Title
A Comparison of Two Quantile Models With Endogeneity

Journal
Journal of Business and Economic Statistics, 38(2)

ISSN
0735-0015

Author
Wüthrich, Kaspar

Publication Date
2020-04-02

DOI
10.1080/07350015.2018.1514307

Peer reviewed

A comparison of two quantile models with endogeneity

Kaspar Wüthrich
Department of Economics, UC San Diego

August 16, 2018

Abstract

This paper studies the relationship between the two most-used quantile models with endogeneity: the instrumental variable quantile regression (IVQR) model (Chernozhukov and Hansen, 2005) and the local quantile treatment effects (LQTE) model (Abadie et al., 2002). The key condition of the IVQR model is the rank similarity assumption, a restriction on the evolution of individual ranks across treatment states, under which population quantile treatment effects (QTE) are identified. By contrast, the LQTE model achieves identification through a monotonicity assumption on the selection equation but only identifies QTE for the subpopulation of compliers. This paper shows that, despite these differences, there is a close connection between both models: (i) the IVQR estimands correspond to QTE for the compliers at transformed quantile levels and (ii) the IVQR estimand of the average treatment effect is equal to a convex combination of the local average treatment effect and a weighted average of integrated QTE for the compliers. These results do not rely on the rank similarity assumption and therefore provide a characterization of IVQR in settings where this key condition is violated. Underpinning the analysis are novel closed-form representations of the IVQR estimands. I illustrate the theoretical results with two empirical applications.

Keywords: Instrumental variables, quantile treatment effects, local quantile treatment effects, rank similarity

1 Introduction

Quantile regression methods have become popular tools for analyzing the heterogenous impact of policy variables beyond simple averages. When the policy variable of interest is endogenous, classical quantile regression (Koenker and Bassett, 1978) is inconsistent for estimating quantile treatment effects (QTE). To overcome this problem, different instrumental variable (IV) approaches have been proposed. This paper studies the relationship between the two most-used IV quantile models: the instrumental variable quantile regression (IVQR) model (Chernozhukov and Hansen, 2005) and the local quantile treatment effect (LQTE) model (Abadie et al., 2002).

The key condition underlying the IVQR model is rank similarity, a restriction on the evolution of individual ranks across treatment states. By virtue of this assumption, the IVQR model identifies QTE for the overall population. While rank similarity has substantial identifying power, it is a strong and controversial condition that substantially restricts treatment effect heterogeneity (e.g., Heckman and Vytlacil, 2007). In the LQTE model identification is achieved through a monotonicity assumption on the selection equation. As this model allows for unrestricted treatment effect heterogeneity, only the QTE for the subpopulation that responds to the instrument—the compliers—are identified. The LQTE model has been criticized because in many policy evaluation problems other subpopulations such as the treated or the overall population are of primary interest; see for instance the controversial discussion by Imbens (2010), Deaton (2010), and Heckman and Urzua (2010). Thus, researchers face a fundamental trade-off between the LQTE model that allows for unrestricted treatment effect heterogeneity but only identifies treatment effects for the compliers, and the IVQR model that restricts treatment effect heterogeneity but identifies population treatment effects.

On the surface, the IVQR and the LQTE model do not seem to be connected. The estimands of both models are different and the underlying assumptions are non-nested, non-contradictory, and concern different aspects of the models. Because neither model is more general than the other, Chernozhukov and Hansen (2013) have described the IVQR model and the LQTE model as complementary approaches for estimating heterogeneous treatment effects and suggested a further investigation of their similarities and differences.

Furthermore, comparisons of both models have been used as specification checks for the underlying assumptions (e.g., Chernozhukov and Hansen, 2004).

The main contribution of this paper is to characterize the treatment effect estimands based on the IVQR model under the LQTE assumptions. I show that irrespectively of the validity of the rank similarity assumption, the IVQR QTE estimands are equal to LQTE at transformed quantile levels. Moreover, the IVQR estimand of the average treatment effect (ATE) can be decomposed into a convex combination of the local average treatment effect (LATE) and a weighted average of integrated LQTE. Thus, the IVQR estimands are not completely arbitrary when the rank similarity assumption is violated, but correspond to well-defined (functions of) causal effects for the compliers. Differences between the estimands of both models are determined by discrepancies between the potential outcome distributions of always-takers, never-takers, and compliers and the relative size of these subpopulations. The theoretical analysis further implies analytical characterizations of the bias of the IVQR estimands when rank similarity is violated. I show that the key observable determinant of the bias is the size of the instrument first stage. Underpinning the theoretical analysis are novel closed-form solutions of the IVQR estimands, which may be of independent interest.

On the one hand, this paper confirms that with unrestricted treatment effect heterogeneity all the information about the treatment effects has to come from the compliers. On the other hand, it shows how IVQR extrapolates from the compliers to the whole population. This motivates the use of IVQR as an approach to extrapolation based on the LQTE model; see, for example, Angrist and Fernandez-Val (2013) and the references therein for alternative approaches to external validity. Constructive identification through extrapolation based on the IVQR model has been studied in independent and concurrent work by Vuong and Xu (2017). However, in sharp contrast to the present paper, their results rely on the IVQR assumptions and thus do not provide a characterization of IVQR in the absence rank similarity, which is one of the main contributions of this paper. Moreover, they do not provide closed-form solutions for the IVQR estimands, but implicit characterizations based on counterfactual mappings.

The analysis is extended to more general settings that allow for failures of the LQTE monotonicity assumption, nonbinary instruments, and covariates. I show that the main results describing the relationship between the IVQR estimands and their counterparts in the LQTE model have intuitive analogues in these more general settings.

I illustrate the theoretical results using two empirical applications. In the first application, I re-examine the causal effect of Job Training Partnership Act (JTPA) training programs on the distribution of subsequent earnings. In the second application, I consider the problem of estimating the structural effect of Vietnam veteran status on civilian wages using draft lottery data.

This paper contributes to the extensive literatures on identification and estimation based on both models. The IVQR model has been introduced by Chernozhukov and Hansen (2004, 2005, 2006). Estimation and inference in linear conditional quantile models have been analyzed by Chernozhukov and Hong (2003), Chernozhukov and Hansen (2006), Chernozhukov et al. (2007), Chernozhukov and Hansen (2008), Chernozhukov et al. (2009), Kaplan and Sun (2017) and Chen and Lee (2018). Nonparametric estimation has been studied by Chernozhukov et al. (2007), Horowitz and Lee (2007), Chen and Pouzo (2009), Chen and Pouzo (2012), Gagliardini and Scaillet (2012), and Su and Hoshino (2016). Recent surveys of the IVQR model are provided by Chernozhukov and Hansen (2013) and Chernozhukov et al. (2017).

The LQTE model, introduced by Abadie et al. (2002), extends the LATE framework (Imbens and Angrist, 1994; Imbens and Rubin, 1997) to the analysis of conditional LQTE using the weighting theorem of Abadie (2003). In subsequent work, Frandsen et al. (2012) have analyzed estimation of LQTE based on regression discontinuity frameworks, Frölich and Melly (2013) have studied nonparametric identification and estimation of unconditional LQTE with covariates, De Chaisemartin (2017a,b) has analyzed the LQTE framework under a weaker version of the monotonicity assumption, and Belloni et al. (2017) have derived the properties of regression-based estimators for unconditional LQTE after selection among high-dimensional controls. A recent survey of the LQTE model is provided by Melly and Wüthrich (2017).

The remainder of the paper is organized as follows. Section 2 introduces the basic notation and reviews both models. In Section 3, I characterize the IVQR treatment effect estimands under the LQTE assumptions. Section 4 generalizes these results to setups that allow for failures of the LQTE monotonicity assumption, nonbinary instruments, and covariates. In Section 5, I present two empirical applications. Section 6 concludes. The appendix contains all proofs, some additional results, and implementation details for the applications.

2 Setup and models

I consider a setup with an absolutely continuous outcome variable Y, a binary treatment D, and a binary instrument Z. Let $F_{Y|D=d,Z=z}$ and $f_{Y|D=d,Z=z}$ denote the cumulative distribution function (CDF) and the density function of $Y|D=d, Z=z$ and define $p(d|z) \equiv P(D=d|Z=z)$. Covariates are omitted for deriving the main results of the paper. Section 4 presents extensions that incorporate covariates and nonbinary instruments. The analysis is developed within the potential outcomes framework (cf. Rubin, 1974). Potential outcomes and potential treatments are denoted by Y_d, $d \in \{0,1\}$, and D_z, $z \in \{0,1\}$. Observed outcomes and observed treatments are given by $Y = DY_1 + (1-D)Y_0$ and $D = ZD_1 + (1-Z)D_0$.

Based on their potential treatments (D_0, D_1), individuals can be categorized by four types, $T \in \{a, n, c, f\}$ (e.g., Angrist et al., 1996):

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

Henceforth, for each type $T=t$, let π_t, $F_{Y_d|t}$, $f_{Y_d|t}$, and $Q_{Y_d|t}$ denote its proportion, the CDF of Y_d, the density function of Y_d, and the quantile function (QF) of Y_d.

In this paper, I focus on estimands of the CDF and the QF of Y_d, F_{Y_d} and Q_{Y_d}, the τ-QTE, $\delta(\tau) \equiv Q_{Y_1}(\tau) - Q_{Y_0}(\tau)$, and the ATE, $\Delta \equiv \mathbb{E}(Y_1 - Y_0) = \int_0^1 \delta(\tau)d\tau$. Without loss of generality, I consider the following structural QF, $q(d,\tau) = Q_{Y_d}(\tau)$:

$$q(d,\tau) = d\delta(\tau) + Q_{Y_0}(\tau). \tag{1}$$

By the Skorohod representation of random variables, potential outcomes can be related to the structural QF as follows (Chernozhukov and Hansen, 2005):

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

Similarly, observed outcomes can be expressed as $Y = q(D, U)$, where $U \equiv U_D$. This representation is essential for the IVQR model.

