A comparison of two quantile models with endogeneity*

Kaspar Wüthrich

September 19, 2015

Abstract

This paper compares the two most used instrumental variable models for estimating quantile treatment effects (QTE): the instrumental variable quantile regression (IVQR) model and the local quantile treatment effects (LQTE) model. On the surface, these methods do not seem to be connected. Both models rely on different non-nested assumptions and identify treatment effects for different populations. The IVQR model identifies population QTE by imposing rank similarity in the outcome equation, an assumption that restricts treatment effect heterogeneity. In contrast, the LQTE model achieves identification through a monotonicity assumption in the selection equation. However, as this model allows for unrestricted treatment effect heterogeneity, it only identifies LQTE for the compliers. This paper shows that even if the rank similarity assumption fails, the IVQR QTE estimators are consistent for LQTE at transformed quantile levels. Moreover, the IVQR estimand of the average treatment effect can be decomposed into a convex combination of the local average treatment effect and a weighted average of QTE for the compliers. These results establish a formal relationship between both models and provide a characterization of the IVQR estimands under misspecification. Underpinning the analysis are novel closed form representations of the IVQR estimands of the potential outcome distributions.

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, Andreas Bachmann, Stefan Boes, Daniel Burkhard, Victor Chernozhukov, Iván Fernández-Val, Tobias Müller, Klaus Neusser, and seminar participants at the University of Bern (2014), the MIT Econometrics Lunch (2014), and the Ski & Labor Workshop in Laax (2015) for helpful comments. I would like to thank Alberto Abadie for sharing the data for the empirical application. The research was supported by the Swiss National Science Foundation (Doc.Mobility Project P1BEP1_155467). All errors are my own. The usual disclaimer applies.

†University of Bern, Department of Economics, Schanzeneckstrasse 1, CH-3001 Bern, Switzerland, phone: +41 79 830 7302, email: kaspar.wuethrich@vwi.unibe.ch.
1 Introduction

The ability of quantile treatment effects (QTE) to characterize the impact of policy variables on distributional outcomes beyond simple averages makes them appealing in many applications. Just as with ordinary least squares, endogeneity of the policy variables renders quantile regression inconsistent for estimating QTE. One approach to addressing this problem are instrumental variable (IV) methods. The two most popular IV quantile models are the local quantile treatment effect (LQTE) model (Abadie et al., 2002)\(^1\) and the instrumental variable quantile regression (IVQR) model (Chernozhukov and Hansen, 2005). In the LQTE model identification is achieved through a monotonicity assumption in 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. However, the LQTE model has been attacked 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 in Imbens (2010), Deaton (2010), and Heckman and Urzua (2010). The key condition underlying the IVQR model is the rank similarity assumption, a restriction on the evolution of individual ranks across treatment states. In sharp contrast to the LQTE model, the IVQR model point identifies the QTE for the overall population. While rank similarity has substantial identifying power, it is a strong and controversial assumption that substantially restricts the treatment effect heterogeneity. For instance, it excludes scenarios where individuals select into treatment based on private information about comparative advantages (Frandsen and Lefgren, 2015). Thus, researchers face a fundamental trade-off between the LQTE model that allows for unrestricted treatment effect heterogeneity but only identifies the causal effects for the compliers on the one hand, and the IVQR model that substantially restrict treatment effect heterogeneity but identifies population treatment effects on the other hand.

On the surface, the LQTE and the IVQR 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. Therefore, Chernozhukov and Hansen (2013) have described the LQTE and the IVQR model as complementary approaches for estimating heterogeneous treatments 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).

This paper establishes a close relationship between the estimands of both models by characterizing the IVQR estimands under the LQTE assumptions under which treatment effects are unrestricted. I show that even in the absence of the underlying assumptions, the IVQR estimators of the QTE are consistent for LQTE at transformed quantile levels. This result has interesting implications for the connection between both models: (i) if the LQTE estimands are constant across quantiles, the estimates of both models converge to the same true effect, (ii) if the LQTE estimands are positive (or negative) at all quantiles, then the sign of the QTE will be the

\(^1\)The main contributions of Abadie et al. (2002) pertain to the issues that arise when including covariates in a linear-in-parameters model. Identification and estimation in the absence of covariates has already been studied by Imbens and Rubin (1997) and Abadie (2002).
same in both models, and (iii) monotonicity of the LQTE function (which is implied, for example, by a location-scale shift model for the compliers) implies monotonicity of the IV QR function. 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 LQTE. Underpinning these results are novel closed form representations of the IVQR estimands of the potential outcome cumulative distribution functions (cdf). These closed form solutions are also useful for estimation as they suggest direct analog estimators for the IVQR estimands. Such estimators have computational advantages over existing criterion-based approaches and I analyze them in a companion paper (Wüthrich, 2015).

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 clarifies how the IVQR model extrapolates from the compliers to the whole population. As this extrapolation is based on the restrictive rank similarity assumption, it is important to understand the properties of the IVQR estimands when treatment effects are unrestricted. To this end, the results imply that the IVQR estimands are quite robust: they preserve sign and monotonicity of the LQTE estimands whenever these properties are invariant across quantiles. If the LQTE is constant, the estimands of both models coincide. Furthermore, these findings suggest that the IVQR estimates are not arbitrary under misspecification but identify well-defined (functions of) causal effects for the compliers.

The analysis is extended to more general setups 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 treatment effect estimands and their counterparts in the LQTE model have intuitive analogues in these more general setups.

I illustrate the theoretical results using two empirical applications. Special attention is devoted to the factors that determine the similarities and differences between the estimates of both models, because these factors may interact in complicated ways due to the nonlinearity of the problem. In the first application, I examine the causal effect of Job Training Partnership Act (JTPA) training programs on the distribution of subsequent earnings. I show that both models yield quantitatively similar results. This finding can primarily be attributed to the magnitude of the first stage that outweights the differences between the potential outcome distributions of never-takers and untreated compliers. In the second application, I study estimation of the structural effect of Vietnam veteran status on civilian wages using draft lottery data. I show that the substantial numerical differences between the estimates of both models are due to the relatively small first stage combined with the large heterogeneity in the LQTE across quantiles.

This paper contributes to the extensive literature on identification and estimation in 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. (2007a), Chernozhukov and Hansen (2008), and Chernozhukov et al. (2009). Nonparametric estimation has been studied by Chernozhukov et al. (2007b), Horowitz and Lee (2007), Chen and Pouzo (2009), Chen and Pouzo (2012), Gagliardini and Scaillet (2012), Su and Hosino (2013), Vuong
and Xu (2014), and Belloni et al. (2014). For a survey on the IVQR model including references to empirical applications, see Chernozhukov and Hansen (2013).

The LQTE model, introduced by Abadie et al. (2002), extends the LATE framework (Imbens and Angrist, 1994) to the analysis of conditional LQTE for the compliers using the weighting theorem by 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, Belloni et al. (2014) have derived the properties of regression-based estimators for unconditional LQTE after selection among high-dimensional controls, and De Chaisemartin (2014a,b) has analyzed the LQTE framework under a weaker version of the LQTE monotonicity assumption.

The remainder of the paper is organized as follows. Section 2 introduces the basic notation and reviews both models. In Section 3, I characterize treatment effect estimands based on the IVQR model 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 proofs and a further discussion of the relationship between the LQTE assumptions and identifying conditions in the IVQR model.

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 $F_{Y|D=d,Z=z}(y)$ and $f_{Y|D=d,Z=z}(y)$ denote the cdf and the density function of $Y|D=d, Z=z$ respectively and define $p(d|z) \equiv P(D = d|Z = z)$. Covariates are omitted for deriving the main results of the paper. In Section 4, I consider extensions that incorporate covariates and nonbinary instruments. The analysis is developed in the potential outcomes framework (Rubin, 1974). Denote potential outcomes by $Y_1$ and $Y_0$ (indexed by $D$) and potential treatments by $D_1$ and $D_0$ (indexed by $Z$). Observed outcomes and observed treatments are given by $Y = DY_1 + (1-D)Y_0$ and $D = ZD_1 + (1-Z)D_0$ respectively. Based on their potential treatments, individuals can be categorized by four types, $T \in \{a, n, c, f\}$, (e.g., Angrist et al., 1996):

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

Let $\pi_t$, $f_{Y_{d|t}}(y)$, and $Q_{Y_{d|t}}(\tau)$ denote the proportion, density, cdf, and quantile function for type $T = t$, respectively.

In this paper, I focus on estimands of the cdf and the quantile function of $Y_d$, $F_{Y_{d}}(y)$ and $Q_{Y_{d}}(\tau)$, the $\tau$-QTE, $\delta(\tau) \equiv Q_{Y_1}(\tau) - Q_{Y_0}(\tau)$, and the related ATE, $\Delta \equiv E(Y_1 - Y_0) = \int_0^1 \delta(\tau)d\tau$.

