Testing for Rank Invariance or Similarity in Program Evaluation

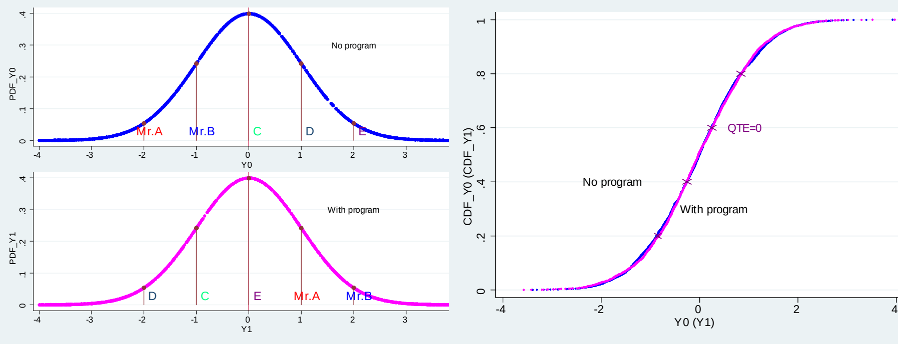
Yingying Dong and Shu Shen

UC Irvine & UC Davis

Motivation - Program Evaluation

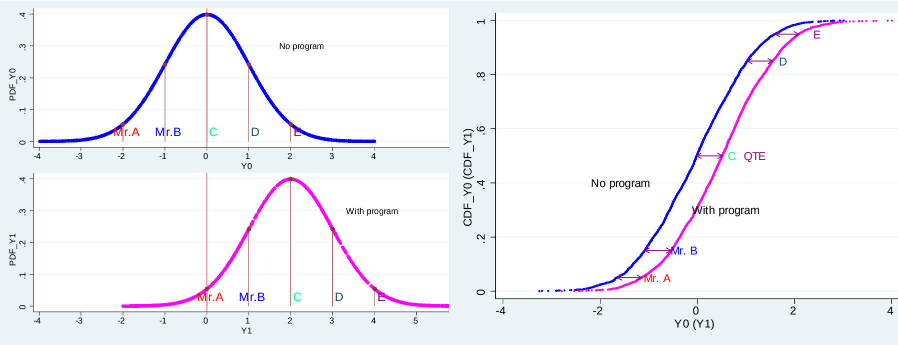
- Distributional effects have been increasingly popular among researchers in program evaluation.
  - E.g., Heckman, Smith, and Clements (1997), Bitler, Gelbach and Hoynes (2006, 2008), Dammert (2008), Djebbari and Smith (2008), Eren and Ozbeklik (2014), and Bitler and Hoynes (2014).
- *Distributional effects* (how a program changes the outcome distribution) $\neq$ *Individual causal effects* (the distribution of program impacts).
- *Rank preservation* is required to interpret the *distributional effects* of a program as *individual causal effects*.

Outcome distributions with or without a program:



- Distributional effects (represented by quantile differences, or QTEs) are zero.
- True causal effects on individuals are not zero.

Outcome distributions with or without a program:



- When invidiual ranks are preserved, quantile differences of the marginal distributions give the distribution of program effects.

Example: the JTPA Training Program

No effects on ***male trainees****?*

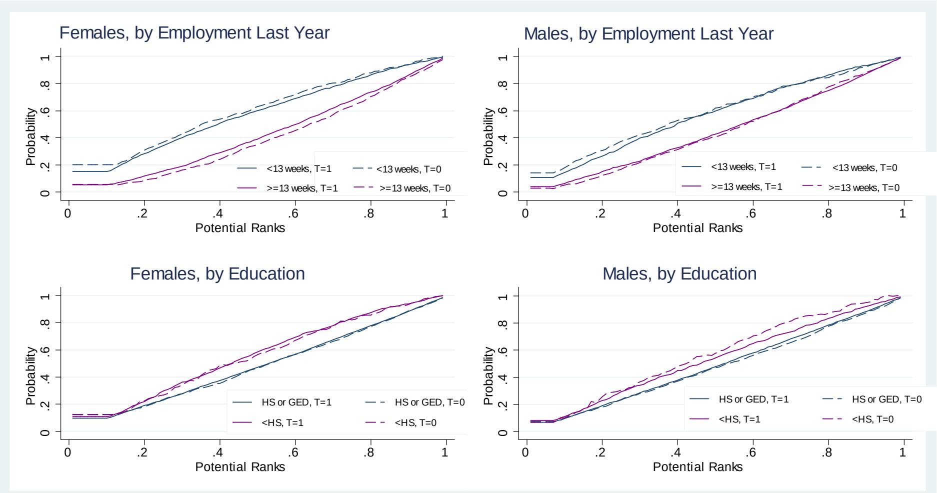
Table: Estimated Distributional Effects of JTPA Training Program on Earnings

	Female			Male		
Quantile	Y_0	QTE		Y_0	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)**

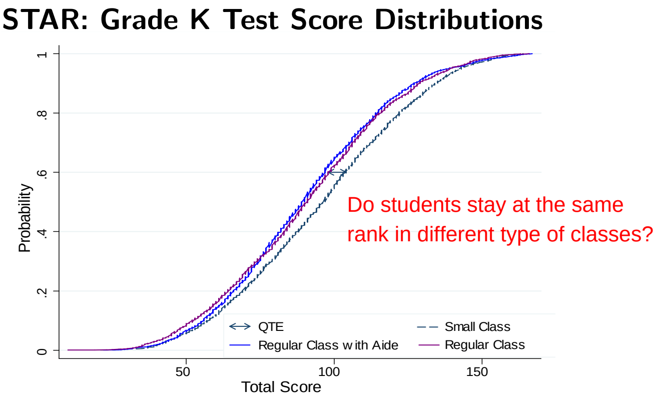
Note: Standard errors are in the parentheses; All estimates control for covariates including dum-
mies for black, Hispanic, high-school graduates (including GED holders), marital status, whether

JTPA: Distribution of Potential Ranks for Sub-groups

Trainees' ranks are not preserved (some are relatively better off, while others worse off)



Example: Project STAR



- Attending a small class greatly improves the score distribution.
- Attending a regular class with a teacher aide have no much impact.

Motivation - QTE literature

The popular QTE literature:

Rank invariance or rank similarity is required

- for identification:
  - e.g., IVQR model of Chernozhukov and Hansen (2005, 06, 08), Chesher (2003), Chernozhukov, Imbens, and Newey (2007), Horowitz and Lee (2007) etc.
- or for interpretation:
  - e.g., LQTE framework (Abadie, Angrist and Imbens, 2002), Firpo (2007), Imbens and Newey (2009), and Frolich and Melly (2013) etc.

Contributions

This paper

- discusses testable implications of rank invariance and rank similarity.
- provides nonparametric identification of the entire counterfactual distribution of potential ranks (or features of the distribution) among observationally equivalent individuals.
- proposes powerful nonparametric tests applicable to both assumptions.
- Shows usefulness in empirical settings: JTPA training, Project STAR.