2.1 IVQR model

The IVQR model consists of the following main conditions (e.g., Chernozhukov and Hansen, 2005, 2013):

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: $D \equiv \rho(Z, V)$ for some unkown function $\rho(\cdot)$ and random vector V.*
4. *Rank similarity: conditional on (Z, V), $\{U_d\}$ are identically distributed.*

Assumption 1.1 restricts the outcome to be nonatomic conditional on the instrument. The independence condition in Assumption 1.2 states that potential outcomes are independent of the instrument. In Assumption 1.3, the random vector V leads to differences in treatment choices among observationally identical individuals. Assumption 1.4 is the key condition of the IVQR model. It requires that individual ranks are constant across potential outcome distributions up to unsystematic deviations from a common rank level U. Rank similarity is considered to be a strong restriction in many applications. For instance, as noted by Heckman and Vytlacil (2007), rank similarity precludes scenarios in which agents self-select based on their individual effects and does not allow for effect heterogeneity as generated by the generalized Roy model. For in-depth discussions of the IVQR model the interested reader is referred to Chernozhukov and Hansen (2005) and the reviews by Chernozhukov and Hansen (2013) and Chernozhukov et al. (2017).

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

$$P\left(Y \leq D\delta(\tau) + Q_{Y_0}(\tau)|Z\right) = \tau. \tag{2}$$

Chernozhukov and Hansen (2005) prove point identification of $\delta(\tau)$ and $Q_{Y_0}(\tau)$ under an additional full rank condition on the Jacobian of (2); see Section 1 of the appendix. The conditional moment restriction (2) justifies the following unconditional moment equations for estimation:

$$\mathbb{E}\left(\left(\tau - 1\left[Y \leq D\delta(\tau) + Q_{Y_0}(\tau)\right]\right) f(Z)\right) = 0, \tag{3}$$

where $1[\cdot]$ is the indicator function and $f(Z)$ is a vector of (transformations of) instruments. The introduction provides references to different estimation procedures.

Note that the main theoretical results of this paper do not rely on Assumption 1. Instead, the idea is to take the unconditional IVQR moment conditions as given and to characterize the solutions for $\delta(\tau)$ and $Q_{Y_0}(\tau)$ under the LQTE assumptions.

2.2 LQTE model

The LQTE model is based on the following set of assumptions (e.g., Abadie et al., 2002).

Assumption 2.

1. *Independence:* (Y_1, Y_0, D_1, D_0) *are jointly independent of* Z*.*

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

3. *First-stage:* $p(1|1) - p(1|0) > 0$*.*

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

Assumption 2.1 states that both potential outcomes and potential treatments are independent of the instrument. This assumption is stronger than the corresponding independence assumption in the IVQR model. Assumption 2.2 requires that the instrument assignment is nontrivial. Assumption 2.3 is a standard first stage assumption. Note that, under Assumption 2, the size of the first stage corresponds to the fraction of compliers, $\pi_c = p(1|1) - p(1|0)$. Hence, Assumption 2.3 can alternatively be stated as $\pi_c > 0$. The monotonicity Assumption 2.4 rules out the presence of defiers. Consequently, always-takers, never-takers, and compliers exhaustively partition the whole population. If the instrument satisfies one-sided non-compliance as in the application of Section 5.2, Assumption 2.4 holds by design. Furthermore, Assumption 2.4 is often plausible in field experiments, where the presence of defiers would imply counter-intuitive behavior to the experimental protocol (Huber and Wüthrich, 2018). By contrast, in quasi-experimental settings, Assumption 2.4 is not innocuous and is likely to be violated in many instances (e.g., De Chaisemartin, 2017b). Vytlacil (2002) shows that Assumption 2 is equivalent to a selection model in which selection into the program is modeled by a latent index crossing a threshold. I refer to Melly and Wüthrich (2017) for a recent review of the LQTE model.

Under Assumption 2, the potential outcome QFs and the LQTE for the compliers, $\delta_c(\tau) \equiv Q_{Y_1|c}(\tau) - Q_{Y_0|c}(\tau)$, are identified from the following weighted population quantile regression (Abadie et al., 2002; Abadie, 2003):

$$\left(Q_{Y_0|c}(\tau), \delta_c(\tau)\right) = \arg \min_{\left(Q_{Y_0|c}, \delta_c\right)} \mathbb{E}\left(\kappa \cdot \rho_\tau\left(Y - \delta_c D - Q_{Y_0|c}\right)\right), \tag{4}$$

where $\rho_\tau(u) = u\left(\tau - 1[u<0]\right)$ is the usual check function and $\kappa = 1 - \frac{D(1-Z)}{1-P(Z=1)} - \frac{(1-D)Z}{P(Z=1)}$. The introduction reviews different approaches for estimating conditional and unconditional QTE based on the LQTE model.

3 IVQR estimands under the LQTE assumptions

In this section, I characterize the IVQR estimands under the LQTE assumptions. I focus on the following unconditional IVQR moment conditions:

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

where $\delta^*(\tau)$ and $Q^*_{Y_0}(\tau)$ are referred to as the *IVQR estimands* of $\delta(\tau)$ and $Q_{Y_0}(\tau)$. When the instrument Z is binary, $\delta^*(\tau)$ and $Q^*_{Y_0}(\tau)$ also solve the conditional moment restrictions (3); see Section 2 of the appendix. The IVQR estimands of the potential outcome CDFs, $F^*_{Y_1}$ and $F^*_{Y_0}$, are given by the inverses of the corresponding QFs, and the IVQR estimand of the ATE is obtained as $\Delta^* = \int_0^1 \delta^*(\tau) d\tau$. Except for Section 3.5, the analysis in this paper does not rely on the IVQR assumptions. Hence, the IVQR estimands will generally be biased. For instance, $\delta^*(\tau)$ will generally differ from $\delta(\tau)$.

When interpreting the theoretical results, it is important to note that the IVQR and the LQTE assumptions are non-nested and neither set of assumptions is more general than the other. As discussed in Section 2.2, the LQTE assumptions are not innocuous and are likely to be violated in many applications. Section 4.1 therefore extends the analysis to settings where the LQTE monotonicity assumption is violated.

To derive the main results, I impose the following additional assumption. Let $\mathcal{S}(A)$ denote the support of a random variable A and define $\mathcal{S}_d \equiv \left[\underline{y}_d, \overline{y}_d\right]$, where $-\infty < \underline{y}_d < \overline{y}_d < \infty$ for $d \in \{0, 1\}$.

Assumption 3. *For $(d,t) \in \{0,1\} \times \{a,c,n\}$:*

1. *Regularity: $f_{Y_d|t}$ is continuous.*
2. *Support: $\mathcal{S}(Y_d|T = t) = \mathcal{S}(Y_d)$, where $\mathcal{S}(Y_d) = \mathcal{S}_d$.*

Assumption 3.1 is a standard regularity condition. Assumption 3.2 imposes full support of the subpopulation potential outcome distributions. This full support assumption is the key condition for deriving the closed-form solutions underlying the main results of this paper.

Remark 1. *Section 1 of the appendix shows that, under Assumptions 1 and 2, Assumption 3 is implied by the identification conditions in Chernozhukov and Hansen (2005), including full rank and continuity of the Jacobian of (2). Hence, Assumption 3 does not impose any additional restrictions relative to the analysis in Chernozhukov and Hansen (2005).*

Remark 2. *Assumption 3.2 can be weakened if one is only interested in a shorter range of quantiles; for instance, $\tau \in [0.1, 0.9]$. However, to compute the ATE (cf. Section 3.3), Assumption 3.2 is required.*

3.1 QTE

Theorem 1 presents the first main result of the paper.

Theorem 1. *Suppose that Assumptions 2 and 3 hold and that the IVQR estimands solve the moment conditions (5). Then, for all $\tau \in (0,1)$,*

$$\delta^*(\tau) = \delta_c\left(F_{Y_0|c}\left(Q^*_{Y_0}(\tau)\right)\right) = \delta_c\left(F_{Y_1|c}\left(Q^*_{Y_1}(\tau)\right)\right).$$

Theorem 1 relies on closed-form solutions for $\delta^*(\tau)$ in terms of the joint distribution of (Y, D, Z), which are derived from the IVQR moment conditions (5). They key condition underlying these closed-form solutions is Assumption 3. It ensures that, for $d \in \{0,1\}$, $F_{Y_d|c}$ and thus $Q_{Y_d|c}$ and $Q^*_{Y_d}$ are strictly monotonic, which implies that the moment conditions (5) have a unique solution.

Theorem 1 shows that the IVQR QTE estimands are equivalent to LQTE for the compliers at transformed quantile levels. These transformations reflect differences between the IVQR estimands of the potential outcome CDFs, $F^*_{Y_1}$ and $F^*_{Y_0}$, and the corresponding CDFs for the compliers, $F_{Y_1|c}$ and $F_{Y_0|c}$, as measured by the probability-probability transforms $F_{Y_0|c} \circ Q^*_{Y_0}$ and $F_{Y_1|c} \circ Q^*_{Y_1}$. Theorem 2 below derives analytical expressions for $F^*_{Y_1}$ and $F^*_{Y_0}$ in terms of the identified potential outcome distribution and quantile functions of always-takers, never-takers, and compliers. While the IVQR estimand $\delta^*(\tau)$ will generally differ from the true QTE, $\delta(\tau)$, when Assumption 1 is violated, Theorem 1 shows that $\delta^*(\tau)$ is not completely arbitrary but corresponds to a well-defined causal effect for the compliers.

Theorem 1 has interesting implications for the connection between both models, which are summarized in the following corollary.

