UNIVERSITÄT BERN

Faculty of Business, Economics and Social Sciences

Department of Economics

A Comparison of two Qantile Models with Endogeneity

Kaspar Wüthrich

14-08

November 2014

DISCUSSION PAPERS

Schanzeneckstrasse 1
Postfach 8573
CH-3001 Bern, Switzerland
http://www.vwi.unibe.ch

A Comparison of two Quantile Models with Endogeneity*

Kaspar Wüthrich†
University of Bern

November 18, 2014

Abstract

This paper analyzes estimators based on the instrumental variable quantile regression (IVQR) model (Chernozhukov and Hansen, 2004, 2005, 2006) under the local quantile treatment effects (LQTE) framework (Abadie et al., 2002). I show that the quantile treatment effect (QTE) estimators in the IVQR model are equivalent to LQTE for the compliers at transformed quantile levels. This transformation adjusts for differences between the subpopulation-specific potential outcome distributions that are identified in the LQTE model. Moreover, the IVQR estimator of the average treatment effect (ATE) corresponds to a convex combination of the local average treatment effect (LATE) and a weighted average of LQTE for the compliers. I extend the analysis to more general setups that allow for partial failures of the LQTE assumptions, non-binary instruments, and covariates. The results are illustrated with two empirical applications.

JEL Classification: C14, C21, C26

Keywords: Endogeneity, instrumental variables, quantile treatment effect, local quantile treatment effect, average treatment effect, local average treatment effect, rank similarity

*Acknowledgments: This paper has benefited from numerous discussions with Blaise Melly. I am grateful to Alberto Abadie, Josh Angrist, Stefan Boes, Daniel Burkhard, Victor Chernozhukov, Tobias Müller, and seminar participants at the University of Bern (2014) and the MIT Econometrics Lunch (2014) for helpful comments. The research was supported by the Swiss National Science Foundation (Doc.Mobility Project P1BEP1_155467). The usual disclaimer applies.

†University of Bern, Department of Economics, Schanzeneckstrasse 1, CH-3001 Bern, Switzerland, phone: +41 31 631 4089, email: kaspar.wuethrich@vwi.unibe.ch.

1 Introduction

Since the seminal work of Koenker and Bassett (1978), quantile regression methods have become important tools for analyzing effect of policy variables on distributional outcomes beyond simple averages. In many economic applications, the policy variables of interest are endogenous, rendering standard quantile regression inconsistent for recovering structural quantile treatment effects (QTE).

One approach to addressing this problem is to use instrumental variable (IV) methods. In this paper, I study the relationship between two competing IV models for estimating QTE: the instrumental variable quantile regression (IVQR) model (Chernozhukov and Hansen, 2004, 2005, 2006) and the local quantile treatment effects (LQTE) model (Abadie et al., 2002). Apart from instrument validity, both models rely on different sets of assumptions and identify different quantities. The IVQR model is based on the rank similarity assumption, a condition that restricts the evolution of individual ranks in the potential outcome distributions across treatment states. Under rank similarity, the IVQR model identifies the QTE for the whole population. The LQTE model exploits the monotonicity assumption introduced by Imbens and Angrist (1994) to identify the LQTE for the compliers, the subpopulation whose treatment status is affected by the instrument. The two sets of assumptions are generally non-nested and neither model is more general than the other (Chernozhukov and Hansen, 2013). Despite these differences, the two models often yield similar results in empirical applications (e.g., Chernozhukov and Hansen, 2004).

The goal of this paper is to formalize the relationship between the estimates of both models by characterizing estimators based on the IVQR model under the LQTE assumptions. First, I show that the IVQR estimators of the potential outcome cumulative distribution functions (CDF) can be expressed as functions of the potential outcome CDFs for never takers, always takers, and compliers. The key ingredients for deriving these results are closed form solutions for the IVQR estimands that are derived from the IVQR moment conditions (e.g., Chernozhukov and Hansen, 2005, 2006). Moreover, I establish a close relationship between the IVQR and the changes-in-changes (CIC) model (Athey and Imbens, 2006) that may be of interest in its own right. Second, I show that the IVQR estimands of the QTE correspond to LQTE evaluated at transformed quantile levels. This transformation adjusts for the difference between the distributions of the potential outcomes in the treated state for always takers and compliers as well as for the difference between the distributions of potential outcomes in the untreated state for never takers and compliers. Third, I show that the IVQR estimate of the ATE corresponds to a convex combination of the local average treatment effect (LATE) and a weighted average of LQTE. Consequently, differences between the estimates of both models are uniquely determined by two factors: between-subpopulation differences in the potential outcome distributions that are identified under the LQTE assumptions and the relative size of the respective subpopulations, which depends on the strength of the instrument.

These results have several important implications: First, comparisons between the estimates of both models are not indicative of the validity of the rank similarity assumption. Second, the sensitivity of the IVQR estimates to deviations from rank similarity decreases with the strength of the instrument. Third, IVQR estimates capture particular causal effects for the compliers irrespective of the validity of the rank similarity assumption so long as the LQTE assumptions hold. Forth, the IVQR estimates preserve sign and monotonicity of the LQTE estimates provided that these properties are invariant across quantiles. Fifth, the IVQR model estimates the QTE and the ATE by extrapolating from the compliers to the whole population based on the rank similarity assumption. Therefore, it constitutes an alternative approach to extrapolation in the LATE framework.[1]

The analysis is extended to more general setups that allow for failures of the LQTE monotonicity assumption, non-binary instruments, and covariates. I show that the main results describing the relationship between the IVQR and the LQTE estimates have intuitive analogues in these more general setups.

The results are illustrated using two application. In the first application, I examine the causal effect of JTPA training programs on the distribution of subsequent earnings. I find that both models yield similar results, which can be attributed to the strength of the instrument that outweighs the differences between the potential outcome distributions of never takers and compliers. In the second application, I study estimation of the structural effect of veteran status on civilian wages using draft lottery data. The substantial numerical differences between the estimates of both models can be attributed to a relatively weak instrument combined with a large treatment effect heterogeneity for the compliers.

This paper is related to the extensive literature on both models. The IVQR model is introduced by Chernozhukov and Hansen (2004, 2005, 2006) and recently surveyed by Chernozhukov and Hansen (2013). Linear conditional quantile models are analyzed by Chernozhukov and Hong (2003), Chernozhukov and Hansen (2006), Chernozhukov et al. (2007a), Chernozhukov and Hansen (2008), and Chernozhukov et al. (2009). Nonparametric estimation of the IVQR model is studied

[1] There are several approaches to extrapolation in the LATE framework: Heckman et al. (2001, 2003) and Angrist (2004) use parametric models latent index models, Chamberlain (2011) analyzes a Bayesian semiparametric approach, and Angrist and Fernandez-Val (2013) consider covariate-based extrapolation methods.

by Chernozhukov et al. (2007b), Horowitz and Lee (2007), and Gagliardini and Scaillet (2012).

The LQTE model is introduced by Abadie et al. (2002) and extends the LATE framework (Imbens and Angrist, 1994; Angrist et al., 1996; Imbens and Rubin, 1997; Abadie, 2002) to the analysis of conditional LQTE for the compliers using the weightig theorem by Abadie (2003). In subsequent work, Frandsen et al. (2012) analyze estimation and inference of LQTE in regression discontinuity frameworks, Frölich and Melly (2013) study nonparametric identification and estimation of marginal LQTE with covariates, and Frölich and Melly (2013) analyze identification of marginal LQTE under one-sided non-compliance.

Two additional papers are related to mine. First, the analysis of the IVQR estimands under the LQTE assumptions without monotonicity nests the framework analyzed by De Chaisemartin (2014a,b) and imposing his compliers-defiers condition is helpful for interpreting the IVQR estimands in this case. Seond, my paper is related to Yu (2014), who extends the (MTE) framework (e.g., Heckman and Vytlacil, 2005) to marginal QTE as a means to unify different quantile treatment effects. The author provides a detailed discussion of the relationship between his framework and the IVQR and LQTE model that complements my analysis.

The remainder of the paper is organized as follows. In Section 2, I introduce the basic notation and review the IVQR and the LQTE model. In Section 3, I characterize the IVQR model under the LQTE assumptions. Section 4 generalizes these results to setups that allow for failures of the LQTE monotonicity assumption, non-binary instruments, and covariates. In Section 5, I illustrate the results in this paper using two empirical examples. Section 6 concludes. All proofs and additional results are collected in the appendix.

2 Setup and Models

The data consist of a random sample of N observations on a continuous outcome Y, a binary treatment D, and a binary instrument Z. In Section 4, I generalize this setup to incorporate multivalued instruments and covariates X. Throughout the paper, I assume that $Y|Z, D$ is absolutely continuous with respect to the Lebesgue measure on the support $\mathcal{Y}$. This technical assumption is made for expositional convenience. The analysis is developed in the potential outcomes framework (Rubin, 1974). Let Y_1 and Y_0 (indexed by D) denote the potential outcomes with and without the treatment. The fundamental problem of causal inference is that we only observe one potential outcome for each individual. Formally, the observed outcome is given by $Y = DY_1 + (1 - D)Y_0$. Similarly, let D_1 and D_0 (indexed by Z) denote the potential treatments that are related to the observed treatment as $D = ZD_1 + (1 - Z)D_0$. Based on the potential treatment status, the population can be categorized by four types, $\mathcal{T} \in \{a, n, c, f\}$, (Angrist et al., 1996):

Definition 1. *(a) Compliers ($\mathcal{T} = c$): the subpopulation with $D_1 = 1$ and $D_0 = 0$. (b) Always takers ($\mathcal{T} = a$): the subpopulation with $D_1 = D_0 = 1$. (c) Never takers ($\mathcal{T} = n$): the subpopulation with $D_1 = D_0 = 0$. (d) Defiers ($\mathcal{T} = f$): the subpopulation with $D_1 = 0$ and $D_0 = 1$*

where I use f instead of d to denote defiers to distinguish the type from the treatment status d. The objects of interest is the quantile function of Y_d, $Q_{Y_d}(\tau) = q(d, \tau)$, of the CDF $F_{Y_d}(y)$,

$$Q_{Y_d}(\tau) = \inf\{y : F_{Y_d}(y) \geq \tau\}.$$

If potential outcomes are continuous, we have $Q_{Y_d}(F_{Y_d}(y)) = y$ and $Q_{Y_d}(\tau) = F^{-1}_{Y_d}(\tau)$. The potential outcomes can be related to the structural quantile functions by the Skorohod representation of random variables.

$$Y_d = q(d, U_d), \text{ where } U_d \sim U(0, 1)$$

Similarly, observed outcomes can be expressed as $Y = q(d, U)$, where $Y \equiv Y_D$ and $U \equiv U_D$. This representation is essential for the IVQR model. I am also interested in the τ-QTE,

$$\delta(\tau) \equiv Q_{Y_1}(\tau) - Q_{Y_0}(\tau),$$

and in the related ATE, $\Delta \equiv \mathbb{E}(Y_1 - Y_0) = \int_0^1 \delta(\tau) d\tau$.

2.1 The IVQR Model

The IVQR model consists of the following main conditions.[2]

Assumption 1. *The following conditions hold jointly with probability one:*

1. *Monotonicity: $q(d, \tau)$ is strictly increasing in τ*
2. *Independence: For each d, U_d is independent of Z.*
3. *Selection: $D \equiv \rho(Z, V)$, where $\rho(\cdot)$ is an unknown function.*
4. *Rank similaritiy: Conditional on (Z, V), $\{U_d\}$ are identically distributed.*

[2]These conditions are taken from Chernozhukov and Hansen (2005) and Chernozhukov and Hansen (2013) with modifications.

Assumption 1.1 requires Y to be non-atomic conditional on Z.[3] The independence condition in Assumptions 1.2 and 1.3 states that potential outcomes are independent of the instrument. Note that this assumption is weaker than the conventional assumption that the disturbances in the outcome and the selection equation are jointly independent of Z. Assumption 1.4 is arguably the most important assumption of the IVQR model. It requires that individual ranks are constant across potential outcome distributions up to unsystematic deviations. Chernozhukov and Hansen (2005) present a more detailed discussion of the IVQR model.

The main statistical implication of Assumption 1 is the following nonlinear conditional moment restriction (Chernozhukov and Hansen, 2005, Theorem 1).

$$P\left(Y \leq q(D,\tau)|Z\right) = \tau \tag{1}$$

Under additional full rank and completeness conditions, the conditional moment restriction (1) point identifies the structural quantile function $q(D,\tau)$ (Chernozhukov and Hansen, 2005, 2013).[4] The conditional moment restriction (1) justifies the following unconditional moment equations for estimating $q(D,\tau)$.

$$\mathbb{E}\left(\left(\tau - 1\left[Y \leq q(D,\tau)\right]\right) Z\right) = 0, \tag{2}$$

where $1[\cdot]$ is the indicator function and Z is a vector of (transformations of) instruments. Estimation based on (2) is challenging because the sample analogue of the GMM objective function is non-smooth and non-convex. A non-exhaustive list of references that address this problem is given in the introduction.

In this paper, I consider the following specification for $q(D,\tau)$, which is fully nonparametric given that D is binary.

$$q(D,\tau) = D\delta(\tau) + Q_{Y_0}(\tau) \tag{3}$$

For linear-in-parameters models, which nest model (3), a computationally attractive approach for estimating $\delta(\tau)$ and $Q_{Y_0}(\tau)$ is the inverse quantile regression proposed by Chernozhukov and Hansen (2006). The idea behind this approach is that at the true coefficient on D, $\delta(\tau)$, the τ-quantile regression of $Y - D\delta(\tau)$ on a constant and Z would yield a zero coefficient on Z by

[3] Chernozhukov and Hansen (2013) discuss the IVQR model when Y has atoms conditional on Z, e.g., count or discrete variables.

[4] I further discuss the specific conditions for binary treatments and binary instruments as well as their relationship to the LQTE assumptions in Section 3.2 and in the appendix.

equation (1). This motivates a simple grid search algorithm over $\delta(\tau)$:

1. Define a grid $\{\delta_j, j = 1, ..., J\}$ and estimate the coefficients on the constant, $\hat{Q}_{Y_0}(\alpha_j, \tau)$, and on the instrument, $\hat{\gamma}(\alpha_j, \tau)$, using an ordinary τ-quantile regression of $Y - D\delta_j$ on a constant and the instrument Z.

2. Choose $\hat{\delta}(\tau)$ as the value in $\{\delta_j, j = 1, ..., J\}$ that minimizes $||\hat{\gamma}(\alpha_j, \tau)||$. The estimated coefficient on the constant is then given by $\hat{Q}_{Y_0}(\hat{\delta}(\tau), \tau)$.

Chernozhukov and Hansen (2006) prove consistency of the IVQR estimators obtained form inverse quantile regression and derive functional limit theory for the IVQR process. Moreover, they establish validity of subsampling for estimating the limiting law.

2.2 The LQTE Model

The LQTE model is based on the following set of assumptions.[5]

Assumption 2. *The following conditions hold jointly with probability one:*

1. *Monotonicity:* $P(D_1 \geq D_0) = 1$

2. *Independence:* (U_1, U_0, D_1, D_0) *is jointly independent of* Z

3. *Nontrivial assignment:* $0 < P(Z = 1) < 1$

4. *First-stage:* $P(D = 1|Z = 1) > P(D = 1|Z = 0)$

The monotonicity Assumption 2.1 rules out the presence of defiers. Consequently, always takers, never takers, and compliers exhaustively partition the whole population. The independence Assumption 2.2 states that both, potential outcomes and potential treatments, are independent of the instrument. Assumptions 2.3 and 2.4 require that the instrument assignment is non-trivial and that the instrument affects the treatment status.[6]

Under Assumption 2, the potential outcome quantile functions and the LQTE for the compliers, $Q_{Y_0|c}(\tau)$, $Q_{Y_1|c}(\tau)$, and $\delta_c(\tau) \equiv Q_{Y_1|c}(\tau) - Q_{Y_0|c}(\tau)$, are determined by the following weighted quantile regression objective function.

$$\left(Q_{Y_0|c}(\tau), \delta_c(\tau)\right) = \text{argmin}_{\left(Q_{Y_0|c}, \delta_c\right)} \mathbb{E}\left[\kappa \cdot \rho_\tau\left(Y - \delta_c D - Q_{Y_0|c}\right)\right], \tag{4}$$

[5] The assumptions are taken from Abadie et al. (2002) with modifications.