---

2Defiers are denoted by $f$ instead of $d$ to distinguish the type from the treatment status.
Without loss of generality, I consider the following structural quantile function, \( q(d, \tau) \):

\[
q(d, \tau) = d\delta(\tau) + QY_0(\tau). \tag{1}
\]

Potential outcomes can be related to the structural quantile function by the Skorohod representation of random variables.

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

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:

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

1. **Monotonicity:** \( q(d, \tau) \) is strictly increasing in \( \tau \).
2. **Independence:** for each \( d \), \( U_d \) is independent of \( Z \).
3. **Selection:** \( D \equiv \rho(Z, V) \) for some unknown function \( \rho(\cdot) \) and random vector \( V \).
4. **Rank similarity:** conditional on \((Z, V)\), \( \{U_d\} \) are identically distributed.

Assumption 1.1 restricts \( Y \) to be nonatomic conditional on \( Z \).\(^4\) 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 across 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 \( U \). For a more detailed discussion of the IVQR model, I refer to Chernozhukov and Hansen (2005, 2013).

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

\[
P(Y \leq D\delta(\tau) + QY_0(\tau)|Z) = \tau. \tag{2}
\]

Chernozhukov and Hansen (2005) prove point identification of \( \delta(\tau) \) and \( QY_0(\tau) \) under an additional full rank condition on the Jacobian of (2); see Appendix A for a detailed discussion.

The conditional moment restriction (2) justifies the following unconditional moment equations for estimation:

\[
E((\tau - 1)[Y \leq D\delta(\tau) + QY_0(\tau)]f(Z)) = 0, \tag{3}
\]

\(^3\) These conditions are taken from Chernozhukov and Hansen (2005) and Chernozhukov and Hansen (2013) with modifications.

\(^4\) See Chesher (2010), Chesher and Smolinski (2012), and Chernozhukov and Hansen (2013) for extensions where \( Y \) has atoms conditional on \( Z \).
where \(1[\cdot]\) is the indicator function and \(f(Z)\) is a vector of (transformations of) instruments. The introduction contains references to different estimation procedures.

### 2.2 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:** \((Y_1, Y_0, D_1, D_0)\) are jointly independent of \(Z\).
3. **Nontrivial assignment:** \(0 < P(Z = 1) < 1\).
4. **First-stage:** \(p(1|1) > p(1|0)\).

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. This assumption is stronger than the corresponding independence assumption in the IVQR model.\(^6\) Assumptions 2.3 and 2.4 require that the instrument assignment is nontrivial and that it affects the treatment status.

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 identified from the following weighted quantile regression objective function (Abadie et al., 2002; Abadie, 2003):

\[
(Q_{Y_0|c}(\tau), \delta_c(\tau)) = \arg\min_{(Q_{Y_0|c}, \delta_c)} \mathbb{E}(\kappa \cdot \rho_\tau(Y - \delta_c D - Q_{Y_0|c})),
\]

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

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

The introduction reviews different estimation approaches for conditional and unconditional QTE based on the LQTE model.

### 3 Estimands under the LQTE framework

I focus on the following IVQR moment condition:

\[
\mathbb{E}((\tau - 1 [Y \leq \delta^*(\tau)D + Q_{Y_0}^*(\tau)]) (1, Z)^t) = 0,
\]

where \(\delta^*(\tau)\) and \(Q_{Y_0}^*(\tau)\) denote the IVQR estimands of \(\delta(\tau)\) and \(Q_{Y_0}(\tau)\). Since the analysis in this section does not rely on the IVQR assumptions, \(\delta^*(\tau)\) and \(Q_{Y_0}^*(\tau)\) will generally differ.

\(^5\)These assumptions are taken from Abadie et al. (2002) with modifications.

\(^6\)Kitagawa (2009) analyzes and compares the identifying power of both types of independence assumptions in a partial identification framework.
from $\delta(\tau)$ and $Q_{Y_{0}}(\tau)$. The IVQR estimates of the potential outcome cdfs are identified as the inverses of the corresponding quantile functions, and the IVQR estimand of the ATE is obtained as $\Delta^* = \int_{0}^{1} \delta^*(\tau)d\tau$.

To develop the results in this paper, let the region of interest $\mathcal{Y}$ be a compact interval in $\mathbb{R}$ and impose the following additional assumptions.

**Assumption 3.** For all $y \in \mathcal{Y}$ and all $(d, t) \in \{0, 1\} \times \{a, n, c, f\}$ for which $F_{Y|d}(y)$ is identified under the maintained assumptions and $\pi_t > 0$:

1. **Support:** $\text{Supp}(Y_d|T = t) = \mathcal{Y}$
2. **Regularity:** $f_{Y_d|t}(y)$ is continuous.

Under the LQTE framework, Assumption 3 is closely related to the conditions under which Chernozhukov and Hansen (2005) establish point identification of the IVQR estimands as discussed in Appendix B. Most importantly, $f_{Y_d|c}(y) > 0$ for $d \in \{0, 1\}$ implies full rank of the Jacobian of (2), which is the key identifying condition. This suggests that full support of the complier distributions constitutes a transparent and easily interpretable sufficient condition for point identification in the IVQR model.

### 3.1 Cumulative 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 characterize the IVQR estimands of the QTE and the ATE under the LQTE assumptions.

Under the LQTE assumptions, one can decompose the cdfs of $Y_0$ and $Y_1$ as

$$F_{Y_0}(y) = \pi_{n} F_{Y|a}(y) + \pi_{n} F_{Y|n}(y) + \pi_{c} F_{Y|c}(y),$$

$$F_{Y_1}(y) = \pi_{a} F_{Y|a}(y) + \pi_{n} F_{Y|n}(y) + \pi_{c} F_{Y|c}(y).$$

Imbens and Rubin (1997) show that under Assumption 2 the following potential distributions are identified from the data:

$$F_{Y_{0}|n}(y) = F_{Y|D=0,Z=1}(y),$$

$$F_{Y_{0}|c}(y) = \frac{p(0|0) F_{Y|D=0,Z=0}(y) - p(0|1) F_{Y|D=0,Z=1}(y)}{p(1|1) - p(1|0)},$$

$$F_{Y_{1}|a}(y) = F_{Y|D=1,Z=0}(y),$$

$$F_{Y_{1}|c}(y) = \frac{p(1|1) F_{Y|D=1,Z=1}(y) - p(1|0) F_{Y|D=1,Z=0}(y)}{p(1|1) - p(1|0)},$$

Moreover, the proportions of all 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, $F_{Y_{0}}(y)$ and $F_{Y_{1}}(y)$, are unidentified under the LQTE assumptions. This is in sharp contrast to the IVQR model where these quantities are identified as the inverses of the corresponding quantile functions. Therefore, the key question is how the IVQR model imputes the unidentified quantities $F_{Y_{1}|n}(y)$
and \( F_{Y_0\mid a}(y) \) using the rank similarity assumption. I answer this question in Theorem 1 by characterizing the IVQR estimands of the potential outcome cdfs under the LQTE assumptions.

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

\[
\begin{align*}
F_{Y_1}(y) &= \pi_a F_{Y_1\mid a}(y) + \pi_c F_{Y_1\mid c}(y) + \pi_n F_{Y_0\mid n} \left( Q_{Y_0\mid c} \left( F_{Y_1\mid c}(y) \right) \right), \\
F_{Y_0}(y) &= \pi_a F_{Y_0\mid n}(y) + \pi_c F_{Y_0\mid c}(y) + \pi_n F_{Y_1\mid a} \left( Q_{Y_1\mid c} \left( F_{Y_0\mid c}(y) \right) \right).
\end{align*}
\]

The proof of Theorem 1 proceeds in two steps. First, I solve (5) to obtain analytic closed form solutions \( F_{Y_1}(y) \) and \( F_{Y_0}(y) \) in terms of the distribution of \((Y, D, Z)\). Second, I express these closed form solutions in terms of expressions for never-takers, always-takers, and compliers.

Theorem 1 shows that the IVQR estimands can be decomposed into convex combinations of the potential outcome cdfs that are identified under the LQTE assumptions and the IVQR estimands of \( F_{Y_1\mid n}(y) \), \( F_{Y_1\mid c}(y) \equiv F_{Y_0\mid n} \left( Q_{Y_0\mid c} \left( F_{Y_1\mid c}(y) \right) \right) \), and \( F_{Y_0\mid a}(y) \), \( F_{Y_0\mid c}(y) \equiv F_{Y_1\mid a} \left( Q_{Y_1\mid c} \left( F_{Y_0\mid c}(y) \right) \right) \).