Features of the Tests

Nice features of the proposed tests:

- allow treatment to be endogenous (exogenous treatment is a special case).
- can handle IVs that are valid conditional on covariates or valid unconditionally.
- essentially do not require any additional assumptions, other than those used to identify and estimate (L)QTEs.

Basic Idea of the Proposed Tests

- Let Y_t for $t = 0, 1$ be potential outcomes under no treatment or treatment.
- Let F_t, for $t = 0, 1$ be the distributions of counterfactuals Y_t.
- Define $U_t \equiv F_t\left(Y_t\right)$ for $t = 0, 1$.
- By construction, $U_t \sim U(0, 1)$ for $t = 0, 1$.
- Rank invariance holds if U_0 and U_1 are the same random variable.
- We can't observe if this is true, because we can't observe U_0 and U_1 for the same person - can only estimate U_0 for untreated people and U_1 for treated people.

Basic Idea of the Proposed Tests

So how can we test?

- Let $\mathbf{X}$ be covariates. If U_0 and U_1 are the same random variables, then the conditional distribution of $U_0|\mathbf{X}$ must be the same as the conditional distribution of $U_1|\mathbf{X}$.
- That these conditional distributions of $U_0|\mathbf{X}$ and $U_1|\mathbf{X}$ are the same functions is a testable implication of rank invariance. It also holds under the weaker condition of rank similarity.

Rank Invariance

$U_t \equiv F_t(Y_t) \sim U(0,1)$ is the rank in the *unconditional* distribution of potential outcomes Y_t, $t = 0, 1$.

- Look at *unconditional potential ranks* first: $Y_t = q(t, U_t)$ for $U_t \sim U(0,1)$.
- Then extend the idea to testing conditional potential ranks.

Definition

Rank invariance is the condition $U_0 = U_1$.

Rank Similarity

- Rank invariance is restrictive – does not allow for random slippages in potential ranks (e.g., caused by luck).
- E.g., $Y_t = g(t, \mathbf{X}, V, S_t)$, $t = 0, 1$, where Y_t = test score, $\mathbf{X}$ (gender, age etc) and V (ability) determine the common rank level, S_t (luck) is a random shock responsible for the random slippages.
  - Assume $Y_t = g(\mathbf{X}, V, S_t)$. $U_0 \neq U_1$.
  - If S_t, $t = 0, 1$ are mutually i.i.d,.then given X and V, $U_0 \sim U_1$.

Definition

Rank similarity is the condition $U_0 \mid \mathbf{X} = \mathbf{x}, V = v \sim U_1 \mid \mathbf{X} = \mathbf{x}, V = v$ for any $(\mathbf{x}, v) \in \mathcal{W}$, where $\mathbf{X}$ and V are observable and unobservable determinants or 'shifters' of the common rank level.

Testable Implications of Rank Similarity

Invariance: U_0, U_1 are the same random variable; **Similarity**: Conditional on $\mathbf{X} = \mathbf{x}$, $V = v$, U_0, U_1 have the same distribution.

Focus on rank similarity (rank invariance is a special case)

Lemma (1.2)

(Main Testable Implication) Given rank similarity,
$F_{U_0|\mathbf{X}}(\tau|\mathbf{x}) = F_{U_1|\mathbf{X}}(\tau|\mathbf{x})$ *for all* $\tau \in (0,1)$ *and* $\mathbf{x} \in \mathcal{X}$.

- The distribution of ranks is the same under treatment or no treatment among observationally equivalent individuals.

Identification: Endogenous Treatment

- $T = 0, 1$ is the treatment
  - e.g., $T =$ training in the JTPA program.
- $Z = 0, 1$ is an IV
  - e.g., $Z =$ random assignment to training.
- Let C be the set of compliers (those who comply with the random assignment, i.e., $T = Z$).
- Interested in testing rank similarity among compliers.
- Let $q_{t|C}(\tau) \equiv F_{t|C}^{-1}(\tau)$ and $QTE_C(\tau) \equiv q_{1|C}(\tau) - q_{0|C}(\tau)$ (Doksum 1974, Lehmann 1974).

Identification: Endogenous Treatment

Identifying assumption:

Assumption 1 Let (Y_t, T_z, X, Z), $t, z = 0, 1$ be random variables mapped from the common probability space $(\Omega, \mathrm{F}, \mathrm{P})$. The following conditions hold jointly with probability one.

1. Independence: $(Y_0, Y_1, T_0, T_1) \perp Z|\mathbf{X}$.
2. First stage: $E(T_1) \neq E(T_0)$.
3. Monotonicity: $\Pr(T_1 \geq T_0) = 1$.
4. Nontrivial assignment: $0 < \Pr(Z = 1|\mathbf{X} = \mathbf{x}) < 1$ for all $\mathbf{x} \in \mathcal{X}$.

- Standard LQTE identifying assumption (Abadie, Angrist and Imbens, 2002, Abadie 2003 etc.), except for **a weaker first-stage** here.

Key Identification Results

Theorem (1)

Let $I(\tau) \equiv 1\left(Y \leq \left(Tq_{1|C}(\tau) + (1-T)\, q_{0|C}(\tau)\right)\right)$. *Given Assumption 1, for all* $\tau \in (0,1)$, $\mathbf{x} \in \mathcal{X}_C$, *and* $t = 0, 1$, $F_{U_t|C,\mathbf{x}}(\tau|\mathbf{x})$ *is identified by*

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

$F_{U_1|C,\mathbf{x}}(.|\mathbf{x}) = F_{U_0|C,\mathbf{x}}(.|\mathbf{x})$ *for* $\mathbf{x} \in \mathcal{X}_C$ *if and only if for all* $\tau \in (0,1)$ *and* $\mathbf{x} \in \mathcal{X}$

$$E\left[I(\tau)|Z=1,\mathbf{X}=\mathbf{x}\right] = E\left[I(\tau)|Z=0,\mathbf{X}=\mathbf{x}\right]. \tag{2}$$

- Notice the change from $\mathbf{x} \in \mathcal{X}_C$ to $\mathbf{x} \in \mathcal{X}$.
- eq.(2) is a reduced-form (conditioning on Z instead of T).
- Exogenous T is a special case ($T = Z$; everyone is a complier).

Test the Distribution of Ranks

Eq. (1) can be used to *estimate the counterfactual distribution* of potential ranks for subgroup with $\mathbf{X} = \mathbf{x}$.

Eq. (2) can be used to *test* H_0: $F_{U_0|C,\mathbf{X}}(.|\mathbf{x}) = F_{U_1|C,\mathbf{X}}(.|\mathbf{x})$.

1. Estimate $q_{0|C}(\tau)$ and $q_{1|C}(\tau)$.
2. Test $E[I(\tau)|Z = 1, \mathbf{X} = \mathbf{x}] = E[I(\tau)|Z = 0, \mathbf{X} = \mathbf{x}]$ for $\tau \in (0, 1)$ *and any* $\mathbf{x} \in \mathcal{X}$, replacing $q_{0|C}(\tau)$ and $q_{1|C}(\tau)$ with their estimates.

Can also test *a particular quantile* (median) or *a subset of quantiles of interest*.