[6] Vytlacil (2002) shows that Assumption 2 is equivalent, to a class of latent index models, $D_z = 1[\nu(z) > V]$, where V is a scalar disturbance.

where $\rho_\tau(\cdot)$ is the usual check function and the weights κ are given by

$$\kappa = 1 - \frac{D(1-Z)}{1-P(Z=1)} - \frac{(1-D)Z}{P(Z=1)}.$$

Because κ is negative when $D \neq Z$, the sample counterpart of (4) is typically globally nonconvex. The circumvent this problem, Abadie et al. (2002) propose to modify the objective function by taking conditional expectation given (Y, D), which amounts to use a different set of nonnegative weights. Under appropriate regularity conditions (e.g., Abadie et al., 2002), the estimators based on the sample analogue of this weighted quantile regression objective function are consistent and asymptotically normal.

2.3 Comments on the Difference between both Models

Here I briefly summarize and highlight the most important differences between the IVQR and the LQTE model.

First, the models are based on two different and non-nested sets of assumptions.[7] The IVQR model relies on rank similarity in the outcome equation, whereas the LQTE model requires the selection equation to be weakly monotonic in a scalar disturbance. Yu (2014) shows that while allowing for a general selection equation, the rank similarity assumption of the IVQR model imposes strong restrictions on the treatment effect heterogeneity. In contrast, the LQTE model imposes restrictions on the selection equation but allows for essential treatment effect heterogeneity. Furthermore, the LQTE model relies on a stronger independence assumption than the IVQR model that requires not only potential outcomes but also potential treatments to be independent of the instrument.

Second, both models identify different quantities. The IVQR model identifies the QTE and the ATE for the whole population, arguably the more interesting objects than the LQTE and the LATE that are identified under the LQTE assumptions.

Third, the IVQR model accommodates arbitrary numbers and types of instruments and treatment variables. This sharply contrasts the LQTE model that has not been extended beyond the case of one binary instrument and one binary treatment variable (see e.g. the discussion in Imbens, 2007).

[7]This discussion draws from the exposition in Chernozhukov and Hansen (2013).

3 IVQR Estimands under the LQTE Framework

In this section, I study estimators based on the IVQR model under the LQTE assumptions.

3.1 Setup

Let $\delta^{IVQR}(\tau) \equiv Q^{IVQR}_{Y_1}(\tau) - Q^{IVQR}_{Y_0}(\tau)$, and $Q^{IVQR}_{Y_0}(\tau)$ denote the IVQR estimands that are given by the IVQR moment conditions.

$$\mathbb{E}\left(\left(\tau - 1\left[Y \leq \delta^{IVQR}(\tau)D + Q^{IVQR}_{Y_0}(\tau)\right]\right)(1, Z)'\right) = 0 \tag{5}$$

Under appropriate regularity conditions (e.g., Chernozhukov and Hansen, 2006), the IVQR estimators obtained from the inverse quantile regression described in Section 2.1 are consistent for $\delta^{IVQR}(\tau)$ and $Q^{IVQR}_{Y_0}(\tau)$. It should be noted that $\delta^{IVQR}(\tau)$ and $Q^{IVQR}_{Y_0}(\tau)$ can differ from the structural QTE and the structural quantile functions, $\delta(\tau)$ and $Q_{Y_0}(\tau)$, because I do not impose the IVQR assumptions in this section. To simplify the exposition, I impose the following common support assumption.

Assumption 3. *$Y_d|\mathcal{T} = t$ is continuously distributed with support $\mathcal{Y}$ for all $d \in \{0, 1\}$ and $t \in \{a, n, c, f\}$.*

Assumption 3 plays a similar role as the support assumptions in Athey and Imbens (2006) and is essential for point identification in the IVQR model as I discuss below.

3.2 Point Identification under the LQTE Assumptions

The conditional moment equations (1) do not point identify the IVQR estimands absent additional full rank and completeness conditions as pointed out in Section 2.1. In the appendix, I show in more detail that the LQTE assumptions and the common support assumption are sufficient for these additional assumptions. In particular, they imply the following monotone likelihood condition (Chernozhukov and Hansen, 2005, 2013).

$$\frac{f_{Y,D|Z}(y, D = 1|Z = 1)}{f_{Y,D|Z}(y, D = 0|Z = 1)} > \frac{f_{Y,D|Z}(y, D = 1|Z = 0)}{f_{Y,D|Z}(y, D = 0|Z = 0)}, \tag{6}$$

where $f_{Y,D|Z}(y, D = d, Z = z)$ is the joint probability density function of (Y, D) given Z. Condition (6) implies that the Jacobian of the moment equations (1) is of full rank, which is essential for point identification.

3.3 Cummulative Distribution Functions

Here I characterize the estimands of the potential outcome CDFs in the IVQR model under the LQTE assumptions. In subsequent sections, these results allow me to analyze the IVQR estimands of the QTE and the ATE under the LQTE assumptions.

For all types $t \in \{a, n, c, f\}$, let $F_{Y_d|t}(y) \equiv F_{Y_d|\mathcal{T}=t}(y)$, and $Q_{Y_d|t}(\tau) \equiv Q_{Y_d|\mathcal{T}=t}(\tau)$ denote the CDF and quantile functions associated with potential outcome Y_d and let $\pi_t \equiv P(\mathcal{T} = t)$ denote the proportion of type t. Under the LQTE assumptions, one can decompose the CDFs of Y_0 and Y_1 as

$$
\begin{aligned}
F_{Y_0}(y) =& \pi_a F_{Y_0|a}(y) + \pi_n F_{Y_0|n}(y) + \pi_c F_{Y_0|c}(y) \\
F_{Y_1}(y) =& \pi_a F_{Y_1|a}(y) + \pi_n F_{Y_1|n}(y) + \pi_c F_{Y_1|c}(y).
\end{aligned}
$$

Imbens and Rubin (1997) show that the following potential outcome distributions are identified from the data.

$$
\begin{aligned}
F_{Y_0|n}(y) =& F_{Y|D=0,Z=1}(y) \\
F_{Y_0|c}(y) =& \frac{p(0|0)F_{Y|D=0,Z=0}(y) - p(0|1)F_{Y|D=0,Z=1}(y)}{p(1|1) - p(1|0)} \\
F_{Y_1|a}(y) =& F_{Y|D=1,Z=0}(y) \\
F_{Y_1|c}(y) =& \frac{p(1|1)F_{Y|D=1,Z=1}(y) - p(1|0)F_{Y|D=1,Z=0}(y)}{p(1|1) - p(1|0)}
\end{aligned}
$$

where $F_{Y|D=d,Z=z}(y) \equiv F_{Y|D,Z}(y_d|D = d, Z = z)$ and $p(d|z) \equiv P(D = d|Z = z)$. However, $F_{Y_1|n}(y)$ and $F_{Y_0|a}(y)$, and consequently, $F_{Y_0}(y)$ and $F_{Y_1}(y)$, are unidentified under the LQTE assumptions. In contrast, these quantities are identified under the IVQR assumptions and can be computed by inverting the corresponding quantile functions. The key question is how the IVQR model imputes the unidentified quantities $F_{Y_1|n}(y)$ and $F_{Y_0|a}(y)$ using the rank similarity assumption. I answer this question in Theorem 1 by characterizing the IVQR estimands of the potential outcome CDFs, $F_{Y_1}^{IVQR}(y)$ and $F_{Y_0}^{IVQR}(y)$, under the LQTE assumptions.

Theorem 1. *Suppose that Assumptions 2 and 3 hold and that the IVQR estimands are given by (5). Then*

$$
\begin{aligned}
F_{Y_1}^{IVQR}(y) =& \pi_a F_{Y_1|a}(y) + \pi_c F_{Y_1|c}(y) + \pi_n F_{Y_0|n}\left(Q_{Y_0|c}\left(F_{Y_1|c}(y)\right)\right) \\
F_{Y_0}^{IVQR}(y) =& \pi_n F_{Y_0|n}(y) + \pi_c F_{Y_0|c}(y) + \pi_a F_{Y_1|a}\left(Q_{Y_1|c}\left(F_{Y_0|c}(y)\right)\right)
\end{aligned}
$$

The proof of Theorem 1 proceeds by equating the IVQR moment conditions after iterating expectations over Z and D. This yields a relationship between conditional CDFs of $Y|Z,D$ and the IVQR estimands. Under the LQTE assumptions these conditional CDFs correspond to CDFs for compliers, never takers, and always takers as discussed before.

Theorem 1 shows that the IVQR model imputes $F_{Y_1|n}(y)$ and $F_{Y_0|a}(y)$ as

$$
\begin{aligned}
F^{IVQR}_{Y_1|n}(y) &= F_{Y_0|n}\left(Q_{Y_0|c}\left(F_{Y_1|c}(y)\right)\right) \\
F^{IVQR}_{Y_0|a}(y) &= F_{Y_1|a}\left(Q_{Y_1|c}\left(F_{Y_0|c}(y)\right)\right).
\end{aligned}
$$

These formulas reveal a close analogy between the IVQR and the CIC model (Athey and Imbens, 2006).[8] This analogy relates to the fact that both models rely on conditions that restrict the evolution of individual ranks in the outcome distributions. In the IVQR model it is assumed individual ranks that are constant (up to random deviations) accross treatment states, whereas the CIC model requires individual ranks to be constant across time periods. Moreover, Theorem 1 highlights the importance of the common support assumption (Assumption 3) that guarantees that $F^{IVQR}_{Y_1|n}(y)$ and $F^{IVQR}_{Y_0|a}(y)$ are well-defined.

3.4 Quantile Treatment Effects

Here I derive an explicit relationship between the QTE estimands of the IVQR model and the LQTE. Based on Theorem 1, the QTE estimands in the IVQR model can be expressed as LQTE for the compliers evaluated at transformed quantile levels.

Theorem 2. *Suppose that Assumptions 2 and 3 hold and that the IVQR estimands are given by (5). Then*

$$
\begin{aligned}
\delta^{IVQR}(\tau) &= \delta_c\left(F_{Y_0|c}\left(Q^{IVQR}_{Y_0}(\tau)\right)\right) = \delta_c\left(F_{Y_1|c}\left(Q^{IVQR}_{Y_1}(\tau)\right)\right) \\
\delta^{IVQR}_n(\tau) &= \delta_c\left(F_{Y_0|c}\left(Q_{Y_0|n}(\tau)\right)\right) \\
\delta^{IVQR}_a(\tau) &= \delta_c\left(F_{Y_1|c}\left(Q_{Y_1|a}(\tau)\right)\right)
\end{aligned}
$$

where $\delta^{IVQR}_n(\tau) \equiv Q^{IVQR}_{Y_1|n}(\tau) - Q_{Y_0|n}(\tau)$ *and* $\delta^{IVQR}_a(\tau) \equiv Q_{Y_1|a}(\tau) - Q^{IVQR}_{Y_0|a}(\tau)$.

The results in Theorem 2 are implications from the relationship between the IVQR estimands and their counterparts for the compliers established in Theorem 1.

Theorem 2 shows that the IVQR estimands of the QTE can be expressed as LQTE for the compliers at transformed quantile levels. The transformation adjusts for the respective difference to

[8] The formulas also exhibit similarities with the counterfactual mapping by Vuong and Xu (2014), who show that under rank invariance or rank preservation (instead of rank similarity) and monotonicity the individual treatment effects can be identified using this mapping.

the complier distributions as measured by the rank functions $F_{Y_0|c}\left(Q_{Y_0}^{IVQR}(\cdot)\right)$, $F_{Y_1|c}\left(Q_{Y_1}^{IVQR}(\cdot)\right)$, $F_{Y_0|c}\left(Q_{Y_0|n}(\cdot)\right)$, and $F_{Y_1|c}\left(Q_{Y_1|a}(\cdot)\right)$. Moreover, it follows from Theorem 1 that, ceteris paribus, the absolute value of the difference between the arguments at which the functions $\delta^{IVQR}(\cdot)$ and $\delta_c(\cdot)$ are evaluated decreases with the strength of the instrument as measured by the first stage, $p(1|1) - p(1|0)$, which equals the fraction of the compliers.[9] Consequently, the difference between the QTE estimates based on the IVQR model and the LQTE is uniquely determined by differences between the potential outcome distributions that are identified under the LQTE assumptions and the strength of the instrument. This result has two important implications. First, comparisons of estimates between both models do not provide robustness or reality checks for the IVQR model and the rank similarity assumption, and second, the sensitivity of the IVQR estimates to deviations from the rank similarity assumption is decreasing in the strength of the instrument.

Theorem 2 has three corollaries that further describe the relationship between the IVQR estimates and the LQTE for the compliers.

Corollary 1. *Suppose that Assumptions 2 and 3 hold and that the IVQR estimands are given by (5). Then: (i) If $\delta_c(\tau) \geq 0$ for all $\tau \in (0,1)$, then $\delta^{IVQR}(\tau) \geq 0$ for all $\tau \in (0,1)$. (ii) If $\delta_c(\tau) \leq 0$ for all $\tau \in (0,1)$, then $\delta^{IVQR}(\tau) \leq 0$ for all $\tau \in (0,1)$.*

Corollary 1 shows that the sign of the IVQR estimates corresponds to the sign of the LQTE estimates, whenever the sign of the LQTE does not change as a function of the quantile level. Now suppose that the LQTE is constant across quantile levels, i.e. $\delta_c(\tau) = \delta_c$, as in a location model, $Q_{Y_d|c}(\tau) = (d,1)(\delta_c,\beta)' + Q_{U_d|c}(\tau)$. Corollary 2 shows that if the LQTE is constant, the estimates of both models are equivalent.

Corollary 2. *Suppose that Assumptions 2 and 3 hold, that the IVQR estimands are given by (5), and that $\delta_c(\tau) = \delta_c$ for all $\tau \in (0,1)$. Then $\delta^{IVQR}(\tau) = \delta_c(\tau) = \delta_c$ for all $\tau \in (0,1)$.*

Furthermore, if the LQTE is monotonically increasing or decreasing in the quantile level as in a location location-scale model, $Q_{Y_d|c}(\tau) = (d,1)(\delta_c,\beta)' + (d,1)(\gamma_1,\gamma_2)' \cdot Q_{U_d|c}(\tau)$, where $\delta_c(\tau) = \delta_c + \gamma_1 \cdot Q_{U_d|c}(\tau)$, such monotonicity is preserved by the IVQR estimates.

Corollary 3. *Suppose that Assumptions 2 and 3 hold, that the IVQR estimands are given by (5), and that $\delta_c(\tau)$ monotonically increasing (decreasing) in τ. Then $\delta^{IVQR}(\tau)$ is monotonically increasing (decreasing) in τ.*

[9] However, this does not necessarily imply that the discrepancy between $\delta^{IVQR}(\tau)$ and $\delta_c(\tau)$ is reduced for a given quantile level τ.

In other words, if the treatment increases (decreases) the variance of the potential outcomes for the compliers, the IVQR estimates yield the same conclusion.

Taken together, Theorem 2 and Corollaries 1, 2, and 3 show that there is a close relationship between the estimates of the IVQR and the LQTE model. Because the analysis does not rely on the validity of the rank similarity assumption, these results shed new light on the interpretation of the IVQR estimates when this assumption fails. Moreover, these results can be interpreted as robustness properties of the IVQR model to deviations from the underlying assumptions. In particular, the model captures a causal effect and inherits desirable properties of the LQTE estimates irrespective of the validity of the rank similarity assumption.

