TESTING FOR RANK INVARIANCE OR SIMILARITY IN PROGRAM EVALUATION

Yingying Dong and Shu Shen*

Abstract—This paper discusses testable implications of rank invariance or rank similarity, assumptions that are common in program evaluation and in the quantile treatment effect (QTE) literature. We nonparametrically identify, estimate, and test the counterfactual distribution of potential ranks, or features of the distribution. The proposed tests allow treatment to be endogenous, with exogenous treatment following as a special case. The tests essentially do not require any additional assumptions other than those to identify and estimate QTEs. We apply the proposed tests to investigate whether the Job Training Partnership Act training causes trainees to systematically change their ranks in the earnings distribution.

I. Introduction

IN the last decade or so, researchers have increasingly sought to identify and estimate distributional effects of social programs or public policies. Rank invariance or rank preservation is required either as a key identifying assumption or to interpret the identified distributional effects as individual causal effects.[1]

This paper discusses testable implications of rank invariance or a weaker condition, rank similarity, and proposes nonparametric tests that work for both. It is well known that without functional form restrictions, one cannot identify the joint distribution of potential outcomes. As a result, any conditions imposed on the joint distribution are not directly testable. Our tests instead draw on the implications of the conditional distribution of potential outcomes conditional on observable covariates, that is, among observationally equivalent individuals.

We nonparametrically identify, estimate, and test the counterfactual distribution of potential ranks or features of the distribution (e.g., mean, median, or any particular quantile) among observationally equivalent individuals. We focus on ranks in the unconditional distribution of potential outcomes. The proposed test can be readily extended to testing conditional potential ranks or ranks in the conditional distribution of potential outcomes.[2] In an independent and contemporaneous work, Frandsen and Lefgren (2015) also leverage observable covariates and propose a regression-based test for rank similarity.

The proposed tests allow treatment to be endogenous. Exogenous treatment follows as a special case. Except for mild regularity conditions, our tests do not require any additional assumptions other than those used to identify and estimate QTEs. Covariates are permitted in estimating the unconditional QTEs, so these tests can handle instrumental variables regardless of whether they are valid conditional on covariates or valid unconditionally. We apply our tests to evaluate rank invariance or similarity in the Job Training Partnership Act (JTPA) training program. We show that this training program causes trainees to systematically change their ranks in the earnings distribution. The impacts of the JTPA training on individual trainees can be very different from the distributional effects of this program.

The rest of the paper proceeds as follows. Section II discusses the testable implication of rank invariance and rank similarity, as well as the corresponding conditional moment restrictions. Also discussed is the identification of the counterfactual distribution of potential ranks among observationally equivalent individuals. Section III provides our test statistic along with its asymptotic distribution. Section IV presents the empirical application. Section V concludes. Monte Carlo simulations, proofs, and several extensions are provided in the online appendixes.

II. Model Setup and Identification

A. *Rank Invariance, Rank Similarity, and Their Testable Implications*

We first define rank invariance and rank similarity imposed on the unconditional distribution of potential outcomes and then discuss their implications.

Consider the standard potential outcome framework. Let T be the binary treatment indicator. $T = 1$ if an individual is treated and 0 otherwise. Let Y_t for $t = 0, 1$ be the potential outcome under no treatment or treatment. The observed

Received for publication October 23, 2015. Revision accepted for publication February 8, 2017. Editor: Bryan S. Graham.

* Dong: University of California, Irvine; Shen: University of California, Davis.

We thank seminar participants at UCSD, UCLA, UC Riverside, McGill University for helpful comments. Thanks also go to participants at the 11th World Congress of the Econometric Society, the California Econometrics Conference, the Interactions workshop at University of Chicago, and the 2016 North American Winter Meeting of the Econometric Society. All errors are our own. The second author acknowledges financial support from the Hellman Foundation.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/REST_a_00686.

[1] For the former see, e.g., the instrumental variable quantile regression (IVQR) model by Chernozhukov and Hansen (2005, 2006, 2008) or the nonparametric IV quantile regression models of Chernozhukov, Imbens, and Newey (2007) and Horowitz and Lee (2007), which implicitly impose rank invariance by imposing a scalar disturbance. For the latter, see, e.g., the LQTE framework of Abadie, Angrist, and Imbens (2002), Frolich and Melly (2013), Firpo (2007), Imbens and Newey (2009), and Firpo and Pinto (2015).

[2] Note that we do not view the proposed tests as tests for the identifying assumption of the Chernozhukov and Hansen (2005, 2006, 2008) IVQR model or the nonparametric IV quantile regression model of Chernozhukov et al. (2007) and Horowitz and Lee (2007), since these models impose rank invariance or similarity on conditional potential ranks, after conditioning on all relevant observables. Similarity or invariance of conditional ranks is more plausible when conditioning on a rich set of covariates, as discussed in Chernozhukov and Hansen (2006). Our proposed tests utilize the predictive power of observable covariates for potential ranks. These tests are useful in testing rank invariance or preservation when not all relevant covariates are included in the conditioning set of the conditional potential ranks.

The Review of Economics and Statistics, March 2018, 100(1): 78–85

doi:10.1162/REST_a_00686

outcome is then $Y = Y_0(1-T) + Y_1 T$. Assume that Y is continuous. Let $F_t(.)$ and $q_t(.)$ for $t = 0, 1$ be the cumulative distribution function and quantile function of Y_t. Following Doksum (1974) and Lehmann (1974), we define the unconditional QTE as $QTE(\tau) = q_1(\tau) - q_0(\tau)$ for $\tau \in (0, 1)$. Let $U_t = F_t(Y_t)$, $t = 0, 1$ be the potential rank; $U_t \sim U(0, 1)$ by construction.

Definition 1. *Rank invariance is the condition that* $U_0 = U_1$.

Under rank invariance, an individual's potential rank with or without treatment remains the same, so U_0 and U_1 are the same random variable. For example, assume $Y_t = g(t, \mathbf{X}, V)$, $t = 0, 1$ for some observables $\mathbf{X}$ and unobservables V. V may be a vector. The potential rank is then given by $U_t = F_t(g(t, \mathbf{X}, V))$. Rank invariance holds if and only if $U_0|(\mathbf{X} = \mathbf{x}, V = v) = U_1|(\mathbf{X} = \mathbf{x}, V = v)$ for all $(\mathbf{x}, v)$. Immediately it implies $U_0|(\mathbf{X} = \mathbf{x}) \sim U_1|(\mathbf{X} = \mathbf{x})$ for all $\mathbf{x}$, where $\sim$ means "follows the same distribution as."

