000
SOUTH66	3613	0.4126764	0.4923837	0	1.0000000
WORK76	3613	0.8350401	0.3711957	0	1.0000000
WORK78	3613	0.7351232	0.4413287	0	1.0000000
LWAGE76	**3010**	**6.2618319**	**0.4437977**	**4.6051702**	**7.7848893**
LWAGE78	2639	6.3291080	0.4442450	4.6965200	8.2409240
FAMED	3613	5.9128148	2.6504318	1.0000000	9.0000000
BLACK	3613	0.2300028	0.4208925	0	1.0000000
SMSA76R	3613	0.6947135	0.4605924	0	1.0000000
SMSA78R	3319	0.6929798	0.4613273	0	1.0000000
REG76R	3613	0.3996679	0.4898978	0	1.0000000
REG78R	3319	0.3968063	0.4893089	0	1.0000000
REG80R	3227	0.4028509	0.4905473	0	1.0000000
SMSA66R	3613	0.6426792	0.4792768	0	1.0000000
WAGE76	**3017**	**576.0888300**	**263.8199090**	**25.0000000**	**2404.00**
WAGE78	2656	724.5591114	526.1991520	17.0000000	17628.00
WAGE80	2520	869.8940476	492.1729068	27.0000000	13857.00
NOINT78	3613	0.0813728	0.2734447	0	1.0000000
NOINT80	3613	0.1068364	0.3089479	0	1.0000000
ENROLL76	3613	0.0946582	0.2927827	0	1.0000000
ENROLL78	3317	0.0654206	0.2473038	0	1.0000000
ENROLL80	3220	0.0583851	0.2345066	0	1.0000000
KWW	3543	33.4891335	8.6918079	0	56.0000000
IQ	2470	102.5878543	15.4450703	50.0000000	156.0000000
MARSTA76	3604	2.3571032	2.1096377	1.0000000	6.0000000
MARSTA78	3319	2.2136186	2.0058342	1.0000000	6.0000000
MARSTA80	3227	2.1041215	1.9088835	1.0000000	6.0000000
LIBCRD14	3598	0.6717621	0.4696372	0	1.0000000

Table 2: Probit regression of instrument Z on X

```
. probit nearc4 kww exp76 exp762  $listev
Probit regression                                 Number of obs   =       3543
                                                  LR chi2(29)     =     953.90
                                                  Prob > chi2     =     0.0000
Log likelihood = -1753.1777                       Pseudo R2       =     0.2139
------------------------------------------------------------------------------
      nearc4 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         kww |   .0094064   .0033342     2.82   0.005     .0028716    .0159413
       exp76 |  -.0107292   .0237444    -0.45   0.651    -.0572673    .0358089
      exp762 |  -.0001186   .0011236    -0.11   0.916    -.0023208    .0020835
       black |   .2748305   .0730832     3.76   0.000     .1315902    .4180709
     smsa76r |   .2206061   .0675998     3.26   0.001     .0881129    .3530993
      reg76r |  -.0669019    .092463    -0.72   0.469    -.2481261    .1143222
     smsa66r |   .9840224   .0662247    14.86   0.000     .8542245     1.11382
      reg662 |   .0459755   .1539718     0.30   0.765    -.2558036    .3477547
      reg663 |  -.5406716   .1447889    -3.73   0.000    -.8244527   -.2568905
      reg664 |  -.3624839     .15834    -2.29   0.022    -.6728245   -.0521433
      reg665 |  -.6388701   .1644876    -3.88   0.000    -.9612599   -.3164804
      reg666 |  -1.006969   .1747037    -5.76   0.000    -1.349382   -.6645563
      reg667 |  -.9341539   .1739213    -5.37   0.000    -1.275033   -.5932744
      reg668 |   -.770222   .1859383    -4.14   0.000    -1.134654   -.4057896
      reg669 |   -.241757   .1597352    -1.51   0.130    -.5548322    .0713183
       daded |   .0096197   .0153433     0.63   0.531    -.0204527    .0396921
       momed |  -.0117358   .0146098    -0.80   0.422    -.0403705     .016899
     nodaded |  -.1937773   .1928149    -1.00   0.315    -.5716875     .184133
     nomomed |  -.0784467   .1258471    -0.62   0.533    -.3251024     .168209
          f1 |  -.0416213   .2844138    -0.15   0.884    -.5990621    .5158196
          f2 |   -.314627   .2581221    -1.22   0.223    -.8205371    .1912831
          f3 |  -.1896419   .2373843    -0.80   0.424    -.6549065    .2756228
          f4 |  -.0248458     .16387    -0.15   0.879    -.3460252    .2963335
          f5 |  -.2643152   .2345548    -1.13   0.260    -.7240343    .1954038
          f6 |  -.2811006   .2241565    -1.25   0.210    -.7204393    .1582382
          f7 |  -.3104026    .238763    -1.30   0.194    -.7783694    .1575642
          f8 |  -.1938829    .206183    -0.94   0.347    -.5979941    .2102283
    momdad14 |  -.0287816    .094051    -0.31   0.760    -.2131181     .155555
    sinmom14 |   .1059404   .1294821     0.82   0.413    -.1478398    .3597207
       _cons |   .2952601   .3249659     0.91   0.364    -.3416613    .9321816
------------------------------------------------------------------------------
```