Corollary 1. *Under the assumptions of Theorem 1,*

1. *$\delta_c(\tau) \geq 0$ for all $\tau \in (0,1)$ implies that $\delta^*(\tau) \geq 0$ for all $\tau \in (0,1)$ and $\delta_c(\tau) \leq 0$ for all $\tau \in (0,1)$ implies that $\delta^*(\tau) \leq 0$ for all $\tau \in (0,1)$.*

2. *if δ_c is monotonically increasing (decreasing) then δ^* is monotonically increasing (decreasing).*

3. *$\delta_c(\tau) = \delta_c$ for all $\tau \in (0,1)$ implies that $\delta^*(\tau) = \delta_c(\tau) = \delta_c$ for all $\tau \in (0,1)$.*

Corollary 1 shows that the IVQR QTE estimand, $\delta^*(\tau)$, inherits various qualitative features from the LQTE, which is the QTE for the the largest subpopulation for which this effect is identified. As a consequence, $\delta^*(\tau)$ exhibits several desirable *robustness properties* in settings where the IVQR assumptions are violated. Specifically, $\delta^*(\tau)$ does not suffer from undesirable sign reversals: if the QTE is positive (negative) for all subpopulations and all quantiles, $\delta^*(\tau)$ will not be negative (positive) at any quantile. Similarly, if the QTE is monotonically increasing (decreasing) for all subpopulations, δ^* will not be constant or monotonically decreasing (increasing). Corollary 1 further shows that $\delta^*(\tau)$ is equal to $\delta_c(\tau)$ at all quantiles if δ_c is constant, as implied, for example, by a location-shift model for the compliers. As a consequence, the IVQR QTE estimand exhibits effect heterogeneity across quantiles if and only if there is effect heterogeneity in the LQTE.

3.2 Potential outcome CDFs and QTE for subpopulations

To further discuss and interpret Theorem 1, it is useful to analyze closed-form characterizations for the IVQR estimands of the potential outcome CDFs.

Note that under Assumption 2, one can decompose $F_{Y_0}(y)$ and $F_{Y_1}(y)$ as

$$F_{Y_1}(y) = \pi_c F_{Y_1|c}(y) + \pi_a F_{Y_1|a}(y) + \pi_n F_{Y_1|n}(y), \tag{6}$$

$$F_{Y_0}(y) = \pi_c F_{Y_0|c}(y) + \pi_a F_{Y_0|a}(y) + \pi_n F_{Y_0|n}(y). \tag{7}$$

Imbens and Rubin (1997) and Abadie (2002) show that the following potential outcome distributions are identified from the data:

$$F_{Y_1|c}(y) = \tilde{F}_{Y_1}(y),\ F_{Y_1|a}(y) = F_{Y|D=1,Z=0}(y),\ F_{Y_0|c}(y) = \tilde{F}_{Y_0}(y),\ \text{and}\ F_{Y_0|n}(y) = F_{Y|D=0,Z=1}(y),$$

where

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

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

Moreover, the proportions of the three subpopulations are identified as $\pi_c = p(1|1) - p(1|0)$, $\pi_a = p(1|0)$, and $\pi_n = p(0|1)$. However, $F_{Y_1|n}(y)$ and $F_{Y_0|a}(y)$, and, consequently, $F_{Y_1}(y)$ and $F_{Y_0}(y)$, are not point identified under the LQTE assumptions. By contrast, these quantities are identified under the IVQR assumptions. The following theorem shows how IVQR imputes the unidentified quantities $F_{Y_1|n}(y)$ and $F_{Y_0|a}(y)$ using the rank similarity assumption. To describe the results, let $\mathcal{Y}_1 \equiv \left\{Q^*_{Y_1}(\tau) : \tau \in (0,1)\right\}$ and $\mathcal{Y}_0 \equiv \left\{Q^*_{Y_0}(\tau) : \tau \in (0,1)\right\}$ denote the regions of interest.

Theorem 2. *Suppose that Assumptions 2 and 3 hold and that the IVQR estimands solve the moment conditions (5). Then, for all $(y_1, y_0) \in \mathcal{Y}_1 \times \mathcal{Y}_0$,*

$$\begin{aligned} F^*_{Y_1}(y_1) &= \pi_c F_{Y_1|c}(y_1) + \pi_a F_{Y_1|a}(y_1) + \pi_n F_{Y_0|n}\left(Q_{Y_0|c}\left(F_{Y_1|c}(y_1)\right)\right), \\ F^*_{Y_0}(y_0) &= \pi_c F_{Y_0|c}(y_0) + \pi_a F_{Y_1|a}\left(Q_{Y_1|c}\left(F_{Y_0|c}(y_0)\right)\right) + \pi_n F_{Y_0|n}(y_0). \end{aligned}$$

Theorem 2 shows that the IVQR estimands can be decomposed into mixtures of the potential outcome CDFs that are identified under the LQTE assumptions and the IVQR estimands of $F_{Y_1|n}(y)$ and $F_{Y_0|a}(y)$:

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

The additive mixture structures of $F^*_{Y_0}(y)$ and $F^*_{Y_1}(y)$ allow separating the determinants of the discrepancies between the estimands of both models. Consider the difference between $F^*_{Y_0}(y)$ and $F_{Y_0|c}(y)$:

$$F^*_{Y_0}(y) - F_{Y_0|c}(y) = \pi_a\left(F_{Y_1|a}\left(Q_{Y_1|c}\left(F_{Y_0|c}(y)\right)\right) - F_{Y_0|c}(y)\right) + \pi_n\left(F_{Y_0|n}(y) - F_{Y_0|c}(y)\right). \quad (8)$$

The difference in equation (8) is determined by two factors: (i) discrepancies between the potential outcome distributions of compliers and always-takers (through $F_{Y_1|a} \circ Q_{Y_1|c}$) and discrepancies between the potential outcome distributions of compliers and never-takers (through $F_{Y_0|n}(y) - F_{Y_0|c}(y)$) and (ii) the fractions of always-takers and never-takers, which are related to the first stage as $\pi_a + \pi_n = 1 - \pi_c$. The analysis of the difference between $F^*_{Y_1}(y)$ and $F_{Y_1|c}(y)$ is symmetric and thus omitted. It is important to note that discrepancies between the estimands of both models cannot be used to assess the validity of the rank similarity assumption, because rank similarity is fundamentally untestable with binary instruments (Kim and Park, 2017).

Next, I use the closed-form solutions in Theorem 2 to analyze the bias of the IVQR estimands in settings where the underlying assumptions are violated. Consider the bias of

$F^*_{Y_0}(y)$:

$$B_{F^*_{Y_0}}(y) \equiv F^*_{Y_0}(y) - F_{Y_0}(y).$$

The analysis of the bias of $F^*_{Y_1}(y)$ is similar and thus omitted. Theorem 2 implies that

$$B_{F^*_{Y_0}}(y) = \pi_a \left(F^*_{Y_0|a}(y) - F_{Y_0|a}(y) \right) = \pi_a \left(F_{Y_1|a} \left(Q_{Y_1|c} \left(F_{Y_0|c}(y) \right) \right) - F_{Y_0|a}(y) \right).$$

Thus, the bias is determined by discrepancies between the IVQR estimand $F^*_{Y_0|a}(y)$ and $F_{Y_0|a}(y)$, which result from violations of the IVQR assumptions, and the proportion of always-takers, π_a. Since $F_{Y_0|a}(y)$ is fundamentally unidentified, the key observable determinant of the bias is π_a, which is related to the instrument first stage as $\pi_a = 1 - \pi_c - \pi_n$. Thus, for fixed π_n, the bias decreases if the size of the first stage increases. Violations of the IVQR assumptions can be modeled similar to Masten and Poirier (2017). Suppose that, for $t \in [0, 1]$,

$$F_{Y_0|a}(y) = tF_{Y_1|a} \left(Q_{Y_1|c} \left(F_{Y_0|c}(y) \right) \right) + (1 - t)G(y), \tag{9}$$

where G is an unrestricted CDF. In equation (9), the mixing factor t captures the degree of failure of the IVQR assumptions. For $t = 1$, the IVQR assumptions hold and $F_{Y_0|a}(y) = F_{Y_1|a} \left(Q_{Y_1|c} \left(F_{Y_0|c}(y) \right) \right)$, while for $t = 0$ the IVQR assumptions fail in a completely unrestricted way and $F_{Y_0|a}(y) = G(y)$. The case of $t \in (0, 1)$ captures partial failures of the IVQR assumptions. Based on equation (9), $B_{F^*_{Y_0}}(y)$ be expressed as a function of the fraction of always-takers and the degree of violation of the IVQR assumptions:

$$B_{F^*_{Y_0}}(y) = \pi_a(1 - t) \left(F_{Y_1|a} \left(Q_{Y_1|c} \left(F_{Y_0|c}(y) \right) \right) - G(y) \right). \tag{10}$$

Equations (9) and (10), combined with similar expressions for $F_{Y_1|n}(y)$ and $B_{F^*_{Y_1}}(y)$, can be used to derive bounds for the IVQR estimands of the CDFs which imply bounds for the QFs, the QTE, and the ATE; see for instance Manski (2003).

Theorems 1 and 2 imply explicit characterizations of the IVQR QTE estimands for the always-takers and never-takers. Let $Q^*_{Y_1|n}$ and $Q^*_{Y_0|a}$ denote the QFs associated with $F^*_{Y_1|n}$ and $F^*_{Y_0|a}$ and define $\delta^*_n(\tau) \equiv Q^*_{Y_1|n}(\tau) - Q_{Y_0|n}(\tau)$ and $\delta^*_a(\tau) \equiv Q_{Y_1|a}(\tau) - Q^*_{Y_0|a}(\tau)$.

Corollary 2. *Under the assumptions of Theorem 1, for all $\tau \in (0, 1)$,*

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

Corollary 2 shows that the differences between the LQTE and the IVQR QTE estimands for the never-takers and always-takers are determined by differences between the identified potential outcome distributions as measured by the probability-probability transforms $F_{Y_0|c} \circ Q_{Y_0|n}$ and $F_{Y_1|c} \circ Q_{Y_1|a}$, but do not depend on the relative size of the respective subpopulations.