Rank invariance may be restrictive in practice. Consider the following thought experiment. A test is given to a random sample of students and their clones. Let the treatment T be a binary indicator for being a clone. Y_0 and Y_1 are then the potential test scores for a student and her clone, respectively. This treatment is supposed to have no effects on test scores. However, due to random chance or luck, a student and her clone may not have the same test score and hence the same rank. Nevertheless, if we repeat this experiment infinitely many times, the student and her clone will have the same distribution of ranks.[3]

Rank similarity relaxes rank invariance by allowing for random deviations, or slippages in one's rank away from some common level (Chernozhukov & Hansen, 2005). Consider again the above thought experiment. Assume that potential test score is determined by $Y_t = g(\mathbf{X}, V, S_t)$, where $\mathbf{X}$ (e.g., education) and V (e.g., ability) determine the expected common rank level for a student and her clone, and S_t, $t = 0, 1$, are idiosyncratic shocks (e.g., luck) mutually independent and identically distributed across the two treatment states. S_t is responsible for random slippages from the common rank level. Then $U_t|(\mathbf{X} = \mathbf{x}, V = v)$, $t = 0, 1$ have the same distribution. Formally define rank similarity as follows.

Definition 2. *Rank similarity is the condition that* $U_0|(\mathbf{X} = \mathbf{x}, V = v) \sim U_1|(\mathbf{X} = \mathbf{x}, V = v)$ *for any* $(\mathbf{x}, v)$ *in its support* $\mathscr{W}$, *where* $\mathbf{X}$ *and* V *are the observable and unobservable determinants or shifters of the common rank level under treatment or no treatment.*

While rank invariance requires U_0 and U_1 to be the same random variable, rank similarity assumes only that they have the same conditional distribution. Chernozhukov and Hansen (2005) consider a weaker condition: rank similarity for conditional potential ranks. Rank similarity among unconditional potential ranks implies rank similarity among conditional potential ranks.[4]

Rank invariance is a special case of rank similarity, where the distribution of potential ranks is degenerate. The following discussion therefore focuses on rank similarity. All the conclusions hold trivially for rank invariance.

Lemma 1. *Given rank similarity, for all* $\tau \in (0, 1)$,

1. $F_{\mathbf{X},V|U_0}(\mathbf{x}, v|\tau) = F_{\mathbf{X},V|U_1}(\mathbf{x}, v|\tau)$ *for all* $(\mathbf{x}, v) \in \mathscr{W}$;
2. *(Main testable implication)* $F_{U_0|\mathbf{X}}(\tau|\mathbf{x}) = F_{U_1|\mathbf{X}}(\tau|\mathbf{x})$ *for all* $\mathbf{x}$ *in its support* $\mathscr{X}$.

Part 1 of lemma 1 follows immediately from the definition of rank similarity and Bayes' rule. It states that given rank similarity, at the same rank of the potential outcome distributions, the distribution of all relevant observables and unobservables remains the same. To investigate rank preservation, empirical researchers in program evaluation often check covariate similarity at the same quantile of the treatment and control outcome distributions (Bitler, Gelbach, & Hoynes, 2006, 2008). Here we show that the distribution of covariates is also identical at the same potential rank under this weaker condition, rank similarity.

Part 2 of lemma 1 states that under rank similarity, the distribution of potential ranks among observationally equivalent individuals is the same across treatment states. This implication is what we directly test. Without further assumptions, this condition is only a necessary condition for rank similarity. Below we provide an assumption under which this testable implication is also a sufficient condition.

Lemma 2. *If* $F_{V|\mathbf{X},U_0}(v|\mathbf{x},\tau) = F_{V|\mathbf{X},U_1}(v|\mathbf{x},\tau)$ *for all* $\tau \in (0, 1)$ *and* $\mathbf{x} \in \mathscr{X}$, *rank similarity holds if and only if* $F_{U_0|\mathbf{X}}(\tau|\mathbf{x}) = F_{U_1|\mathbf{X}}(\tau|\mathbf{x})$ *for all* $\tau \in (0, 1)$ *and* $\mathbf{x} \in \mathscr{X}$.

Lemma 2 assumes that given observables $\mathbf{X} = \mathbf{x}$, the distribution of unobservables is the same at the same rank of the potential outcome distribution. Note that this assumption does not assume away unobservables. In particular, it does not imply $F_{U_t|\mathbf{X},V}(.|\mathbf{x},v) = F_{U_t|\mathbf{X}}(.|\mathbf{x})$, $t = 0, 1$. This assumption resembles in spirit the unconfoundedness assumption that is popular in program evaluation (e.g., in various matching estimators). It is essentially not testable, just like unconfoundedness.

[3] Note that luck is plausibly mutually i.i.d. across treatment states (i.e., for a student and her clone).

[4] To see the difference, by the Skorohod representation of potential outcomes, one can write $Y_t = q(t, U_t)$, where $q(t, U_t)$ is the quantile function of Y_t. U_t is then the unconditional potential rank when $T = t$. In contrast, Chernozhukov and Hansen (2005, 2006, 2008) define a conditional potential rank using $Y_t \equiv q(t, \mathbf{x}, \widetilde{U}_t)$, so $\widetilde{U}_t$ represents the conditional potential rank conditional on $\mathbf{X} = \mathbf{x}$. $\widetilde{U}_t$ is responsible for the heterogeneity of outcomes among individuals with the same observed characteristics $\mathbf{X} = \mathbf{x}$ in treatment state t. Their rank similarity then assumes $\widetilde{U}_0 \sim \widetilde{U}_1$ conditional on $\mathbf{X}$ and the treatment model unobservables V.

Given rank similarity, $E[Y_1 - Y_0|\mathbf{X} = \mathbf{x}, V = v] = \int_0^1 QTE(\tau) dF_{U|\mathbf{X},V}(\tau|\mathbf{x}, v)$ for all $(\mathbf{x}, v) \in \mathscr{W}$, where $F_{U|\mathbf{X},V}(.|\mathbf{x}, v) \equiv F_{U_t|\mathbf{X},V}(.|\mathbf{x}, v)$, $t = 0, 1$. Identification of $E[Y_1 - Y_0|\mathbf{X} = \mathbf{x}, V = v]$ then relies on researchers' ability to identify the QTEs and the conditional distribution of ranks $F_{U|\mathbf{X},V}(\tau|\mathbf{x}, v)$. That is, under rank similarity, one loses the ability to nonparametrically identify treatment effects for particular individuals (see, e.g., Imbens & Newey, 2009).

B. *Identification of the Potential Rank Distribution*

This section discusses identification of the potential rank distributions and, further, identification of the conditional moment restrictions implied by rank invariance or rank similarity. We focus on endogenous treatment, with exogenous treatment following as a special case.