Figure 1: Distribution of P(Z=1|X) in the Z=0 and Z=1 subpopulation

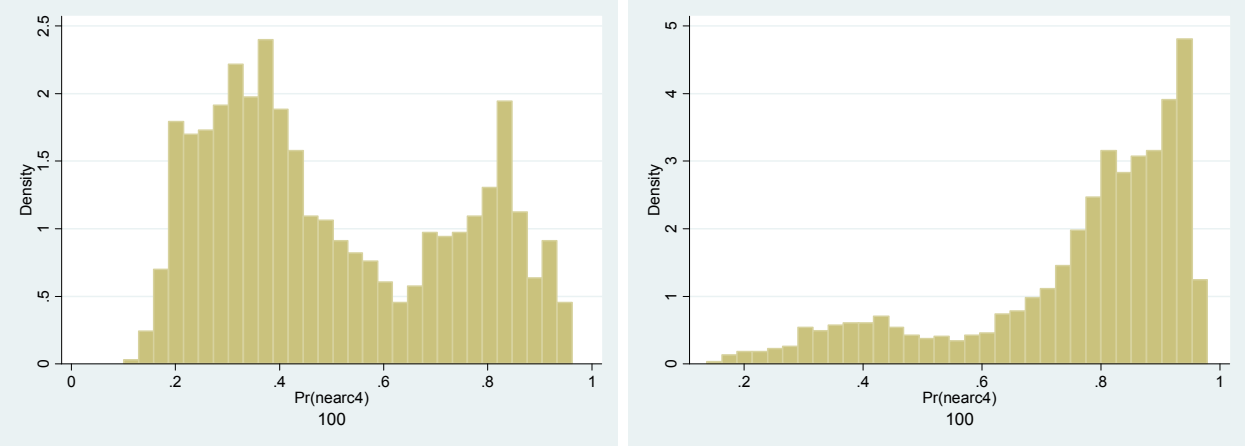

Figure 2: Nonparametric IV estimators of the QTE of college attendance