Test Moments of Ranks

Can test features of the potential rank distribution:

- E.g., test moments of ranks

  $F_{U_1|C,\mathbf{X}}(\tau|\mathbf{x}) = F_{U_0|C,\mathbf{X}}(\tau|\mathbf{x})$ implies $E[U_1^p|C, \mathbf{X} = \mathbf{x}] = E[U_0^p|C, \mathbf{X} = \mathbf{x}]$

  for some $p > 0$. When $p = 1$, **a mean test** for rank similarity.

- Analogous to Theorem 1, given rank similarity,

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

  where $U \equiv TU_1 + (1 - T)U_0 =$
  $\int_0^1 1\left(Tq_{1|C}(\tau) + (1 - T)q_{0|C}(\tau) \leq Y\right) d\tau = 1 - \int_0^1 I(\tau)\, d\tau$.
- Individual rank U is identified since $I(\tau)$ is identified.

The Distributional Test: Null Hypothesis

Let $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_J\}$.

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

where $m_j^z(\tau_k) \equiv E\left[1(\tau_k) | Z = z, \mathbf{X} = \mathbf{x}_j\right]$, $z = 0, 1$;
Linear restriction: $\sum_{j=1}^J m_j^0(\tau_k) \Pr(\mathbf{X} = \mathbf{x}_j) = \sum_{j=1}^J m_j^1(\tau_k) \Pr(\mathbf{X} = \mathbf{x}_j)$.

- Estimate $q_{0|C}(\tau_k)$ and $q_{1|C}(\tau_k)$ following Frolich and Melly (2013).

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

where $\rho_{\tau_k}(u) \equiv u(\tau_k - 1(u \leq 0))$,
$\hat{\omega}_i \equiv \left(\frac{Z_i}{\widehat{\pi}(\mathbf{X}_i)} - \frac{1-Z_i}{1-\widehat{\pi}(\mathbf{X}_i)}\right)(2T_i - 1)$ and $\widehat{\pi}(\mathbf{x}) = \hat{P}r(Z = 1 | \mathbf{X} = \mathbf{x})$.

- Then estimate $m_j^z(\tau_k)$ by

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

Assumption for Asymptotic Properties

Assumption 2

1. i.i.d. data: the data $(Y_i, T_i, Z_i, \mathbf{X}_i)$ for $i = 1, ..., n$ is a random sample of size n from $(Y, T, Z, \mathbf{X})$.
2. For all $\tau \in \Omega = \{\tau_1, \tau_2, ..., \tau_K\}$, the random variable Y_1 and Y_0 are continuously distributed with positive density in a neighborhood of $q_{0|C}(\tau)$ and $q_{1|C}(\tau)$ in the subpopulation of compliers.
3. For all $j = 1, ..., J$, $\hat{\pi}(\mathbf{x}_j)$ is consistent, or $\hat{\pi}(\mathbf{x}_j) \xrightarrow{p} \pi(\mathbf{x}_j)$.
4. Let $f_{Y|T,Z,\mathbf{X}}$ be the conditional density of Y given T, Z and $\mathbf{X}$. For all $t, z = 0, 1$, $j = 1, ..., J$ and $\tau \in \Omega$, $f_{Y|T,Z,\mathbf{X}}(y|t, z, \mathbf{x}_j)$ has a bounded first derivative with respect to y in a neighborhood of $q_{t|C}(\tau)$. Let $f_{Y|\mathbf{X}}(y|\mathbf{x})$ be the conditional density of Y given $\mathbf{X}$. For all $\tau \in \Omega$ and $j = 1, ..., J$, $f_{Y|\mathbf{X}}(.|\mathbf{x}_j)$ is positive and bounded in a neighborhood of $q_{t|C}(\tau)$.

Let $\hat{\mathbf{m}}^0$ and $\hat{\mathbf{m}}^1$ be $K(J-1)\times 1$ vectors of estimated $\hat{m}_j^0(\tau_k)$ and $\hat{m}_j^1(\tau_k)$ for $k=1,\ldots K$ and $j=1,\ldots,J-1$.

Theorem (2)

Given Assumptions 1 and 2,

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

where $\mathbf{V}$ *is* $K(J-1)\times K(J-1)$ *asymptotic variance covariance matrix; the* $\left(\sum_{j=1}^{J-1}K(j-1)+k,\sum_{j'=1}^{J-1}K(j'-1)+k'\right)$*-th element is* $E\left[\left(\phi_j^1(\tau_k)-\phi_j^0(\tau_k)\right)\left(\phi_{j'}^1(\tau_{k'})-\phi_{j'}^0(\tau_{k'})\right)\right]$, *where*

$$\phi_j^z(\tau_k)=\frac{I(\tau_k)-m_j^z(\tau_k)}{p_{Z,X}(z,\mathbf{x}_j)}1(Z=z,\mathbf{X}=\mathbf{x}_j)-\frac{f_{Y,T|Z,\mathbf{X}}(q_{0|C}(\tau_k),0|z,\mathbf{x}_j)}{P_c f_{0|C}(q_{0|C}(\tau_k))}\psi_0(Y,T,Z,\mathbf{X})$$
$$-\frac{f_{Y,T|Z,\mathbf{X}}(q_{1|C}(\tau_k),1|z,\mathbf{x}_j)}{P_c f_{1|C}(q_{1|C}(\tau_k))}\psi_1(Y,T,Z,\mathbf{X}).$$

Test Statistic and Asymptotic Properties

Propose a Wald-type test statistic

$$W \equiv n\left(\hat{\mathbf{m}}^1 - \hat{\mathbf{m}}^0\right)' \hat{\mathbf{V}}^{-1}\left(\hat{\mathbf{m}}^1 - \hat{\mathbf{m}}^0\right) \sim \chi^2(K(J-1)).$$

Bootstrap $\hat{\mathbf{V}}$ in practice.

The critical value c_α is the $(1-\alpha)\times 100$-th percentile of the $\chi^2(K(J-1))$ distribution.

Corollary (1)

Given Assumptions 2 and 3, and the decision rule "reject the null hypothesis H_0 if $W > c_\alpha$," we have

1. *If H_0 is true, $\lim_{n\to\infty} P(\text{reject the null}) = \alpha$.*
2. *If H_0 is false, $\lim_{n\to\infty} P(\text{reject the null}) = 1$.*

The Mean Test for Rank Similarity

Define $\bar{m}_j^z \equiv E[U|Z = z, \mathbf{X} = \mathbf{x}_j]$ for $z = 0, 1$. $\overline{\mathbf{m}}^z \equiv \left(\bar{m}_1^z, ..., \bar{m}_{J-1}^z\right)'$ is then the $(J-1) \times 1$ vector.

$$H_{0,mean} : \overline{\mathbf{m}}_j^0 = \overline{\mathbf{m}}_j^1, \text{for all } j = 1, ..., J-1.$$

- First estimate individual rank $\hat{U}_i \equiv T\hat{U}_{1i} + (1-T)\hat{U}_{0i}$ by simulation.