3.3 ATE

As a consequence of the results in the previous sections, the IVQR estimand of the ATE can be expressed as a convex combination of the LATE, Δ_c, and IVQR estimands of the ATE for always-takers and never-takers, which correspond to integrated LQTE at transformed quantile levels.

Theorem 3. *Suppose that Assumptions 2 and 3 hold and that the IVQR estimands solve the moment conditions* (5)*. Then*

$$\Delta^* = \pi_c \Delta_c + \pi_a \Delta_a^* + \pi_n \Delta_n^*,$$

where $\Delta_a^* \equiv \int_0^1 \delta_c \left(F_{Y_1|c} \left(Q_{Y_1|a}(\tau) \right) \right) d\tau$ *and* $\Delta_n^* \equiv \int_0^1 \delta_c \left(F_{Y_0|c} \left(Q_{Y_0|n}(\tau) \right) \right) d\tau$.

Theorem 3 implies that the difference between the IVQR ATE estimand and the LATE is equal to a weighted average of the differences between the IVQR ATE estimands for always-takers and never-takers and the LATE:

$$\Delta^* - \Delta_c = \pi_a \left(\Delta_a^* - \Delta_c \right) + \pi_n \left(\Delta_n^* - \Delta_c \right).$$

The differences $\Delta_a^* - \Delta_c$ and $\Delta_n^* - \Delta_c$ reflect differences between the subpopulation potential outcome distributions and are weighted by the proportions of always-takers and never-takers.

Moreover, Theorem 3 implies the following expression for the bias of the IVQR ATE estimand:

$$B_{\Delta^*} \equiv \Delta^* - \Delta = \pi_a \left(\Delta_a^* - \Delta_a \right) + \pi_n \left(\Delta_n^* - \Delta_n \right).$$

Thus, the bias B_{Δ^*} is equal to a weighted average of the differences between the IVQR estimands of the always-taker and never-taker ATE and the corresponding true ATEs.

Such differences are caused by violations of the IVQR assumptions; see Section 3.2 for a further discussion. The weights are equal to the proportions of always-takers and never-takers, which are the key observable determinants of the bias.

The IVQR estimand of the ATE inherits some of the robustness properties outlined in Corollary 1. Specifically, the sign of the IVQR estimand coincides with the sign of the LATE if the LQTE is strictly positive or strictly negative at all quantiles and the IVQR estimand is equal to the LATE if the LQTE is constant across quantiles.

3.4 One-sided non-compliance

The results in the previous sections have interesting implications when the instrument satisfies one-sided non-compliance as for example in empirical application in Section 5.2. Under one-sided non-compliance, only individuals with $Z = 1$ can choose $D = 1$.

Assumption 4. *One-sided non-compliance:* $p(1|0) = 0$.

Assumption 4 rules out the existence of both defiers and always-takers, which implies that the monotonicity Assumption 2.4—one of the main conditions underlying the analysis in this paper—holds by design. The key implication of Assumptions 2 and 4 is that $Q_{Y_0}(\tau)$ is identified as

$$Q_{Y_0}(\tau) = Q_{Y|D=0,Z=0}(\tau).$$

This allows for deriving more explicit analytical characterizations of the IVQR QTE estimands in Theorem 1, which shed light on the nonlinear interactions between the different determinants of the discrepancies between both models and the bias of IVQR.

Corollary 3. *Under the assumptions of Theorem 1 and Assumption 4, for all* $\tau \in (0, 1)$*,*

$$\begin{aligned} \delta^*(\tau) &= \delta_c\left(F_{Y_0|c}\left(Q_{Y_0}(\tau)\right)\right) \\ &= Q_{Y_1|c}\left(F_{Y_0|c}\left(Q_{Y_0}(\tau)\right)\right) - Q_{Y_0}(\tau) \\ &= Q_{Y_1|c}\left(\frac{\tau - (1-\pi_c)F_{Y_0|n}\left(Q_{Y_0}(\tau)\right)}{\pi_c}\right) - Q_{Y_0}(\tau) \end{aligned}$$

Corollary 3 shows that, under Assumption 4, the IVQR estimand $Q^*_{Y_0}$ is equal to the true QF, Q_{Y_0}, irrespective of the validity of the IVQR assumptions. As a consequence, the bias of $\delta^*(\tau)$ is uniquely determined by differences between $Q^*_{Y_1}$ and Q_{Y_1}:

$$B_{\delta^*}(\tau) \equiv \delta^*(\tau) - \delta(\tau) = Q_{Y_1|c}\left(\frac{\tau - (1-\pi_c)F_{Y_0|n}\left(Q_{Y_0}(\tau)\right)}{\pi_c}\right) - Q_{Y_1}(\tau).$$

The next corollary analyzes the implications of Theorem 2 under one-sided non-compliance.

Corollary 4. *Under the assumptions of Theorem 2 and Assumption 4, for all* $(y_1, y_0) \in \mathcal{Y}_1 \times \mathcal{Y}_0$,

$$\begin{aligned} F^*_{Y_1}(y_1) &= \pi_c F_{Y_1|c}(y_1) + \pi_n F_{Y_0|n}\left(Q_{Y_0|c}\left(F_{Y_1|c}(y_1)\right)\right), \\ F^*_{Y_0}(y_0) &= \pi_c F_{Y_0|c}(y_0) + \pi_n F_{Y_0|n}(y_0) = F_{Y_0}(y_0). \end{aligned}$$

Corollary 4 shows that, under one-sided non-compliance, the IVQR estimand $F^*_{Y_0}$ is unbiased and equal to the true CDF, F_{Y_0}. It is instructive to relate this result to the analysis in Section 3.2, where I derive the following expression for the bias of $F^*_{Y_0}$:

$$B_{F^*_{Y_0}}(y) = \pi_a\left(F_{Y_1|a}\left(Q_{Y_1|c}\left(F_{Y_0|c}(y)\right)\right) - F_{Y_0|a}(y)\right).$$

Under one-sided non-compliance, there are no always-takers, $\pi_a = p(1|0) = 0$, which implies that $B_{F^*_{Y_0}}(y) = 0$ and thus that $F^*_{Y_0}(y) = F_{Y_0}(y)$.

Finally, based on the previous results, one can derive the following corollary to Theorem 3, which provides alternative characterizations of the IVQR estimands of the ATE under one-sided non-compliance.

Corollary 5. *Under the assumptions of Theorem 3 and Assumption 4,*

$$\begin{aligned} \Delta^* &= \pi_c \Delta_c + \pi_n \Delta^*_n \\ &= \int_0^1 Q_{Y_1|c}\left(F_{Y_0|c}\left(Q_{Y_0}(\tau)\right)\right) d\tau - \mathbb{E}(Y_0) \\ &= \int_0^1 Q_{Y_1|c}\left(\frac{\tau - (1-\pi_c)F_{Y_0|n}\left(Q_{Y_0}(\tau)\right)}{\pi_c}\right) d\tau - \mathbb{E}(Y_0). \end{aligned}$$

3.5 Constructive identification through extrapolation

The results in the previous sections provide characterizations of the IVQR estimands in settings where the underlying assumptions are potentially violated. Here I discuss the implications of the main results for settings where the IVQR assumptions hold.

Under the IVQR assumptions, the key implications of the previous results are the following:

$$F_{Y_1|n}(y) = F_{Y_0|n}\left(Q_{Y_0|c}\left(F_{Y_1|c}(y)\right)\right) \quad \text{and} \quad F_{Y_0|a}(y) = F_{Y_1|a}\left(Q_{Y_1|c}\left(F_{Y_0|c}(y)\right)\right), \tag{11}$$

or, equivalently,

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

Thus, the quantile-quantile transforms

$$P_{01|t}(y) \equiv Q_{Y_0|t}\left(F_{Y_1|t}(y)\right) \quad \text{and} \quad P_{10|t}(y) \equiv Q_{Y_1|t}\left(F_{Y_0|t}(y)\right)$$

do not depend on the type $T = t$. Hence, one can use $P_{01|c}$ and $P_{10|c}$, both of which are identified under the LQTE assumptions, to impute the missing potential outcome distributions for always-takers and never-takers. This is exactly how the IVQR model achieves identification. Specifically, one first looks for the rank in the same-state outcome distribution of compliers, and then uses the same quantile of the complier distribution in the other treatment state to impute the missing counterfactuals. In other words, to obtain point identification for the whole population, IVQR extrapolates from the compliers based on the rank similarity assumption. Identification of the population treatment effects can be described as a two-step procedure:

1. Obtain the proportions of the three types π_c, π_n, and π_a as well as the subpopulation potential outcome CDFs $F_{Y_1|c}$, $F_{Y_1|a}$, $F_{Y_0|c}$, and $F_{Y_0|n}$ as described in Section 3.2 and compute the quantile-quantile transforms $P_{01|c}$ and $P_{10|c}$.

2. Impute missing counterfactual distributions $F_{Y_1|n}$ and $F_{Y_0|a}$ based on equation (11) and obtain F_{Y_1} and F_{Y_0} based on equations (6) and (7). Obtain Q_{Y_1} and Q_{Y_0} by inverting F_{Y_1} and F_{Y_0} and compute the τ-QTE and the ATE as $\delta(\tau) = Q_{Y_1}(\tau) - Q_{Y_0}(\tau)$ and $\Delta = \int_0^1 \delta(\tau) d\tau$.

Step 1 exploits the LQTE assumptions. Step 2 is based on the IVQR assumptions and specifically the rank similarity assumption. This suggests that IVQR constitutes a natural and constructive rank-based approach to external validity in the LQTE model; see

Angrist and Fernandez-Val (2013) and the references therein for alternative extrapolation approaches.

This discussion provides further intuition as to why the support Assumption 3.2 is essential for deriving the results in this paper. Without this condition, the quantile-quantile transforms $P_{01|c}$ and $P_{10|c}$ are not necessarily one-to-one and, consequently, $F_{Y_1|n}$ and $F_{Y_0|a}$ and thus F_{Y_1} and F_{Y_0} are no longer point identified. Hence, full support of the complier potential outcome distributions is the key condition for point identification. The rank-based extrapolation in Step 2 will generally fail if the supports of the complier distributions do not nest the supports of the same-state distributions of the always-takers and never-takers.