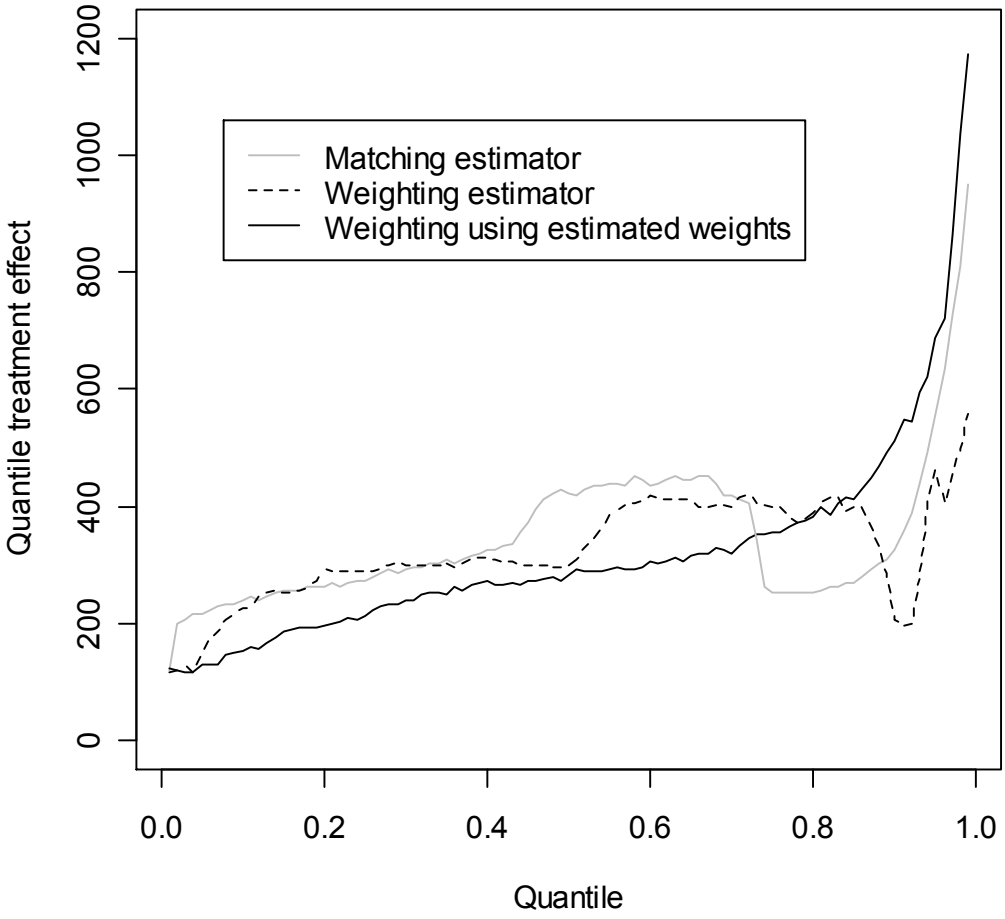

Figure 3: Nonparametric estimation of QTE with and without endogeneity

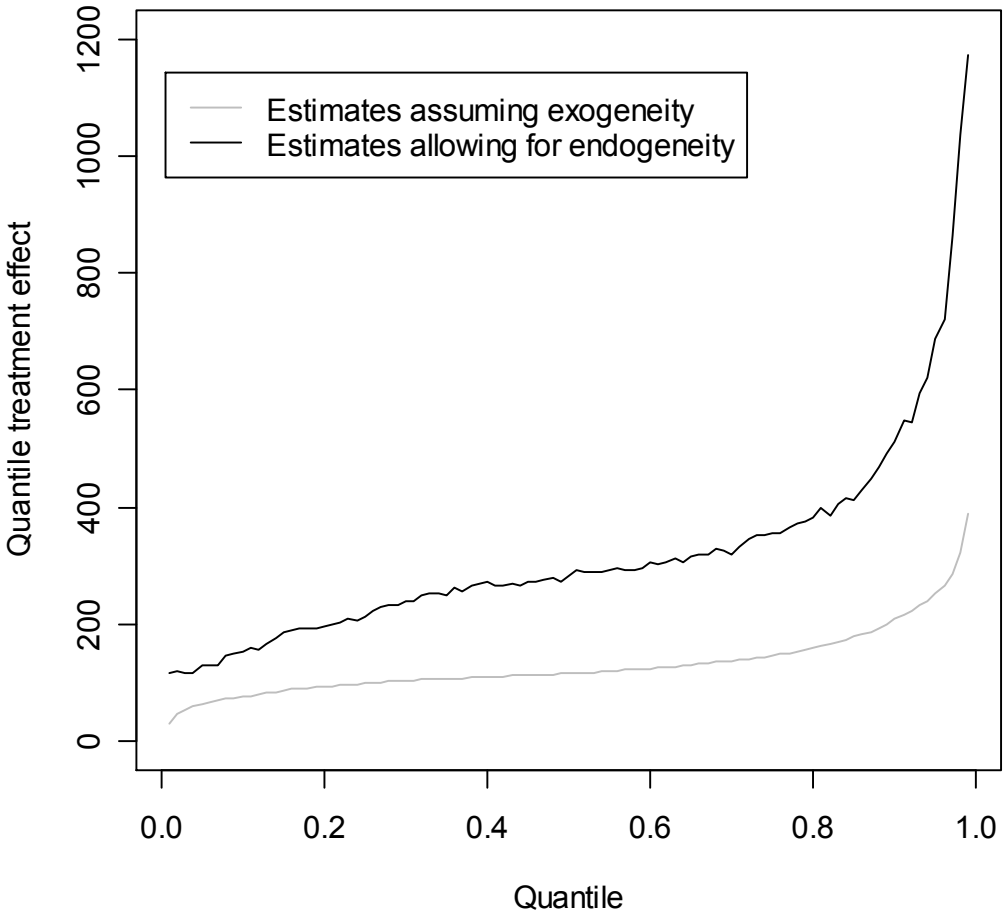