To test our main testable implication, different frameworks may be adopted. As is evident from theorem 1 and its proof, what is required is a valid instrumental variable (or the unconfoundedness assumption) that allows one to identify and estimate (a) the quantile function of the potential outcome distribution or equivalent and (b) conditioning on some observable covariates, the effects of treatment on the distribution of potential ranks (constructed using the estimated quantiles in the first step).

Here we adopt the LQTE framework. The LQTE identifying assumptions are particularly suitable for our empirical application of the JTPA training program (see the discussion in Abadie et al., 2002). The LQTE framework permits unrestricted heterogeneity of treatment effects and consequently identifies QTEs only for compliers, which are the largest subpopulation for which QTEs can be point identified without further restrictions. Testing for rank similarity is then relevant only among compliers. If assumptions are made to identify unconditional QTEs for the whole population, one can analogously test for rank similarity for the whole population. A brief discussion on this is provided at the end of this section.

Let Z be a binary instrumental variable (IV). Further let $T_z, z = 0, 1$, be the potential treatment status if $Z = z$. The observed treatment status is then given by $T = T_0(1 - Z) + T_1 Z$. We make the following identifying assumptions.

Assumption 1. *Let* $(Y_0, Y_1, T_0, T_1, X, Z)$ *be random variables mapped from the common probability space* $(\Omega, \mathsf{F}, \mathsf{P})$. *Assume that the data-generating process satisfies the following conditions: (a) Independence:* $(Y_0, Y_1, T_0, T_1) \perp Z|\mathbf{X}$. *(b) First stage:* $E(T_1) \neq E(T_0)$. *(c) Monotonicity:* $\Pr(T_1 \geq T_0) = 1$. *(d) Nontrivial assignment:* $0 < \Pr(Z = 1|\mathbf{X} = \mathbf{x}) < 1$ *for all* $\mathbf{x} \in \mathscr{X}$.

Assumption 1 is the standard LQTE identifying assumption used in Abadie et al. (2002) and Abadie (2003), except that we allow for a weaker first stage. In particular, we do not require compliers to exist at every value of $\mathbf{X}$. The reason is that we identify and estimate unconditional potential ranks instead of conditional potential ranks (see the discussion in Frolich & Melly, 2013) and that our test examines whether rank similarity is violated at any potential ranks.

When Z is a random assignment indicator, as in our empirical application, the independence restriction is valid without conditioning on covariates $\mathbf{X}$. Including covariates can remove any chance association between T and $\mathbf{X}$ or improve efficiency.

Define compliers as individuals with $T_1 > T_0$ (Angrist, Imbens, & Rubin, 1996). Let C denote the set of compliers. Define the cumulative distribution function of Y_t among compliers as $F_{t|C}(y) = \Pr(Y_t \leq y|T_1 > T_0)$ for $t = 0, 1$. We are interested in testing for rank similarity among compliers. For notational convenience, unless stated otherwise in the following, we use U_t to refer to potential ranks among compliers only, that is, $U_t \equiv U_{t|C} = F_{t|C}(Y_t)$ for $t = 0, 1$. Let $\mathscr{X}_C = \{\mathbf{x} \in \mathscr{X} : \Pr[T_1 > T_0|\mathbf{X} = \mathbf{x}] > 0\}$. Analogous to part 2 of lemma 1, rank similarity among compliers implies

$$F_{U_1|C,\mathbf{X}}(\tau|\mathbf{x}) = F_{U_0|C,\mathbf{X}}(\tau|\mathbf{x}) \text{ for all } \tau \in (0, 1) \text{ and } \mathbf{x} \in \mathscr{X}_C. \tag{1}$$

Let $q_{t|C}(\tau)$ for $t = 0, 1$ be the τ quantile of potential outcomes Y_t among compliers. The following theorem provides identification of equation (1):

Theorem 1. *Define* $I(\tau) \equiv \mathbf{1}(Y \leq (Tq_{1|C}(\tau) + (1 - T) q_{0|C}(\tau)))$. *Given assumption 1, for all* $\tau \in (0, 1)$, $\mathbf{x} \in \mathscr{X}_C$, *and* $t = 0, 1$,

$$F_{U_t|C,\mathbf{X}}(\tau|\mathbf{x}) = \frac{E[I(\tau)\mathbf{1}(T = t)|Z = 1, \mathbf{X} = \mathbf{x}] - E[I(\tau)\mathbf{1}(T = t)|Z = 0, \mathbf{X} = \mathbf{x}]}{E[\mathbf{1}(T = t)|Z = 1, \mathbf{X} = \mathbf{x}] - E[\mathbf{1}(T = t)|Z = 0, \mathbf{X} = \mathbf{x}]}. \tag{2}$$

Further, equation (1) *holds if and only if*

$$E[I(\tau)|Z = 1, \mathbf{X} = \mathbf{x}] = E[I(\tau)|Z = 0, \mathbf{X} = \mathbf{x}] \text{ for all } \tau \in (0, 1) \text{ and } \mathbf{x} \in \mathscr{X}. \tag{3}$$

$I(\tau)$ is identified because $q_{t|C}(\tau)$ for $t = 0, 1$ is identified following Frolich and Melly (2013) under assumption 1. Note that equation (3) holds trivially for any $\mathbf{x} \in \mathscr{X}/\mathscr{X}_C$, and hence under rank similarity, equation (3) holds for all $\mathbf{x} \in \mathscr{X}$. Note also that theorem 1 nests exogenous treatment as a special case. When T is exogenous, $Z = T$ and everyone is a complier.

Equation (2) in theorem 1 provides a convenient way to estimate the entire distribution of potential ranks among those with $\mathbf{X} = \mathbf{x}$ and thereby quantify the degree of violation in rank similarity at different quantiles or different covariate values. For testing purposes only, one can simply use the reduced-form equation (3). In practice, the covariates $\mathbf{X}$ need to be nontrivial in order for a test based on equation (3) to have any power. If $F_{t|C,\mathbf{X}}(\mathbf{x}) = F_{t|C}$ for $t = 0, 1$ and all $\mathbf{x} \in \mathscr{X}_C$, then equation (3) would hold by construction.

Theorem 1 suggests testing for rank similarity by a two-step procedure: first, estimate the unconditional quantiles $q_{0|C}(\tau)$ and $q_{1|C}(\tau)$; second, test whether equation (3) holds for all $\tau \in (0, 1)$ and $\mathbf{x} \in \mathscr{X}$, after replacing $q_{0|C}(\tau)$ and $q_{1|C}(\tau)$ with their estimators.