  Let $\{\tau^s\}_{s=1}^S$ be S random draws from *uniform* (0,1) distribution. $\hat{U}_i$ for $i = 1, ..., n$, can be estimated by

$$\hat{U}_i = \frac{1}{S}\sum_{s=1}^{S} 1\left(\left(T\hat{q}_{1|C}\left(\tau^s\right) + (1-T)\hat{q}_{0|C}\left(\tau^s\right)\right) \leq Y_i\right),$$

- Then estimate $\bar{m}_j^z$ for $j = 1, ..., J$ by

$$\ddot{m}_j^z = \frac{1}{n_j^z}\sum_{Z_i = z, \mathbf{X}_i = \mathbf{x}_j} \hat{U}_i.$$

The Mean Test for Rank Similarity

$\ddot{\mathbf{m}}^z \equiv (\ddot{m}_1^z, ..., \ddot{m}_{J-1}^z)'$ is $(J-1) \times 1$ vector of estimated conditional mean ranks.

Corollary (2)

Suppose that Assumptions 1 and 2 hold for $\Omega = (0,1)$. Under the null hypothesis, when $S, n \to \infty$

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

where $\mathbf{V}_{mean}$ is the$(J-1) \times (J-1)$ asymptotic variance-covariance matrix. The (j,j')-th element of $\mathbf{V}_{mean}$ is $E\left[\left(\int_0^1 \phi_j^1(\tau)d\tau - \int_0^1 \phi_j^0(\tau)d\tau\right)\left(\int_0^1 \phi_{j'}^1(\tau)d\tau - \int_0^1 \phi_{j'}^0(\tau)d\tau\right)\right]$.

Can again construct a Wald-type test statistic

$$W_{mean} \equiv n\left(\ddot{\mathbf{m}}^1 - \ddot{\mathbf{m}}^0\right)' \ddot{\mathbf{V}}^{-1}\left(\ddot{\mathbf{m}}^1 - \ddot{\mathbf{m}}^0\right) \sim \chi^2(J-1),$$

where $\ddot{\mathbf{V}}$ is a consistent estimator of $\mathbf{V}_{mean}$. Bootstrap $\ddot{\mathbf{V}}$ in practice.

Simulation

DGPs:

$$Y_0 = X + V + S_0,$$
$$Y_1 = X + V + (1 - bXV) + S_1 \text{ for } b = 0, 2, 3$$
$$Y = Y_1 T + Y_0 (1 - T),$$
$$\Pr(X = 0.4j) = 1/5 \text{ for } j = 1, ..., 5,$$
$$V,\ S_0,\ S_1 \sim N(0, 1).$$

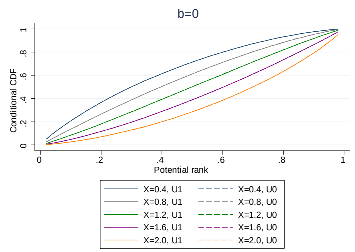
- Rank similarity holds when $b = 0$ but not when $b = 2, 3$. Greater b leads to greater violation.
- Exogenous treatment:

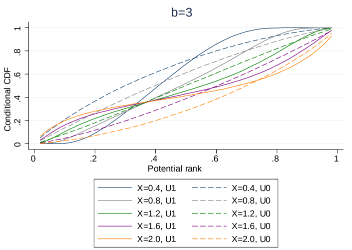
  $$\Pr(T = t) = 0.5 \text{ for } t = 0, 1.$$

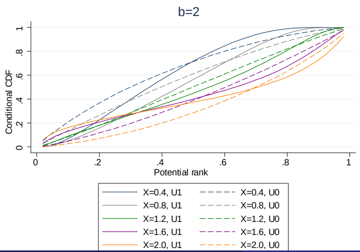
- Endogenous treatment:

  $$T = 1\left(0.15(Y_1 - Y_0) + Z - 0.5 > 0\right),$$
  $$\Pr(Z = z) = 0.5 \text{ for } z = 0, 1.$$

Illustration of DGPs: Exogenous treatment





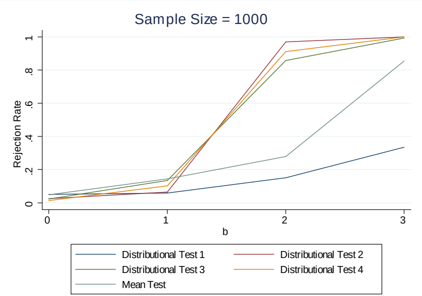


Size and Power Properties: Exogenous Treatment

N	500	1000	1500	2000	2500
			$b = 0$		
Test 1: $\Omega = \{0.5\}$	0.034	0.039	0.051	0.040	0.053
Test 2: $\Omega = \{0.2, 0.3, 0.4\}$	0.013	0.013	0.025	0.021	0.023
Test 3: $\Omega = \{0.5, 0.6, 0.7, 0.8\}$	0.014	0.014	0.023	0.023	0.018
Test 4: $\Omega = \{0.2, 0.3, ..., 0.8\}$	0.006	0.010	0.013	0.013	0.013
Test 5: Mean Test	0.051	0.044	0.048	0.041	0.067
			$b = 2$		
Test 1: $\Omega = \{0.5\}$	0.074	0.150	0.232	0.303	0.388
Test 2: $\Omega = \{0.2, 0.3, 0.4\}$	0.269	0.776	0.968	0.994	1.000
Test 3: $\Omega = \{0.5, 0.6, 0.7, 0.8\}$	0.151	0.581	0.857	0.962	0.991
Test 4: $\Omega = \{0.2, 0.3, ..., 0.8\}$	0.287	0.910	0.996	1.000	1.000
Test 5: Mean Test	0.103	0.213	0.278	0.424	0.500
			$b = 3$		
Test 1: $\Omega = \{0.5\}$	0.143	0.335	0.512	0.640	0.800
Test 2: $\Omega = \{0.2, 0.3, 0.4\}$	0.817	0.999	1.000	1.000	1.000
Test 3: $\Omega = \{0.5, 0.6, 0.7, 0.8\}$	0.306	0.880	0.992	1.000	1.000
Test 4: $\Omega = \{0.2, 0.3, ..., 0.8\}$	0.836	0.999	1.000	1.000	1.000
Test 5: Mean Test	0.340	0.659	0.853	0.941	0.971

Table: Size and power property of the proposed tests: exogenous treatment

Small Sample Performance: Exogenous Treatment



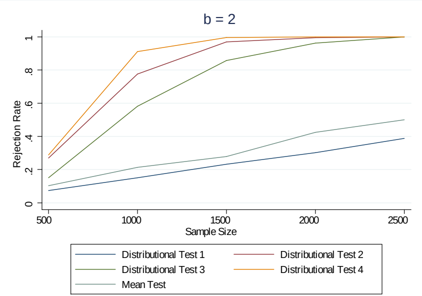
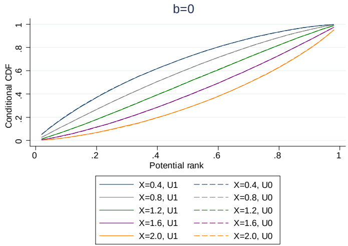
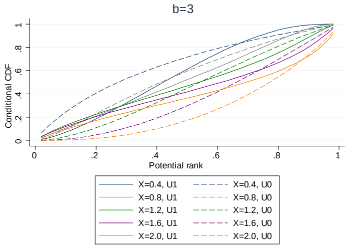
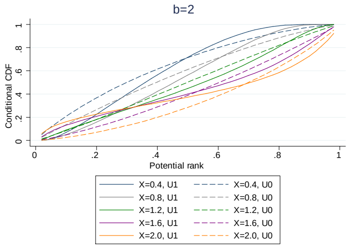