The results in Theorem 1 are important ingredients for characterizing the QTE and ATE estimands based on the IVQR model.

**Remark 1.** While the results in Theorem 1 do not rely on Assumption 1, it is interesting to discuss the implications of the results under this assumption. Under Assumption 1, Theorem 1 suggests that \( F_{Y_1\mid n}(y) \) and \( F_{Y_0\mid a}(y) \) are given by

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

A comparison of these expressions with Theorem 3.1 in Athey and Imbens (2006) reveals a close connection between the IVQR and the CIC model: (i) both models impute counterfactual quantities by restricting the evolution of unobservables across treatment states (IVQR model) and time (CIC model) respectively. (ii) In both models, identification is achieved by means of a specific subpopulation for which the potential outcome distribution of interest is identified in both treatment states (IVQR model) and time periods (CIC model) respectively. To summarize, Theorem 1 shows that to obtain point identification for the whole population, the IVQR model extrapolates from the compliers based on the rank similarity assumption. This intuition is also formalized in independent and simultaneous work by Vuong and Xu (2014).

**Remark 2.** The analogy to the CIC model discussed in Remark 1 shows that Assumption 3 is related to the support restriction (Assumption 3.4) in Athey and Imbens (2006). Without it, we have no information about \( F_{Y_1\mid n}(y) \) \( F_{Y_0\mid a}(y) \) outside the support of \( F_{Y_1\mid c}(y) \) \( F_{Y_0\mid c}(y) \). Consequently, it is no longer possible to point identify \( F_{Y_1\mid n}(y) \) and \( F_{Y_0\mid a}(y) \) for all \( y \in \mathcal{Y} \).

### 3.2 Quantile treatment effects

The next theorem characterizes the relationship between the QTE estimands based on the IVQR model and the LQTE.
**Theorem 2.** Suppose that Assumptions 2 and 3 hold and that the IVQR estimands are given by (5). Then,

$$
\delta^*(\tau) = \delta_c\left(F_{Y_i|c}\left(Q_{Y_i}(\tau)\right)\right) = \delta_c\left(F_{Y_i|c}\left(Q_{Y_i}^* (\tau)\right)\right)
$$

Theorem 2 shows that the QTE estimands based on the IVQR model are equivalent to LQTE for the compliers at transformed quantile levels. These transformations adjust for differences between the IVQR estimands of the potential outcome cdfs and the corresponding potential outcome cdfs for the compliers as measured by the quantile-quantile transforms $F_{Y_i|c}\left(Q_{Y_i}(\cdot)\right)$ and $F_{Y_i|c}\left(Q_{Y_i}^* (\cdot)\right)$. In conjunction with Theorem 1, Theorem 2 implies that the difference between the QTE estimates based on the IVQR model and the LQTE model is uniquely determined by two factors: (i) differences between treated compliers and always-takers, and untreated compliers and never-takers and (ii) the relative size of the subpopulations, which depends on the first stage. However, these two factors interact in complicated ways because of the inherent nonlinearity of the problem. Section 5 illustrates this point based on two empirical applications.

**Remark 3.** Let $Q_{Y_{1|n}}^*(\tau)$ and $Q_{Y_{0|a}}^*(\tau)$ denote the quantile functions associated with $F_{Y_{1|n}}^*(y)$ and $F_{Y_{0|a}}^*(y)$ and define the QTE estimands for the never-takers and always-takers based on the IVQR model as

$$
\delta_n^*(\tau) \equiv Q_{Y_{1|n}}^*(\tau) - Q_{Y_{0|n}}(\tau) \quad \text{and} \quad \delta_a^*(\tau) \equiv Q_{Y_{1|a}}(\tau) - Q_{Y_{0|a}}(\tau).
$$

Using the same arguments as in the proof of Theorem 2, one can show that

$$
\delta_n^*(\tau) = \delta_c\left(F_{Y_i|c}\left(Q_{Y_i|n}(\tau)\right)\right) \quad \text{and} \quad \delta_a^*(\tau) = \delta_c\left(F_{Y_i|c}\left(Q_{Y_i|a}(\tau)\right)\right).
$$

These results show that the differences between the LQTE and the IVQR QTE estimands for the never-takers and the always-takers are determined by differences between the identified potential outcome distributions as measured by the quantile-quantile transforms $F_{Y_i|c}\left(Q_{Y_{1|n}}(\tau)\right)$ and $F_{Y_i|c}\left(Q_{Y_{1|a}}(\tau)\right)$ but do not depend on the relative size of the respective subpopulations.

As the results do not rely on Assumption 1, Theorem 2 provides a characterization of the IVQR estimands under misspecification. The results suggest that even in the absence of the rank similarity assumption, the estimates based on the IVQR model are not completely arbitrary but capture particular causal effects for the compliers. Moreover, Theorem 2 has interesting consequences for the connection between both models which can be interpreted as robustness properties:

**Corollary 1.** Suppose that the assumptions of Theorem 2 hold. Then $\delta_c(\tau) \geq 0$ 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 $\tau \in (0,1)$.

**Corollary 2.** Suppose that the assumptions of Theorem 2 hold and that $\delta_c(\tau) = \delta_c$ for all $\tau \in (0,1)$. Then $\delta^*(\tau) = \delta_c(\tau) = \delta_c$ for all $\tau \in (0,1)$. 

9
Corollary 3. Suppose that the assumptions of Theorem 2 hold and that $\delta_\tau(\tau)$ is monotonically increasing (decreasing) in $\tau$. Then $\delta^*(\tau)$ is monotonically increasing (decreasing) in $\tau$.

3.3 Average treatment effects

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, $\Delta_c$, and estimands of the ATE for the never-takers and always-takers, which correspond to average LQTE at transformed quantile levels.

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

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

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

As for the QTE, differences between the estimands of both models are determined by differences between the potential outcome distributions for compliers, always-takers, and never-takers, as well as the relative size of these subpopulations, which depends on the magnitude of the first stage. Moreover, the IVQR estimand of the ATE inherits the robustness properties outlined in Corollaries 1 and 2. Namely, the sign of the IVQR estimand coincides with the sign of the LATE if the LQTE is uniformly positive or negative across all quantiles, and the IVQR estimand equals the LATE if the LQTE is constant across quantiles.

4 Extensions

Here I discuss three extensions of the main results. In the conclusion, I mention additional extensions that are left for future research. Throughout this section, I focus on the IVQR estimands of the potential outcome cdfs and QTE. Results for the ATE can be derived using similar arguments as in the previous sections and are thus omitted.

4.1 LQTE assumptions without monotonicity

The monotonicity assumption of the LQTE model is not innocuous and is likely to be violated in many applications (e.g., De Chaisemartin, 2014b). In contrast, under the rank similarity assumption, the IVQR model identifies causal effects irrespective of the validity of the monotonicity assumption. Motivated by this observation, I analyze a setting where only Assumptions 2.2-2.4 and 3 are maintained. Under these assumptions neither the proportion nor the potential outcome distribution for any single subpopulation are identified. In particular, the quantities that correspond to potential outcome cdfs for the compliers under monotonicity are equal to

---

7Note that independence between the potential treatments and the instrument is not required in the IVQR model but aids the exposition and the interpretation of the results.
Assumption S2 in the supplementary material states that $P(\mathcal{T} = f) = P(\mathcal{T} = f)$ and $Y_d|\mathcal{T} = c_f \sim Y_d|\mathcal{T} = f$. Under these assumptions, the compliers cancel with the defiers and weighted differences between compliers and defiers:

$$\frac{p(0|0)F_{Y|D=0,Z=0}(y) - p(0|1)F_{Y|D=0,Z=1}(y)}{p(1|1) - p(1|0)} = \frac{\pi_c F_{Y|c}(y) - \pi_f F_{Y|f}(y)}{\pi_c - \pi_f} = F_{Y|c-f}(y),$$

$$\frac{p(1|1)F_{Y|D=1,Z=1}(y) - p(1|0)F_{Y|D=1,Z=0}(y)}{p(1|1) - p(1|0)} = \frac{\pi_c F_{Y|f}(y) - \pi_f F_{Y|f}(y)}{\pi_c - \pi_f} = F_{Y|c-f}(y).$$

I denote the mixture subpopulation corresponding to this mixture distribution as the compliers-defiers. The inverses of $F_{Y|c-f}(y)$ and $F_{Y|f-c}(y)$ are denoted as $Q_{Y|c-f}(\tau)$ and $Q_{Y|f-c}(\tau)$ and the $\tau$-QTE for the compliers-defiers is denoted as $\delta_{c-f}(\tau) \equiv Q_{Y|c-f}(\tau) - Q_{Y|f-c}(\tau)$. De Chaisemartin (2014a,b) has shown that under a weaker version of the LQTE monotonicity assumption, the compliers-defiers correspond to a well-defined subpopulation which he calls the convivors.8

The following theorem characterizes the estimands of the potential outcome cdfs based on the IVQR model 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 $F_{Y|c-f}(y)$ and $F_{Y|f-c}(y)$ are strictly increasing and well-defined cdfs. Then,

$$F_{Y}^*(y) = \pi_c F_{Y|c}(y) + \pi_f F_{Y|f}(y),$$

$$F_{Y|c}^*(y) = \pi_c F_{Y|c}(y) + \pi_f F_{Y|f}(y),$$

Theorem 4 constitutes a natural generalization of Theorem 1. Instead of the compliers, the IVQR model imputes the unidentified quantities using the compliers-defiers. Without the LQTE monotonicity assumption, $F_{Y|c-f}(y)$ and $F_{Y|f-c}(y)$ are not necessarily monotonic. However, assuming that $F_{Y|c-f}(y)$ and $F_{Y|f-c}(y)$ are well-defined and strictly increasing is essential for deriving the closed form solutions of the IVQR estimands (cf. Lemma 1) and implies full rank of the Jacobian of (2) as discussed in Appendix B.