If desired, one may test a particular quantile, such as the median, or a subset of quantiles. One can also test functionals, such as any moments of the potential rank distribution. Equation (1) implies that

$$E[U_1^p|C, \mathbf{X} = \mathbf{x}] = E[U_0^p|C, \mathbf{X} = \mathbf{x}], \tag{4}$$

for all $\mathbf{x} \in \mathscr{X}_C$ and some integer $p \geq 1$. When $p = 1$, equation (4) suggests a mean test for rank similarity. Let $U \equiv TU_1 + (1-T)U_0 = \int_0^1 1((Tq_{1|C}(\tau) + (1-T)q_{0|C}(\tau)) < Y)d\tau = 1 - \int_0^1 I(\tau)d\tau$. Analogous to theorem 1, equation (4) holds if and only if for all $\mathbf{x} \in \mathscr{X}$,

$$E\left[U^p|Z = 1, \mathbf{X} = \mathbf{x}\right] = E\left[U^p|Z = 0, \mathbf{X} = \mathbf{x}\right].$$

This discussion assumes that a valid instrument is available. If instead unconfoundedness holds given some covariates, one can first apply the estimator of Firpo (2007) or Donald and Hsu (2014) to estimate unconditional quantiles for the whole population, and then test rank similarity for the whole population using these covariates.[5]

III. Testing

This section constructs nonparametric tests for the testable implication discussed in section B. The vector of observables $\mathbf{X}$ is assumed to be discrete with finite support. In practice, one typically has a limited number of covariates, and one can always discretize covariates. Tests with continuous covariates or discrete covariates with large support are discussed in online appendix C.

A. The Distributional Test for Rank Similarity

Let $\mathscr{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_J\}$ be the support of $\mathbf{X}$ and $\Omega = \{\tau_1, \tau_2, \ldots, \tau_K\}$ be the set of quantiles of interest. For any $j = 1, \ldots, J$, $z = 0, 1$ and $k = 1, \ldots, K$, define $m_j^z(\tau_k) \equiv E\left[I(\tau_k)|Z = z, \mathbf{X} = \mathbf{x}_j\right]$. We are interested in testing the null and alternative hypotheses:

$$H_0\text{: } m_j^0(\tau_k) = m_j^1(\tau_k), \text{ for all } j = 1, \ldots, J-1 \text{ and } k = 1, \ldots, K,$$

$$H_a\text{: } H_0 \text{ is not true.}$$

Note that only $J - 1$ values of $\mathbf{X}$ are included in the null hypothesis, since $\sum_{j=1}^J m_j^1(\tau) \Pr\left(\mathbf{X} = \mathbf{x}_j\right) = \sum_{j=1}^J m_j^0(\tau) \Pr\left(\mathbf{X} = \mathbf{x}_j\right)$ for all $\tau \in (0, 1)$, as is shown in online appendix B.

Let $\{Y_i, T_i, Z_i, \mathbf{X}_i\}_{i=1}^n$ be a sample of i.i.d. draws of size n from $(Y, T, Z, \mathbf{X})$. Let $\hat{q}_{t|C}(\tau_k)$ be the $\sqrt{n}$-consistent estimators of $q_{t|C}(\tau_k)$ proposed by Frolich and Melly (2013), which have the advantage of allowing for covariates while estimating unconditional quantiles.

$$\left(\hat{q}_{0|C}(\tau_k), \hat{q}_{1|C}(\tau_k)\right) = \arg\min_{q_0, q_1} \frac{1}{n} \sum_{i=1}^n \rho_{\tau_k}(Y_i - q_0(1 - T_i) - q_1 T_i)\hat{\omega}_i,$$

where $\rho_{\tau_k}(u) = u(\tau_k - \mathbf{1}(u < 0))$ is the standard check function, $\hat{\omega}_i = \left(\frac{Z_i}{\widehat{\pi}(\mathbf{X}_i)} - \frac{1-Z_i}{1-\widehat{\pi}(\mathbf{X}_i)}\right)(2T_i - 1)$ and $\widehat{\pi}(\mathbf{x}) = \frac{\sum_{i=1}^n \mathbf{1}(Z_i=1, \mathbf{X}_i=\mathbf{x})}{\sum_{i=1}^n \mathbf{1}(\mathbf{X}_i=\mathbf{x})}$ estimates $\pi(\mathbf{x}) \equiv P(Z_i = 1|\mathbf{X}_i = \mathbf{x})$.[6]

Define $n_j^z \equiv \sum_{i=1}^n \mathbf{1}(Z_i = z, \mathbf{X}_i = \mathbf{x}_j)$. The nonparametric estimator of $m_j^z(\tau_k)$ is then

$$\hat{m}_j^z(\tau_k) = \frac{1}{n_j^z} \sum_{Z_i = z, \mathbf{X}_i = \mathbf{x}_j} \times \mathbf{1}\left(Y_i \leq \left(T_i \hat{q}_{1|C}(\tau_k) + (1 - T_i)\hat{q}_{0|C}(\tau_k)\right)\right).$$

Let $\hat{\mathsf{m}}_j^z = \left[\hat{m}_j^z(\tau_1)\ \hat{m}_j^z(\tau_2)\ \cdots\ \hat{m}_j^z(\tau_K)\right]'$ and $\mathsf{m}_j^z = \left[m_j^z(\tau_1)\ m_j^z(\tau_2)\ \cdots\ m_j^z(\tau_K)\right]'$ be $K \times 1$ vectors, and $\hat{\mathbf{m}}^z = \left[(\hat{\mathsf{m}}_1^z)'\ (\hat{\mathsf{m}}_2^z)' \cdots\ (\hat{\mathsf{m}}_{J-1}^z)'\right]'$ and $\mathbf{m}^z = \left[(\mathsf{m}_1^z)'\ (\mathsf{m}_2^z)'\ \cdots\ (\mathsf{m}_{J-1}^z)'\right]'$ be $K(J-1) \times 1$ vectors. Let $\hat{\mathbf{V}}$ be a consistent estimator of the asymptotic variance-covariance matrix of $\sqrt{n}\left(\hat{\mathbf{m}}^1 - \hat{\mathbf{m}}^0\right)$ under the null hypothesis H_0. We propose a Wald-type test statistic:

$$W = n\left(\hat{\mathbf{m}}^1 - \hat{\mathbf{m}}^0\right)' \hat{\mathbf{V}}^{-1}\left(\hat{\mathbf{m}}^1 - \hat{\mathbf{m}}^0\right).$$

To derive asymptotic properties of the test statistic W, we make following assumptions.