3.5 Average Treatment Effects

Based on the results obtained in the previous sections, the IVQR estimand of the ATE, $\Delta^{IVQR} \equiv \int_0^1 \delta^{IVQR}(\tau)d\tau$, can be expressed as a convex combination of the LATE, Δ_c, and weighted averages of LQTE for the compliers.

Theorem 3. *Suppose that Assumptions 2 and 3 hold and that the IVQR estimands are given by (5). Then*

$$\Delta^{IVQR} = \pi_c\Delta_c + \pi_a\Delta_a^{IVQR} + \pi_n\Delta_n^{IVQR}$$

where

$$\Delta_n^{IVQR} \equiv \int_{0,1} \delta_c\left(F_{Y_0|c}\left(Q_{Y_0|n}(\tau)\right)\right)d\tau$$
$$\Delta_a^{IVQR} \equiv \int_{0,1} \delta_c\left(F_{Y_1|c}\left(Q_{Y_1|a}(\tau)\right)\right)d\tau$$

Theorem 3 shows that the IVQR model estimates a weighted average of the standard LATE for the compliers and imputed ATE for the never takers and always takers that are both averages of rank function adjusted LQTE for the compliers. As in Theorem 2, differences between the estimates of both models are determined by differences between the potential outcome distributions of compliers, always takers, and never takers as well as the relative size of these three subpopulations, which is determined by the strength of the instrument. The IVQR estimate of the ATE inherits the properties of the LQTE outlined in Corollaries 1 and 2. If the LQTE is positive (negative) for all τ, then the sign of IVQR estimand equals the sign of the LATE. Moreover, the IVQR estimate is equivalent to the LATE if the LQTE is constant as in a location model.

An interesting implication of Theorems 1, 2, and 3 is that the IVQR model estimates treatment effects by extrapolating from the compliers to the whole population. Theorefore, the IVQR model can alternatively be considered as an approach to extrapolation in the LATE framework based on restrictions of the evolution of individual ranks.

4 Generalizations

4.1 LQTE Assumptions without Monotonicity

The monotonicity assumption of the LQTE model is not innocuous and may be questionable in many contexts; see e.g. the examples discussed by De Chaisemartin (2014b). Here I analyze the IVQR estimands under the LQTE assumptions without monotonicity. Throughout this section, I maintain Assumptions 2.2-2.4. It should be noted that independence between the potential treatments (D_0, D_1) and the instrument Z is not required in the IVQR model but aids the exposition and the interpretation of the results. Theorem 4 characterizes the IVQR estimates of the potential outcome CDFs absent the LQTE monotonicity assumption.

Theorem 4. *Suppose that Assumptions 2.2-2.4 and 3 hold, that the IVQR estimands are given by (5), and that*

$$\begin{aligned} F_{Y_1|c-f}(y) \equiv& \frac{\pi_c F_{Y_1|c}(y) - \pi_f F_{Y_1|f}(y)}{\pi_c - \pi_f} \\ F_{Y_0|c-f}(y) \equiv& \frac{\pi_c F_{Y_0|c}(y) - \pi_f F_{Y_0|f}(y)}{\pi_c - \pi_f} \end{aligned}$$

are well-defined and strictly increasing CDFs. Then

$$\begin{aligned} F_{Y_1}^{IVQR}(y) =& \pi_a F_{Y_1|a}(y) + \pi_c F_{Y_1|c}(y) + \pi_n F_{Y_0|n}\left(Q_{Y_0|c-f}\left(F_{Y_1|c-f}(y)\right)\right) \\ &+ \pi_f F_{Y_0|f}\left(Q_{Y_0|c-f}\left(F_{Y_1|c-f}(y)\right)\right) \\ F_{Y_0}^{IVQR}(y) =& \pi_n F_{Y_0|n}(y) + \pi_c F_{Y_0|c}(y) + \pi_a F_{Y_1|a}\left(Q_{Y_1|c-f}\left(F_{Y_0|c-f}(y)\right)\right) \\ &+ \pi_f F_{Y_1|f}\left(Q_{Y_1|c-f}\left(F_{Y_0|c-f}(y)\right)\right) \end{aligned}$$

where $Q_{Y_1|c-f}(\tau) \equiv F^{-1}_{Y_1|c-f}(\tau)$ and $Q_{Y_0|c-f}(\tau) \equiv F^{-1}_{Y_0|c-f}(\tau)$.

Theorem 4 shows that the IVQR model imputes the (mixtures of) distributions that are not directly identified, $\pi_n F_{Y_1|n}(y) + \pi_f F_{Y_1|f}(y)$ and $\pi_a F_{Y_0|a}(y) + \pi_f F_{Y_0|f}(y)$, using CIC-type arguments. The key distribution is $F_{Y_d|c-f}(y)$, a weighted difference between the distributions of compliers and defiers. Under the assumptions of Theorem 4, this is the only subpopulation for which both potential outcome distributions are identified. Under the compliers-defiers assumption (Assumption S2) of De Chaisemartin (2014a,b), $F_{Y_d|c-f}(y)$ corresponds to the CDF for the well-defined subpopulation of the compliers referred to as the comvivors.[10]

It should be noted that contrarily to the analysis under monotonicity, Assumptions 2.2-2.4 and 3 do not imply that $F_{Y_0|c-f}(y)$ and $F_{Y_1|c-f}(y)$ are well-defined and strictly increasing CDFs. As suggested by the discussion in Section 3.2 and in the appendix, the assumption that $F_{Y_0|c-f}(y)$ and $F_{Y_1|c-f}(y)$ are well-defined and strictly increasing is closely related to the conditions that are required for point identification in the IVQR model.

The next theorem characterizes the QTE estimands in the IVQR model under the LQTE framework without monotonicity.

Theorem 5. *Suppose that Assumptions 2.2-2.4 and 3 hold, that the IVQR estimands are given by (5), and that $F_{Y_1|c-f}(y)$ and $F_{Y_0|c-f}(y)$ are well-defined and strictly increasing. Then*

$$\delta^{IVQR}(\tau) = \delta_{c-f}\left(F_{Y_0|c-f}\left(Q_{Y_0}^{IVQR}(\tau)\right)\right) = \delta_{c-f}\left(F_{Y_1|c-f}\left(Q_{Y_1}^{IVQR}(\tau)\right)\right)$$

where $\delta_{c-f}(\tau) \equiv Q_{Y_1|c-f}(\tau) - Q_{Y_0|c-f}(\tau)$.

Theorem 5 shows that the IVQR estimands can be expressed as QTE for the compliers-defiers mixture population at transformed quantile levels. In contrast to before, this transformation additionally takes into account differences between the compliers-defiers mixture population and the defiers.

4.2 Multivalued Instruments

Suppose that instead of being binary, the instrument Z takes values in a discrete set $\mathcal{Z} = \{z_1, z_2, ..., z_K\}$ with $0 \leq z_1 < z_2 < ... < z_K$. The following assumption extends the LQTE model defined by Assumption 2 to the case with multivalued instruments.

Assumption 4. *The following conditions hold jointly with probability one:*

1. *Monotonicity:* $P(D_z \geq D_{z'}) = 1$ *for any values* $z, z' \in \mathcal{Z}$*, where* $z > z'$.

2. *Independence:* $(U_1, U_0, \{D_z\}_{z\in\mathcal{Z}})$ *is jointly independent of* Z

3. *Nontrivial assignment:* $0 < P(Z = z) < 1$ *for all* $z \in \mathcal{Z}$

[10] De Chaisemartin (2014a,b) splits the compliers in two groups: the comvivors (c_V) and the comfiers (c_f). He assumes that $P(\mathcal{T} = c_F) = P(\mathcal{T} = f)$ and $Y_d|\mathcal{T} = c_F \sim Y_d|\mathcal{T} = f$. Under these assumptions, the comfiers cancel with the compliers and $F_{Y_d|c-f}(y) = F_{Y_d|c_V}(y)$.

4. *First-stage:* $P(D=1|Z=z) > P(D=1|Z=z')$ *for any two values* $z, z' \in \mathcal{Z}$*, where* $z > z'$.

The corresponding common support assumption reads:

Assumption 5. $Y_d|\mathcal{T} = t$ *is continuously distributed with support* $\mathcal{Y}$ *for all* $d \in \{0,1\}$ *and* $t \in \{a, n, \{c_{z_j}\}_{j=2}^{K}\}$.

Under Assumption 4, there are $K+1$ types that are characterized by a unique value of the instrument, $z_k \in \mathcal{Z}$, where the treatment for that type switches from zero to one (Imbens, 2007). I denote individuals with $D_z = 0$ for all $z \in \mathcal{Z}$ as never takers and individuals with $D_z = 1$ for all $z \in \mathcal{Z}$ as always takers. In addition, there are now $K-1$ different types of compliers that are indexed by instrument value where their treatment status switches from zero to one. I denote compliers who switch between z_{j-1} and z_j by $\mathcal{T} = c_{z_j}$ for $j = 2, ..., K$. Furthermore, for all types $t \in \{a, n, \{c_{z_j}\}_{j=2}^{K}\}$ let $F_{Y_d|t}(y)$ denote the CDF of Y_d and let π_t denote the proportion of type t.

Under Assumption 4, the data generating process is informative about the fraction of all subpopulations as well as the distributions of Y_1 for always takers and compliers and the distribution of Y_0 for never takers and compliers (Imbens, 2007). In particular, the conditional probabilities $p(d|z_k) \equiv P(D = d|Z = z_k)$ can be related to the proportions of types as

$$
\begin{aligned}
p(1|z_k) =& \pi_a + \sum_{j=2}^{K} \pi_{c_{z_j}} 1(z_j \leq z_k) \\
p(0|z_k) =& \pi_n + \sum_{j=2}^{K} \pi_{c_{z_j}} 1(z_j > z_k).
\end{aligned}
$$

Moreover, the observed conditional CDFs, $F_{Y|D=d,Z=z_k}(y)$, can be related to the potential outcome CDFs as

$$
\begin{aligned}
F_{Y|D=1,Z=z_k}(y) &= \frac{\pi_a F_{Y_1|a}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y) 1(z_j \leq z_k)}{\pi_a + \sum_{j=2}^{K} \pi_{c_{z_j}} 1(z_j \leq z_k)} \\
F_{Y|D=0,Z=z_k}(y) &= \frac{\pi_n F_{Y_0|n}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y) 1(z_j > z_k)}{\pi_n + \sum_{j=2}^{K} \pi_{c_{z_j}} 1(z_j > z_k)}.
\end{aligned}
$$

Imbens (2007) presents an illuminating discussion of the simple example with $K = 3$.

I consider IVQR estimands $\delta^{IVQR}(\tau)$ and $Q_{Y_0}^{IVQR}(y)$ that are given by the following moment equations,

$$
\mathbb{E}\left(\left(\tau - 1\left[Y \leq \delta^{IVQR}(\tau) D + Q_{Y_0}^{IVQR}(\tau)\right]\right)(1, Z)'\right) = 0. \tag{7}
$$

It should be noted that the unconditional moment equations (7) are not the only possible moment restrictions that can be used to estimate $\delta^{IVQR}(\tau)$ and $Q_{Y_0}^{IVQR}(y)$ because the conditional moment restriction (1) gives rise to infinitely many unconditional moment restrictions. Theorem 6 generalizes Theorem 1 to multivalued instruments.

Theorem 6. *Suppose that Assumptions 4 and 5 hold and that the IVQR estimands are given by (7). Then*

$$
\begin{aligned}
F_{Y_1}^{IVQR}(y) =& \pi_a F_{Y_1|a}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y) P(Z \geq z_j) \\
& + \pi_n F_{Y_0|n}\left(\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)\right) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}\left(\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)\right) P(Z < z_j) \\
F_{Y_0}^{IVQR}(y) =& \pi_a F_{Y_1|a}\left(\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)\right) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}}\left(\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)\right) \frac{\sum_{k=j}^{K} z_k P(Z = z_k)}{\mathbb{E}(Z)} \\
& + \pi_n F_{Y_0|n}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y) \frac{\sum_{k=1}^{j-1} z_k P(Z = z_k)}{\mathbb{E}(Z)}
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{F}_{Y_1}(y) \equiv& \frac{\sum_{j=2}^{K} w_j \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y)}{\sum_{j=2}^{K} w_j \pi_{c_{z_j}}} \\
\tilde{F}_{Y_0}(y) \equiv& \frac{\sum_{j=2}^{K} w_j \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y)}{\sum_{j=2}^{K} w_j \pi_{c_{z_j}}}
\end{aligned}
$$

with $w_j = \left(\frac{\mathbb{E}(Z|Z \geq z_j)}{\mathbb{E}(Z)} - 1\right) P(Z \geq z_j)$, $\tilde{Q}_{Y_1}(\tau) \equiv \tilde{F}_{Y_1}^{-1}(\tau)$, *and* $\tilde{Q}_{Y_0}(\tau) \equiv \tilde{F}_{Y_0}^{-1}(\tau)$.

Theorem 6 shows that the basic mechanism described in Theorem 1 pertains when Z is multivalued. The key functions, $\tilde{F}_{Y_1}(y)$ and $\tilde{F}_{Y_0}(y)$ are convex combinations of CDFs for the now $K-1$ different compliers. The respective weight of $F_{Y_d|c_{z_j}}(y)$ is determined by two components: the weighting function w_j and the size of the respective compliant subpopulation, $\pi_{c_{z_j}}$. Note that the weights w_j are strictly positive because $\mathbb{E}(Z|Z \geq z_j) > \mathbb{E}(Z)$ for $j > 1$. Moreover, consider the difference between two adjacent weights:

$$
w_{j+1} - w_j = \left(1 - \frac{z_j}{\mathbb{E}(Z)}\right) P(Z = z_j)
$$

Hence, the weighting function is increasing for z_j is smaller than the mean and decreasing for z_j larger than the mean such that complier who switch their treatment status at an instrument value z_j near the mean receive more weight than those who switch further away from the mean. It is interesting to note that weighting function w_j bears similarities to the weights that linear IV gives to the LATE for compliers $\mathcal{T} = c_{z_j}$ (e.g., Heckman and Vytlacil, 2005, 2007).

Differently to the before, the IVQR estimands use the generalized rank functions $\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(\cdot)\right)$ and $\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(\cdot)\right)$ not only for imputing the distributions of Y_1 for the always takers and Y_0 for the never takers but also for imputing the distributions of Y_1 and Y_0 for mixtures of all compliant subpopulations.

Theorem 7 characterizes the IVQR estimands of the QTE under the generalized LQTE model.

Theorem 7. *Suppose that Assumptions 4 and 5 hold and that the IVQR estimands are given by (7). Then*

$$\delta^{IVQR}(\tau) = \tilde{\delta}\left(\tilde{F}_{Y_0}\left(Q_{Y_0}^{IVQR}(\tau)\right)\right) = \tilde{\delta}\left(\tilde{F}_{Y_1}\left(Q_{Y_1}^{IVQR}(\tau)\right)\right)$$

where $\tilde{\delta}(\tau) \equiv \tilde{Q}_{Y_1}(\tau) - \tilde{Q}_{Y_0}(\tau)$.

Under the generalized LQTE assumptions, the transformation of the quantile level adjusts for differences between the potential outcome distributions of the average compliant subpopulation and never takers, always takers, and the $K-1$ types compliers.

In the appendix (Lemma 2), I present closed form solutions for the IVQR estimands that accommodate arbitrary instruments and/or transformations of instruments $g(Z)$, where $g(\cdot)$ is a measurable function. These results can be used to extend the analysis in this section in many interesting directions.