The next theorem characterizes the IVQR QTE estimands 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 $F_{Y|c-f}(y)$ and $F_{Y|f-c}(y)$ are strictly increasing and well-defined cdfs. Then,

$$\delta^*(\tau) = \delta_{c-f} \left( F_{Y|c-f} \left( Q_{Y|f}(\tau) \right) \right) = \delta_{c-f} \left( F_{Y|f-c} \left( Q_{Y|c}(\tau) \right) \right).$$

Theorem 5 shows that the IVQR estimands are equivalent to QTE for the compliers-defiers at transformed quantile levels.

---

8 De Chaisemartin (2014a,b) splits the compliers in two groups: the convivors ($c_v$) and the compliers ($c_f$). His Assumption S2 in the supplementary material states 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 compliers cancel with the defiers and $F_{Y|c-f}(y) = F_{Y|c|c}(y)$. 
4.2 Multivalued instruments

Suppose that instead of being binary, the instrument \( Z \) takes values in a finite set \( Z = \{z_1, z_2, ..., z_K\} \) with \( 0 < z_1 < z_2 < ... < z_K < \infty \). The following assumption extends the LQTE framework defined by Assumption 2 to multivalued instruments.

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

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

Under Assumption 4, there are \( K + 1 \) different types. Consistent with standard LQTE model, I denote individuals with \( D_{z_k} = 0 \) for all \( z_k \in Z \) as never-takers and individuals with \( D_{z_k} = 1 \) for all \( z_k \in Z \) 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 \( \pi_t, f_{Y|a}(y), f_{Y|a}(y), \) and \( Q_{Y|a}(\tau) \) denote the proportion, density, cdf, and quantile function of \( Y_0|T = t \).

Under Assumption 4, the conditional probabilities \( p(d|z_k) \equiv P(D = d|Z = z_k) \) are related to the proportions of types as

\[
p(1|z_k) = \pi_a + \sum_{j=2}^{K} \pi_{c_{z_j}} 1(z_j \leq z_k) \quad \text{and} \quad p(0|z_k) = \pi_n + \sum_{j=2}^{K} \pi_{c_{z_j}} 1(z_j > z_k),
\]

and the observed conditional cdfs \( F_{Y|D=d,Z=z_k}(y) \) are related to the potential outcome CDFs of the \( K + 1 \) types as

\[
F_{Y|D=1,Z=z_k}(y) = \frac{\pi_a f_{Y_1|a}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} f_{Y_1|c_{z_j}}(y) 1(z_j \leq z_k)}{\pi_a + \sum_{j=2}^{K} \pi_{c_{z_j}} 1(z_j \leq z_k)},
\]

\[
F_{Y|D=0,Z=z_k}(y) = \frac{\pi_n f_{Y_0|n}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} f_{Y_0|c_{z_j}}(y) 1(z_j > z_k)}{\pi_n + \sum_{j=2}^{K} \pi_{c_{z_j}} 1(z_j > z_k)}.
\]

Hence, 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 distributions of \( Y_0 \) for never-takers and compliers (e.g., Imbens, 2007).

In this section, I study IVQR estimands \( \delta^*(\tau) \) and \( Q_{Y_0}^*(\tau) \) that are given by the following moment equations:

\[
\mathbb{E} \left( (\tau - 1 \left[ Y \leq \delta^*(\tau) D + Q_{Y_0}^*(\tau) \right]) (1, Z)' \right) = 0. \tag{6}
\]

\(^9\)Similar frameworks are considered by Imbens (2007) and Frölich (2007).
\(^{10}\)It is important to note that the moment equations (6) are not the only possible unconditional moment restrictions that can be used to estimate \( \delta^*(\tau) \) and \( Q_{Y_0}^*(\tau) \).
Theorem 6. Suppose that Assumptions 4 and 5 hold and that the IVQR estimands are given by (6). Then,

\[ F_{Y_1}(y) = F_{Y_1|a}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_1|c_{z_j}}(y) P(Z \geq z_j) \]
\[ + \pi_n F_{Y_0|a}(y) \left( \tilde{Q}_{Y_0} \left( \tilde{F}_{Y_1}(y) \right) \right) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}} \left( \tilde{Q}_{Y_0} \left( \tilde{F}_{Y_1}(y) \right) \right) P(Z < z_j), \]
\[ F_{Y_0}(y) = F_{Y_1|a}(y) \left( \tilde{Q}_{Y_1} \left( \tilde{F}_{Y_0}(y) \right) \right) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}} \left( \tilde{Q}_{Y_1} \left( \tilde{F}_{Y_0}(y) \right) \right) \frac{E(Z|Z \geq z_j)}{E(Z)} P(Z \geq z_j) \]
\[ + \pi_n F_{Y_0|a}(y) + \sum_{j=2}^{K} \pi_{c_{z_j}} F_{Y_0|c_{z_j}} \left( y \right) \frac{E(Z|Z < z_j)}{E(Z)} P(Z < z_j), \]

where

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

and \( w_j \equiv \left( \frac{E(Z|Z \geq z_j)}{E(Z)} - 1 \right) P(Z \geq z_j), \tilde{Q}_{Y_1}(\tau) \equiv \tilde{F}_{Y_1}^{-1}(\tau), \) 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 cdfs, \( \tilde{F}_{Y_1}(y) \) and \( \tilde{F}_{Y_0}(y) \), are convex combinations of cdfs for the \( K - 1 \) different compliers. The weight attached to the cdf of compliant type \( T = c_j, F_{Y_1|c_{z_j}}(y) \), 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 \( E(Z|Z \geq z_j) > E(Z) \) for \( j \geq 2 \). To gain some intuition about the shape of the weighting function, consider the difference between two adjacent weights:

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

Hence, the weighting function is increasing whenever \( z_j \) is smaller than the mean and decreasing whenever \( z_j \) is larger than the mean.

Remark 4. The weighting function \( w_j \) is 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 next theorem shows that under the generalized LQTE framework the IVQR QTE estimands can be expressed as QTE for the mixture of all compliant types, \( \delta(\tau) \equiv \tilde{Q}_Y(\tau) - \tilde{Q}_{\tilde{Y}_0}(\tau) \), at transformed quantile levels.

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

\[
\delta^*(\tau) = \tilde{\delta} \left( \tilde{F}_{\tilde{Y}_0} \left( Q_{\tilde{Y}_0}(\tau) \right) \right) = \tilde{\delta} \left( \tilde{F}_{\tilde{Y}_1} \left( Q_{\tilde{Y}_1}(\tau) \right) \right).
\]

In contrast to the previous results, the transformation of the quantile levels does 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 Covariates

Including 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.\textsuperscript{11} Third, even if the instrument and the rank similarity assumption are unconditionally valid, it might be interesting to consider conditional QTE.\textsuperscript{12}

With discrete covariates, the previous analysis results remain 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 approximations to the true potentially nonlinear conditional quantile functions of \( Y_d | X = x \) for the overall population and for the compliers respectively. Because the results obtained in the previous sections are fully nonparametric, they can be expected to hold approximately. This approximation can be improved upon by choosing richer specifications (e.g. through interactions, polynomials, or B-splines).

### 5 Empirical applications

#### 5.1 Practical implementation details

Here I briefly discuss some practical implementation details. The quantile functions in the IVQR model are estimated using the inverse quantile regression procedure proposed by Chernozhukov and Hansen (2006). This approach is based on the fact that at the true coefficient on \( D \), \( \delta(\tau) \), the \( \tau \)-quantile regression of \( Y - D\delta(\tau) \) on a constant and \( Z \) would yield a zero coefficient on \( Z \) by equation (2), motivating a simple grid search algorithm over \( \delta(\tau) \):

\textsuperscript{11}For example, Chernozhukov and Hansen (2004) assume that 401(k) eligibility is exogenous conditional on income (and further covariates).