Finally, the analysis in this section shows that there is a close connection between IVQR and the chances-in-changes (CIC) model of Athey and Imbens (2006). Both models impute counterfactual quantities by restricting the evolution of the unobservables across treatment states (IVQR) and time (CIC). Moreover, in both models, identification is achieved by means of a specific subpopulation for which the potential outcome distributions of interest are identified in both treatment states (IVQR) and time periods (CIC).

4 Extensions

This section presents three extensions of the main results. I focus on the IVQR estimands of the QTE and the potential outcome CDFs. Results for ATE can be derived using similar arguments as in Section 3.3 and are thus omitted.

4.1 LQTE assumptions without monotonicity

The monotonicity assumption of the LQTE model is restrictive and likely to be implausible in many settings (e.g., De Chaisemartin, 2017b). By contrast, IVQR identifies causal effects irrespective of the validity of the monotonicity assumption. Motivated by this observation, I analyze a setting where only Assumptions 2.1–2.3 and 3 are maintained. Under these assumptions neither the proportion nor the potential outcome distribution of any single subpopulation are identified. In particular, the quantities that correspond to potential outcome CDFs for the compliers under monotonicity are equal to weighted differences between compliers and defiers:

$$\tilde{F}_{Y_1}(y) = \frac{\pi_c F_{Y_1|c}(y) - \pi_f F_{Y_1|f}(y)}{\pi_c - \pi_f} \equiv F_{Y_1|c-f}(y),$$
$$\tilde{F}_{Y_0}(y) = \frac{\pi_c F_{Y_0|c}(y) - \pi_f F_{Y_0|f}(y)}{\pi_c - \pi_f} \equiv F_{Y_0|c-f}(y).$$

Henceforth, I refer to the mixture subpopulation corresponding to this mixture distribution as the *compliers-defiers*. De Chaisemartin (2017a,b) has shown that under a weaker version of the LQTE monotonicity assumption, $F_{Y_d|c-f} = F_{Y_d|cV}$, where $F_{Y_d|cV}$ is the CDF of Y_d for a subpopulation of the compliers that he refers to as the *comvivors*. Let $f_{Y_1|c-f}$ and $f_{Y_0|c-f}$ denote the derivatives of $F_{Y_1|c-f}$ and $F_{Y_0|c-f}$, let $Q_{Y_1|c-f}$ and $Q_{Y_0|c-f}$ denote the inverses of $F_{Y_1|c-f}$ and $F_{Y_0|c-f}$, and define the τ-QTE for the compliers-defiers as $\delta_{c-f}(\tau) \equiv Q_{Y_1|c-f}(\tau) - Q_{Y_0|c-f}(\tau)$.

The following assumption generalizes Assumption 3 to allow for failures of monotonicity.

Assumption 5. *For $(d,t) \in \{0,1\} \times \{a,c,n,f\}$:*

1. *Regularity: $f_{Y_d|t}$ is continuous.*

2. *Support: $\mathcal{S}(Y_d|T=t) = \mathcal{S}(Y_d)$ and $f_{Y_d|c-f}(y_d) > 0$ for $y_d \in \mathcal{S}(Y_d)$, where $\mathcal{S}(Y_d) = \mathcal{S}_d$.*

Under Assumptions 2.1–2.3, $F_{Y_1|c-f}$ and $F_{Y_0|c-f}$ are not guaranteed to be monotonic. However, assuming that $F_{Y_1|c-f}$ and $F_{Y_0|c-f}$ are strictly increasing is essential for deriving the closed-form solutions for the IVQR estimands and is implied by the identification conditions in Chernozhukov and Hansen (2005); see Section 1 of the appendix.

The next theorem characterizes the IVQR QTE estimands under the LQTE assumptions without monotonicity.

Theorem 4. *Suppose that Assumptions 2.1–2.3 and 5 hold and that the IVQR estimands solve the moment conditions (5). Then, for all $\tau \in (0,1)$,*

$$\delta^*(\tau) = \delta_{c-f}\left(F_{Y_0|c-f}\left(Q^*_{Y_0}(\tau)\right)\right) = \delta_{c-f}\left(F_{Y_1|c-f}\left(Q^*_{Y_1}(\tau)\right)\right).$$

Theorem 4 shows that the IVQR estimands are equivalent to QTE for the compliers-defiers at transformed quantile levels. The next theorem provides closed-form solutions for the IVQR estimands of the potential outcome CDFs.

Theorem 5. *Suppose that Assumptions 2.1–2.3 and 5 hold and that the IVQR estimands solve the moment conditions (5). Then, for all* $(y_1, y_0) \in \mathcal{Y}_1 \times \mathcal{Y}_0$,

$$
\begin{aligned}
F^*_{Y_1}(y_1) &= \pi_c F_{Y_1|c}(y_1) + \pi_a F_{Y_1|a}(y_1) + \pi_n F_{Y_0|n}\left(Q_{Y_0|c-f}\left(F_{Y_1|c-f}(y_1)\right)\right) \\
&\quad + \pi_f F_{Y_0|f}\left(Q_{Y_0|c-f}\left(F_{Y_1|c-f}(y_1)\right)\right), \\
F^*_{Y_0}(y_0) &= \pi_c F_{Y_0|c}(y_0) + \pi_a F_{Y_1|a}\left(Q_{Y_1|c-f}\left(F_{Y_0|c-f}(y_0)\right)\right) + \pi_n F_{Y_0|n}(y_0) \\
&\quad + \pi_f F_{Y_1|f}\left(Q_{Y_1|c-f}\left(F_{Y_0|c-f}(y_0)\right)\right).
\end{aligned}
$$

Theorems 4 and 5 constitute natural generalizations of Theorems 1 and 2. Without monotonicity, the role of the compliers is played by the compliers-defiers.

4.2 Multivalued instruments

Suppose that instead of being binary, the instrument Z takes values in a finite set $\mathcal{Z} = \{z_1, z_2, \ldots, z_K\}$ with $0 < z_1 < z_2 < \cdots < z_K < \infty$. The following assumption extends the LQTE model as defined by Assumption 2 to multivalued instruments (e.g., Imbens, 2007; Frölich, 2007).

Assumption 6.

1. *Monotonicity:* $P(D_{z_k} \geq D_{z_j}) = 1$ *for any two values* $(z_k, z_j) \in \mathcal{Z} \times \mathcal{Z}$ *with* $z_k > z_j$.
2. *Independence:* $(Y_1, Y_0, \{D_{z_k}\}_{z_k \in \mathcal{Z}})$ *are jointly independent of* Z.
3. *Nontrivial assignment:* $0 < P(Z = z_k) < 1$ *for all* $z_k \in \mathcal{Z}$.
4. *First-stage:* $P(D = 1|Z = z_k) > P(D = 1|Z = z_j)$ *for any two values* $(z_k, z_j) \in \mathcal{Z} \times \mathcal{Z}$ *with* $z_k > z_j$.

Under Assumption 6, there are $K + 1$ different types. Consistent with the standard LQTE model, I denote individuals with $D_{z_k} = 0$ for all $z_k \in \mathcal{Z}$ as never-takers and individuals with $D_{z_k} = 1$ for all $z_k \in \mathcal{Z}$ as always-takers. In addition, there are $K - 1$ different types of compliers, $T = c_{z_j}$, indexed by a unique instrument value z_j at which their treatment status switches from zero to one. For each type $T = t$, $t \in \{a, n, \{c_{z_j}\}_{j=2}^K\}$, let π_t, $f_{Y_d|t}$, $F_{Y_d|t}$, and $Q_{Y_d|t}$ denote its proportion, the density function of Y_d, the CDF of Y_d, and the QF of Y_d.

Under Assumption 6, the conditional probabilities $p(d|z_k) \equiv P(D = d|Z = z_k)$ are related to the proportions of types as

$$p(1|z_k) = \pi_a + \sum_{j=2}^{k} \pi_{c_{z_j}} \quad \text{and} \quad p(0|z_k) = \pi_n + \sum_{j=k+1}^{K} \pi_{c_{z_j}}$$

and the observed conditional CDFs $F_{Y|D=d,Z=z_k}$ are related to the potential outcome CDFs of the $K+1$ types as (provided that these quantities are well defined for all $z_k \in \mathcal{Z}$)

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

Hence, the data are informative about the fraction of all subpopulations as well as the distributions of Y_1 for always-takers and compliers and the distributions of Y_0 for never-takers and compliers (e.g., Imbens, 2007).

I study IVQR estimands $\delta^*(\tau)$ and $Q^*_{Y_0}(\tau)$ that solve the following moment equations:

$$\mathbb{E}\left(\left(\tau - 1\left[Y \leq \delta^*(\tau)D + Q^*_{Y_0}(\tau)\right]\right)(1, Z)'\right) = 0, \quad \tau \in (0,1). \tag{12}$$

Remark 3. *Equation (12) is not the only possible moment condition that could be used with multivalued instruments. All instrument of the form $1[Z = z_k]$ are valid instruments and any combination of instruments identifies the same parameter under the IVQR assumptions. This testable implication can be used to develop overidentication-type tests of the IVQR model and the rank similarity assumption with nonbinary instruments (e.g., Kim and Park, 2017; Yu, 2017; Wüthrich, 2017); see Section 2 of the appendix.*

The following assumption extends Assumption 3 to multivalued instruments.