4.3 Covariates

Including a covariates X into the analysis can be desirable for at least three reasons. First, conditioning on a set of covariates may be important to achieve rank similarity as pointed out by Chernozhukov and Hansen (2005). Second, the instrument may only be valid conditional on appropriate covariates. For example, Chernozhukov and Hansen (2004) assume that 401(k) eligibility is exogenous conditional on income (and further covariates). Third, even if the instrument and the rank similarity assumption are unconditionally valid, it might be interesting to consider conditional QTE (see e.g., Firpo (2007) or Frölich and Melly (2013) for a discussion of the differences between conditional and unconditional QTE).

When covariates are discrete, the previous analysis remains valid conditionally and the analysis can proceed in subsamples defined by $X = x$. Alternatively, one can consider fully saturated models for the conditional quantiles. When X contains continuous elements, the fully saturated approach is obviously not feasible. In this case, it is common to work with linear-in-parameters IVQR and LQTE models (e.g., Abadie et al., 2002; Chernozhukov and Hansen, 2006). Such models can be interpreted as sieve-approximations to the true potentially nonlinear conditional quantile function of $Y_d|X$ and $Y_d|X, \mathcal{T} = c$.[11] Because the results obtained in the previous sections are fully nonparametric, they can be expected to hold approximately. This approximation can be improved by choosing richer specifications (e.g. through interactions, polynomials, or B-splines).

5 Empirical Applications

In this section, I present two empirical applications. The goal here is to compare the QTE and ATE estimates of both models and to shed new light on the similarities and differences between both approaches and, in particular, on the role of the rank similarity assumption underlying the IVQR model.

5.1 Implementation

Here I briefly discuss the estimation and inference methods used in the empirical applications. The quantile functions in the IVQR model can be estimated using the inverse quantile regression procedure outlined in Section 2.1. The corresponding CDFs can be obtained by inverting the quantile functions. The quantile functions in the LQTE model can be estimated from the weighted quantile regression described in Section 2.2. The CDFs for the compliers can be obtained by either inverting the quantile functions or directly using the sample analogues of $F_{Y_1|c}(y)$ and $F_{Y_0|c}(y)$:

$$\hat{F}_{Y_0|c}(y) = \frac{\hat{p}(0|0)\hat{F}_{Y|D=0,Z=0}(y) - \hat{p}(0|1)\hat{F}_{Y|D=0,Z=1}(y)}{\hat{p}(1|1) - \hat{p}(1|0)}$$
$$\hat{F}_{Y_1|c}(y) = \frac{\hat{p}(1|1)\hat{F}_{Y|D=1,Z=1}(y) - \hat{p}(1|0)\hat{F}_{Y|D=1,Z=0}(y)}{\hat{p}(1|1) - \hat{p}(1|0)}$$

Neither of these estimators explicitly imposes monotonicity of the quantile and distribution functions. Nonmonotonicity can arise due either failures of the underlying assumptions or sampling variation.[12] I address this problem by rearranging the original estimates as proposed by Chernozhukov et al. (2010). The rearrangement procedure is easy to implement and has a number of desirable properties (Chernozhukov et al., 2010).[13]

[11] Sieve-estimation of the IVQR model is considered by Su and Hosino (2013). The importance of a flexible specification is highlighted for example by Yu (2014, Section 6.2).

[12] Formal tests for the validity of the underlying assumptions that take sampling variation into account have been proposed for example by Chernozhukov et al. (2010) and Kitagawa (2014) for the LQTE model.

[13] Chernozhukov et al. (2010) show that rearranged curve is closer to the original curve in finite samples and that rearrangement outperforms isotonization procedure in a simulation study designed to match closely the second empirical application. The generic results given by Chernozhukov et al. (2010) provide functional central limit theory for the rearranged quantile functions, given that a functional central limit theorem applies to the original estimators. This is the case for the IVQR estimators as discussed in Section 2.1. Moreover, the functional delta method for the bootstrap

To estimate the average effects in the IVQR model, I use the following trimmed versions of the formulas in Theorem 3 to avoid the estimation of tail quantiles:

$$\hat{\Delta}^{IVQR} \equiv \int_{\varepsilon}^{1-\varepsilon} \hat{\delta}^{IVQR(\tau)} d\tau$$
$$\hat{\Delta}_n^{IVQR} \equiv \int_{\varepsilon}^{1-\varepsilon} \hat{\delta}_c \left(\hat{F}_{Y_0|c} \left(\hat{Q}_{Y_0|n}(\tau)\right)\right) d\tau$$
$$\hat{\Delta}_a^{IVQR} \equiv \int_{\varepsilon}^{1-\varepsilon} \hat{\delta}_c \left(\hat{F}_{Y_1|c} \left(\hat{Q}_{Y_1|a}(\tau)\right)\right) d\tau$$

for some small constant $\varepsilon > 0$.[14] Moreover, I directly estimate the LATE, $\hat{\Delta}_c$, using two-stage least squares instead of averaging over the LQTE. Simulation exercises suggest that trimming combined with direct estimation of the LATE can substantially improve the numerical accuracy of the estimates in samples of moderate size.

5.2 JTPA

I consider estimation of the causal effect of JTPA training programs on subsequent earnings. I use the same data set as Abadie et al. (2002) restricting the analysis to the subsample of men. As described for example in Bloom et al. (1997) and Abadie et al. (2002), the JTPA was a largely publicly-funded federal training program that started in October 1983 and lasted up until the late 1990's. An important part of the JTPA were training programs for the economically disadvantaged (classroom training, on-the-job training, job search assistance, etc.). The JTPA also included a mandate for a large-scale randomized training evaluation study that collected data from about 20,000 participants in 16 different sites. Because the assignment (Z) was randomized, it can be used as an instrument for estimating the causal effect of actual participation in training programs of the sum of earnings in the 30 month after the random assignment (Y) without further conditioning. About 38% of the men in the sample, who received a training offer, chose not to participate in the training program. Only about 1% of the individuals participated in the program despite the fact that they did not receive an offer, implying that Z satisfies one-sided non-compliance almost perfectly. For the purpose of illustration, I drop the observations violating this condition from the sample, which yields a total number of observations of 5,083. Abadie et al. (2002) give additional information about the dataset and presented descriptive statistics.

One-sided non-compliance rules out the existence of both, always takers and defiers, such that

implies implies validity of the bootstrap for estimating the limiting law of the rearranged curve (Chernozhukov et al., 2010).

[14] In the apllications, I use $\varepsilon = 0.01$.

the results in Theorem 1 simplify to

$$\begin{aligned}F_{Y_1}^{IVQR}(y) =&\pi_c F_{Y_1|c}(y) + \pi_n F_{Y_0|n}\left(Q_{Y_0|c}\left(F_{Y_1|c}(y)\right)\right)\\ F_{Y_0}^{IVQR}(y) =&\pi_c F_{Y_0|c}(y) + \pi_n F_{Y_0|n}(y).\end{aligned}$$

Consequently, the IVQR estimand $F_{Y_0}^{IVQR}(y)$ equals the true potential outcome CDF, $F_{Y_0}(y)$, irrespectively of the validity of the rank similarity assumption so long as the LQTE assumptions hold.

Figure 1 plots potential outcome CDFs for compliers and never takers (panel A) and compares QTE estimates from the IVQR and the LQTE model (panel B). The estimates from the IVQR model are obtained from an inverse quantile regression with a grid search over $\{-2500, -2495, ..., 15000\}$ and the LQTE are estimated by inverting rearranged versions of the estimated potential outcome CDFs for the compliers.

[Figure 1 about here.]

Panel A shows that the distributions of earnings absent the treatment, Y_0, for compliers and never takers exhibit substantial differences at the lower quantiles. In contrast, the differences between Y_0 and Y_1 for the compliers are generally of smaller magnitude and more pronounced at the upper quantiles. Panel B compares the QTE estimated by both models. The results show a number of interesting features. First, both model yield qualitatively and quantitatively similar results. Second, both models indicate substantial heterogeneity in the effect of the training programs at different quantiles of the earnings distribution. Third, the QTE estimates are overall increasing in the quantile level ranging from values close or below zero up to 6000 USD.

In Figure 2, I further explore the determinants of the similarities of difference between the estimates of both models.

[Figure 2 about here.]

Panel A plots the rank function $F_{Y_0|c}\left(Q_{Y_0|n}(\cdot)\right)$ as the relevant measure of the difference between $F_{Y_0|c}(y)$ and $F_{Y_0|n}(y)$. The pronounced differences at the lower quantiles in 1 (panel A) translates in the difference between the rank function and the 45-degree line. Moreover, I plot the IVQR estimate of the QTE for the never takers against the LQTE in panel B. Although qualitatively similar, the IVQR model estimates the QTE for the never takers to be smaller than the LQTE at most quantiles. The reason for this finding is the combination of the the increasing LQTE combined with the shape of the rank function that suggests that the τ-QTE for the never taker corresponds to the τ'-LQTE with $\tau' > \tau$. Taken together, Figures 1 and 2 suggest that the small differences between the estimates of both models can be attributed to the strength of the instrument (first stage:

$p(1|1) - p(1|0) = 0.68$), which outweighs the difference between the distributions of Y_0 between never takers and compliers.

Finally, I use the results from Theorem 3 to decompose the IVQR estimand of the ATE into a convex combination of the LATE for the compliers and the IVQR estimate of the ATE for the never takers, which corresponds to a weighted average of rank function adjusted LQTE.

$$\underbrace{\Delta^{IVQR}}_{1697.02} = \underbrace{\pi_c}_{0.68} \cdot \underbrace{\Delta_c}_{1715.956} + \underbrace{\pi_n}_{0.32} \cdot \underbrace{\Delta_n^{IVQR}}_{1665.76}$$

The LATE estimate is slightly larger than the IVQR estimates for the overall ATE and the ATE for the never takers. The reason for this finding is displayed in Figure 2 (panel B) showing that the QTE for the never takers lie below the LQTE at most of the quantiles.

5.3 Veteran Status and Earnings

I consider estimation of the causal effects of Vietnam veteran status D on the distribution of annual labor earnings, Y. Because veteran status is likely to be endogenous, I follow Angrist (1990) and use the U.S. draft lottery as an instrument Z that takes the value one if someone was eligible for draft and zero otherwise. I use the same dataset as Abadie (2002) and Chernozhukov et al. (2010). The dataset contains information about 11,637 white men, born in 1950–1953, from the Current Population Surveys of 1979, and 1981–1985; 2461 are Vietnam veterans and 3234 are eligible for military service. In total, there are 18% always takers, 71% never takers, and 11% compliers. Abadie (2002) gives more information on the dataset.

Figure 3 (panel A) shows that there is a pronounced difference between the distributions of Y_0 for compliers and never takers at the lower quantiles but essentially no difference at higher quantiles as well as between the distributions of Y_1 for always takers and compliers. In Figure 3 (panel B) compares the QTE estimates from the IVQR model to the LQTE estimates. The IVQR estimates are computed using a grid search over a fine grid of $\{-10000, -9995, ..., 5000\}$ and the LQTE estimates are obtained from inverting the corresponding CDFs. Some features of the results deserve comments. Unlike in the JTPA example, there are quantitatively substantive differences between the estimates of both models at the quantiles below the median and, in particular, at the lowest quantiles. Yet, both models qualitatively point at a pronounced treatment effect heterogeneity. In particular, both models yield large negative effects at the lower quantiles of the wage distribution and small positive impacts at higher quantiles.

[Figure 3 about here.]

Figure 4 sheds some light on determinants of the these results. Panel A plots the rank functions function that measure the difference between the CDFs in displayed in Figure 3 (panel A). The substantive difference between compliers and never takers combined with the substantial effect heterogeneity in the LQTE translates into a large differences in between the IVQR estimand of the QTE for never takers and the corresponding LQTE as shown in panel B. Together with the relatively large fraction of never takers, this finding explains the differences between the estimates of both models in Figure 3 (panel B).

[Figure 4 about here.]

Finally, I decompose the IVQR estimand of the ATE into a convex combination of the LATE and the IVQR estimates for never takers and always takers.

$$\underbrace{\Delta^{IVQR}}_{-1458.32} = \underbrace{\pi_c}_{0.11} \cdot \underbrace{\Delta_c}_{-1277.78} + \underbrace{\pi_n}_{0.71} \cdot \underbrace{\Delta_n^{IVQR}}_{-1378.42} + \underbrace{\pi_a}_{0.18} \cdot \underbrace{\Delta_a^{IVQR}}_{-1930.12}$$

The LATE is smaller (in absolute values) than the ATE estimated by the IVQR model. Figure 3 suggests that the results are mainly driven by the large differences at the lower tail quantiles.

6 Conclusion

In this paper, I characterize the treatment effects estimators based on the IVQR model under the assumptions of the LQTE model. I show that the IVQR model estimates LQTE at transformed quantile levels. The transformation adjusts for differences between the subpopulation-specific distributions of Y_1 and Y_0 that are identified in the LQTE model and the relative size of these subpopulations. Moreover, the IVQR estimand of the ATE can be expressed as a convex combination of the LATE and weighted averages of LQTE at transformed quantile levels. Consequently, differences between the estimates of both models are uniquely determined by two factors: the differences between the subpopulation-specific potential outcome distributions and the relative size of these subpopulations, which depends on the strength of the instrument. The analysis is generalized to incorporate failures of the LQTE monotonicity assumption, non-binary instruments, and covariates. I illustrate the results with two empirical applications.

I conclude by summarizing the main implications of this paper for applied empirical research. First, comparisons of both models are fundamentally uninformative about the validity of the rank similarity assumption. Second, the sensitivity of the IVQR estimates to deviations from the rank similarity assumption is a decreasing function of the fraction of compliers. Third, even in the absence of the underlying assumptions, the IVQR model captures particular causal effects so long as the LQTE assumptions hold. Forth, the IVQR estimates have a number of desirable properties under the LQTE assumptions. In particular, the estimates preserve sign and monotonicity of the LQTE estimates whenever these properties are invariant across quantile levels.

References

Abadie, A., 2002. Bootstrap tests for distributional treatment effects in instrumental variable models. Journal of the American Statistical Association 97 (457), pp. 284–292.

Abadie, A., 2003. Semiparametric instrumental variable estimation of treatment response models. Journal of Econometrics 11, pp. 231–263.

Abadie, A., Angrist, J., Imbens, G., 2002. Instrumental variable estimates of the effect of subsidized training on the qquantile of trainee earnings. Econometrica 70 (1), pp. 91–117.

Angrist, J., Fernandez-Val, I., 2013. ExtrapoLATE - ing: External validity and overidentification in the LATE framework. In: Acemoglu, D., Arellano, M., Dekel, E. (Eds.), Advances in Economics and Econometrics. Cambridge University Press.

Angrist, J. D., 1990. Lifetime earnings and the vietnam era draft lottery: Evidence from social security administrative records. The American Economic Review 80 (3), pp. 313–336.

Angrist, J. D., 2004. Treatment effect heterogeneity in theory and practice. The Economic Journal 114 (494), pp. 52–C83.

Angrist, J. D., Imbens, G. W., Rubin, D. B., 1996. Identification of causal effects using instrumental variables. Journal of the American Statistical Association 91 (434), pp. 444–455.

Athey, S., Imbens, G. W., 2006. Identification and inference in nonlinear difference-in-differences models. Econometrica 74 (2), pp. 431–497.

Balke, A., Pearl, J., 1997. Bounds on treatment effects from studies with imperfect compliance. Journal of the American Statistical Association 92 (439), pp. 1171–1176.

Bloom, H. S., Orr, L. L., Bell, S. H., Cave, G., Doolittle, F., Lin, W., Bos, J. M., 1997. The benefits and costs of JTPA title II-A programs: Key findings from the national job training partnership act study. The Journal of Human Resources 32 (3), pp. 549–576.

Chamberlain, G., 2011. Bayesian Aspects of Treatment Choice. Oxford University Press, New York, pp. pp. 11–39.