\textsuperscript{12}I refer to Firpo (2007) and Frölich and Melly (2013) for a discussion of the differences between conditional and unconditional QTE.)
1. Define a grid \( \{ \delta_j, j = 1, \ldots, J \} \) and estimate the coefficients on the constant, \( \hat{Q}_Y(\delta_j, \tau) \), and on the instrument, \( \hat{\gamma}(\delta_j, \tau) \), using an ordinary \( \tau \)-quantile regression of \( Y - D\delta_j \) on a constant and the instrument \( Z \).

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

Absent additional covariates, the quantile functions for the compliers can be estimated from a weighted quantile regression as discussed in Section 2.2 or by inverting the sample analogues of \( F_{Y|c}(y) \) and \( F_{Y_0|c}(y) \), which is the approach taken here.

To deal with the potential lack of monotonicity in the estimation of the cdfs and quantile functions, I rearrange the original estimates as suggested by Chernozhukov et al. (2010). The rearrangement procedure is easy to implement and has a number of desirable properties.\(^{13}\) The cdfs and quantile function for the never-takers and always-takers are estimated directly in the subsamples with \((D = 0, Z = 1)\) and \((D = 1, Z = 0)\) respectively.

To avoid the estimation of tail quantiles, I use trimmed sample analogues of the expressions in Theorem 3 to estimate the average effects for the always-takers and never-takers in the IV QR model.\(^{14}\) Moreover, I directly estimate the LATE using two-stage least squares (2SLS) instead of averaging over the LQTE. The 2SLS estimate is also used to compute the overall IV QR ATE estimate.

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.\(^{15}\) 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 \((D)\) on the sum of earnings in the 30 months 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

\(^{13}\)Chernozhukov et al. (2010) show that the rearranged curve is closer to the original curve in finite samples and that rearrangement outperforms the isotonization procedure in a simulation study designed to match closely the second empirical application. Their generic results provide functional limit theory for the rearranged quantile functions, given that a functional central limit theorem applies to the original estimators, which is the case for the IVQR estimators as discussed in Section 2.1. Moreover, the functional delta method for the bootstrap implies validity of the bootstrap for estimating the limiting law of the rearranged curve.

\(^{14}\)In the applications, I use a trimming constant of \( \varepsilon = 0.05 \).

\(^{15}\)The data were obtained from the Angrist Data Archive (last accessed January 14, 2015): economics.mit.edu/faculty/angrist/data1/data/abangim02.
drop the observations violating this condition from the sample, which yields a total number of N=5,083 observations. Abadie et al. (2002) give additional information about the dataset and present descriptive statistics.

One-sided non-compliance rules out the existence of both always-takers and defiers, such that the results in Theorem 1 simplify to

\[
F^*_Y (y) = \pi_c F_{Y1|c}(y) + \pi_n F_{Y0|n}\left(Q_{Y0|n}(y)\right),
\]

\[
F^*_Y (y) = \pi_c F_{Y0|c}(y) + \pi_n F_{Y0|n}(y).
\]

Hence, the IVQR estimand \(F^*_Y (y)\) equals the true potential outcome cdf, \(F_Y (y)\), irrespective of the validity of the rank similarity assumption.

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 \((-2500, -2495, ..., 15000)\). Panel A shows that the distributions of \(Y_0\) 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 estimates from the IVQR and the LQTE model. Both models yield qualitatively and quantitatively similar results that are characterized by substantial effect heterogeneity across quantiles and overall increasing QTE estimates ranging from values below zero up to around 6000 USD.

In Figure 2, I explore the determinants of the similarities and differences between the estimates of both models. Panel A plots the estimated quantile-quantile transform \(F_{Y0|c}\left(Q_{Y0|n}(\cdot)\right)\). The pronounced differences at the lower quantiles in Figure 1 (panel A) translate into deviations...
Figure 2: The left panel plots the estimated quantile-quantile transform and the right panel compares the LQTE estimates with the IVQR estimates for the never-takers. All estimates are computed over $\tau = \{0.05, 0.06, ..., 0.95\}$. N=5,083.

of the quantile-quantile transform from the 45-degree line. In panel B, I plot the IVQR estimate of the QTE for the never-takers against the LQTE. Although qualitatively similar, the IVQR estimates for the never-takers 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 the quantile-quantile transform, implying that the $\tau$-QTE for the never-taker corresponds to the $\tau'$-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 rather large first stage, $p(1|1) − p(1|0) = 0.62$, that outweighs the difference between the distributions of $Y_0$ for never-takers and compliers.

Finally, I estimate the corresponding average effects. The IVQR estimates are $\Delta^*_n = 1228.42$ and $\Delta^* = 1532.07$. These effects are smaller than the corresponding LATE, $\Delta_c = 1715.96$, as the QTE for the never-takers is smaller than the LQTE at most quantiles as discussed before.

5.3 Veteran status and earnings

In this section, I consider the 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 is eligible for draft and zero otherwise. I use the same dataset as Abadie (2002) and Chernozhukov et al. (2010). The dataset contains information about N=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 12% compliers. Abadie (2002) gives more information about the dataset.

16The fractions to not sum to one because they correspond to the original fractions rounded up to two decimals.
Figure 3: The left panel shows estimated cdfs for the different subpopulations. The right panel contains estimates of the QTE from the IVQR model and the LQTE model. The QTE are computed over $\tau = \{0.05, 0.06, \ldots, 0.95\}$. N=11,637.

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, while the differences at higher quantiles as well as between the distributions of $Y_1$ for always-takers and compliers are substantially smaller. Figure 3 (panel B) compares the QTE estimates based on the IVQR model and the LQTE estimates. The IVQR estimates are computed using a grid search over a fine grid of $\{-10000, -9995, \ldots, 5000\}$. The QTE estimates of both models display similar overall patterns that indicate a 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. However, for given quantiles there are large numerical differences between the QTE estimates of both models up to the third quartile.

Figure 4 sheds some light on determinants of these results. Panel A plots estimates of the quantile-quantile transforms $F_{Y_1|c}(Q_{Y_1|a}(\cdot))$ and $F_{Y_0|c}(Q_{Y_0|n}(\cdot))$ that measure the differences between the cdfs displayed in Figure 3 (panel A). Panel B compares the IVQR estimates of the QTE for the never-takers and always-takers with the LQTE. Taken together panels A and B suggest that the large numerical differences between the QTE estimates in Figure 3 (panel B) are driven by the interplay of three main factors: (i) the substantive differences between the distributions of $Y_0$ for never-takers and compliers that lead to horizontal shifts between the QTE curves, (ii) the substantive heterogeneity in the LQTE that translates these shifts into large vertical differences, and (iii) the large fraction of never-takers that determines the extent of (i) and (ii). The differences between always-takers and treated compliers are less important. This is because there are only small differences between the distributions of $Y_1$ for the always-takers and compliers at the lower quantiles where the variability of the LQTE is most pronounced. The somewhat larger differences at the higher quantiles only lead to small differences between
the QTE estimates because the LQTE is more stable at these quantiles.

Finally, I estimate the corresponding average effects. The IVQR estimates are $\Delta_n^* = -1481.59$, $\Delta_a^* = -1767.51$, and $\Delta_n^* = -1509.30$. These estimates are smaller than the corresponding LATE, $\Delta_c = -1277.78$, which reflects the differences between the corresponding QTE estimates reported in Figure 3 (panel B) and Figure 4 (panel B).

6 Conclusion

In this paper, I characterize the estimands of estimators based on the IVQR model under the LQTE framework under which the treatment effect heterogeneity is unrestricted. I show that even if the rank similarity assumption fails, the IVQR QTE estimands are equivalent to LQTE at transformed quantile levels. Moreover, the IVQR estimate of the ATE can be expressed as a convex combination of the LATE and weighted averages of LQTE at transformed quantile levels. Underpinning these results are closed form representations for the IVQR estimands of the potential outcome cdfs. In a companion paper (Wüthrich, 2015), I further exploit these analytic representations to construct analog estimators of the IVQR estimands. These estimators have computational advantages over existing approaches such as the inverse quantile regression algorithm (Chernozhukov and Hansen, 2006).