Illustration of DGPs: Endogenous Treatment





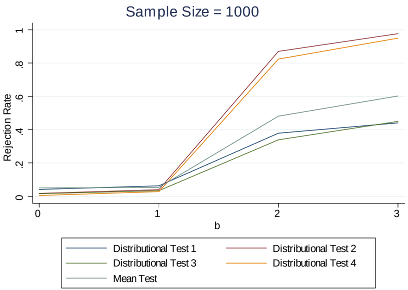


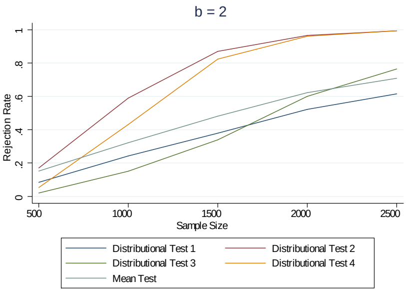
Size and Power Properties: Endogenous Treatment

N	500	1000	1500	2000	2500
			$b = 0$		
Test 1: $\Omega = \{0.5\}$	0.025	0.036	0.041	0.038	0.057
Test 2: $\Omega = \{0.2, 0.3, 0.4\}$	0.012	0.012	0.018	0.017	0.025
Test 3: $\Omega = \{0.5, 0.6, 0.7, 0.8\}$	0.006	0.013	0.016	0.022	0.015
Test 4: $\Omega = \{0.2, 0.3, ..., 0.8\}$	0.002	0.010	0.006	0.010	0.008
Test 5: Mean Test	0.054	0.050	0.051	0.045	0.057
			$b = 2$		
Test 1: $\Omega = \{0.5\}$	0.084	0.242	0.379	0.522	0.615
Test 2: $\Omega = \{0.2, 0.3, 0.4\}$	0.170	0.589	0.870	0.965	0.993
Test 3: $\Omega = \{0.5, 0.6, 0.7, 0.8\}$	0.021	0.150	0.340	0.600	0.764
Test 4: $\Omega = \{0.2, 0.3, ..., 0.8\}$	0.053	0.431	0.823	0.960	0.993
Test 5: Mean Test	0.152	0.322	0.481	0.622	0.709
			$b = 3$		
Test 1: $\Omega = \{0.5\}$	0.113	0.293	0.441	0.617	0.700
Test 2: $\Omega = \{0.2, 0.3, 0.4\}$	0.284	0.783	0.975	1.000	1.000
Test 3: $\Omega = \{0.5, 0.6, 0.7, 0.8\}$	0.020	0.198	0.450	0.704	0.865
Test 4: $\Omega = \{0.2, 0.3, ..., 0.8\}$	0.093	0.634	0.949	1.000	1.000
Test 5: Mean Test	0.191	0.441	0.602	0.772	0.843

Table: Size and power property of the proposed tests: endogenous treatment

Small Sample Performance: Endogenous Treatment





Empirical Application: JTPA

Investigate rank preservation in the Job Training Partner Act (JTPA) training

- A large publicly funded program (1 million participants a year, an annual cost of about 1.6 billion dollars in early 1990's).

Experimental data from the National JTPA Study (Abadie, Angrist and Imbens, 2002 and others):

- Sample size: 5,102 for males, 6,102 for females.
- Y = 30 months' earnings following assignment,
- T = receiving training services,
- Z = random assignment indicator,
- $\mathbf{X}$ = black, Hispanic, HS or GED, married, worked at least 12 weeks the year before, AFDC receipt (for women only) and 5 age category dummies.

JTPA: The Distributional Test

Table: The distributional test for rank similarity

	Female				Male			
	I		II		I		II	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
	Panel A: Dependent Var. Earnings							
χ^2	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
	Panel B: Falsification test (Dependent Var. Age)							
χ^2	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

Note: Results are based on the Chi-squared test in Theorem 2; Variance-covariance matrices are bootstrapped with 2,000 replications; P-values are in the parentheses; Columns I report a joint test at equally-spaced 15 quantiles from 0.15 to 0.85; Columns II reports 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 5 observations when either $Z = 0$ or $Z = 1$ are not used in the test to ensure the common support assumption.

JTPA: The Individual Quantile Test

Table: The individual quantile test for rank similarity

	Panel A: Dependent Var. Earnings				Panel B: Falsification test (Dependent Va			
	Female		Male		Female		Male	
Quantile	χ^2		χ^2		χ^2		χ^2	
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)

Note: Results are based on the Chi-squared test in Theorem 2; Variance-covariance matrices are bootstrapped with 2,000 replications; P-values are in the parentheses; Covariates are controlled

JTPA: the Mean Test

Table: The mean test for rank similarity

	Female				Male			
	(1)		(2)		(1)		(2)	
	Panel A: Dependent Var. Earnings							
χ^2	123.1	(0.098)	123.1	(0.098)	115.2	(0.009)	115.2	(0.009)
d.f.	104		104		82		82	
	Panel B: Falsification test (Dependent Var. Age)							
χ^2	30.6	(0.683)	30.6	(0.683)	18.4	(0.736)	18.4	(0.736)
d.f.	35		35		23		23	

Note: Results are based on the Chi-squared test for the mean ranks only; Variance-covariance matrices are bootstrapped with 2,000 replications; P-values are in the parentheses; (1) controls for covariates in the first-stage unconditional QTE estimation, while (2) does not; X values with fewer than 5 observations when either $Z = 1$ or $Z = 0$ are not used in the test to ensure the common support assumption.

JTPA: Conclusions

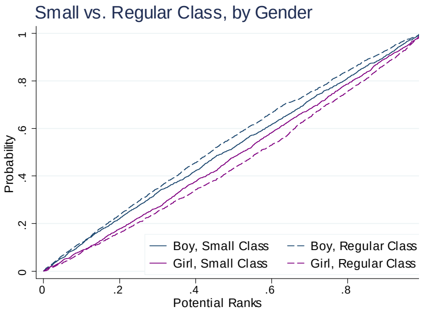
- Training causes some individuals to systemically change their ranks in the earnings distribution.
  - For female trainees, at the lower tail of the distribution
  - For male trainees, throughout the distribution
- Program impacts are more complicated than what would be suggested by the standard QTEs.
- Cannot equate the impacts of the JTPA training on the distribution of earnings with the true effects on individual trainees.