Chernozhukov, V., Fernandez-Val, I., Galichon, A., 2010. Quantile and probability curves without crossing. Econometrica 78 (3), pp. 1093–1125.

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

Chernozhukov, V., Hansen, C., 2005. An IV model of quantile treatment effects. Econometrica 73 (1), pp. 245–261.

Chernozhukov, V., Hansen, C., 2006. Instrumental quantile regression inference for structural and treatment effects models. Journal of Econometrics 132, pp. 491–525.

Chernozhukov, V., Hansen, C., 2008. Instrumental variable quantile regression: A robust inference approach. Journal of Econometrics 142 (1), pp. 379–398.

Chernozhukov, V., Hansen, C., 2013. Quantile models with endogeneity. Annual Review of Economics 5 (1), pp. 57–81.

Chernozhukov, V., Hansen, C., Jansson, M., 2007a. Inference approaches for instrumental variable quantile regression. Economics Letters 95 (2), pp. 272–277.

Chernozhukov, V., Hansen, C., Jansson, M., 2009. Finite sample inference for quantile regression models. Journal of Econometrics 152 (2), pp. 93–103.

Chernozhukov, V., Hong, H., 2003. An {MCMC} approach to classical estimation. Journal of Econometrics 115 (2), pp. 293–346.

Chernozhukov, V., Imbens, G. W., Newey, W. K., 2007b. Instrumental variable estimation of nonseparable models. Journal of Econometrics 139 (1), pp. 4–14.

De Chaisemartin, C., 2014a. Supplementary material to: "tolerating defiance: Local average treatment effects without monotonicity". Unpublished Manuscript.

De Chaisemartin, C., 2014b. Tolerating defiance: Local average treatment effects without monotonicity. Unpublished Manuscript.

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

Frandsen, B. R., Frölich, M., Melly, B., 2012. Quantile treatment effects in the regression discontinuity design. Journal of Econometrics 168 (2), 382 – 395.

Frölich, M., Melly, B., 2013. Identification of treatment effects on the treated with one-sided noncompliance. Econometric Reviews 32 (3), pp. 384–414.

Frölich, M., Melly, B., 2013. Unconditional quantile treatment effects under endogeneity. Journal of Business & Economic Statistics 31 (3), pp. 346–357.

Gagliardini, P., Scaillet, O., 2012. Nonparametric instrumental variable estimation of structural quantile effects. Econometrica 80 (4), pp. 1533–1562.

Heckman, J., Tobias, J. L., Vytlacil, E., 2001. Four parameters of interest in the evaluation of social programs. Southern Economic Journal 68 (2), pp. 210–223.

Heckman, J., Tobias, J. L., Vytlacil, E., 2003. Simple estimators for treatment parameters in a latent-variable framework. The Review of Economics and Statistics 85 (3), pp. 748–755.

Heckman, J. J., Vytlacil, E., 2005. Structural equations, treatment effects, and econometric policy evaluation. Econometrica 73 (3), pp. 669–738.

Heckman, J. J., Vytlacil, E. J., 2007. Chapter 71 econometric evaluation of social programs, part ii: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. Vol. 6, Part B of Handbook of Econometrics. Elsevier, pp. 4875 – 5143.

Horowitz, J. L., Lee, S., 2007. Nonparametric instrumental variables estimation of a quantile regression model. Econometrica 75 (4), pp. 1191–1208.

Huber, M., 2014. Sensitivity checks for the local average treatment effect. Economics Letters 123, pp. 220–223.

Imbens, G. W., 2007. Nonadditive models with endogenous regressors. In: Blundell, R., Newey, W., Persson, T. (Eds.), Advances in Economics and Econometrics. Vol. 3. Cambridge University Press, pp. 17–46, cambridge Books Online.

Imbens, G. W., Angrist, J. D., 1994. Identification and estimation of local average treatment effects. Econometrica 62 (2), pp. 467–475.

Imbens, G. W., Rubin, D. B., 1997. Estimating outcome distributions for compliers in instrumental variables models. The Review of Economic Studies 64 (4), pp. 555–574.

Kitagawa, T., 2014. A test for instrument validity. Cenmap Working Paper.

Koenker, R., Bassett, Gilbert, J., 1978. Regression quantiles. Econometrica 46 (1), pp. 33–50.

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

Su, L., Hosino, T., 2013. Sieve instrumental variable quantile regression estimation of functional coefficient models. Unpublished Manuscript.

Vuong, Q., Xu, H., 2014. Counterfactual mapping and individual treatment effects in nonseparable models with discrete endogeneity. Unpublished Manuscript.

Vytlacil, E., 2002. Independence, monotonicity, and latent index models: An equivalence result. Econometrica 70 (1), pp. 331–341.

Yu, P., 2014. Marginal quantile treatment effect and counterfactual analysis. Unpublished Manuscript.

Point Identification under the LQTE Assumption

The moment conditions (5) point identify the IVQR estimands under additional full rank conditions. Here I show LQTE assumptions are sufficient for these full rank conditions. The results in this section are related to Yu (2014) as I discuss in further detail at the end of this section.

To state the result, some additional notation is required. I keep the notation as close as possible to Chernozhukov and Hansen (2005) and Chernozhukov and Hansen (2013). Define $q(d) \equiv Q^{IVQR}_{Y_d}(\tau)$. The function $q(\cdot)$ can be represented by a vector of its values $q = (q(0), q(1))'$. For a vector $y = (y_0, y_1)'$, define the vector of moment equations $\Pi(y)$ as

$$\Pi(y) \equiv (P(Y \leq y_D|Z=0) - \tau, P(Y \leq y_D|Z=1) - \tau)'$$

where $y_D = (1-D) \cdot y_0 + D \cdot y_1$. q is said to be identified in some parameters space, $\mathcal{L}$, if $y = q$ is the only solution to $\Pi(y) = 0$ among all $y \in \mathcal{L}$. Here, I show that the LQTE assumptions imply the simple sufficient condition for Theorem 2 in Chernozhukov and Hansen (2013).[15] These conditions are (Chernozhukov and Hansen, 2013, Comment 3.1 and Appendix A.4.):

$$\frac{f_{Y,D|Z}(y, D=1|Z=1)}{f_{Y,D|Z}(y, D=0|Z=1)} > \frac{f_{Y,D|Z}(y, D=1|Z=0)}{f_{Y,D|Z}(y, D=0|Z=0)} \text{ for all } y = (y_0, y_1) \in \mathcal{L}, \tag{8}$$

and

$$f_{Y,D|Z}(y, D=1|Z=1) > 0, \;\; f_{Y,D|Z}(y, D=0|Z=0) > 0 \text{ for all } y = (y_0, y_1) \in \mathcal{L}. \tag{9}$$

where the parameter space $\mathcal{L}$ is defined as either $\mathcal{L} = q + C$ (a cube centered at q), or $\mathcal{L} = (q + C) \cap H$ (the intersection of that cube with the halfspace H).

I now show that Assumptions 2 and 3 imply conditions (8) and (9). First, note that Assumption 2 implies

$$\begin{aligned}
f_{Y,D|Z}(y, D=1|Z=1) - f_{Y,D|Z}(y, D=1|Z=0) &= \pi_c f_{Y_1|c}(y) \\
f_{Y,D|Z}(y, D=0|Z=0) - f_{Y,D|Z}(y, D=0|Z=1) &= \pi_c f_{Y_0|c}(y)
\end{aligned}$$

where $f_{Y_d|c}(y)$ denotes the probability densitiy function of Y_d for the compliers. Consequently, $\pi_c f_{Y_1|c}(y)$ and $\pi_c f_{Y_0|c}(y)$ must be nonnegative. This observations dates back to Imbens and Rubin (1997) and can be used to construct tests for the LQTE assumptions (e.g., Kitagawa, 2014). By

[15] Theorem 2 in Chernozhukov and Hansen (2013) is a generalization of the corresponding Theorem 2 in Chernozhukov and Hansen (2005).

Assumptions 2 and 3, $\pi_c f_{Y_1|c}(y) > 0$ and $\pi_c f_{Y_0|c}(y) > 0$ for all $y \in \mathcal{Y}$. Hence,

$$f_{Y,D|Z}(y, D = 1|Z = 1) > f_{Y,D|Z}(y, D = 1|Z = 0)$$
$$f_{Y,D|Z}(y, D = 0|Z = 0) > f_{Y,D|Z}(y, D = 0|Z = 1)$$

Combining both inequalities implies condition (8). Condition (9) follows directly from Assumptions 2 and 3.

The result here is related to the result in Section 6.1 of Yu (2014), who shows that in the MTE framework, the sufficient condition (8) is fulfilled if $p(1|1) > p(1|0)$ (which corresponds to Assumption 4.4 and generalizes Assumption 2.4 in my paper). This assumption implies the existence of compliers, which is the essential element for the arguments of this section.

Auxilliary Results

This section contains two auxiliary lemmas that provide closed form solutions for the IVQR estimands with binary and general instruments.

Lemma 1. *Suppose that* $0 < P(Z = 1) < 1$ *and that*

$$\tilde{F}_{Y_1}(y) \equiv \frac{p(1|1)F_{Y|D=1,Z=1}(y) - p(1|0)F_{Y|D=1,Z=0}(y)}{p(1|1) - p(1|0)}$$
$$\tilde{F}_{Y_0}(y) \equiv \frac{p(0|0)F_{Y|D=0,Z=0}(y) - p(0|1)F_{Y|D=0,Z=1}(y)}{p(1|1) - p(1|0)}$$

are strictly increasing and well-defined CDFs. Then: (i)

$$Q^{IVQR}_{Y_1}\left(F^{IVQR}_{Y_0}(y)\right) = \tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)$$
$$Q^{IVQR}_{Y_0}\left(F^{IVQR}_{Y_1}(y)\right) = \tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)$$

and (ii)

$$F^{IVQR}_{Y_1}(y) = (p(1|1) - p(1|0))\tilde{F}_{Y_1}(y) + p(1|0)F_{Y|D=1,Z=0}(y) + p(0|1)F_{Y|D=0,Z=1}\left(\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)\right)$$
$$F^{IVQR}_{Y_0}(y) = (p(1|1) - p(1|0))\tilde{F}_{Y_0}(y) + p(0|1)F_{Y|D=0,Z=1}(y) + p(1|0)F_{Y|D=1,Z=0}\left(\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)\right)$$

where $\tilde{Q}_{Y_0}(y) = \tilde{F}^{-1}_{Y_0}(y)$ *and* $\tilde{Q}_{Y_1}(y) = \tilde{F}^{-1}_{Y_1}(y)$*.*

Lemma 1 can be deduced from the more general result in Lemma 2, but it is instructive to give a direct proof to illustrate the mechanics behind the main results.

Proof of Lemma 1.
Part (i) The moment conditions of the IVQR model read:

$$\mathbb{E}\left(\left(\tau - 1\left[Y \leq Q^{IVQR}_{Y_D}(\tau)\right]\right)1\right) = 0$$
$$\mathbb{E}\left(\left(\tau - 1\left[Y \leq Q^{IVQR}_{Y_D}(\tau)\right]\right)Z\right) = 0$$

By the law of iterated expectations,

$$\begin{aligned}
\mathbb{E}\left(\left(\tau - 1\left[Y \leq Q^{IVQR}_{Y_D}(\tau)\right]\right)1\right) =& \mathbb{E}\left(\tau - 1\left[Y \leq Q^{IVQR}_{Y_D}(\tau)\right]|Z=1\right)P(Z=1) \\
&+ \mathbb{E}\left(\tau - 1\left[Y \leq Q^{IVQR}_{Y_D}(\tau)\right]|Z=0\right)P(Z=0) = 0 \\
\mathbb{E}\left(\left(\tau - 1\left[Y \leq Q^{IVQR}_{Y_D}(\tau)\right]\right)Z\right) =& \mathbb{E}\left(\tau - 1\left[Y \leq Q^{IVQR}_{Y_D}(\tau)\right]|Z=1\right)P(Z=1) = 0.
\end{aligned}$$

Because $0 < P(Z = 1) < 1$ by assumption, $Q^{IVQR}_{Y_D}$ solves

$$\begin{aligned}
\mathbb{E}\left(\tau - 1\left[Y \leq Q^{IVQR}_{Y_D}(\tau)\right]|Z=0\right) =& 0 \\
\mathbb{E}\left(\tau - 1\left[Y \leq Q^{IVQR}_{Y_D}(\tau)\right]|Z=1\right) =& 0.
\end{aligned}$$

By the law of iterated expectation,

$$\mathbb{E}\left(\tau - 1\left[Y \leq Q_{Y_1}^{IVQR}(\tau)\right] | D = 1, Z = 1\right) p(1|1) + \mathbb{E}\left(\tau - 1\left[Y \leq Q_{Y_0}^{IVQR}(\tau)\right] | D = 0, Z = 1\right) p(0|1) = 0$$
$$\mathbb{E}\left(\tau - 1\left[Y \leq Q_{Y_1}^{IVQR}(\tau)\right] | D = 1, Z = 0\right) p(1|0) + \mathbb{E}\left(\tau - 1\left[Y \leq Q_{Y_0}^{IVQR}(\tau)\right] | D = 0, Z = 0\right) p(0|0) = 0.$$

Rewriting both equations using the definition of conditional CDFs yields

$$p(1|1)F_{Y|D=1,Z=1}\left(Q_{Y_1}^{IVQR}(\tau)\right) + p(0|1)F_{Y|D=0,Z=1}\left(Q_{Y_0}^{IVQR}(\tau)\right) = \tau \tag{10}$$
$$p(1|0)F_{Y|D=1,Z=0}\left(Q_{Y_1}^{IVQR}(\tau)\right) + p(0|0)F_{Y|D=0,Z=0}\left(Q_{Y_0}^{IVQR}(\tau)\right) = \tau. \tag{11}$$

Equating equations (10) and (11) and rearranging terms gives

$$p(1|1)F_{Y|D=1,Z=1}\left(Q_{Y_1}^{IVQR}(\tau)\right) - p(1|0)F_{Y|D=1,Z=0}\left(Q_{Y_1}^{IVQR}(\tau)\right) =$$
$$p(0|0)F_{Y|D=0,Z=0}\left(Q_{Y_0}^{IVQR}(\tau)\right) - p(0|1)F_{Y|D=0,Z=1}\left(Q_{Y_0}^{IVQR}(\tau)\right).$$

Dividing by $p(1|1) - p(1|0)$ on both sides, we have

$$\tilde{F}_{Y_1}\left(Q_{Y_1}^{IVQR}(\tau)\right) = \tilde{F}_{Y_0}\left(Q_{Y_0}^{IVQR}(\tau)\right)$$

by definition of $\tilde{F}_{Y_d}(y)$ and $\tilde{Q}_{Y_d}(\tau)$. Because $\tilde{F}_{Y_1}(y)$ and $\tilde{F}_{Y_0}(y)$ are strictly increasing and well-defined, we can apply $\tilde{Q}_{Y_1}(y) \equiv \tilde{F}_{Y_1}^{-1}(y)$ on both sides:

$$Q_{Y_1}^{IVQR}(\tau) = \tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}\left(Q_{Y_0}^{IVQR}(\tau)\right)\right)$$

Finally, note that $F_{Y_0}^{IVQR}(y)$ is strictly increasing by our assumption that $\tilde{F}_{Y_1}(y)$ and $\tilde{F}_{Y_0}(y)$ are strictly increasing and well-defined. Hence, we can substitute $\tau = F_{Y_0}^{IVQR}(y)$ to obtain

$$Q_{Y_1}^{IVQR}\left(F_{Y_0}^{IVQR}(y)\right) = \tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)$$

which implies that $Q_{Y_0}^{IVQR}\left(F_{Y_1}^{IVQR}(y)\right) = \tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)$. This completes the proof of part (i).
Part (ii): Consider equation (10). Substituting $\tau = F_{Y_1}^{IVQR}(y)$ and and adding and subtracting $p(1|0)F_{Y|D=1,Z=0}(y)$ yields:

$$F_{Y_1}^{IVQR}(y) = (p(1|1) - p(1|0))\tilde{F}_{Y_1}(y) + p(1|0)F_{Y|D=1,Z=0}(y) + p(0|1)F_{Y|D=0,Z=1}\left(Q_{Y_0}^{IVQR}\left(F_{Y_1}^{IVQR}(y)\right)\right)$$

Similar arguments applied to equation (11) yield:

$$F_{Y_0}^{IVQR}(y) = (p(1|1) - p(1|0))\tilde{F}_{Y_0}(y) + p(0|1)F_{Y|D=0,Z=1}(y) + p(1|0)F_{Y|D=1,Z=0}\left(Q_{Y_1}^{IVQR}\left(F_{Y_0}^{IVQR}(y)\right)\right)$$

The result in part (ii) of the Lemma now follows from the results in part (i). □

I now present general closed form solutions for the IVQR moment conditions that can be used to characterize the IVQR model under more general setups than those considered in the main text. I consider the following IVQR moment equations,

$$\mathbb{E}\left(\left(\tau - 1\left[Y \leq \delta^{IVQR}(\tau)D + Q_{Y_0}^{IVQR}(\tau)\right]\right) g(Z)\right) = 0,$$

where Z is a general instrument and $g(Z) \equiv (g_1(Z), g_2(Z))'$ is a measurable function, where $g_0(Z)$ and $g_1(Z)$ are linearly independent and $\mathbb{E}(g(Z)) \neq 0$. Let $f_Z(z)$ be the density function of Z if Z

is continuous and $f_Z(z) = P(Z = z)$ if Z is discrete.