I focus on IVQR estimands with binary treatments and binary instrument as well as multivalued instruments, but I do not cover multivalued and continuous treatments and continuous instruments. Analyzing the IVQR estimands with multivalued and continuous treatments would be particularly insightful as there is no straightforward way to generalize the LQTE model to accommodate non-binary treatment variables. However, such an extension cannot be based on
the closed form solutions presented in this paper as they rely on the binary nature of the treatment variable and is left for future research. In contrast, it is conceptually straightforward to analyze the IVQR estimands with continuous instruments by combining the general closed form solutions in Lemma 2 with the analysis in Yu (2014).
References


23
A  Point identification in the IVQR model

Chernozhukov and Hansen (2005, Section 2.4) provide a set of additional assumptions under which the conditional moment restriction (2) point identifies \((Q_{y_0}(\tau), Q_{Y_1}(\tau))^\prime\).

For vectors \(y = (y_0, y_1)\) and the vector of moment conditions

\[
\Pi(y) \equiv \begin{pmatrix} P(Y \leq y_D | Z = 0) - \tau \\ P(Y \leq y_D | Z = 1) - \tau \end{pmatrix}
\]

where \(y_D = D y_1 + (1 - D)y_0\), point identification requires that \((y_0, y_1) = (Q_{y_0}(\tau), Q_{Y_1}(\tau))\) uniquely solves \(\Pi(y) = 0\). For small constants \(\varepsilon > 0\) and \(f > 0\), define the parameter space \(\mathcal{L}\) as a closed rectangle containing all vectors \(y = (y_0, y_1)\), satisfying

for each \(z\), \(P(Y \leq y_D | Z = z) \in [\tau - \varepsilon, \tau + \varepsilon] \quad (7)\)

for each \(d, y_d \in s_d \equiv \{\lambda : f_{Y|D=d,Z=z}(\lambda) \geq f \text{ for all } z \text{ with } p(d|z) > 0\} \quad (8)\)

Condition (7) defines \(\mathcal{L}\) as a set of potential solutions of \(\Pi(y) = 0\) and condition (8) requires the solutions to be in the support of the response variable. Define the Jacobian of \(\Pi(y)\) with respect to \((y_0, y_1)\) as

\[
\Pi'(y) \equiv \begin{pmatrix} f_{Y|D=0,Z=0}(y)p(0|0) & f_{Y|D=1,Z=0}(y)p(1|0) \\ f_{Y|D=0,Z=1}(y)p(0|1) & f_{Y|D=1,Z=1}(y)p(1|1) \end{pmatrix}
\]

Chernozhukov and Hansen (2005) establish point identification under the following conditions:

**Assumption 6.** For all sets \(\mathcal{L}\) and \(s_d\) defined in (7) and (8), assume

1. **Continuity:** \(\Pi'(y)\) is continuous for all \((y_0, y_1) \in \mathcal{L}\).
2. **Inclusion:** \(Q_{Y_d}(\tau) \in s_d\) for each \(d\).
3. **Full rank:** \(\Pi'(y)\) is of full rank for all \(y \in \mathcal{L}\).

Assumption 6.1 imposes continuity of the \(f_{Y|D,Z}(y|D = d, Z = z)\). Assumption 6.2 requires that \(Y\) takes on values around \(Q_{Y_d}(\tau)\) and implies that \((Q_{Y_d}(\tau), Q_{Y_1}(\tau)) \in \mathcal{L}\). Assumption 6.3 is the key assumption for point identification and insures that \((y_0, y_1) = (Q_{y_0}(\tau), Q_{Y_1}(\tau))\) uniquely solves \(\Pi(y) = 0\).

B  Relationship between Assumption 3 and Assumption 6

Here I discuss the relationship between Assumption 3 and Assumption 6 under Assumption 2.

Assumption 3.1 requires (i) that \(f_{y_0|a}(y) > 0\) if \(\pi_n > 0\) and \(f_{Y_1|a}(y) > 0\) if \(\pi_a > 0\), and (ii) that \(f_{y_d|e}(y) > 0\) for \(d \in \{0, 1\}\) (as \(\pi_e > 0\) under Assumption 2). Condition (i) corresponds to (8) because under Assumption 2, by Imbens and Rubin (1997),

\[
f_{y_0|a}(y) = f_{Y|D=0,Z=1}(y), \quad \pi_n = p(0|1), \quad f_{Y_1|a}(y) = f_{Y|D=1,Z=0}(y), \quad \text{and} \quad \pi_a = p(1|0).
\]

Condition (ii) implies full rank of \(\Pi'(y)\). To see this, note that by Imbens and Rubin (1997)

\[
\pi_e f_{Y_1|e}(y) = p(1|1)f_{Y|D=1,Z=1}(y) - p(1|0)f_{Y|D=1,Z=0}(y),
\]

\[
\pi_e f_{y_0|e}(y) = p(0|0)f_{Y|D=0,Z=0}(y) - p(0|1)f_{Y|D=0,Z=1}(y).
\]

Hence, \(f_{y_0|e}(y) > 0\) and \(\pi_e > 0\) implies that

\[
p(1|1)f_{Y|D=1,Z=1}(y) > p(1|0)f_{Y|D=1,Z=0}(y),
\]

\[
p(0|0)f_{Y|D=0,Z=0}(y) > p(0|1)f_{Y|D=0,Z=1}(y),
\]

which in turn implies that \(\det(\Pi'(y)) \neq 0\) and thus that rank \(\Pi'(y)\) is full. This suggest that the rank condition of Chernozhukov and Hansen (2005) is implied by a support condition for the compliers; see Remark 1 for a further discussion and interpretation of this condition.
Assumption 3.1 imposes continuity of the potential outcome densities. This requirement corresponds to Condition 6.1 because under Assumption 2, the densities \( f_{Y|D}(y) \) correspond to (functions of) observed densities \( f_{Y|D=d,Z=z}(y) \) (Imbens and Rubin, 1997).

C Proofs

C.1 Auxiliary Lemmata

This section contains two auxiliary lemmata that provide closed form solutions for the IVQR estimands with binary (Lemma 1) and general instruments (Lemma 2). I consider the following IVQR moment condition:

\[
\mathbb{E} ((\tau - 1 | Y \leq \delta^*(\tau) D + Q_{Y_0}^*(\tau))) (1, Z') = 0,
\]

(9)

The instrument \( Z \) with support \( Z \) can be discrete or continuous. 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. In the statements of the lemmata, it is understood that all relevant moments exist and are finite. The region of interest \( Y \) is a compact interval in \( \mathbb{R} \).

To derive the closed form results, the following additional assumptions are imposed.

**Assumption 7.** For all \((y, z) \in Y \times Z\):

1. Monotonicity:

\[
\tilde{F}_{Y_1}(y) \equiv \frac{\mathbb{E} (F_{Y|D=1,Z}(y) p(1|Z) \left(\frac{z}{\mathbb{E}[Z]} - 1\right))}{\mathbb{E} (p(1|Z) \left(\frac{z}{\mathbb{E}[Z]} - 1\right))}
\]

and

\[
\tilde{F}_{Y_0}(y) \equiv \frac{\mathbb{E} (F_{Y|D=0,Z}(y) p(0|Z) \left(1 - \frac{z}{\mathbb{E}[Z]}\right))}{\mathbb{E} (p(0|Z) \left(1 - \frac{z}{\mathbb{E}[Z]}\right))}
\]

are well-defined and strictly increasing.

2. Non-trivial assignment: \( f_Z(z) > 0 \).

Lemma 1 can be deduced from the general results in Lemma 2. However, given the importance of binary instruments, I state and prove this result separately.

**Lemma 1.** Suppose that \( Z = \{0, 1\} \), that Assumption 7 holds, and that the IVQR estimands given by (5). Then

(i) \( \tilde{Q}_{Y_1}(F_{Y_0}(y)) = \tilde{Q}_{Y_1}(\tilde{F}_{Y_0}(y)) \) and \( \tilde{Q}_{Y_0}(F_{Y_1}(y)) = \tilde{Q}_{Y_0}(\tilde{F}_{Y_1}(y)) \),

and

(ii) \( F_{Y_1}(y) = (p(1|1) - p(1|0))F_{Y_1}(y) + p(1|0)F_{Y|D=1,Z=0}(y) + p(0|1)F_{Y|D=0,Z=1}(y) \),

\( F_{Y_0}(y) = (p(1|1) - p(1|0))F_{Y_0}(y) + p(0|1)F_{Y|D=0,Z=1}(y) + p(1|0)F_{Y|D=1,Z=0}(y) \),

where \( \tilde{Q}_{Y_0}(\tau) \equiv \tilde{F}_{Y_0}^{-1}(\tau) \) and \( \tilde{Q}_{Y_1}(\tau) \equiv \tilde{F}_{Y_1}^{-1}(\tau) \).