Assumption 7. *For $(d,t) \in \{0,1\} \times \left\{a, \{c_{z_j}\}_{j=2}^K, n\right\}$:*

1. *Regularity: $f_{Y_d|t}$ is continuous.*

2. *Support: $\mathcal{S}(Y_d|T = t) = \mathcal{S}(Y_d)$, where $\mathcal{S}(Y_d) = \mathcal{S}_d$.*

With multivalued instruments, the IVQR estimands will be functions of the following mixtures distributions:

$$\tilde{F}_{Y_1}(y) \equiv \frac{\sum_{j=2}^{K} w_j \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y)}{\sum_{j=2}^{K} w_j \pi_{c_{z_j}}} \quad \text{and} \quad \tilde{F}_{Y_0}(y) \equiv \frac{\sum_{j=2}^{K} w_j \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y)}{\sum_{j=2}^{K} w_j \pi_{c_{z_j}}},$$

where

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

Let $\tilde{Q}_{Y_1}$ and $\tilde{Q}_{Y_0}$ denote the inverses of $\tilde{F}_{Y_1}$ and $\tilde{F}_{Y_0}$. The mixture weight attached to the CDF of compliant type $T = c_j$, $F_{Y_d|c_{z_j}}$, is determined by two components: (i) the weighting function w_j and (ii) 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 $j \geq 2$. To gain some intuition about the shape of the weighting function, consider the difference between two adjacent weights:

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

Thus, the weighting function is increasing whenever z_j is smaller than $\mathbb{E}(Z)$ and decreasing whenever z_j is larger than $\mathbb{E}(Z)$.

Remark 4. *The weights w_j are similar to the weights that linear IV models attach to the LATE for compliers $T = c_{z_j}$ (e.g., Heckman and Vytlacil, 2005, 2007).*

The following theorem shows that, under the generalized LQTE assumptions, the IVQR QTE estimands can be expressed as QTE for the mixture of all compliant types, $\tilde{\delta}(\tau) \equiv \tilde{Q}_{Y_1}(\tau) - \tilde{Q}_{Y_0}(\tau)$, at transformed quantile levels.

Theorem 6. *Suppose that Assumptions 6 and 7 hold and that the IVQR estimands solve the moment conditions (12). Then, for all $\tau \in (0,1)$,*

$$\delta^*(\tau) = \tilde{\delta}\left(\tilde{F}_{Y_0}\left(Q^*_{Y_0}(\tau) \right) \right) = \tilde{\delta}\left(\tilde{F}_{Y_1}\left(Q^*_{Y_1}(\tau) \right) \right).$$

Theorem 7 generalizes Theorem 2 to multivalued instruments.

Theorem 7. *Suppose that Assumptions 6 and 7 hold and that the IVQR estimands solve the moment conditions (12). Then, for all $(y_1, y_0) \in \mathcal{Y}_1 \times \mathcal{Y}_0$,*

$$\begin{aligned}
F^*_{Y_1}(y_1) &= \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y_1) P(Z \geq z_j) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}} \left(\tilde{Q}_{Y_0} \left(\tilde{F}_{Y_1}(y_1) \right) \right) P(Z < z_j) \\
&\quad + \pi_a F_{Y_1|a}(y_1) + \pi_n F_{Y_0|n} \left(\tilde{Q}_{Y_0} \left(\tilde{F}_{Y_1}(y_1) \right) \right), \\
F^*_{Y_0}(y_0) &= \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}} \left(\tilde{Q}_{Y_1} \left(\tilde{F}_{Y_0}(y_0) \right) \right) \frac{\mathbb{E}(Z|Z \geq z_j)}{\mathbb{E}(Z)} P(Z \geq z_j) \\
&\quad + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}}(y_0) \frac{\mathbb{E}(Z|Z < z_j)}{\mathbb{E}(Z)} P(Z < z_j) + \pi_a F_{Y_1|a} \left(\tilde{Q}_{Y_1} \left(\tilde{F}_{Y_0}(y_0) \right) \right) + \ \pi_n F_{Y_0|n}(y)
\end{aligned}$$

Theorem 7 shows that the basic mechanism described in Theorem 2 pertains when the instrument is multivalued. The key CDFs, $\tilde{F}_{Y_1}$ and $\tilde{F}_{Y_0}$, are convex combinations of the CDFs of the $K-1$ different compliers. In contrast to the previous results, Theorem 7 implies that the transformations of the quantile levels in Theorem 6 do not only reflect differences between the potential outcome distributions of the untreated compliers and never-takers and treated compliers and always-takers, but also differences between the potential outcome distributions of the $K-1$ different compliers and their mixture.

4.3 Incorporating covariates into the analysis

Including additional covariates X into the analysis can be important for at least three reasons. First, conditioning on a set of covariates may be crucial 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 may be interesting to consider conditional QTE; see for instance Frölich and Melly (2013) for a discussion of the differences between conditional and unconditional QTE. All the results in this paper can be taken to hold conditional on covariates. With discrete X the analysis can proceed within 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 as in Chernozhukov and Hansen (2006) and Abadie et al. (2002). Such models can be interpreted as approximations to the true potentially nonlinear conditional QFs. Because the results obtained in the previous sections are fully nonparametric, they can be expected to hold approximately. The quality of these approximations can be improved by choosing richer specifications (e.g., through interactions, polynomials, or splines).

5 Empirical applications

This section illustrates the theoretical results using two empirical applications.

5.1 Implementation details

Here I provide a brief overview over the estimation and inference procedures used in the empirical applications. Section 3 of the appendix contains more details.

- The IVQR QTE are estimated using the plug-in approach by Wüthrich (2017).
- The complier QFs and the LQTE are estimated by inverting the sample analogs of the complier CDFs. To deal with the potential lack of monotonicity, I rearrange the original estimates as suggested by Chernozhukov et al. (2010).
- Pointwise confidence intervals for the IVQR QTE estimands and the LQTE are obtained using the empirical bootstrap. Validity of the bootstrap follows from the results in Wüthrich (2017).
- In both applications, I employ Kolmogorov–Smirnov (KS) and Cramér–von Mises (CM) tests to formally assess the null hypothesis that the estimands of both models are equivalent:

$$H_0: \ \delta^*(\tau) = \delta_c(\tau) \quad \text{for each} \quad \tau \in \mathcal{T}, \tag{13}$$

  where $\mathcal{T}$ denotes a set of quantile indices. Critical values are computed using the empirical bootstrap.

- The CDFs and QFs for the always-takers and never-takers are computed using the empirical distribution and quantile functions in the subsamples with $(D = 1, Z = 0)$ and $(D = 0, Z = 1)$. The probability-probability transforms and the IVQR QTE for the always-takers and never-takers are computed using the sample analogs of the corresponding expressions in Theorem 2 and Corollary 2.
- To estimate the IVQR ATE for the always-takers and never-takers, I use trimmed versions of the sample analogs of the expressions in Theorem 3. The LATE is estimated using 2SLS.

5.2 JTPA

I consider the 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 20000 participants in 16 different sites. I use the randomized assignment as an instrument (Z) for estimating the causal effect of actual participation in training programs (D) on the sum of earnings in the 30 months after the random assignment (Y). About 38% of the men in the sample who received a training offer chose not to participate in the training program. Less than 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 (Assumption 4) almost perfectly. For the purpose of illustration, I drop these observations from the sample, which yields a total of $N = 5083$ observations. There are 62% compliers and 38% never-takers. Abadie et al. (2002) provide additional information about the dataset and present descriptive statistics. In this application, the LQTE assumptions are plausible by design since the instrument is randomly assigned and satisfies one-sided non-compliance, while the IVQR assumptions are unlikely to hold as discussed below.

Figure 1 presents the empirical results. Panel A compares the QTE estimates based on the IVQR and the LQTE model. The estimates are very similar and are characterized by substantial effect heterogeneity across quantiles and overall increasing QTE estimates ranging from around zero to over 4000 USD. The null hypothesis (13) cannot be rejected based on the KS and the CM test: $T_N^{KS} = 6.24 \cdot 10^4$ ($p = 0.996$) and $T_N^{CM} = 7.06 \cdot 10^8$ ($p = 0.991$).

The key condition of the IVQR model is the rank similarity assumption. In this application, rank similarity requires that the individual ranks in the potential wage distributions with and without the training program are the same up to unsystematic deviations from a common rank level, postulating a "one-factor" model in which counterfactual earnings are determined by a single unobservable "ability" index. Dong and Shen (2018) statistically reject rank similarity based on a similar dataset. It follows from the analysis in Section 3.2 and Corollaries 3–5 that the key observable determinant of the bias of IVQR in the absence of rank similarity is the size of the never-taker subpopulation, which is related to the first stage as $\pi_n = 1 - \pi_c$. The estimated proportion of never-takers and the first stage are $\hat{\pi}_n = 0.38$ and $\hat{\pi}_c = 0.62$, suggesting that IVQR is relatively robust to violations of rank similarity in this application.

Corollaries 3–5 show that, under one-sided non-compliance, differences between both models are determined by the proportions of never-takers and compliers and differences between the distributions of Y_0 of never-takers and compliers. Panel B shows that the estimated CDFs of Y_0 of never-takers and compliers exhibit substantial differences at the lower quantiles. These differences lead to deviations of the estimated probability-probability transform $\hat{F}_{Y_0|c} \circ \hat{Q}_{Y_0|n}$ from the 45°-line in Panel C. Panel D compares the IVQR QTE estimate for the never-takers to the LQTE. Although qualitatively similar, the never-taker QTE are smaller than the corresponding LQTE at most quantiles. The reason for this finding is the combination of the increasing LQTE and the shape of $\hat{F}_{Y_0|c} \circ \hat{Q}_{Y_0|n}$, which implies that the τ-QTE for the never-taker corresponds to the τ'-LQTE, where $\tau' < \tau$.