Lemma 2. *Suppose that $f_Z(z) > 0$ for all $z \in \mathcal{Z}$ and that*

$$\tilde{F}_{Y_1}(y) \equiv \frac{\mathbb{E}\left(F_{Y|D=1,Z}(y)\, p(1|Z)\left(\frac{g_2(Z)}{\mathbb{E}(g_2(Z))} - \frac{g_1(Z)}{\mathbb{E}(g_1(Z))}\right)\right)}{\mathbb{E}\left(p(1|Z)\left(\frac{g_2(Z)}{\mathbb{E}(g_2(Z))} - \frac{g_1(Z)}{\mathbb{E}(g_1(Z))}\right)\right)}$$
$$\tilde{F}_{Y_0}(y) \equiv \frac{\mathbb{E}\left(F_{Y|D=0,Z}(y)\, p(0|Z)\left(\frac{g_1(Z)}{\mathbb{E}(g_1(Z))} - \frac{g_2(Z)}{\mathbb{E}(g_2(Z))}\right)\right)}{\mathbb{E}\left(p(1|Z)\left(\frac{g_2(Z)}{\mathbb{E}(g_2(Z))} - \frac{g_1(Z)}{\mathbb{E}(g_1(Z))}\right)\right)}$$

are strictly increasing and well-defined. Then: (i)

$$Q_{Y_1}^{IVQR}\left(F_{Y_0}^{IVQR}(y)\right) = \tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)$$
$$Q_{Y_0}^{IVQR}\left(F_{Y_1}^{IVQR}(y)\right) = \tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)$$

and (ii)

$$F_{Y_1}^{IVQR}(y) = \frac{\mathbb{E}\left(g_1(Z)\left(F_{Y|D=1,Z}(y)\, p(1|Z) + F_{Y|D=0,Z}\left(\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)\right) p(0|Z)\right)\right)}{\mathbb{E}(g_1(Z))}$$
$$F_{Y_0}^{IVQR}(y) = \frac{\mathbb{E}\left(g_2(Z)\left(F_{Y|D=1,Z}\left(\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)\right) p(1|Z) + F_{Y|D=0,Z}(y)\, p(0|Z)\right)\right)}{\mathbb{E}(g_2(Z))}$$

where $\tilde{Q}_{Y_0}(y) = \tilde{F}_{Y_0}^{-1}(y)$ and $\tilde{Q}_{Y_1}(y) = \tilde{F}_{Y_1}^{-1}(y)$.

Proof of Lemma 2.
Part (i): The moment equations of the IVQR model read

$$\mathbb{E}\left(\left(\tau - 1\left[Y \leq Q_{Y_D}^{IVQR}(\tau)\right]\right) g_1(Z)\right) = 0$$
$$\mathbb{E}\left(\left(\tau - 1\left[Y \leq Q_{Y_D}^{IVQR}(\tau)\right]\right) g_2(Z)\right) = 0.$$

By the law of iterated expectations,

$$\mathbb{E}\left(\mathbb{E}\left(\left(\tau - 1\left[Y \leq Q_{Y_D}^{IVQR}(\tau)\right]\right) g_1(Z)|Z\right)\right)$$
$$\mathbb{E}\left(\mathbb{E}\left(\left(\tau - 1\left[Y \leq Q_{Y_D}^{IVQR}(\tau)\right]\right) g_2(Z)|Z\right)\right)$$

Expressing both equations in terms of conditional CDFs, we have

$$\mathbb{E}\left(g_1(Z) F_{Y|Z}\left(Q_{Y_D}^{IVQR}(\tau)\right)\right) = \mathbb{E}(g_1(Z))\tau$$
$$\mathbb{E}\left(g_2(Z) F_{Y|Z}\left(Q_{Y_D}^{IVQR}(\tau)\right)\right) = \mathbb{E}(g_2(Z))\tau$$

By the law of iterated expectations, we get

$$\mathbb{E}\left(g_1(Z)\left(F_{Y|D=1,Z}\left(Q_{Y_1}^{IVQR}(\tau)\right) p(1|Z) + F_{Y|D=0,Z}\left(Q_{Y_0}^{IVQR}(\tau)\right) p(0|Z)\right)\right) = \mathbb{E}(g_1(Z))\tau \quad (12)$$
$$\mathbb{E}\left(g_2(Z)\left(F_{Y|D=1,Z}\left(Q_{Y_1}^{IVQR}(\tau)\right) p(1|Z) + F_{Y|D=0,Z}\left(Q_{Y_0}^{IVQR}(\tau)\right) p(0|Z)\right)\right) = \mathbb{E}(g_2(Z))\tau \quad (13)$$

Equating both equations and rearranging terms, we obtain

$$\mathbb{E}\left(F_{Y|D=1,Z}\left(Q_{Y_1}^{IVQR}(\tau)\right)p(1|Z)\left(\frac{g_2(Z)}{\mathbb{E}(g_2(Z))}-\frac{g_1(Z)}{\mathbb{E}(g_1(Z))}\right)\right) =$$
$$\mathbb{E}\left(F_{Y|D=0,Z}\left(Q_{Y_0}^{IVQR}(\tau)\right)p(0|Z)\left(\frac{g_1(Z)}{\mathbb{E}(g_1(Z))}-\frac{g_2(Z)}{\mathbb{E}(g_2(Z))}\right)\right)$$

Dividing both equations by $\mathbb{E}\left(p(1|Z)\left(\frac{g_2(Z)}{\mathbb{E}(g_2(Z))}-\frac{g_1(Z)}{\mathbb{E}(g_1(Z))}\right)\right)$ (which is non-zero by assumption) yields

$$\frac{\mathbb{E}\left(F_{Y|D=1,Z}\left(Q_{Y_1}^{IVQR}(\tau)\right)p(1|Z)\left(\frac{g_2(Z)}{\mathbb{E}(g_2(Z))}-\frac{g_1(Z)}{\mathbb{E}(g_1(Z))}\right)\right)}{\mathbb{E}\left(p(1|Z)\left(\frac{g_2(Z)}{\mathbb{E}(g_2(Z))}-\frac{g_1(Z)}{\mathbb{E}(g_1(Z))}\right)\right)} =$$
$$\frac{\mathbb{E}\left(F_{Y|D=0,Z}\left(Q_{Y_0}^{IVQR}(\tau)\right)p(0|Z)\left(\frac{g_1(Z)}{\mathbb{E}(g_1(Z))}-\frac{g_2(Z)}{\mathbb{E}(g_2(Z))}\right)\right)}{\mathbb{E}\left(p(1|Z)\left(\frac{g_2(Z)}{\mathbb{E}(g_2(Z))}-\frac{g_1(Z)}{\mathbb{E}(g_1(Z))}\right)\right)}$$

The result in part (i) now follows from the same arguments as in the proof of part (i) of Lemma 1.
Part (ii): The proof of part (ii) follows directly from the result in part (i) after substituting $\tau = F_{Y_1}^{IVQR}(y)$ in equation (12) and $F_{Y_0}^{IVQR}(y)$ in equation (13). □

Proofs of the Main Results

Proof of Theorem 1.

By Assumption 2, we have that $0 < P(Z = 1) < 1$. Moreover, Assumptions 2 and 3 imply that $\tilde{F}_{Y_1}(y) = F_{Y_1|c}(y)$ and $\tilde{F}_{Y_0}(y) = F_{Y_0|c}(y)$ are both well-defined and strictly increasing. Hence, by Lemma by Lemma 1,

$$F_{Y_1}^{IVQR}(y) = (p(1|1) - p(1|0))\tilde{F}_{Y_1}(y) + p(1|0)F_{Y|D=1,Z=0}(y) + p(0|1)F_{Y|D=0,Z=1}\left(\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)\right)$$
$$F_{Y_0}^{IVQR}(y) = (p(1|1) - p(1|0))\tilde{F}_{Y_0}(y) + p(0|1)F_{Y|D=0,Z=1}(y) + p(1|0)F_{Y|D=1,Z=0}\left(\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)\right)$$

The result in Theorem 1 now follows from the results in Imbens and Rubin (1997) discussed at the beginning of Section 3 in the main text. □

Proof of Theorem 2.

By Assumption 2, we have that $0 < P(Z = 1) < 1$. Moreover, Assumptions 2 and 3 imply that $\tilde{F}_{Y_1}(y) = F_{Y_1|c}(y)$ and $\tilde{F}_{Y_0}(y) = F_{Y_0|c}(y)$ are both well-defined and strictly increasing. Hence, the results of Lemma 1 apply. To prove the first claim, note that under Assumption 2, part (i) of Lemma 1 implies $F_{Y_1|c}\left(Q_{Y_1}^{IVQR}(\tau)\right) = F_{Y_0|c}\left(Q_{Y_0}^{IVQR}(\tau)\right)$. Applying $Q_{Y_1|c}(\tau)$ on both sides yields $Q_{Y_1}^{IVQR}(\tau) = Q_{Y_1|c}\left(F_{Y_0|c}\left(Q_{Y_0}^{IVQR}(\tau)\right)\right)$ (because $Q_{Y_1|c}(\tau)$ is strictly increasing in τ). Next, consider

$$\begin{aligned}
\delta^{IVQR}(\tau) &\equiv Q_{Y_1}^{IVQR}(\tau) - Q_{Y_0}^{IVQR}(\tau) \\
&= Q_{Y_1|c}\left(F_{Y_0|c}\left(Q_{Y_0}^{IVQR}(\tau)\right)\right) - Q_{Y_0}^{IVQR}(\tau) \\
&= Q_{Y_1|c}\left(F_{Y_0|c}\left(Q_{Y_0}^{IVQR}(\tau)\right)\right) - Q_{Y_0|c}\left(F_{Y_0|c}\left(Q_{Y_0}^{IVQR}(\tau)\right)\right) \\
&= \delta_c\left(F_{Y_0|c}\left(Q_{Y_0}^{IVQR}(\tau)\right)\right) \\
&= \delta_c\left(F_{Y_1|c}\left(Q_{Y_1}^{IVQR}(\tau)\right)\right),
\end{aligned}$$

where the third equality follows from $Q_{Y_0|c}\left(F_{Y_0|c}(y)\right) = y$ because $Q_{Y_0|c}$ and $F_{Y_0|c}(y)$ are strictly increasing; the forth equality is by definition; and the fifth equality follows directly from $F_{Y_1|c}\left(Q_{Y_1}^{IVQR}(\tau)\right) = F_{Y_0|c}\left(Q_{Y_0}^{IVQR}(\tau)\right)$. This proves the first claim. To prove the second and the third claim, note that

$$F_{Y_1|n}^{IVQR}(y) = F_{Y_0|n}\left(Q_{Y_0|c}\left(F_{Y_1|c}(y)\right)\right), \text{ and } F_{Y_0|a}^{IVQR}(y) = F_{Y_1|a}\left(Q_{Y_1|c}\left(F_{Y_0|c}(y)\right)\right)$$

implies that

$$Q_{Y_1|n}^{IVQR}(\tau) = Q_{Y_1|c}\left(F_{Y_0|c}\left(Q_{Y_0|n}(\tau)\right)\right), \text{ and } Q_{Y_0|a}^{IVQR}(\tau) = Q_{Y_0|c}\left(F_{Y_1|c}\left(Q_{Y_1|a}(\tau)\right)\right).$$

The proof of both claims now follows from similar arguments as the proof of the first claim. □

Proof of Theorem 3.

The result in Theorem 3 follows immediately from Theorems 1 and 2 and the relationship between the QTEs and the ATE: $\Delta = \int_0^1 \delta(\tau) d\tau$. □

Proof of Theorem 4.

Under Assumptions 2.2 - 2.4, $p(d|z)$ and $F_{Y|D=d,Z=z}(y)$ can be related to the fractions and potential outcome CDFs of the four different types $T \in \{a, n, c, f\}$ as (e.g., Huber, 2014):

$$
\begin{aligned}
p(1|1) =& \pi_a + \pi_c \\
p(1|0) =& \pi_a + \pi_f \\
p(0|1) =& \pi_n + \pi_f \\
p(0|0) =& \pi_n + \pi_n \\
F_{Y|D=1,Z=1}(y) =& \frac{\pi_a}{\pi_a + \pi_c} F_{Y_1|a}(y) + \frac{\pi_c}{\pi_a + \pi_c} F_{Y_1|c}(y) \\
F_{Y|D=1,Z=0}(y) =& \frac{\pi_a}{\pi_a + \pi_f} F_{Y_1|a}(y) + \frac{\pi_f}{\pi_a + \pi_f} F_{Y_1|f}(y) \\
F_{Y|D=0,Z=1}(y) =& \frac{\pi_n}{\pi_n + \pi_f} F_{Y_0|n}(y) + \frac{\pi_f}{\pi_n + \pi_f} F_{Y_0|f}(y) \\
F_{Y|D=0,Z=0}(y) =& \frac{\pi_n}{\pi_n + \pi_c} F_{Y_0|n}(y) + \frac{\pi_c}{\pi_n + \pi_c} F_{Y_0|c}(y)
\end{aligned}
$$

By Assumption 2.3, we have that $0 < P(Z = 1) < 1$. Moreover, by assumption $F_{Y_1|c-f}(y)$ and $F_{Y_0|c-f}(y)$ are strictly increasing and well-defined. Thus, the results from Lemma 1 apply. Substituting the above expressions in part (ii) of Lemma 1 completes the proof. □

Proof of Theorem 5.

By Assumption 2.3, we have that $0 < P(Z = 1) < 1$. Moreover, by assumption $F_{Y_1|c-f}(y)$ and $F_{Y_0|c-f}(y)$ are strictly increasing and well-defined. Thus, the results from Lemma 1 apply. Under Assumption 2.2 - 2.4, part (i) of Lemma 1 implies $F_{Y_1|c-f}\left(Q^{IVQR}_{Y_1}(\tau)\right) = F_{Y_0|c-f}\left(Q^{IVQR}_{Y_0}(\tau)\right)$. The proof of the theorem now follows from similar arguments as in the proof of Theorem 2. □

Proof of Theorem 6.