Assumption 2. *(a) The data $\{Y_i, T_i, Z_i, \mathbf{X}_i\}_{i=1}^n$ are a random sample from $(Y, T, Z, \mathbf{X})$. (b) For all $\tau \in \Omega = \{\tau_1, \tau_2, \ldots, \tau_K\}$ and $t = 0, 1$, the distribution of Y_t among compliers, or $F_{t|C}$, is absolutely continuous with density function $f_{t|C}$ that is positive and bounded in a neighborhood of $q_{t|C}(\tau)$. (c) For all $j = 1, \ldots, J$, $p_j = \Pr\left(\mathbf{X} = \mathbf{x}_j\right) > 0$. (d) Let $f_{Y|\mathbf{X}}(y|\mathbf{x}_j)$ be the conditional density of Y given $\mathbf{X}$. For $j = 1, \ldots, J$, $f_{Y|\mathbf{X}}(y|\mathbf{x}_j)$ is positive and continuouly differentiable in a neighborhood of $q_{t|C}(\tau)$, for all $t = 0, 1$ and $\tau \in \Omega$.*

Parts a and b of assumption 2 guarantee consistency of $\hat{q}_{t|C}(\tau_k)$ for $t = 0, 1$ and $k = 1, \ldots, K$. Parts c and d ensure that the asymptotic variance-covariance matrix of $\hat{\mathbf{m}}^1 - \hat{\mathbf{m}}^0$

[5] See also Chernozhukov, Fernández-Val, and Melly (2013) for estimation and inference of counterfactual distributions and functionals of the counterfactual distributions without requiring an IV.

[6] The Stata command ivqte can be conveniently used to estimate $\hat{q}_{0|C}(\tau_k)$ and $\hat{q}_{1|C}(\tau_k)$. $\hat{\omega}_i$ in practice is replaced by projected weights projected onto Y and T to make sure that the weights are nonnegative.

is bounded and has full rank. Let $p_{Z,\mathbf{X}}(z,\mathbf{x}_j)$ be the joint probability of $Z=z, \mathbf{X}=\mathbf{x}_j$, $p_{T|Z,\mathbf{X}}(t|z,\mathbf{x}_j)$ be the probability of $T=t$ given $Z=z$ and $\mathbf{X}=\mathbf{x}_j$, let $P_c = E[T|Z=1]-E[T|Z=0]$ be the proportion of compliers, and let $f_{Y|T,Z,\mathbf{X}}$ be the conditional density of Y given T, Z, and $\mathbf{X}$. The following theorem provides the asymptotic distribution of $\hat{\mathbf{m}}^1-\hat{\mathbf{m}}^0$.

Theorem 2. *Given assumptions 1 and 2,*

$$\sqrt{n}\left(\hat{\mathbf{m}}^1-\hat{\mathbf{m}}^0-\left(\mathbf{m}^1-\mathbf{m}^0\right)\right)\Rightarrow N(0,\mathbf{V}).$$

Matrix $\mathbf{V}$ *is* $K(J-1)\times K(J-1)$ *and p.d. Its* $\left(\sum_{j=1}^{J-1}K(j-1)+k, \sum_{j'=1}^{J-1}K(j'-1)+k'\right)$*th element is equal to* $E\left[\left(\phi_j^1(\tau_k)-\phi_j^0(\tau_k)\right)\left(\phi_{j'}^1(\tau_{k'})-\phi_{j'}^0(\tau_{k'})\right)\right]$ *with*

$$\begin{aligned}\phi_j^z(\tau_k) &\equiv \phi_j^z(\tau_k;Y,T,Z,\mathbf{X}) = \left(I(\tau_k)-m_j^z(\tau_k)\right)\\ &\quad\times \mathbf{1}(Z=z,\mathbf{X}=\mathbf{x}_j)/p_{Z,X}(z,\mathbf{x}_j)\\ &\quad-\sum_{t=0,1} f_{Y|T,Z,\mathbf{X}}(q_{t|C}(\tau_k)|t,z,\mathbf{x}_j)p_{T|Z,\mathbf{X}}(t|z,\mathbf{x}_j)\\ &\quad\times \psi_t(Y,T,Z,\mathbf{X})/P_c/f_{t|C}\left(q_{t|C}(\tau_k)\right).\end{aligned}$$

Functions $\psi_0(Y,T,Z,\mathbf{X})$ *and* $\psi_1(Y,T,Z,\mathbf{X})$ *are defined in the proof of theorem 7 in Frolich and Melly (2007) and are restated in the proof of this theorem in online appendix B.*

The last two terms of $\phi_j^z(\tau_k)$ come from the estimation errors of $\hat{q}_{0|C}(\tau_k)$ and $\hat{q}_{1|C}(\tau_k)$. If these quantiles were known, $\phi_j^z(\tau_k)$ would reduce to $\left(I(\tau_k)-m_j^z(\tau_k)\right)\mathbf{1}(Z=z,\mathbf{X}=\mathbf{x}_j)/p_{Z,\mathbf{X}}(z,\mathbf{x}_j)$, and, hence, the $\left(\sum_{j=1}^{J-1}K(j-1)+k, \sum_{j'=1}^{J-1}K(j'-1)+k'\right)$th element of $\mathbf{V}$ would reduce to 0 if $j\neq j'$ and $\sum_{z=0,1}\left(m_j^z(\tau_k\wedge\tau_{k'})-m_j^z(\tau_k)m_j^z(\tau_{k'})\right)$ if $j=j'$. Given the above theorem, it follows immediately that the test statistic W converges in distribution to $\chi^2(K(J-1))$ under the null and explodes under the alternative. Therefore, we will reject the null if W exceeds the $(1-\alpha)\times 100$th percentile of the $\chi^2(K(J-1))$ distribution. The following corollary summarizes the asymptotic properties of the test:

Corollary 1. *Let* c_α *be the* $(1-\alpha)\times 100$*th percentile of the* $\chi^2(K(J-1))$ *distribution. Given assumptions 1 and 2, we have that (a) if* H_0 *is true,* $\lim_{n\to\infty}P(W>c_\alpha)=\alpha$*; (b) if* H_0 *is false,* $\lim_{n\to\infty}P(W>c_\alpha)=1$.

Considering the complicated nature of the asymptotic variance-covariance matrix resulting from the first-stage estimation of the unconditional quantile functions, one may estimate $\mathbf{V}$ by bootstrap. Note that the set $\mathscr{X}$ is finite here. We discuss an extension that allows J to go to infinity with the sample size in online appendix C. There, the estimation error from estimating the unconditional quantile functions does not play a role in the asymptotic distribution of the test statistic and the analytical variance-covariance matrix can be estimated easily.