**Proof of Lemma 1.**

Part (i) Rewrite (5) using the law of iterated expectations as

\[
\mathbb{E} (\tau - 1 | Y \leq \delta^*(\tau) D + Q_{Y_0}^*(\tau)) | Z = 1) = 0,
\]

\[
\mathbb{E} (\tau - 1 | Y \leq \delta^*(\tau) D + Q_{Y_0}^*(\tau)) | Z = 1) = 0.
\]

Because \( 0 < P(Z = 1) < 1 \), \( \delta^*(\tau) \) and \( Q_{Y_0}^*(\tau) \) solve

\[
\mathbb{E} (\tau - 1 | Y \leq \delta^*(\tau) D + Q_{Y_0}^*(\tau)) | Z = 0) = 0,
\]

\[
\mathbb{E} (\tau - 1 | Y \leq \delta^*(\tau) D + Q_{Y_0}^*(\tau)) | Z = 1) = 0.
\]

**Lemma 1** is a special case of Lemma 2. However, given the importance of binary instruments, I state and prove this result separately.

25
By law of iterated expectations and the definition of a conditional cdf,
\[ p(1|1)F_{Y|D=1,Z=1}(Q_{Y_1}^r(\tau)) + p(0|1)F_{Y|D=0,Z=1}(Q_{Y_0}^r(\tau)) = \tau, \]  
\[ p(1|0)F_{Y|D=1,Z=0}(Q_{Y_1}^r(\tau)) + p(0|0)F_{Y|D=0,Z=0}(Q_{Y_0}^r(\tau)) = \tau. \]

Equating (10) and (11) and rearranging terms yields:
\[ p(1|1)F_{Y|D=1,Z=1}(Q_{Y_1}^r(\tau)) - p(1|0)F_{Y|D=1,Z=0}(Q_{Y_1}^r(\tau)) = 
\quad p(0|0)F_{Y|D=0,Z=0}(Q_{Y_0}^r(\tau)) - p(0|1)F_{Y|D=0,Z=1}(Q_{Y_0}^r(\tau)). \]

Dividing by \( p(1|1) - p(1|0) \) on both sides, we have
\[ \tilde{F}_{Y_1}(Q_{Y_1}^r(\tau)) = \tilde{F}_{Y_0}(Q_{Y_0}^r(\tau)), \]  
by definition of \( \tilde{F}_{Y_d}(y) \) and \( \tilde{Q}_{Y_d}(\tau) \) for \( d \in \{0, 1\} \) specialized to the setup of this Lemma.\(^{18}\) It follows from Assumption 7 and equations (10) and (11) that \( \tilde{F}_{Y_d}(y) \) and \( Q_{Y_d}^r(\tau) \) are strictly increasing for \( d \in \{0, 1\} \), and thus one-to-one with strictly increasing inverses. Therefore, we have
\[ Q_{Y_1}^r(\tau) = \tilde{Q}_{Y_1} \left( \tilde{F}_{Y_0}(Q_{Y_0}^r(\tau)) \right) \quad \text{and} \quad Q_{Y_1}^r(F_{Y_0}(y)) = \tilde{Q}_{Y_1} \left( \tilde{F}_{Y_0}(y) \right). \]

Similarly, obtain
\[ Q_{Y_0}^r(\tau) = \tilde{Q}_{Y_0} \left( \tilde{F}_{Y_1}(Q_{Y_1}^r(\tau)) \right) \quad \text{and} \quad Q_{Y_0}^r(F_{Y_1}(y)) = \tilde{Q}_{Y_0} \left( \tilde{F}_{Y_1}(y) \right). \]

This completes the proof of part (i).

Part (ii): From part (i), we have
\[ Q_{Y_0}^r(\tau) = \tilde{Q}_{Y_0} \left( \tilde{F}_{Y_1}(Q_{Y_1}^r(\tau)) \right). \]

Substituting this expression into (10) yields
\[ p(1|1)F_{Y|D=1,Z=1}(Q_{Y_1}^r(\tau)) + p(0|1)F_{Y|D=0,Z=1} \left( \tilde{Q}_{Y_0} \left( \tilde{F}_{Y_1}(Q_{Y_1}^r(\tau)) \right) \right) = \tau. \]

Add and subtract \( p(1|0)F_{Y|D=1,Z=0}(Q_{Y_1}^r(\tau)) \) and substitute \( \tau = F_{Y_1}^r(y) \) to obtain
\[ F_{Y_1}^r(y) = (p(1|1) - p(1|0))\tilde{F}_{Y_1}(y) + p(0|1)F_{Y|D=1,Z=0}(y) + p(0|1)F_{Y|D=0,Z=1} \left( \tilde{Q}_{Y_0} \left( \tilde{F}_{Y_1}(y) \right) \right). \]

Using similar arguments, one can show that
\[ F_{Y_0}^r(y) = (p(1|1) - p(1|0))\tilde{F}_{Y_0}(y) + p(0|1)F_{Y|D=0,Z=1}(y) + p(1|0)F_{Y|D=1,Z=0} \left( \tilde{Q}_{Y_1} \left( \tilde{F}_{Y_0}(y) \right) \right). \]

This completes the proof of part (i). \( \square \)

**Lemma 2.** Suppose that Assumption 7 holds and that the IVQR estimands given by (9). Then (i)
\[ Q_{Y_1}^r(F_{Y_1}^r(y)) = \tilde{Q}_{Y_1} \left( \tilde{F}_{Y_0}(y) \right) \quad \text{and} \quad Q_{Y_0}^r(F_{Y_1}^r(y)) = \tilde{Q}_{Y_0} \left( \tilde{F}_{Y_1}(y) \right), \]
and (ii)
\[ F_{Y_1}^r(y) = E \left( \frac{Z}{E(Z)} \right) \left( \tilde{F}_{Y_1}(y) \right), \]
\[ F_{Y_0}^r(y) = E \left( \frac{Z}{E(Z)} \right) \left( \tilde{Q}_{Y_1} \left( \tilde{F}_{Y_0}(y) \right) \right), \]
where \( Q_{Y_0}(\tau) \equiv F_{Y_1}^{-1}(\tau) \) and \( Q_{Y_1}(\tau) \equiv F_{Y_1}^{-1}(\tau). \)

\(^{18}\)The last two steps share some similarities with the derivations in Vuong and Xu (2014, Section 4). In contrast to this paper, they use these steps to derive a counterfactual mapping.
Proof of Lemma 2.

Part (i): Rewrite (9) using law of iterated expectations and the definition of a conditional cdf as

\[ E \left( F_{Y|Z} (Q_{Y_2}^*(\tau)) \right) = \tau, \]
\[ E \left( ZF_{Y|Z} (Q_{Y_2}^*(\tau)) \right) = E(Z) \tau. \]

By the law of iterated expectations,

\[ E \left( F_{Y|D=1,Z} (Q_{Y_1}^*(\tau)) p(1|Z) \right) + E \left( F_{Y|D=0,Z} (Q_{Y_0}^*(\tau)) p(0|Z) \right) = \tau, \]
\[ E \left( ZF_{Y|D=1,Z} (Q_{Y_1}^*(\tau)) p(1|Z) \right) + E \left( ZF_{Y|D=0,Z} (Q_{Y_0}^*(\tau)) p(0|Z) \right) = E(Z) \tau. \]

(13)

(14)

Solving both equations for \( \tau \) and equating them yields:

\[ E \left( F_{Y|D=1,Z} (Q_{Y_1}^*(\tau)) p(1|Z) \right) \left( \frac{Z}{E(Z)} - 1 \right) = E \left( F_{Y|D=0,Z} (Q_{Y_0}^*(\tau)) p(0|Z) \right) \left( 1 - \frac{Z}{E(Z)} \right), \]

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

\[ F_{Y_1} (Q_{Y_1}^*(\tau)) = F_{Y_0} (Q_{Y_0}^*(\tau)), \]

by definition of \( F_{Y_1}(y) \) and \( Q_{Y_1}^*(\tau) \) for \( d \in \{0,1\} \). The result in part (i) now follows by the same arguments as in the proof of part (i) of Lemma 1.

Part (ii): The proof of part (ii) now follows from the result in part (i) and equations (13) and (14) using similar arguments as in the proof of part (ii) of Lemma 1.

C.2 Proofs

Proof of Theorem 1.

I first verify the conditions of Assumption 7. Under Assumption 2, \( F_{Y_1}(y) = F_{Y_1|c}(y) \) and \( F_{Y_0}(y) = F_{Y_0|c}(y) \), which are well-defined and strictly increasing by Assumption 3. Thus, Assumption 7.1 holds. Assumption 7.2 holds by Assumption 23. The result then follows by rewriting part (ii) of Lemma 1 using the results discussed at the beginning of Section 3.1 in the main text.

Proof of Theorem 2.