By Assumption 4, we have that $0 < P(Z = z) < 1$ for all $z \in \mathcal{Z}$. Hence, to apply the results in Lemma 2, one needs to show that $\tilde{F}_{Y_1}(y)$ and $\tilde{F}_{Y_0}(y)$ are strictly increasing and well-defined. It is helpful to split the proof into four steps. In step 1, I present the implications of Assumption 4 for observed conditional probabilities and CDFs. In step 2, I use these implications to express $\tilde{F}_{Y_1}(y)$ and $\tilde{F}_{Y_0}(y)$ as functions of potential outcome CDFs for the $K + 1$ subpopulation. Step 3 verifies that $\tilde{F}_{Y_1}(y)$ and $\tilde{F}_{Y_0}(y)$ (and hence their inverses) are well-defined and strictly increasing. Finally, step 4 uses the Lemma 2 to prove the results in the theorem.

Step 1: Recall that Assumption 4 implies that the conditional CDFs $F_{Y|D=d,Z=z_k}$ can be written as the sum of the potential outcome distributions of different types:

$$
\begin{aligned}
F_{Y|D=1,Z=z_k}(y) &= \frac{\pi_a F_{Y_1|a}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y) 1(z_j \leq z_k)}{\pi_a + \sum_{j=2}^{K} \pi_{c_{z_j}} 1(z_j \leq z_k)} \\
F_{Y|D=0,Z=z_k}(y) &= \frac{\pi_n F_{Y_0|n}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y) 1(z_j > z_k)}{\pi_n + \sum_{j=2}^{K} \pi_{c_{z_j}} 1(z_j > z_k)}
\end{aligned}
$$

Moreover, the conditional probabilities $p(1|z_k)$ and $p(0|z_k)$ can be expressed as sums of type-specific fractions:

$$
\begin{aligned}
p(1|z_k) =& \pi_a + \sum_{j=2}^{K} \pi_{c_{z_j}} 1(z_j \leq z_k) \\
p(0|z_k) =& \pi_n + \sum_{j=2}^{K} \pi_{c_{z_j}} 1(z_j > z_k)
\end{aligned}
$$

Step 2: This step expresses $\tilde{F}_{Y_1}(y)$ and $\tilde{F}_{Y_0}(y)$ given by Lemma 2 under Assumptions 4. Consider

$$\begin{aligned}
\tilde{F}_{Y_1}(y) =& \frac{\sum_{k=1}^K F_{Y|D=1,Z=z_k}(y)\, p(1|z_k) P(Z=z_k)\left(\frac{z_k}{\mathbb{E}(Z)}-1\right)}{\sum_{k=1}^K p(1|z_k) P(Z=z_k)\left(\frac{z_k}{\mathbb{E}(Z)}-1\right)} \\
=& \frac{\sum_{k=1}^K \left(\pi_a F_{Y_1|a}(y) + \sum_{j=2}^K \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y) 1(z_j \leq z_k)\right) P(Z=z_k)\left(\frac{z_k}{\mathbb{E}(Z)}-1\right)}{\sum_{k=1}^K \left(\pi_a + \sum_{j=2}^K \pi_{c_{z_j}} 1(z_j \leq z_k)\right) P(Z=z_k)\left(\frac{z_k}{\mathbb{E}(Z)}-1\right)} \\
=& \frac{\sum_{k=1}^K \sum_{j=2}^K \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y) 1(z_j \leq z_k) P(Z=z_k)\left(\frac{z_k}{\mathbb{E}(Z)}-1\right)}{\sum_{k=1}^K \sum_{j=2}^K \pi_{c_{z_j}} 1(z_j \leq z_k) P(Z=z_k)\left(\frac{z_k}{\mathbb{E}(Z)}-1\right)} \\
=& \frac{\sum_{j=2}^K \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y) \sum_{k=1}^K 1(z_j \leq z_k) P(Z=z_k)\left(\frac{z_k}{\mathbb{E}(Z)}-1\right)}{\sum_{j=2}^K \pi_{c_{z_j}} \sum_{k=1}^K 1(z_j \leq z_k) P(Z=z_k)\left(\frac{z_k}{\mathbb{E}(Z)}-1\right)} \\
=& \frac{\sum_{j=2}^K \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y) \left(\frac{\sum_{k=j}^K z_k P(Z=z_k)}{\mathbb{E}(Z)} - P(Z \geq z_j)\right)}{\sum_{j=2}^K \pi_{c_{z_j}} \left(\frac{\sum_{k=j}^K z_k P(Z=z_k)}{\mathbb{E}(Z)} - P(Z \geq z_j)\right)} \\
=& \frac{\sum_{j=2}^K w_j \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y)}{\sum_{j=2}^K w_j \pi_{c_{z_j}}},
\end{aligned}$$

where the first equality follows from Lemma 2 spezialized to the setup of the theorem; the second equality is by step 1; the third equality follows from expanding the numerator and denominator and noting that $\pi_a F_{Y_1|a}(y)(\mathbb{E}(Z)/\mathbb{E}(Z) - 1) = \pi_a(\mathbb{E}(Z)/\mathbb{E}(Z) - 1) = 0$; the fourth equality is by changing the order of summation; and the fifth equality is because

$$\begin{aligned}
w_j \equiv& \frac{\sum_{k=j}^K z_k P(Z=z_k)}{\mathbb{E}(Z)} - P(Z \geq z_j) \\
=& \frac{\mathbb{E}(Z|Z \geq z_j) P(Z \geq z_j)}{\mathbb{E}(Z)} - P(Z \geq z_j) \\
=& \left(\frac{\mathbb{E}(Z|Z \geq z_j)}{\mathbb{E}(Z)} - 1\right) P(Z \geq z_j).
\end{aligned}$$

Similarly, consider

$$\begin{aligned}
\tilde{F}_{Y_0}(y) =& \frac{\sum_{k=1}^{K} F_{Y|D=0,Z=z_k}(y)\, p(0|z_k) P(Z=z_k) \left(1 - \frac{z_k}{\mathbb{E}(Z)}\right)}{\sum_{k=1}^{K} p(1|z_k) P(Z=z_k) \left(\frac{z_k}{\mathbb{E}(Z)} - 1\right)} \\
=& \frac{\sum_{k=1}^{K} \left(\pi_n F_{Y_0|n}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y) 1(z_j > z_k)\right) P(Z=z_k) \left(1 - \frac{z_k}{\mathbb{E}(Z)}\right)}{\sum_{k=1}^{K} \left(\pi_n + \sum_{j=2}^{K} \pi_{c_{z_j}} 1(z_j > z_k)\right) P(Z=z_k) \left(\frac{z_k}{\mathbb{E}(Z)} - 1\right)} \\
=& \frac{\sum_{k=1}^{K} \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y) 1(z_j > z_k) P(Z=z_k) \left(1 - \frac{z_k}{\mathbb{E}(Z)}\right)}{\sum_{k=1}^{K} \sum_{j=2}^{K} \pi_{c_{z_j}} 1(z_j > z_k) P(Z=z_k) \left(\frac{z_k}{\mathbb{E}(Z)} - 1\right)} \\
=& \frac{\sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y) \sum_{k=1}^{K} 1(z_j > z_k) P(Z=z_k) \left(1 - \frac{z_k}{\mathbb{E}(Z)}\right)}{\sum_{j=2}^{K} \pi_{c_{z_j}} \sum_{k=1}^{K} 1(z_j > z_k) P(Z=z_k) \left(\frac{z_k}{\mathbb{E}(Z)} - 1\right)} \\
=& \frac{\sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y) \left(P(Z < z_j) - \frac{\sum_{k=1}^{j-1} z_k P(Z=z_k)}{\mathbb{E}(Z)}\right)}{\sum_{j=2}^{K} \pi_{c_{z_j}} \left(\frac{\sum_{k=j}^{K} z_k P(Z=z_k)}{\mathbb{E}(Z)} - P(Z \geq z_j)\right)} \\
=& \frac{\sum_{j=2}^{K} w_j \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y)}{\sum_{j=2}^{K} \pi_{c_{z_j}} w_j},
\end{aligned}$$

where the last equality is because

$$\begin{aligned}
P(Z < z_j) - \frac{\sum_{k=1}^{j-1} z_k P(Z=z_k)}{\mathbb{E}(Z)} =& \frac{\mathbb{E}(Z) - \sum_{k=1}^{j-1} z_k P(Z=z_k)}{\mathbb{E}(Z)} - (1 - P(Z < z_j)) \\
=& \frac{\sum_{k=j}^{K} z_k P(Z=z_k)}{\mathbb{E}(Z)} - P(Z \geq z_j) \\
=& \left(\frac{\mathbb{E}(Z|Z \geq z_j)}{\mathbb{E}(Z)} - 1\right) P(Z \geq z_j) \\
\equiv& w_j.
\end{aligned}$$

Step 3: This step verifies that $\tilde{F}_{Y_1}(y)$ and $\tilde{F}_{Y_0}(y)$ are well-defined and strictly increasing under the assumptions of the theorem. Consider first the numerators of $\tilde{F}_{Y_1}(y)$ and $\tilde{F}_{Y_0}(y)$. By Assumptions 4 and 5, we have that $\pi_{c_{z_j}} > 0$, $F_{Y_1|c_{z_j}}(y) > 0$, and $F_{Y_0|c_{z_j}}(y) > 0$. Furthermore, because $z > 0$ for all $z \in \mathcal{Z}$, we have

$$w_j \equiv P(Z \geq z_j) \left(\frac{\mathbb{E}(Z|Z \geq z_j)}{\mathbb{E}(Z)} - 1\right) > 0.$$

Hence, the numerators of $\tilde{F}_{Y_1}(y)$ and $\tilde{F}_{Y_0}(y)$ are positive. Furthermore, $F_{Y_1|c_{z_j}}(y)$ and $F_{Y_0|c_{z_j}}(y)$ are strictly increasing by Assumption 4. It follows that both numerators are also strictly increasing. Hence, for $\tilde{F}_{Y_1}(y)$ and $\tilde{F}_{Y_0}(y)$ to be strictly increasing and well-defined, we must show that the denominator is non-zero and positive. This follows directly from $\pi_{c_{z_j}} > 0$ and $w_j > 0$ for $j = 2, ..., K$ as shown before. We conclude that $\tilde{F}_{Y_1}(y)$ and $\tilde{F}_{Y_0}(y)$ are well-defined and strictly increasing.

Step 4: This step expresses $F_{Y_1}^{IVQR}(y)$ and $F_{Y_0}^{IVQR}(y)$ given by Lemma 2 under Assumption 4. Consider,

$$
\begin{aligned}
F_{Y_1}^{IVQR}(y) = & \sum_{k=1}^{K} F_{Y|D=1,Z=z_k}(y)\, p(1|z_k) P(Z=z_k) \\
& + \sum_{k=1}^{K} F_{Y|D=0,Z=z_k}\left(\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)\right) p(0|z_k) P(Z=z_k) \\
= & \sum_{k=1}^{K} \left(\pi_a F_{Y_1|a}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y) 1(z_j \leq z_k) \right) P(Z=z_k) \\
& + \sum_{k=1}^{K} \left(\pi_n F_{Y_0|n}\left(\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)\right) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}\left(\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)\right) 1(z_j > z_k) \right) P(Z=z_k) \\
= & \pi_a F_{Y_1|a}(y) + \sum_{k=1}^{K} \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y) 1(z_j \leq z_k) P(Z=z_k) \\
& + \pi_n F_{Y_0|n}\left(\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)\right) + \sum_{k=1}^{K} \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}\left(\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)\right) 1(z_j > z_k) P(Z=z_k) \\
= & \pi_a F_{Y_1|a}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y) \sum_{k=1}^{K} 1(z_j \leq z_k) P(Z=z_k) \\
& + \pi_n F_{Y_0|n}\left(\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)\right) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}\left(\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)\right) \sum_{k=1}^{K} 1(z_j > z_k) P(Z=z_k) \\
= & \pi_a F_{Y_1|a}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y) P(Z \geq z_j) \\
& + \pi_n F_{Y_0|n}\left(\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)\right) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}\left(\tilde{Q}_{Y_0}\left(\tilde{F}_{Y_1}(y)\right)\right) P(Z < z_j)
\end{aligned}
$$

where the first equality follows from Lemma 2 specialized to the setup of the theorem, the second equality is by step 1; the third equality follows from expanding terms; and the fourth equality is by changing the order of summation.

Similarly, consider

$$
\begin{aligned}
F_{Y_0}^{IVQR}(y) =& \frac{1}{\mathbb{E}(Z)} \sum_{k=1}^{K} F_{Y|D=1,Z=z_k}\left(\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)\right) p(1|z_k) z_k P(Z=z_k) \\
&+ \frac{1}{\mathbb{E}(Z)} \sum_{k=1}^{K} F_{Y|D=0,Z=z_k}(y)\, p(0|z_k) z_k P(Z=z_k) \\
=& \frac{1}{\mathbb{E}(Z)} \sum_{k=1}^{K} \left(\pi_a F_{Y_1|a}\left(\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)\right) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}}\left(\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)\right) 1(z_j \le z_k) \right) z_k P(Z=z_k) \\
&+ \frac{1}{\mathbb{E}(Z)} \sum_{k=1}^{K} \left(\pi_n F_{Y_0|n}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y) 1(z_j > z_k) \right) p(0|z_k) z_k P(Z=z_k) \\
=& \pi_a F_{Y_1|a}\left(\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)\right) + \frac{1}{\mathbb{E}(Z)} \sum_{k=1}^{K} \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}}\left(\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)\right) 1(z_j \le z_k) z_k P(Z=z_k) \\
&+ \pi_n F_{Y_0|n}(y) + \frac{1}{\mathbb{E}(Z)} \sum_{k=1}^{K} \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y) 1(z_j > z_k) z_k P(Z=z_k) \\
=& \pi_a F_{Y_1|a}\left(\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)\right) + \frac{1}{\mathbb{E}(Z)} \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}}\left(\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)\right) \sum_{k=1}^{K} 1(z_j \le z_k) z_k P(Z=z_k) \\
&+ \pi_n F_{Y_0|n}(y) + \frac{1}{\mathbb{E}(Z)} \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y) \sum_{k=1}^{K} 1(z_j > z_k) z_k P(Z=z_k) \\
=& \pi_a F_{Y_1|a}\left(\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)\right) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}}\left(\tilde{Q}_{Y_1}\left(\tilde{F}_{Y_0}(y)\right)\right) \frac{\sum_{k=j}^{K} z_k P(Z=z_k)}{\mathbb{E}(Z)} \\
&+ \pi_n F_{Y_0|n}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y) \frac{\sum_{k=1}^{j-1} z_k P(Z=z_k)}{\mathbb{E}(Z)}
\end{aligned}
$$

This completes the proof of the theorem. □

Proof of Theorem 7.