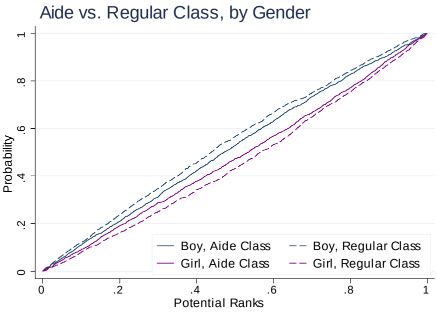
Empirical Application: Project STAR

Analyze students' rank distributions in Project STAR (Student-Teacher Achievement Ratio)

- A large-scale randomized education experiment (11,600 students in 79 schools) in Tennessee in the mid-1980's.
- Students are randomly assigned to 3 types of classes: small class (13-17 students), regular-size class (22-25 students), regular-size class with a full-time teacher aide from grade K to 3.
- Sample size: 5,692 grade K students.
- Y = grade K total test scores,
- $T = Z$ = assigned to a small class or an aide class,
- $\mathbf{X}$ = gender, teacher experience (0-5, 6-10, $>$10 years).

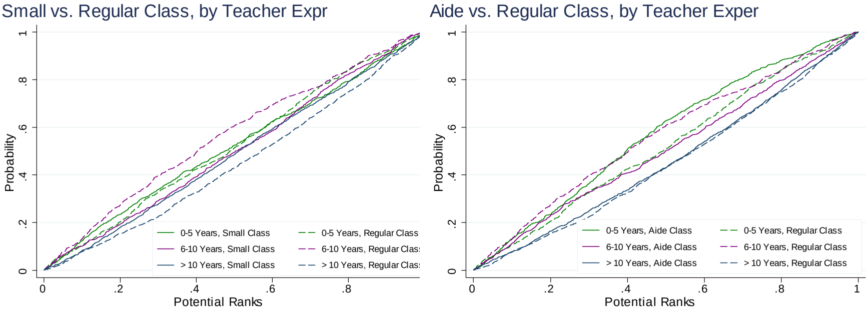
STAR: Gender and Distributions of Potential Ranks





- Attending a small class narrows the gender gap in ranks, i.e., the CDF curves move closer to the (invisible) 45 degree line.
  - Seems more beneficial for relatively better performing boys.
- Having a teacher aide also narrows the gender gap.
  - Seems more helpful for those relatively low performing boys.

STAR: Teacher Expr. and Distributions of Potential Ranks



- Small classes greatly narrow the achievement gaps among students taught by teachers with different levels of experience.
- Assigning a teacher aide to an inexperienced teacher is relatively inefficient and negatively affects students' ranks.

STAR: The Distributional and Mean Tests

Table: The distributional and mean tests for rank similarity - STAR

	Small v.s. Regular		Aide v.s. Regular		Small v.s. Regular and Aid	
	I	II	I	II	I	II
	Panel A: Dependent Var. Total Score					
χ^2	31.90	14.78	25.35	10.75	21.63	7.77
	(0.007)	(0.011)	(0.045)	(0.057)	(0.118)	(0.169)
d.f.	15	5	15	5	15	5
# of clusters	226	226	197	197	324	324
N	3,699	3,699	3,972	3,972	5,688	5,688
	Panel B: Falsification test (Dependent Var. Age)					
χ^2	8.69	3.54	11.43	4.76	11.67	2.35
	(0.893)	(0.617)	(0.722)	(0.445)	(0.704)	(0.799)
d.f.	15	5	15	5	15	5
# of clusters	226	226	197	197	324	324
N	3,699	3,699	3,972	3,972	5,688	5,688

Note: Results are based on the Chi-squared test in Theorem 2 for the special case with $T = Z$; Variance-covariance matrices are bootstrapped with 2,000 replications, clustered at classroom level; P-values are in the parentheses; Columns I report results from the distributional test at quantiles 0.25, 0.50, and 0.75; Columns II reports the results from the mean test.

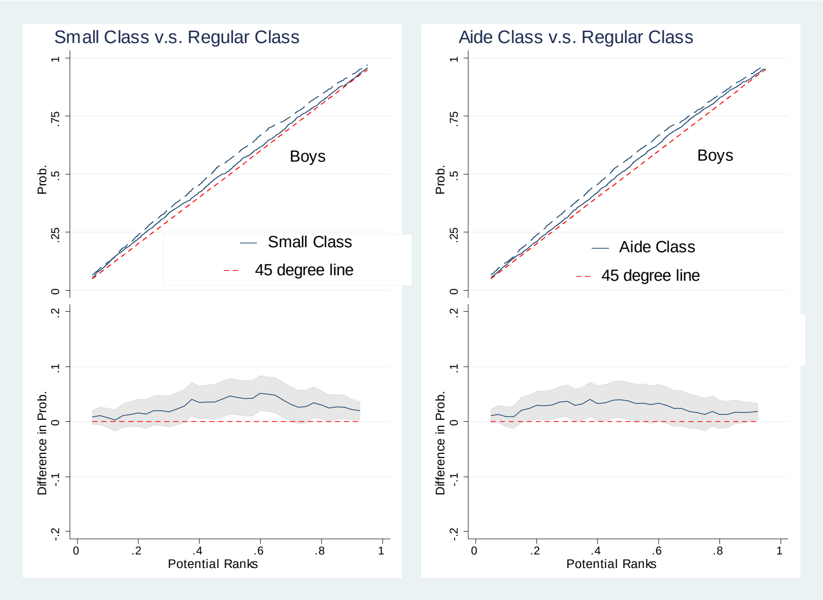
STAR: The Individual Quantile Tests

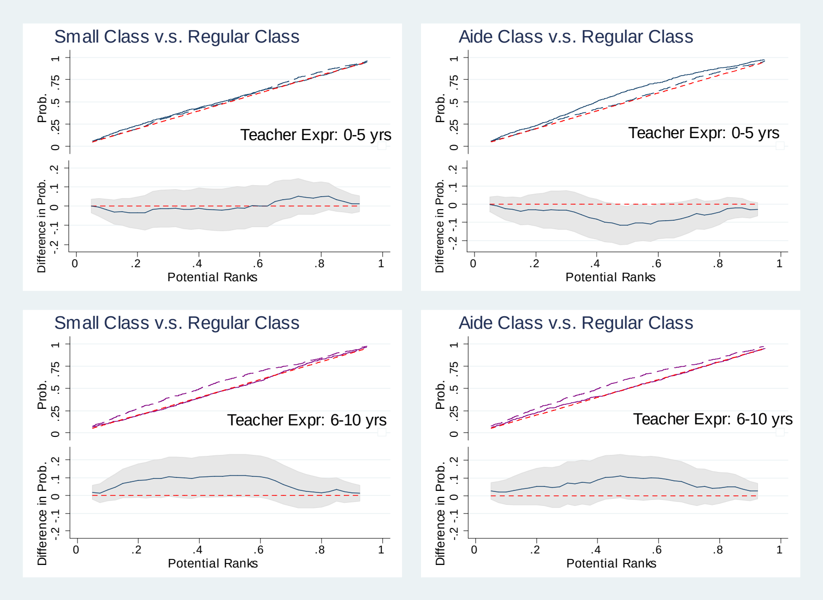
Table: The individual quantile tests for rank similarity - STAR