It follows from the proof of Theorem 1 that Assumptions 2 and 3 imply Assumption 7. To prove the first claim, note that under Assumption 2, part (i) of Lemma 1 implies that \( F_{Y_1|c} (Q_{Y_1}^*(\tau)) = F_{Y_0|c} (Q_{Y_0}^*(\tau)) \). Applying \( Q_{Y_1|c} (\tau) \) on both sides yields \( Q_{Y_1}^*(\tau) = Q_{Y_1|c} (F_{Y_0|c} (Q_{Y_0}^*(\tau))) \). Next, consider

\[ \delta^*(\tau) \equiv Q_{Y_1}^*(\tau) - Q_{Y_0}^*(\tau) \]
\[ = Q_{Y_1|c} (F_{Y_0|c} (Q_{Y_0}^*(\tau))) - Q_{Y_0}^*(\tau) \]
\[ = Q_{Y_1|c} (F_{Y_0|c} (Q_{Y_0}^*(\tau))) - Q_{Y_0|c} (F_{Y_0|c} (Q_{Y_0}^*(\tau))) \]
\[ = \delta_c (F_{Y_0|c} (Q_{Y_0}^*(\tau))) \]
\[ = \delta_c (F_{Y_1|c} (Q_{Y_1}^*(\tau))), \]

where the third equality follows because \( F_{Y_1|c}(y) \) is invertible by Assumption 3; the forth equality is by definition; and the fifth equality follows directly from \( F_{Y_1|c} (Q_{Y_1}^*(\tau)) = F_{Y_0|c} (Q_{Y_0}^*(\tau)) \).

Proof of Theorem 3.

This result is a direct implication of Theorems 1 and 2.

Proof of Theorem 4.

I first verify the conditions of Assumption 7. Under Assumptions 2.2 - 2.4, \( F_{Y_1}(y) = F_{Y_1|c-f}(y) \) and \( F_{Y_0}(y) = F_{Y_0|c-f}(y) \) which are well-defined and strictly increasing by assumption. Assumption 7.2 holds by Assumption...
2.3. Next, under Assumptions 2.2 - 2.4, $p(d, z)$ and $F_{Y|D=d, Z=z}(y)$ for $(d, z) \in \{0, 1\} \times \{0, 1\}$ can be related to the fractions and potential outcome cdfs of types $T \in \{a, n, c, f\}$ as (e.g., Huhber, 2014):

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

Substituting the above expressions into the results in part (ii) of Lemma 1 completes the proof.

**Proof of Theorem 5.**

It follows from the proof of Theorem 4 that Assumption 7 is implied by the assumptions of the theorem. Under Assumptions 2.2 - 2.4, part (i) of Lemma 1 implies that $F_{Y_1|c-f}(Q_1^*(\tau)) = F_{Y_0|c-f}(Q_0^*(\tau))$. The result in the theorem now follows by similar arguments as in the proof of Theorem 2.

**Proof of Theorem 6.**

I split the proof into three steps. In step 1, I reexpress $\tilde{F}_Y(y)$ and $\tilde{F}_Z(y)$ using the relationship between the observed conditional probabilities and cdfs and the shares and potential outcome cdfs for the $K+1$ subpopulations discussed in Section 4.2. Step 2 verifies Assumption 7 and step 3 uses Lemma 2 to prove the results in the theorem.

**Step 1:** This step expresses $\tilde{F}_Y(y)$ and $\tilde{F}_Z(y)$ (defined in Lemma 2) under Assumptions 4. Consider

$$
\tilde{F}_Y(y) = \frac{\sum_{k=1}^{K} F_{Y|D=1, Z=z_k}(y) p(1|z_k) P(Z = z_k) \left( \frac{\hat{z}_k}{\hat{f}(z_k)} - 1 \right)}{\sum_{k=1}^{K} p(1|z_k) P(Z = z_k) \left( \frac{\hat{z}_k}{\hat{f}(z_k)} - 1 \right)}
$$

where the first equality follows from specializing $\tilde{F}_Y(y)$ and $\tilde{F}_Z(y)$ to the setup of the theorem; the second equality is by Assumption 4; the third equality follows from expanding the numerator and denominator and because $\pi_a F_{Y_1|a}(y)(E(Z)/E(Z) - 1) = \pi_a (E(Z)/E(Z) - 1) = 0$; the fourth equality is by changing the order of summation; and the fifth equality is because

$$
w_j \equiv \sum_{k=1}^{K} z_k P(Z = z_k) \frac{E(Z)}{E(Z) - 1} - P(Z \geq z_j)
$$

and

$$
\tilde{F}_Z(z) = \frac{\sum_{j=1}^{J} c_{jz} (\sum_{k=1}^{K} z_k P(Z = z_k) \frac{E(Z)}{E(Z) - 1} - P(Z \geq z_j))}{\sum_{j=1}^{J} c_{jz}}.
$$

The proof of the theorem follows from these expressions.
Similarly, we find
\[
F_{Y_0}(y) = \frac{\sum_{j=2}^K \pi_{cz_j} F_{Y_0|cz_j}(y) \left( P(Z < z_j) - \frac{\sum_{k=1}^{j-1} z_k \mathbb{E}(Z = z_k)}{\mathbb{E}(Z)} \right)}{\sum_{j=2}^K \pi_{cz_j} \left( \frac{\sum_{k=1}^{j-1} z_k P(Z = z_k)}{\mathbb{E}(Z)} - P(Z \geq z_j) \right)}
\]
\[
= \sum_{j=2}^K \frac{w_j \pi_{cz_j} F_{Y_0|cz_j}(y)}{\sum_{j=2}^K \pi_{cz_j} w_j},
\]

because \( P(Z < z_j) - \sum_{k=1}^{j-1} z_k P(Z = z_k) / \mathbb{E}(Z) = w_j \).

**Step 2:** By Assumptions 4 and 5, \( \pi_{cz_j} > 0 \), \( f_{Y_1|cz_j}(y) > 0 \) and \( f_{Y_0|cz_j}(y) > 0 \) for all \( j = 2, ..., K \). Hence, because \( w_j > 0 \) for all \( j = 2, ..., K \), \( \tilde{F}_{Y_1}(y) \) and \( \tilde{F}_{Y_0}(y) \) are strictly increasing and well-defined and Condition 7.1 holds. Assumption 7.2 is directly implied by Assumption 4.3.

**Step 3:** This step expresses \( F_{Y_1}(y) \) and \( F_{Y_0}(y) \) given by Lemma 2 under Assumption 4. By Lemma 2 and Assumption 4,
\[
F_{Y_1}(y) = \sum_{k=1}^K \pi_a F_{Y_1|a}(y) P(Z = z_k) + \sum_{j=2}^K \sum_{k=1}^K \pi_{cz_j} F_{Y_1|cz_j}(y) \mathbb{E}(Y \geq z_j) P(Z = z_k)
\]
\[
+ \left( \sum_{k=1}^K \pi_{cz_k} F_{Y_0|cz_k}(y) \mathbb{E}(Y < z_j) \right) P(Z < z_j).
\]

By similar arguments, we obtain
\[
F_{Y_0}(y) = \sum_{k=1}^K \pi_a F_{Y_1|a}(y) \mathbb{E}(Y \geq z_j) P(Z = z_k) + \sum_{k=1}^K \pi_{cz_k} F_{Y_0|cz_k}(y) \mathbb{E}(Y < z_j) P(Z < z_j).
\]

Changing the order of summation and summarizing terms yields:
\[
F_{Y_1}(y) = \pi_a F_{Y_1|a}(y) + \sum_{j=2}^K \pi_{cz_j} F_{Y_1|cz_j}(y) P(Z \geq z_j)
\]
\[
+ \pi_a F_{Y_0|a}(y) \mathbb{E}(Y \geq z_j) P(Z < z_j).
\]

By similar arguments, we obtain
\[
F_{Y_0}(y) = \pi_a F_{Y_1|a}(y) \mathbb{E}(Y \geq z_j) P(Z = z_k) + \sum_{j=2}^K \pi_{cz_j} F_{Y_0|cz_j}(y) P(Z < z_j).
\]

This completes the proof of the theorem.

**Proof of Theorem 7.**

By step 2 of the proof of Theorem 6, Assumptions 4 and 5 imply Assumption 7. Therefore, it follows from Lemma 2 that \( Q_{Y_1}(F_{Y_0}^*)(y) = \tilde{Q}_{Y_1} \left( \tilde{F}_{Y_0}(y) \right) \) and \( Q_{Y_0}(F_{Y_1}^*)(y) = \tilde{Q}_{Y_0} \left( \tilde{F}_{Y_1}(y) \right) \). The result in the theorem now follows by similar arguments as in the proof of Theorem 2.