Finally, I compute the IVQR estimates of the never-taker ATE and the overall ATE based on Corollary 5. The estimated never-taker ATE is $\hat{\Delta}_n^* = 1275.48$ and thus smaller than the estimated LATE, $\hat{\Delta}_c = 1715.96$. This is because the QTE for the never-takers is

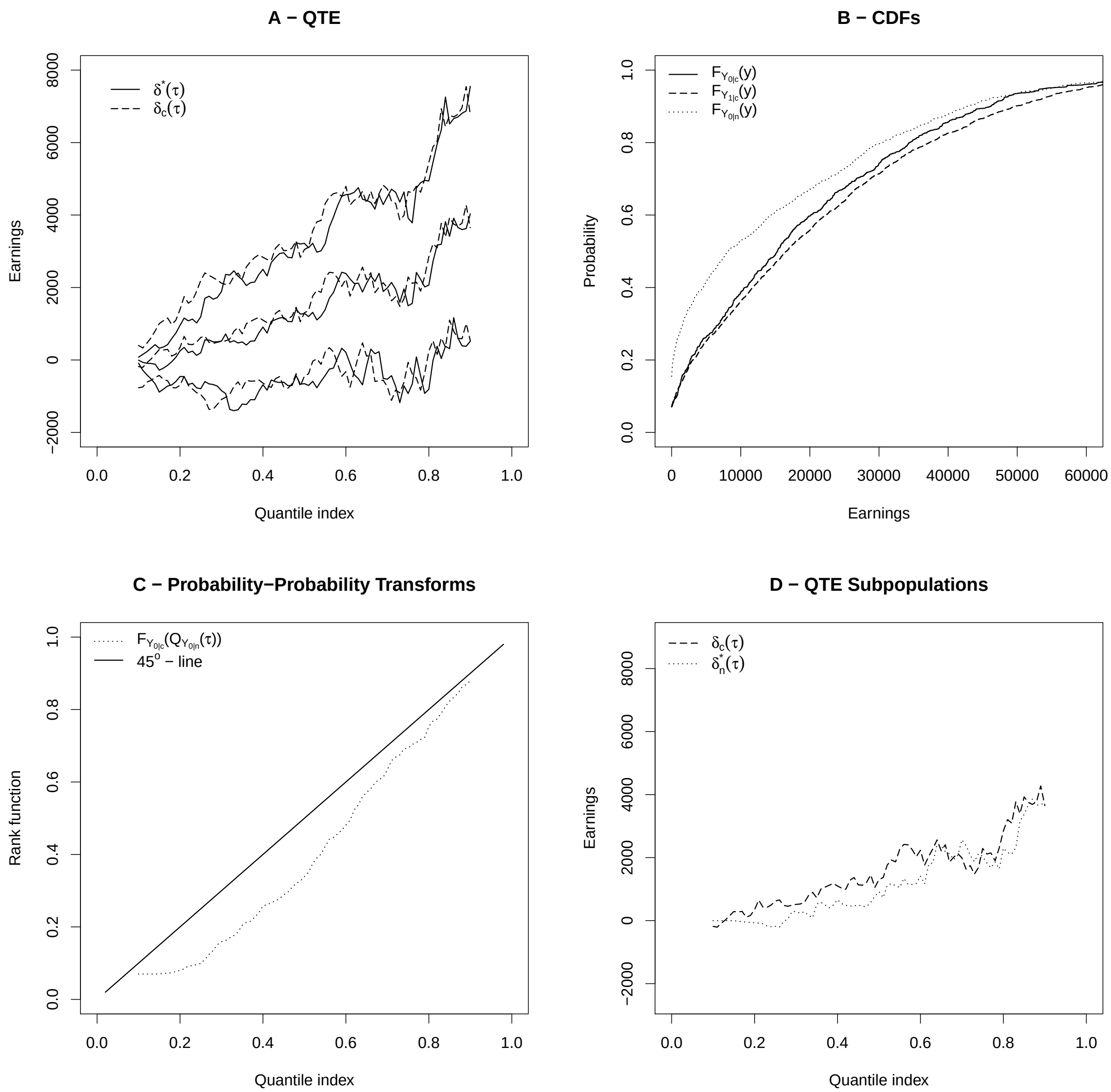

Figure 1: Panel A: QTE estimates including 90% confidence intervals computed using the empirical bootstrap with 1000 repetitions. Panel B: subpopulation CDFs. Panel C: probability-probability transform. Panel D: subpopulation QTE estimates.

smaller than the LQTE at most quantiles in Panel D. The overall ATE estimate based on the IVQR model is

$$\hat{\Delta}^* = \hat{\pi}_c \hat{\Delta}_c + \hat{\pi}_n \hat{\Delta}_n^* = 1549.82.$$

5.3 Veteran status and earnings

Here I study 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 is eligible for being drafted and zero otherwise. I use the same dataset as Abadie (2002) and Chernozhukov et al. (2010). The data contain information about $N = 11637$ 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 12% compliers (fractions are rounded). Abadie (2002) provides more information about the dataset.

Figure 2 presents the empirical results. Panel A compares the QTE estimates based on the IVQR and the LQTE model. The estimates exhibit similar overall patterns indicating that there is substantial effect heterogeneity across quantiles. In particular, there are large negative effects at the lower quantiles of the wage distribution and small positive impacts at higher quantiles. In contrast to the JTPA application in Section 5.2, there are economically significant differences between the QTE estimates up to the third quartile. The null hypothesis (13) is marginally rejected at the 10%-level based on the KS-test ($T_N^{KS} = 7.24 \cdot 10^5$, p-value $= 0.098$), but cannot be rejected based on the CM-test ($T_N^{CM} = 5.15 \cdot 10^{10}$, p-value $= 0.133$).

In this application, rank similarity requires that individual ranks in the potential wage distributions as veterans and non-veterans are the same up to unsystematic deviations from a common rank level. Rank similarity would be violated if there are several unobservable factors that determine wages and are differentially relevant under both treatment states such that a one-factor model is not appropriate. Examples of such wage determinants include motivation, perseverance, intelligence, communication skills, physical skills, resilience, and loyalty. Theorems 1–3 show that, if rank similarity is violated, the key ob-

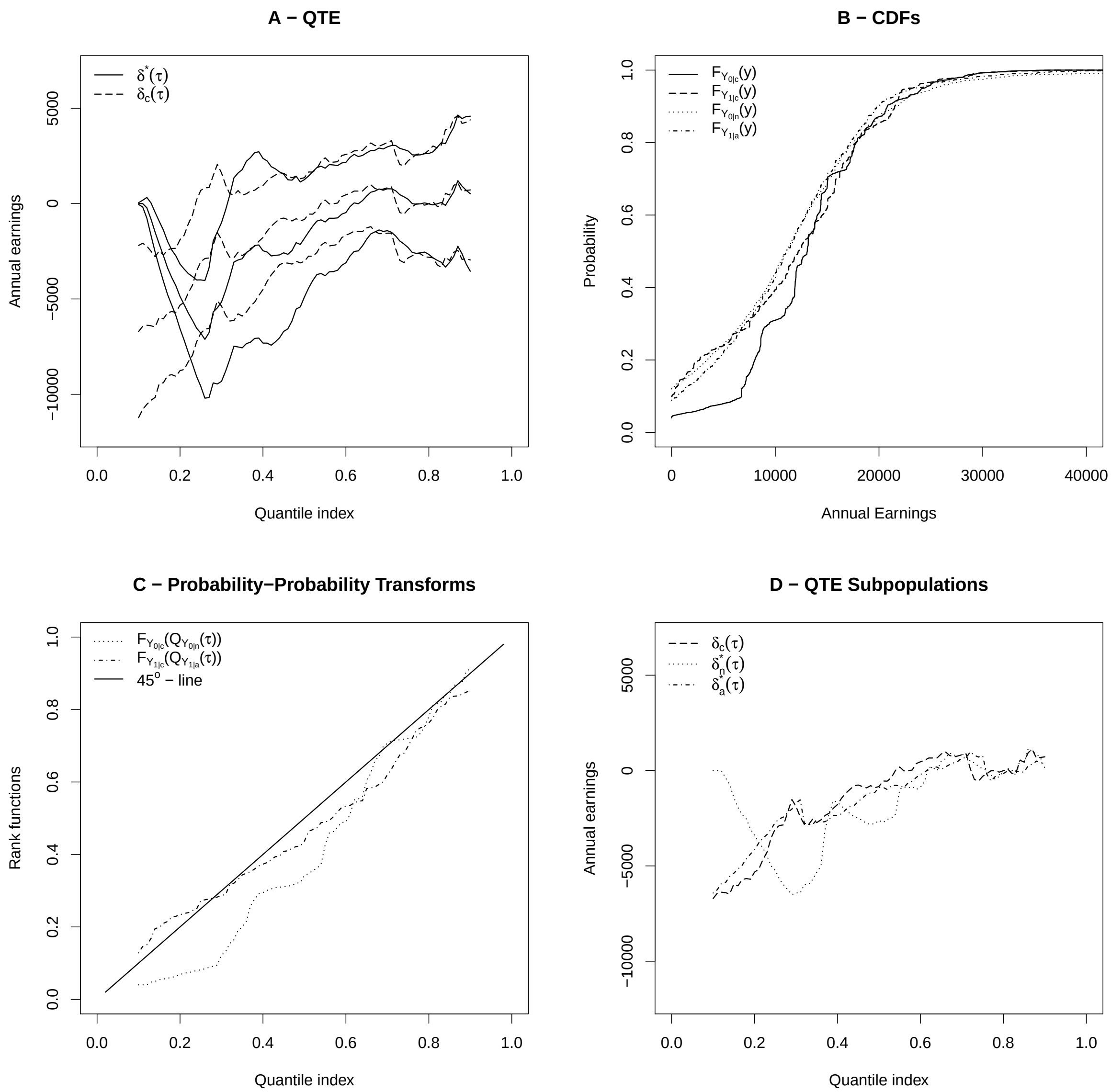

Figure 2: Panel A: QTE estimates including 90% confidence intervals computed using the empirical bootstrap with 1000 repetitions. Panel B: subpopulation CDFs. Panel C: probability-probability transform. Panel D: subpopulation QTE estimates.

servable determinant of the bias of IVQR is the relative size of the three subpopulations. Because the majority of the individuals (89%) are never-takers and always-takers in this application, and, consequently, there are only very few compliers (12%), IVQR cannot be expected to be very robust against violations of rank similarity.