	Panel A: Dependent Var. Test Score				Panel B: Falsification test (Dependent Var			
	Small v.s. Regular		Aide v.s. Regular		Small v.s. Regular		Aide v.s. Regula	
Quantile	χ^2		χ^2		χ^2		χ^2	
0.15	5.62	(0.345)	4.88	(0.431)	3.86	(0.570)	1.07	(0.957)
0.20	8.47	(0.132)	7.53	(0.184)	3.31	(0.652)	1.75	(0.883)
0.25	15.83	(0.007)	10.28	(0.068)	3.89	(0.566)	2.84	(0.724)
0.30	14.86	(0.011)	11.51	(0.042)	2.20	(0.820)	4.16	(0.527)
0.35	16.29	(0.006)	10.98	(0.052)	5.78	(0.328)	4.39	(0.495)
0.40	15.73	(0.008)	10.48	(0.063)	4.15	(0.528)	4.11	(0.534)
0.45	18.14	(0.003)	11.66	(0.040)	2.08	(0.838)	5.08	(0.406)
0.50	13.89	(0.016)	11.02	(0.051)	1.63	(0.898)	2.16	(0.827)
0.55	14.17	(0.015)	8.96	(0.111)	1.21	(0.944)	1.69	(0.891)
0.60	17.37	(0.004)	9.41	(0.094)	1.39	(0.925)	1.68	(0.891)
0.65	12.82	(0.025)	6.26	(0.281)	2.33	(0.801)	1.15	(0.949)
0.70	10.30	(0.067)	5.14	(0.399)	3.01	(0.699)	2.78	(0.735)
0.75	5.24	(0.387)	3.12	(0.682)	2.42	(0.788)	6.68	(0.245)
0.80	7.01	(0.220)	2.04	(0.844)	4.39	(0.495)	11.33	(0.045)
0.85	7.62	(0.179)	2.92	(0.713)	11.23	(0.047)	15.80	(0.007)

Note: Results are based on the Chi-squared test in Theorem 2 for the special case with $T = Z$; Variance-covariance matrices are bootstrapped with 2,000 replications, clustered at classroom level; P-values are in the parentheses.

STAR: Changes in Potential Ranks for Sub-groups





Small Class v.s. Regular Class
Teacher Expr: 0-5 yrs
Aide Class v.s. Regular Class
Teacher Expr: 0-5 yrs
Small Class v.s. Regular Class
Teacher Expr: 6-10 yrs
Aide Class v.s. Regular Class
Teacher Expr: 6-10 yrs
Prob.
Difference in Prob.
Potential Ranks

STAR: Conclusions

- Attending a small class or having a teacher aide improves boys' ranks relative to girls' and narrows the gender performance gap.
- Small classes substantially narrow the performance gaps among students taught by teachers with different experience;
    - Greatest improvement is among students assigned to teachers with 6-10 years' experience.
- Assigning a teacher aide to an inexperienced teacher is relatively inefficient and negatively affects students' ranks.

Extension I: Covariates with Infinite Support

Allow $J \longrightarrow \infty$, as $n \longrightarrow \infty$.

Assumption 3

Assumption

1. *i.i.d. data: the data* $\{Y_i, T_i, Z_i, \mathbf{X}_i\}$ *for* $i = 1, ..., n$ *is a random sample of size n of* $(Y, T, Z, \mathbf{X})$.
2. *For all* $\tau \in \Omega = \{\tau_1, \tau_2, ..., \tau_K\}$, *the random variable* Y_1 *and* Y_0 *are continuously distributed with positive density in a neighborhood of* $q_{0|C}(\tau)$ *and* $q_{1|C}(\tau)$ *in the subpopulation of compliers.*
3. *Let* $n_j = \sum_{i=1}^{n} 1(\mathbf{X} = \mathbf{x_j})$. $n_j \asymp n/J$ *uniformly over* j, *i.e. there exists* $0 < c \leq C < \infty$ *such that* $c\frac{n}{J} \leq n_j \leq C\frac{n}{J}$ *for all* $j = 1, ..., J$.
4. $\hat{\pi}(\mathbf{x}_j)$ *is uniformly consistent, or* $\sup_{j=1,...,J} |\hat{\pi}(\mathbf{x}_j) - \pi(\mathbf{x}_j)| \xrightarrow{p} 0$ *as* $n, J \longrightarrow \infty$.
5. *For all* $t, z = 0, 1$, $j = 1, ..., J$ *and* $\tau \in \Omega$, $f_{Y|T,Z,\mathbf{X}}(.|t, z, \mathbf{x}_j)$ *is bounded in a neighborhood of* $q_{t|C}(\tau)$. *For all* $\tau \in \Omega$ *and* $j = 1, ..., J$, $f_{Y|\mathbf{X}}(.|\mathbf{x}_j)$ *is positive and bounded in a neighborhood of* $q_{t|C}(\tau)$.

Extension I: Covariates with Infinite Support

Let $\hat{\mathbf{m}}_j^z = \left(\hat{m}_j^z(\tau_1), \ldots, \hat{m}_j^z(\tau_K)\right)'$ and $\mathbf{m}_j^z = \left(m_j^z(\tau_1), \ \ldots, \ m_j^z(\tau_K)\right)'$ be $K \times 1$ vector.

Corollary (4)

Given Assumptions 2 and 3, we have

$$\sqrt{\frac{n_j^1 n_j^0}{n_j^1 + n_j^0}} \left(\hat{\mathbf{m}}_j^1 - \hat{\mathbf{m}}_j^0 - \left(\mathbf{m}_j^1 - \mathbf{m}_j^0\right)\right) \Rightarrow \mathbf{Z}_j \sim N(0, \mathbf{V}_j),$$

where $\mathbf{Z}_j$ *for* $j = 1, \ldots, J-1$ *follow independent multivariate normal distributions with mean zero and variance-covariance matrix* $\mathbf{V}_j$*, and the* (k, k')*-th element of* $\mathbf{V}_j$ *is*

$$\begin{aligned} V_{j;k,k'} \quad = \quad & \pi(\mathbf{x}_j) m_j^1(\tau_k \wedge \tau_{k'}) \left(1 - m_j^1(\tau_{k'})\right) \\ & + (1 - \pi(\mathbf{x}_j)) m_j^0(\tau_k \wedge \tau_{k'}) \left(1 - m_j^0(\tau_{k'})\right). \end{aligned}$$

Extension I: Covariates with Inifite Support

For each $j = 1, ..., J-1$, define the Wald-type statistic

$$w_j = \frac{n_j^1 n_j^0}{n_j^1 + n_j^0} \left(\hat{\mathrm{m}}_j^1 - \hat{\mathrm{m}}_j^0\right)' \hat{\mathbf{V}}_j^{-1} \left(\hat{\mathrm{m}}_j^1 - \hat{\mathrm{m}}_j^0\right),$$

where $\hat{\mathbf{V}}_j$ is a consistent estimator of $\mathbf{V}_j$. The (k, k')-th element of $\hat{\mathbf{V}}_j$ is

$$\begin{aligned} \hat{V}_{j;k,k'} &= \frac{n_j^0}{n_j^0 + n_j^1} \hat{m}_j^1(\tau_k \wedge \tau_{k'}) \left(1 - \hat{m}_j^1(\tau_{k'})\right) \\ &\quad + \frac{n_j^1}{n_j^0 + n_j^1} \hat{m}_j^0(\tau_k \wedge \tau_{k'}) \left(1 - \hat{m}_j^0(\tau_{k'})\right). \end{aligned}$$

The test statistic is then

$$W_{largeJ} = \frac{\sum_{j=1}^{J-1} w_j - K(J-1)}{\sqrt{2K(J-1)}} \sim N(0,1).$$

The *one-sided* decision rule of the test is to

"reject the null hypothesis H_0 if $W_{largeJ} > c_\alpha$".