Note that assumption 2 guarantees that the variance-covariance matrix $\mathbf{V}$ has full rank. In practice with a small sample, it is possible that for some small cells defined by values of $\mathbf{X}$ and Z, both $\hat{m}_j^1(\tau_k)$ and $\hat{m}_j^0(\tau_k)$ are degenerate, and, hence, $\hat{\mathbf{V}}$ does not have full rank. The effective number of moment restrictions in H_0 is then the rank of $\hat{\mathbf{V}}$, which should be used as the degrees of freedom for the test statistic.

B. The Mean Test for Rank Similarity

In this section we construct a mean test for rank similarity. Tests for other moments of potential ranks can be constructed similarly and are omitted to save space.

Let $\bar{m}_j^z = E[U|Z=z,\mathbf{X}=\mathbf{x}_j]$ for $z=0,1$. The null hypothesis of interest is

$$\begin{aligned}&H_{0,mean}\colon \bar{m}_j^0=\bar{m}_j^1, \text{ for all } j=1,\ldots,J-1.\\ &H_{a,mean}\colon H_{0,mean} \text{ is not true.}\end{aligned}$$

Let $\left(\tau^1,\ldots\tau^S\right)$ be S random numbers drawn from the uniform distribution $U(0,1)$ that are independent of the data. One can simulate individual i's rank in the complier potential outcome distribution by $\hat{U}_i=\hat{\mathsf{R}}(Y_i,T_i)$ with $\hat{\mathsf{R}}(y,t)=\frac{1}{S}\sum_{s=1}^S\mathbf{1}\left(\left(t\hat{q}_{1|C}(\tau^s)+(1-t)\hat{q}_{0|C}(\tau^s)\right)\leq y\right)$. Let $\ddot{m}_j^z=\frac{1}{n_j^z}\sum_{Z_i=z,\mathbf{X}_i=\mathbf{x}_j}\hat{U}_i$ be the estimator of $\bar{m}_j^z$, for $z=0,1$. Let $\bar{\mathbf{m}}^z=\left[\bar{m}_1^z\ \bar{m}_2^z\ \cdots\ \bar{m}_{J-1}^z\right]'$, $\ddot{\mathbf{m}}^z=\left[\ddot{m}_1^z\ \ddot{m}_2^z\ \cdots\ \ddot{m}_{J-1}^z\right]'$, and let $\ddot{\mathbf{V}}$ be the bootstrapped covariance matrix of $\sqrt{n}\left(\ddot{\mathbf{m}}^1-\ddot{\mathbf{m}}^0\right)$ under the null. Define the test statistic as

$$W_{\text{mean}}\equiv n\left(\ddot{\mathbf{m}}^1-\ddot{\mathbf{m}}^0\right)'\ddot{\mathbf{V}}^{-1}\left(\ddot{\mathbf{m}}^1-\ddot{\mathbf{m}}^0\right).$$

The following corollary summarizes the asymptotic property of $\ddot{\mathbf{m}}^1-\ddot{\mathbf{m}}^0$.

Corollary 2. *Suppose that assumption 2 holds for* $\Omega=(0,1)$*. Given assumptions 1 and 2, under the null hypothesis that* $\bar{\mathbf{m}}^1=\bar{\mathbf{m}}^0$*, when* $S,n\to\infty$,

$$\sqrt{n}\left(\ddot{\mathbf{m}}^1-\ddot{\mathbf{m}}^0\right)\Rightarrow N(0,\mathbf{V}_{mean}).$$

The (j,j')*th element of* $\mathbf{V}_{mean}$ *is* $E\left[\left(\int_0^1\phi_j^1(\tau)d\tau-\int_0^1\phi_j^0(\tau)d\tau\right)\left(\int_0^1\phi_{j'}^1(\tau)d\tau-\int_0^1\phi_{j'}^0(\tau)d\tau\right)\right]$*, with* ϕ_j^z *defined in theorem 2.*

Corollary 2 suggests that W_{mean} converges in distribution to $\chi^2(J-1)$ under the null and explodes under the alternative. This mean test could be useful when testing a large set of quantiles is not practical due to, for example, a small sample size. However, when the sample size is large, the mean test could be computationally demanding as it requires estimating potential ranks for every individual. Also, the mean test could sometimes have lower power than the distributional test as it tests only one particular feature of potential rank distributions, while rank similarity is naturally a distributional concept. We show this point in our empirical application.

IV. Empirical Application

This section applies our proposed tests to investigate whether job training can change trainees' ranking in the earnings distribution. The impact of job training programs on trainee earnings, especially those with low income, is of great interest to both policymakers and economists. Abadie et al. (2002) and Chernozhukov and Hansen (2008) use the JTPA experimental data to evaluate the impact of the JTPA training program on the distribution of trainee earnings. An interesting feature of the JTPA training experiment is that there are almost always no takers, so the estimated LQTEs can be seen as the QTEs for the treated or the trainees. Both Abadie et al. (2002) and Chernozhukov and Hansen (2008) focus on conditional QTEs.

Here we estimate unconditional QTEs, as we are interested in learning how this program affects the unconditional distribution of earnings. We then apply our proposed tests. If rank invariance is plausible, one may then infer from the estimated distributional effects causal impacts of the JTPA training on earnings for individual trainees at different quantiles of the distribution. For example, we show that the program has small and insignificant impacts at the lower tail of the earnings distribution for men. However, can we conclude that the JTPA training has no real impacts for male trainees at the bottom of the earnings distribution?

We use the same data as those used in Abadie et al. (2002). We also have information on age in years (instead of five age categories) to perform falsification tests for our rank similarity tests. The sample consists of 5,102 observations for men and 6,102 observations for women. The data contain information on earnings (Y), training (T), treatment assignment status (Z), and some predetermined individual characteristics ($\mathbf{X}$).[7] Earnings are measured as total earnings over the thirty-month period following the assignment of individuals into the treatment or the control group. Details on the experimental data collection and sample selection criteria are in Abadie et al. (2002).