Figure 4: Comparison of QTE with and without endogeneity: relative comparison

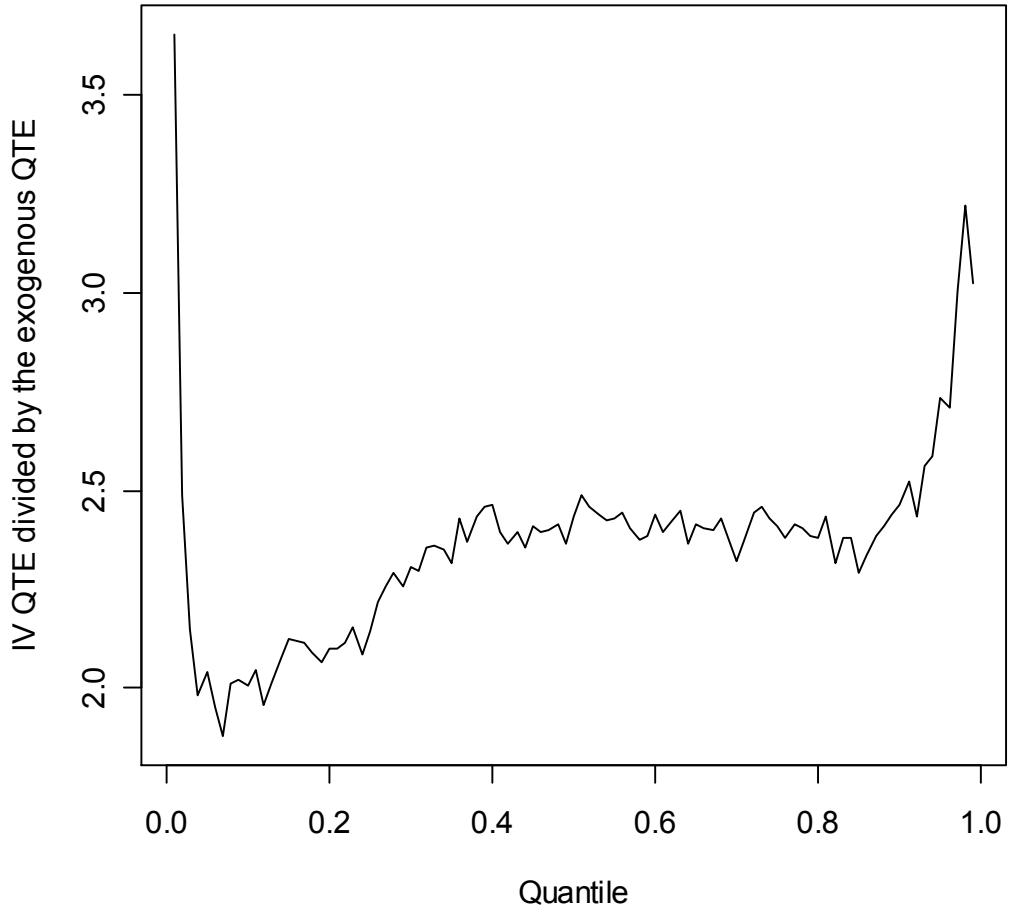

Figure 5: Bootstrap confidence bands for QTE with endogenous college choice