Extension II: Continuous Covariates

Let $m_k^z(\mathbf{x}) = E[I(\tau_k)|Z = z, \mathbf{X} = \mathbf{x}]$ for $z = 0, 1$.

$$H_0 : m_k^1(\mathbf{x}) = m_k^0(\mathbf{x}) \textit{ for all } \mathbf{x} \in \mathcal{X} \textit{ and } k = 1, ..., K,$$

Form Kolmogorov-Smirnov type test statistic:

$$KS = \sup_{k,\mathbf{x}} \left| \frac{\hat{m}_k^1(\mathbf{x}) - \hat{m}_k^0(\mathbf{x})}{s_k(\mathbf{x})} \right|,$$

where $\hat{m}_k^1(\mathbf{x})$ and $\hat{m}_k^0(\mathbf{x})$ are standard local linear estimators, and $s_k(\mathbf{x})$ be the standard error of $\hat{m}_k^1(\mathbf{x}) - \hat{m}_k^0(\mathbf{x})$.

Extension II: Continuous Covariates

- Construct the critical value c_α by Gaussian multiplier bootstrap. Let $\hat{m}_k^*(\mathbf{x})$ be a multiplier process such that

$$\hat{m}_k^*(\mathbf{x}) = \frac{\sum_{Z_i=0} \eta_i \hat{\epsilon}_{k,i} \mathcal{K}_{h_0}(\mathbf{X}_i - \mathbf{x})}{\sum_{Z_i=0} \mathcal{K}_{h_0}(\mathbf{X}_i - \mathbf{x})} - \frac{\sum_{Z_i=1} \eta_i \hat{\epsilon}_{k,i} \mathcal{K}_{h_1}(\mathbf{X}_i - \mathbf{x})}{\sum_{Z_i=1} \mathcal{K}_{h_1}(\mathbf{X}_i - \mathbf{x})}$$

  with $\{\eta_i\}_{i=1}^N$ simulated from i.i.d. $N(0,1)$, independent of data, where $\hat{\epsilon}_{k,i} = 1\left(Y_i \leq \hat{q}_{1|C}(\tau_k) T_i + \hat{q}_{0|C}(\tau_k)(1 - T_i)\right) - \hat{m}_k^1(\mathbf{x}_i) Z_i - \hat{m}_k^0(\mathbf{x}_i)(1 - Z_i)$. c_α is the $(1-\alpha) \times 100$-th percentile of the simulated process $\sup_{k,\mathbf{x}} \left| \frac{\hat{m}_k^*(\mathbf{x})}{s_k(\mathbf{x})} \right|$.
- Reject the null if $KS > c_\alpha$.

Extension III: Testing Conditional Ranks

Two main modifications are required:

1. First, estimate conditional quantiles conditional on some covariates $\mathbf{X}_1$ of interest.
2. Second, use additional covariates $\mathbf{X}_2$ other than $\mathbf{X}_1$ in the first-step to perform the test.

E.g., We estimate quantiles of potential earnings, and perform tests for male and female trainees, so the tests are essentially rank similarity tests for conditional ranks conditional on gender.

Extension III: Testing Conditional Ranks

- Let $q_{t|C,\mathbf{x_1}}(\tau|\mathbf{x}_1) = F^{-1}_{t|C,\mathbf{X}_1}(\tau|\mathbf{x}_1)$ for $t = 0, 1$ and $\tau \in (0, 1)$. If Assumption 1 holds conditional on $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')'$, by Frolich and Melly (2013)

$$\begin{aligned} & \left(q_{0|C,\mathbf{x_1}}(\tau|\mathbf{x}_1), q_{1|C,\mathbf{x_1}}(\tau|\mathbf{x}_1)\right) \\ = & \arg\min_{q_0,q_1} E\left[\rho_\tau(Y - q_0(1-T) - q_1 T)\omega^{FM}|\mathbf{X}_1 = \mathbf{x}_1\right], \quad (3) \end{aligned}$$

where $\omega^{FM} = \left(\frac{Z}{\pi(\mathbf{X})} - \frac{1-Z}{1-\pi(\mathbf{X})}\right)(2T-1)$.

- If Assumption 1 holds conditional on $\mathbf{X}_1$, then may still utilize (3). Or if a linear model for conditional quantiles is assumed, by Abadie, Angrist and Imbens (2002)

$$\begin{aligned} & \left(\tilde{q}_{0|C}(\tau), \tilde{q}_{1|C}(\tau)\right) \\ = & \arg\min_{q_0,q_1} E\left[\rho_\tau(Y - q_0(1-T) - q_1 T - \mathbf{X}_1'\gamma)\omega^{AAI}\right], \quad (4) \end{aligned}$$

where $\omega^{AAI} = 1 - \frac{T(1-Z)}{1-\pi(\mathbf{x}_1)} - \frac{(1-T)Z}{1-\pi(\mathbf{x}_1)}$, $\pi(\mathbf{x}_1) = \Pr(Z = 1|\mathbf{X} = \mathbf{x}_1)$.

Define the rank indicator

$$\tilde{I}(\tau|\mathbf{x}_1) \equiv 1\left(Y \leq \left(Tq_{0|C,\mathbf{x_1}}(\tau|\mathbf{x}_1) + (1-T)\, q_{1|C,\mathbf{x_1}}(\tau|\mathbf{x}_1)\right)\right).$$

- Analogous to Theorem 1, rank similarity for the conditional ranks conditional on $\mathbf{X}_1 = \mathbf{x}_1$ holds *if and only if* for all $\tau \in (0,1)$ and $\mathbf{x}_2 \in \mathcal{X}_2 \equiv Supp(\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1)$,

$$\begin{aligned} & E\left[\tilde{I}(\tau|\mathbf{x}_1)|Z=1, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2\right] \\ = \ & E\left[\tilde{I}\left(\tau|\mathbf{x}_1\right)|Z=0, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2\right]. \end{aligned} \tag{5}$$

- Can test conditional rank invariance or similarity by testing whether equation (5) holds for $\tau \in (0,1)$ and $\mathbf{x}_2 \in \mathcal{X}_2$, replacing $q_{t|\mathbf{x_1},C}(\tau|\mathbf{x}_1)$, $t = 0, 1$ with their estimates.

Conclusions

- Nonparametrically identify and test the counterfactual distribution of potential ranks, or features of the distribution, such as moments, median or any particular quantile.
  - Allow treatment to be endogenous (exogenous treatment follows as a special case).
  - Can handle IVs that are valid conditional on covariates or valid unconditionally.
  - Tests informative regarding at which part of the distribution rank similarity is violated.
  - Good size and power of the proposed tests in small samples.
- Show usefullness of the results in empirical settings: JTPA training and Project STAR.