Table 1 presents the estimated unconditional QTEs at equally spaced quantiles from 0.15 to 0.85 with an increment of 0.05. Also presented are quantiles of the potential earnings without training. Thus, the ratio of the two numbers in each row gives the percentage change in earnings at each quantile. They would represent real impacts of the training on earnings for individuals at each quantile if rank invariance held. Estimates in table 2 show that the JTPA training program has significant impacts at almost every quantile of the earnings distribution for female trainees. The corresponding percentage changes are larger at lower quantiles due to females' very low potential earnings without training at those quantiles. In sharp contrast, the estimated QTEs are

Table 1.—First-Stage Estimates of Unconditional QTEs of Training on Trainee Earnings

Quantile	Female Y_0	Female QTE		Male Y_0	Male QTE	
0.15	195	291	(341.88)	1,462	249	(713.36)
0.20	723	714	(358.31)*	2,733	390	(723.01)
0.25	1,458	1,200	(372.08)***	4,434	489	(746.85)
0.30	2,463	1,380	(399.21)***	6,993	340	(891.74)
0.35	3,784	1,705	(497.01)***	8,836	594	(1, 042.40)
0.40	5,271	1,974	(669.75)***	11,010	723	(1, 104.63)
0.45	6,726	2,451	(766.25)***	13,104	1,069	(1,144.28)
0.50	8,685	2,436	(829.29)***	15,374	1,291	(1,234.59)
0.55	11,007	2,089	(877.56)**	17,357	2,239	(1,295.79)*
0.60	12,618	2,729	(886.96)***	20,409	2,118	(1,418.40)
0.65	14,682	2,943	(920.45)***	23,342	2,319	(1,557.00)
0.70	16,971	2,772	(1,027.14)***	27,169	1,780	(1,606.66)
0.75	20,252	2,106	(1,152.35)*	30,439	2,408	(1,641.47)
0.80	23,064	2,331	(1,149.71)**	34,620	2,800	(1,701.90)*
0.85	26,735	1,762	(1,179.91)	39,233	3,955	(1,886.98)*

Standard errors are in parentheses. All estimates control for covariates including dummies for blacks, Hispanics, high school graduates (including GED holders), marital status, whether the applicant worked at least twelve weeks in the twelve months preceding random assignment, and AFDC receipt (for women only), as well as five age group dummies. Significant at *10%, **5%, and ***1%.

much smaller and insignificant for male trainees at the low quantiles. At the same time, male trainees have much higher potential earnings without training, leading to even small and statistically insignificant percentage changes at the low quantiles. The estimated QTEs for men are larger in absolute terms above the median, but still small in percentage terms.

Panel A of table 2 reports results from the distributional tests for rank similarity with two different sets of quantiles. In columns I, the proposed tests are conducted jointly at $\Omega = \{0.15, 0.20, \ldots, 0.85\}$, while in columns II, the tests are conducted jointly at $\Omega = \{0.20, 0.3, \ldots, 0.80\}$. For both, we either include covariates in the first-stage unconditional QTE estimation or not. To ensure the nontrivial assignment or common support requirement of assumption 1, we drop $\mathbf{X}$ values with fewer than five observations when either $Z = 1$ or $Z = 0$.

Panel B of table 2 reports results from the same tests except that we replace the dependent variable earnings with age in years. Rank similarity holds trivially in this case, since training is supposed to have no effects on age. Further, individual characteristics are correlated with age, so these additional tests can serve as falsification tests for our main tests.

As shown in table 2, rank invariance or the weaker condition, rank similarity, can be strongly rejected among both female and male trainees. The test results are very similar regardless of whether we control for covariates in the first stage. This is due to the fact that assignment to treatment is well randomized in the JTPA experiment, and so the assignment indicator Z is a valid IV regardless of conditioning on covariates. In sharp contrast, when age is used as the dependent variable, rank invariance cannot be rejected in all cases. Not surprisingly, training does not cause individuals to systematically change their ranks in the age distribution.

[7] The set of individual characteristics includes dummies for black or Hispanic applicants, a dummy for high school graduates or GED holders, a dummy for married applicants, whether the applicant worked at least twelve weeks in the twelve months preceding random assignment, a dummy for AFDC receipt (for women only), and five age category dummies.

TABLE 2.—THE DISTRIBUTIONAL TESTS FOR RANK SIMILARITY

	Female				Male			
	I		II		I		II	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
A: Dependent variable is earnings								
W Stat.	7,652.1	7,763.8	1,197.2	1,177.8	2,780.7	2,719.0	886.1	876.8
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
d.f.	1,544	1,544	723	723	1,218	1,218	570	570
B: Falsification test (dependent variable is age)								
W Stat.	478.8	471.9	252.0	259.9	209.3	203.5	124.7	123.0
	(0.926)	(0.953)	(0.366)	(0.245)	(1.000)	(1.000)	(0.977)	(0.982)
d.f.	525	525	245	245	338	338	158	158

Results are based on the chi-squared test in theorem 2; variance-covariance matrices are bootstrapped with 2,000 replications; *p*-values are in parentheses. Columns I report a joint test at equally spaced 15 quantiles from 0.15 to 0.85. Columns II report a joint test at equally spaced 7 quantiles from 0.20 to 0.80; (1) controls for covariates in the first-stage unconditional QTE estimation, while (2) does not; X values with fewer than five observations when either $Z = 0$ or $Z = 1$ are not used in the test to ensure the common support assumption.

TABLE 3.—INDIVIDUAL QUANTILE AND MEAN TESTS FOR RANK SIMILARITY

	A. Earnings				B. Falsification Test (Age)			
	Female		Male		Female		Male	
	W Stat.		*W* Stat.		*W* Stat.		*W* Stat.	
I: Individual quantile test								
0.15	134.4	(0.012)	103.8	(0.045)	43.9	(0.144)	19.4	(0.561)
0.20	143.0	(0.004)	113.3	(0.010)	37.9	(0.340)	22.1	(0.391)
0.25	126.2	(0.060)	107.8	(0.025)	26.0	(0.863)	13.9	(0.907)
0.30	131.9	(0.034)	104.7	(0.039)	26.9	(0.834)	15.0	(0.861)
0.35	147.2	(0.003)	95.8	(0.142)	22.1	(0.956)	17.9	(0.712)
0.40	118.3	(0.160)	88.6	(0.291)	31.1	(0.659)	23.2	(0.447)
0.45	107.5	(0.387)	110.7	(0.019)	32.1	(0.611)	22.4	(0.497)
0.50	110.9	(0.304)	113.6	(0.012)	32.3	(0.599)	19.2	(0.692)
0.55	112.6	(0.266)	110.9	(0.019)	30.8	(0.673)	19.6	(0.664)
0.60	112.1	(0.276)	112.3	(0.015)	32.7	(0.581)	22.3	(0.503)
0.65	121.7	(0.113)	105.0	(0.044)	29.4	(0.734)	18.4	(0.735)
0.70	108.0	(0.375)	106.1	(0.038)	36.7	(0.388)	24.0	(0.402)
0.75	130.4	(0.035)	109.7	(0.018)	45.4	(0.112)	16.5	(0.831)
0.80	118.4	(0.128)	116.5	(0.005)	47.7	(0.074)	17.1	(0.802)
0.85	92.3	(0.697)	118.7	(0.002)	44.7	(0.125)	18.7	(0.716)
II: Mean test								
Mean	123.1	(0.098)	115.2	(0.009)	30.6	(0.683)	18.4	(0.736)

Results are based on the chi-squared test in theorem 2; variance-covariance matrices are bootstrapped with 2,000 replications; *p*-values are in parentheses; covariates are controlled for in the first-stage unconditional QTE estimation. X values with fewer than five observations when either $Z = 1$ or $Z = 0$ are not used in the test to ensure the common support assumption.