By Assumption 4, we have that $0 < P(Z = z) < 1$ for all $z \in \mathcal{Z}$. Moreover, I show in step 3 of the proof of Theorem 6 that $\tilde{F}_{Y_1}(y)$ and $\tilde{F}_{Y_0}(y)$ are strictly increasing under the Assumptions 4 and 5. Therefore, it follows from Lemma 2 that $Q_{Y_1}^{IVQR}\left(F_{Y_0}^{IVQR}(y)\right) = \tilde{Q}_{Y_1}\left(\tilde{G}_{Y_0}(y)\right)$. The proof of the result now follows from identical arguments as in the proof of Theorem 2. □

Figures

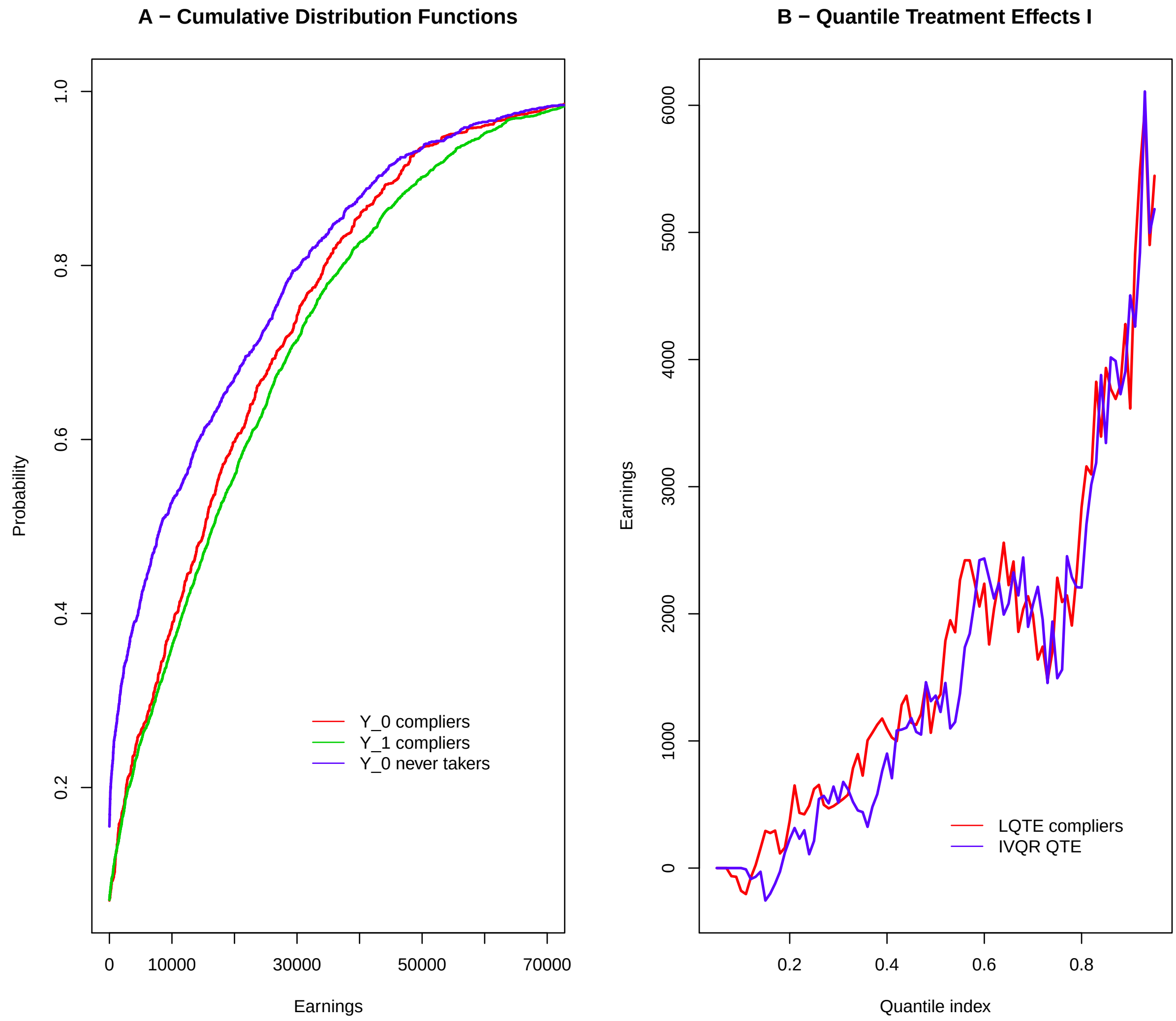

Figure 1: The number of observations is 5,083. The left panel shows estimated CDFs for the different subpopulations. The right panel contains estimates of the QTE form the IVQR model and the LQTE model. The QTE are computed over $\tau = \{0.05, 0.06, ..., 0.95\}$.

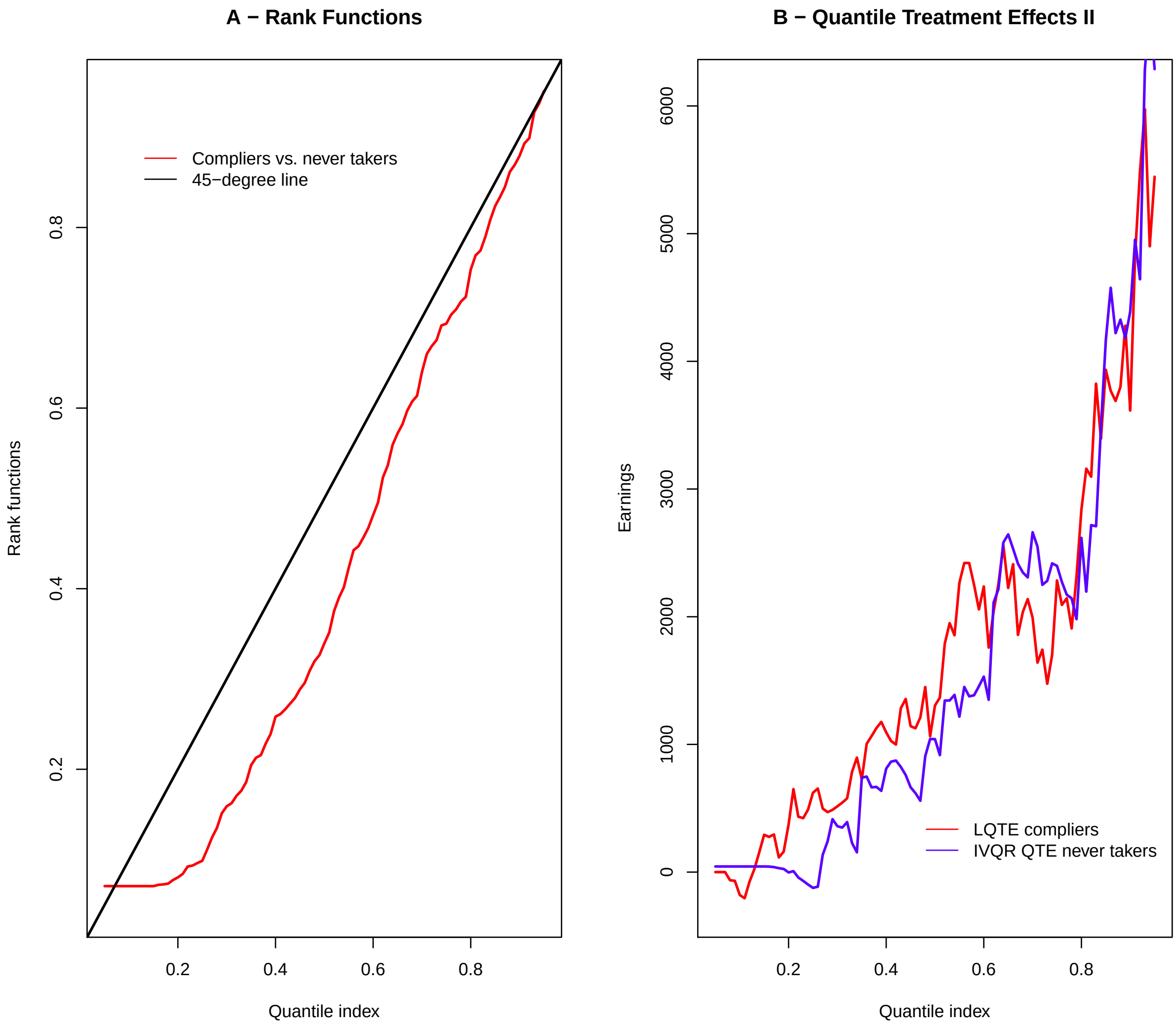

Figure 2: The number of observations is 5,083. The left panel plots the estimated rank functions and the right panel compares the LQTE estimates with the IVQR estimates for the never takers. All estimates are computed over $\tau = \{0.05, 0.06, ..., 0.95\}$.

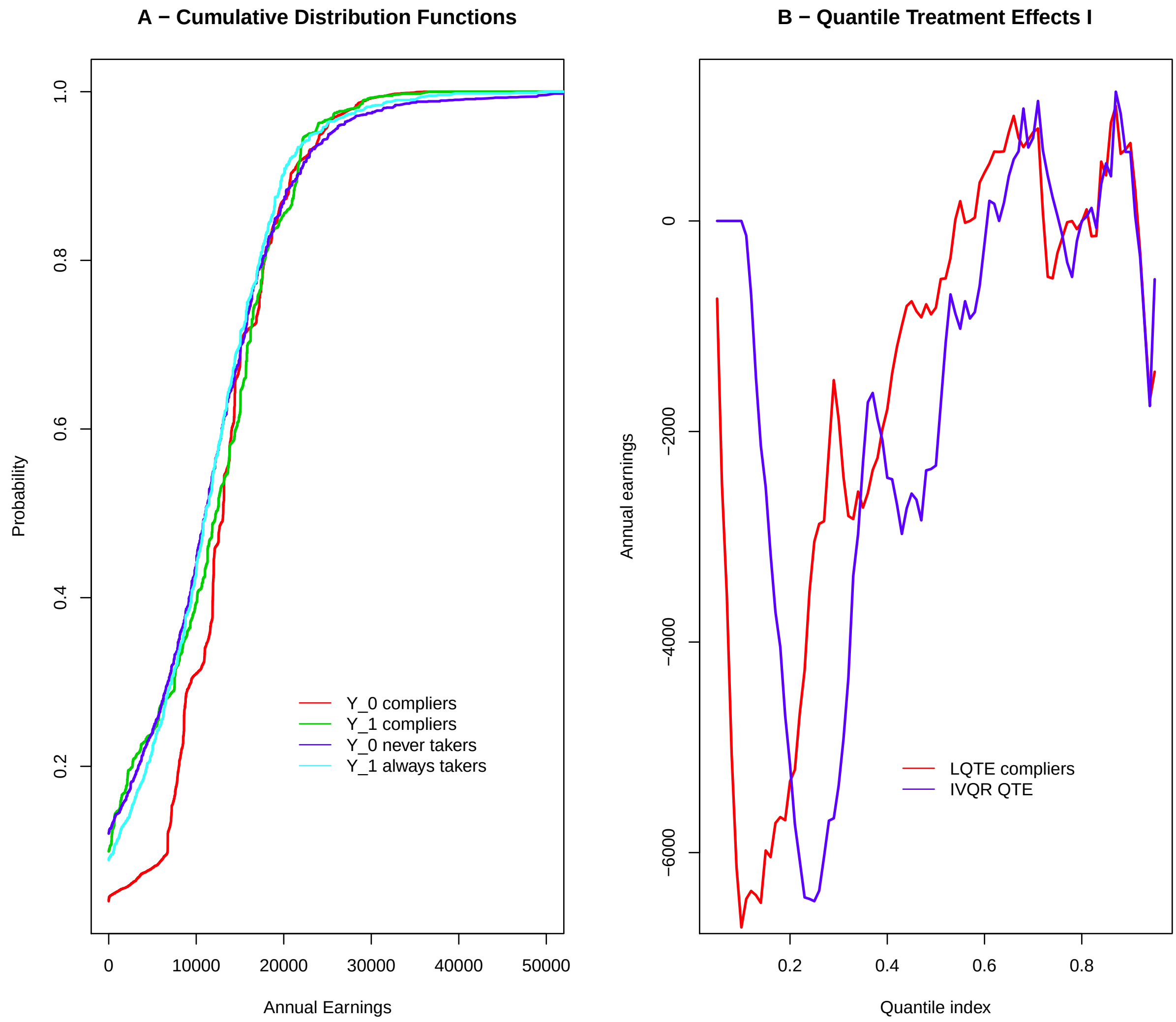

Figure 3: The number of observations is 11,637. The left panel shows estimated CDFs for the different subpopulations. The right panel contains estimates of the QTE form the IVQR model and the LQTE model. The QTE are computed over $\tau = \{0.05, 0.06, ..., 0.95\}$.

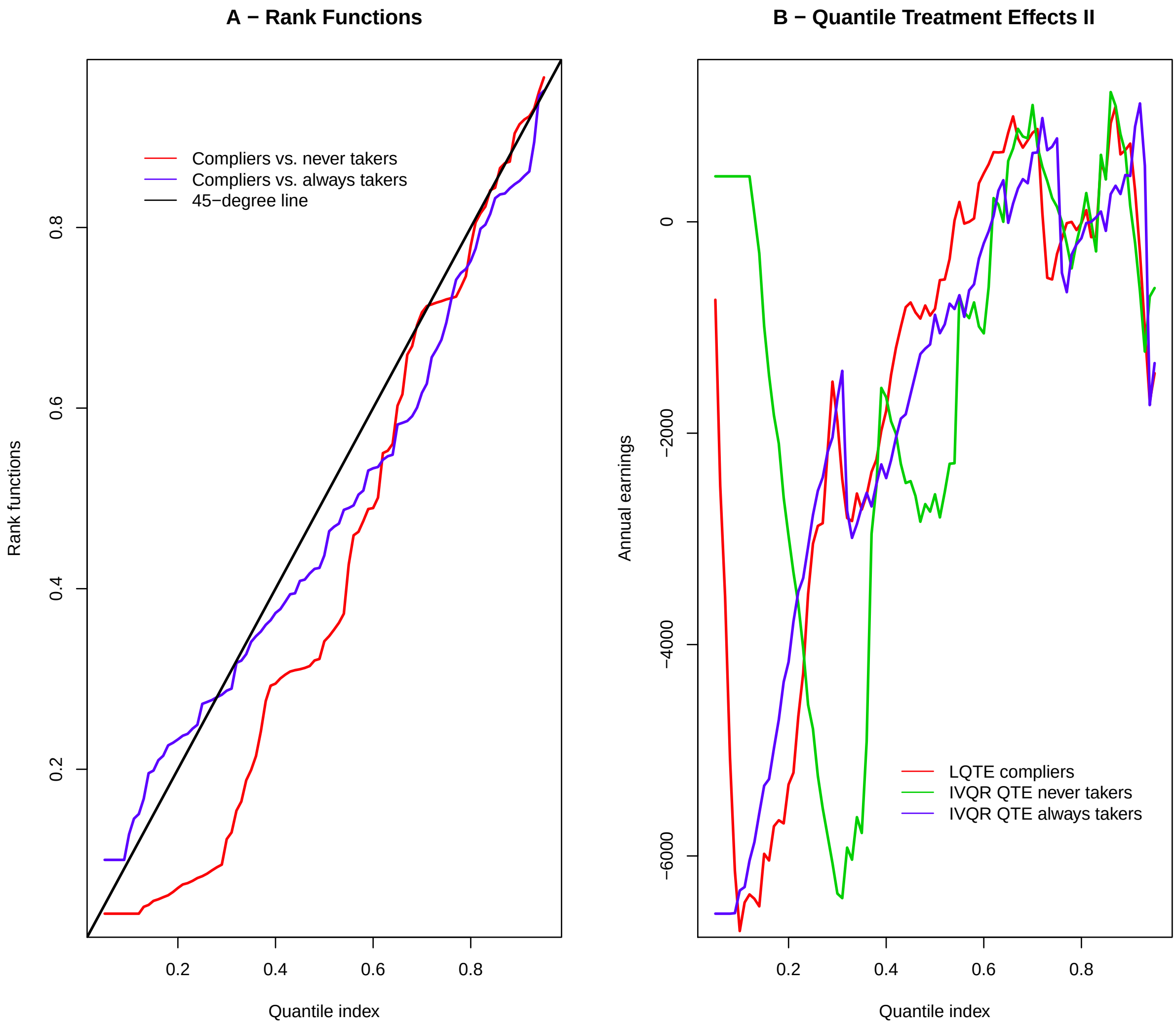

Figure 4: The number of observations is 11,637. The left panel plots the estimated rank functions and the right panel compares the LQTE estimates with the IVQR estimates for the never takers and the always takers. All estimates are computed over $\tau = \{0.05, 0.06, ..., 0.95\}$.