It follows from Theorems 1–2 that the discrepancies between the QTE estimands of both models are determined by the proportions of the three types and differences between the subpopulation potential outcome distributions. Panel B shows that there are pronounced differences between the estimated CDFs of Y_0 of never-takers and compliers at the lower quantiles, while the differences at the higher quantiles are smaller. The differences at the lower quantiles matter because there are 71% never-takers. By contrast, the estimated CDFs of Y_1 of always-takers and compliers are rather similar and their discrepancies are not as important since there are only 18% always-takers. Panel C reports estimates of the probability-probability transforms $F_{Y_1|c} \circ Q_{Y_1|a}$ and $F_{Y_0|c} \circ Q_{Y_0|n}$. Panel D compares the IVQR QTE estimates for the never-takers and always-takers to the LQTE. The discrepancies between the never-taker and complier CDFs in Panel B result in large deviations of $\hat{F}_{Y_0|c} \circ \hat{Q}_{Y_0|n}$ from the 45°-line in Panel C. In conjunction with the substantial heterogeneity in the LQTE at the lower quantiles, this leads to the large differences between the never-taker QTE and the LQTE in Panel D. By contrast, the differences between the always-taker QTE and the LQTE are rather small, reflecting the similarities between the always-taker and complier distributions.

The IVQR ATE estimates for the always-takers and never-takers are $\hat{\Delta}_a^* = -1666.71$ and $\hat{\Delta}_n^* = -1401.26$. Both estimates are larger (in absolute value) than the LATE, $\hat{\Delta}_c = -1277.78$, reflecting the differences between the QTE estimates reported in Panel D of Figure 2. The IVQR estimate of the overall ATE is

$$\hat{\Delta}^* = \hat{\pi}_c \hat{\Delta}_c + \hat{\pi}_a \hat{\Delta}_a^* + \hat{\pi}_n \hat{\Delta}_n^* = -1434.58.$$

When interpreting the results in this application, it is important to note that the validity of the instrument exclusion restriction underlying both models is questionable due to the direct effect of draft lottery on college enrollments (e.g., Card and Lemieux, 2001; Deuchert and Huber, 2017). If the exclusion restriction is violated, $\tilde{F}_{Y_1}$ and $\tilde{F}_{Y_0}$ are not equal to $F_{Y_1|c}$ and $F_{Y_0|c}$, but correspond to mixtures of instrument- and type-specific potential outcome CDFs (e.g., Huber, 2014). Because the closed-form solutions for the IVQR estimands in Lemma 1 of the appendix do not rely on the validity of the exclusion restriction, they can be used to characterize the IVQR estimands when this condition is violated. This analysis suggests that the IVQR QTE estimands correspond to QTE associated with mixture distributions at transformed quantile levels. Discrepancies between the estimands of both models reflect differences between the instrument- and type-specific potential outcome distributions, which are partly driven by violations of the exclusion restriction. A full theoretical analysis of the IVQR and the LQTE estimands under violations of the exclusion restriction is beyond the scope of this paper, but is certainly worth pursuing in future research.

6 Conclusion

In this paper, I characterize the IVQR estimands under the LQTE assumption when treatment effect heterogeneity is unrestricted. I show that even when the rank similarity assumption fails, the IVQR QTE estimands correspond to LQTE at transformed quantile levels. Moreover, the IVQR estimate of the ATE is equal to a convex combination of the LATE and weighted averages of integrated LQTE at transformed quantile levels. Underpinning these results are closed-form representations for the IVQR estimands, which may be of independent interest. For example, Wüthrich (2017) builds on these closed-form solutions to construct plug-in estimators for unconditional QTE.

I analyze IVQR estimands with binary treatments and binary instrument as well as multivalued instruments based on exactly identified GMM problems, but I do not cover nonbinary treatments, continuous instruments, and overidentified GMM. Analyzing the IVQR estimands with nonbinary treatments would be particularly insightful as there is no straightforward way to generalize the LQTE model to accommodate nonbinary treatment variables. Alas, such an extension cannot be based on the closed-form solutions in this paper as they rely on the binary nature of the treatment variable. The arguments used to derive the closed-form solutions further rely on exact identification and do not directly extend to overidentified GMM problems. In contrast, incorporating continuous instruments into the analysis is conceptually straightforward based on Lemma 2 in the appendix; see for example Yu (2016).

For the empirical practitioner, I would like to emphasize several take-aways messages of this paper. First, the IVQR estimands have a meaningful interpretation even when the controversial rank similarity assumption is violated. Second, the key observable determinants of the bias of the IVQR estimands are the proportions of always-takers and never-takers and the instrument first stage. Hence, it is important to report estimates of these key quantities in applications. Third, I show that IVQR achieves identification of population treatment effects by extrapolating from the compliers based on the rank similarity assumption. Hence, IVQR can be viewed as a rank-based approach to external validity in the LQTE model. Fourth, in settings where the assumptions of both models are plausible, one can use the results in this paper to estimate treatment effects for the always-takers and never-takers. Such an analysis provides valuable information about the treatment effect heterogeneity across subpopulations. Finally, based on the results in this paper, Kim and Park (2017) show that rank similarity is fundamentally untestable with binary instruments. Thus, comparisons of the estimands of both models cannot be used to assess the validity of this key condition. By contrast, rank similarity is testable with nonbinary instruments; see Section 2 of the appendix.

Acknowledgements

This paper is based on Chapter 2 of my PhD Dissertation at the University of Bern. I have benefited from numerous discussions with Blaise Melly. I am grateful to Alberto Abadie, Josh Angrist, Andreas Bachmann, Stefan Boes, Daniel Burkhard, Victor Chernozhukov, Iván Fernández-Val, Dennis Kristensen, Tobias Müller, Klaus Neusser, two anonymous referees, the Associate Editor, the Editor, and seminar participants for very helpful comments. I would like to thank Alberto Abadie for sharing the data for the empirical application. This research was supported by the Swiss National Science Foundation (Doc.Mobility Project P1BEP1_155467). All errors are my own.

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

Angrist, J. and I. Fernandez-Val (2013). ExtrapoLATE - ing: External validity and overidentification in the LATE framework. In D. Acemoglu, M. Arellano, and E. Dekel (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., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*(434), pp. 444–455.

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

Belloni, A., V. Chernozhukov, I. Fernandez-Val, and C. Hansen (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica 85*(1), 233–298.

Bloom, H. S., L. L. Orr, S. H. Bell, G. Cave, F. Doolittle, W. Lin, and J. M. Bos (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.

Card, D. and T. Lemieux (2001). Going to college to avoid the draft: The unintended legacy of the vietnam war. *The American Economic Review 91*(2), 97–102.

Chen, L.-Y. and S. Lee (2018). Exact computation of gmm estimators for instrumental variable quantile regression models. *Journal of Applied Econometrics 33*(4), 553–567.

Chen, X. and D. Pouzo (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *Journal of Econometrics 152*(1), pp. 46–60.

Chen, X. and D. Pouzo (2012). Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica 80*(1), pp. 277–321.

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

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

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

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

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

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

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

Chernozhukov, V., C. Hansen, and K. Wüthrich (2017). Instrumental variable quantile regression. In V. Chernozhukov, X. He, R. Koenker, and L. Peng (Eds.), *Handbook of Quantile Regression*, pp. pp. 119–143. CRC Chapman-Hall.

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

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

De Chaisemartin, C. (2017a). Supplement to "tolerating defiance? local average treatment effects without monotonicity".

De Chaisemartin, C. (2017b). Tolerating defiance? local average treatment effects without monotonicity. *Quantitative Economics 8*(2), 367–396.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature 48*(2), pp. 424–55.

Deuchert, E. and M. Huber (2017). A cautionary tale about control variables in IV estimation. *Oxford Bulletin of Economics and Statistics 79*(3), 411–425.

Dong, Y. and S. Shen (2018). Testing for rank invariance or similarity in program evaluation. *The Review of Economics and Statistics 100*(1), 78–85.

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

Frölich, M. (2007). Nonparametric IV estimation of local average treatment effects with covariates. *Journal of Econometrics 139*(1), pp. 35–75.

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

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

Heckman, J. J. and S. Urzua (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics 156*(1), pp. 27–37.

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

Heckman, J. J. and E. J. Vytlacil (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. Volume 6, Part B of *Handbook of Econometrics*, pp. 4875–5143. Elsevier.

Horowitz, J. L. and S. Lee (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.

Huber, M. and K. Wüthrich (2018). Local average and quantile treatment effects under endogeneity: A review. forthcoming at the Journal of Econometric Methods.

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

Imbens, G. W. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature 48*(2), pp. 399–423.

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

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

Kaplan, D. M. and Y. Sun (2017). Smoothed estimating equations for instrumental variables quantile regression. *Econometric Theory 33*(1), 105?157.

Kim, J. H. and B. G. Park (2017). Testing rank similarity in the heterogeneous treatment effect model. Unpublished Manuscript.

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

Manski, C. F. (2003). *Partial Identification of Probability Distributions.* Springer, New York, NY.

Masten, M. and A. Poirier (2017). Inference on breakdown frontiers. Unpublished Manuscript.

Melly, B. and K. Wüthrich (2017). Local quantile treatment effects. In V. Chernozhukov, X. He, R. Koenker, and L. Peng (Eds.), *Handbook of Quantile Regression*, pp. pp. 145–164. CRC Chapman-Hall.

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. and T. Hoshino (2016). Sieve instrumental variable quantile regression estimation of functional coefficient models. *Journal of Econometrics 191*(1), 231 – 254.

Vuong, Q. and H. Xu (2017). Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity. *Quantitative Economics 8*(2), 589–610.

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

Wüthrich, K. (2017). A closed-form estimator for quantile treatment effects with endogeneity. Unpublished Manuscript.

Yu, P. (2016). Treatment effects estimators under misspecification. Unpublished Manuscript.

Yu, P. (2017). Testing conditional rank similarity with and without covariates. Unpublished Manuscript.