To investigate how seriously rank similarity is violated at different parts of the potential earnings distribution, we test rank similarity at each individual quantile from 0.15 to 0.85. Panel A in table 3 presents results from these individual quantile tests. For men, rank similarity can be rejected at almost all quantiles except for the 0.35 and 0.40 quantiles. For women, rank similarity can be rejected strongly at the lower tail of the distribution, but not so near the median or above. Panel B in table 3 presents results from the corresponding falsification tests. Again in sharp contrast, at the 5% significance level, we fail to reject rank invariance or similarity at all quantiles for both women and men when age is the dependent variable.[8]

The bottom part of table 3 reports results from the mean tests for rank similarity. Again we can strongly reject rank similarity among male trainees. The evidence is somewhat weaker for the female sample. The test result is only marginally significant at the 10% level. This is not surprising, since rank similarity is violated largely only at the lower tail of the earnings distribution for women.

The test results show that rank similarity is seriously violated among male trainees. Recall that the estimated distributional effects for men are mostly small and insignificant. These test results are interesting, since they suggest that training causes men to systemically change their ranks in the earnings distribution and that the program effects for male trainees are more complicated than those suggested by the estimated QTEs. For female trainees, the training program seems to mainly cause rank changes at the lower tail of the earnings distribution.

In conclusion, for the JTPA training program, the estimated QTEs can at best capture how the distribution of earnings changes with the training. One should be cautious in equating the distributional effects of the JTPA training

[8] The only time the test is rejected only at the 10% significance level is when looking at the 0.80 quantile for women. There is no systematic evidence of violation of rank similarity otherwise.

with the true impacts on earnings for individual trainees. For example, although the training does not raise the lower tail of the earnings distribution for men, it does not mean that the JTPA training has no real impacts for male trainees at the bottom of the earnings distribution.

Finally, it is worth mentioning that our results are largely consistent with the findings by Heckman, Smith, and Clements (1997). Based on large-scale permutation exercises, they show that "heterogeneity is an important feature of impact distributions" and that "perfect positive dependence across potential outcome distributions produces estimates of impact distributions that are not credible."

V. Conclusion

This paper proposes tests for rank invariance or rank similarity. We nonparametrically identify and test the counterfactual distribution of potential ranks or features of the distribution (such as mean, median, or any particular quantile) among observationally equivalent individuals. These tests can be useful in examining whether some subgroup of particular interest changes its ranks in the outcome distribution under treatment. By testing any particular quantile, the proposed tests are informative regarding at which part of the potential outcome distribution rank similarity is violated.

The proposed tests allow treatment to be endogenous, with exogenous treatment following readily as a special case. Other than mild regularity conditions, the proposed tests do not require any additional assumptions other than those used to identify and estimate the first-stage unconditional QTEs.

The usefulness of our proposed tests is illustrated in evaluating the JTPA training program. We show that while male trainees change their ranks throughout the earnings distribution, female trainees change ranks only at the lower tail of the distribution. Overall evidence suggests that the estimated QTEs capture the impacts of the JTPA training on the distribution of trainee earnings instead of those on individual trainee earnings.

Note that we focus on testing individual ranks in the unconditional distributions of potential outcomes. If desired, the proposed tests can be readily extended to testing for invariance or similarity of ranks in the conditional distribution of potential outcomes. Such tests require additional covariates other than those used in the conditioning set of conditional quantile or rank estimation. In particular, one can first estimate quantiles of the conditional distribution of potential outcomes, conditional on covariates of interest $\mathbf{X}_1$, and then use additional covariates $\mathbf{X}_2$ along with $\mathbf{X}_1$ to perform the tests. For example, in our empirical application, we estimate quantiles of potential earnings and perform the tests separately for male and female trainees. These tests are essentially rank similarity tests for conditional potential ranks conditional on gender.

REFERENCES

Abadie, A., "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics* 113 (2003), 231–263.

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

Angrist, J., G. Imbens, and D. Rubin, "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association* 91 (1996), 444–455.

Bitler, M., J. Gelbach, and H. Hoynes, "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," *American Economic Review* 96 (2006), 988–1012.

——— "Distributional Impacts of the Self-Sufficiency Project," *Journal of Public Economics* 92 (2008), 748–765.

Chernozhukov, V., I. Fernández-Val, and B. Melly, "Inference on Counterfactual Distributions," *Econometrica* 81 (2013), 2205–2269.

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

——— "Instrumental Quantile Regression Inference for Structural and Treatment Effect Models," *Journal of Econometrics* 132 (2006), 491–525.

——— "Instrumental Variable Quantile Regression: A Robust Inference Approach," *Journal of Econometrics* 142 (2008), 379–398.

Chernozhukov, V., G. Imbens, and W. Newey, "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics* 139 (2007), 4–14.

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

Donald, S., and Y. Hsu, "Estimation and Inference for Distribution Functions and Quantile Functions in Treatment Effect Models," *Journal of Econometrics* 178 (2014), 383–397.

Firpo, S., "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica* 75 (2007), 259–276.

Firpo, S., and C. Pinto, "Identification and Estimation of Distributional Impacts of Interventions Using Changes in Inequality Measures," *Journal of Applied Econometrics* 31 (2015), 457–486.

Frandsen, B. R. and L. J. Lefgren, "A Test for Rank Similarity and Partial Identification of the Distribution of Treatment Effects," Brigham Young University working paper (2015).

Frolich, M., and B. Melly, "Unconditional Quantile Treatment Effects under Endogeneity," University of Mannheim working paper (2007).

——— "Unconditional Quantile Treatment Effects under Endogeneity," *Journal of Business and Economic Statistics* 31 (2013), 346–357.

Heckman, J., J. Smith, and N. Clements, "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies* 64 (1997), 487–535.

Horowitz, J., and S. Lee, "Nonparametric Instrumental Variables Estimation of a Quantile Regression Model," *Econometrica* 75 (2007), 1191–1208.

Imbens, G., and W. Newey, "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," *Econometrica* 77 (2009), 1481–1512.

Lehmann, E., *Nonparametrics: Statistical Methods Based on Ranks* (San Francisco: Holden-Day, 1974).