Testing for Rank Invariance or Similarity in Program Evaluation

Yingying Dong* Shu Shen†

First version, February 2015; this version, September 2016

Online Appendices

A Monte Carlo Simulations

This appendix demonstrates the finite sample size and power properties of the proposed tests by Monte Carlo simulations. We first consider exogenous treatment and then endogenous treatment.

For all Monte Carlo simulations, the observed covariate X is generated to take on five values with equal probability. In particular, $\Pr(X = 0.4j) = 0.2$, for $j = 1, ..., 5$. The unobserved covariate V and the idiosyncratic shocks S_0 and S_1 are generated as independent random variables and $V, S_0, S_1 \sim N(0, 1)$. In addition, $Y_0 = X + V + S_0$, $Y_1 = X + 1 - bXV + V + S_1$ and $Y = Y_0 (1 - T) + Y_1 T$. When $b = 0$, rank is invariant to treatment by construction; when $b \neq 0$, rank similarity does not hold. The value of b controls for the degree of violation in rank similarity. We consider $b = 0$, 2 and 3. DGP 1 focuses on exogenous treatment and DGP 2 focuses on endogenous treatment.

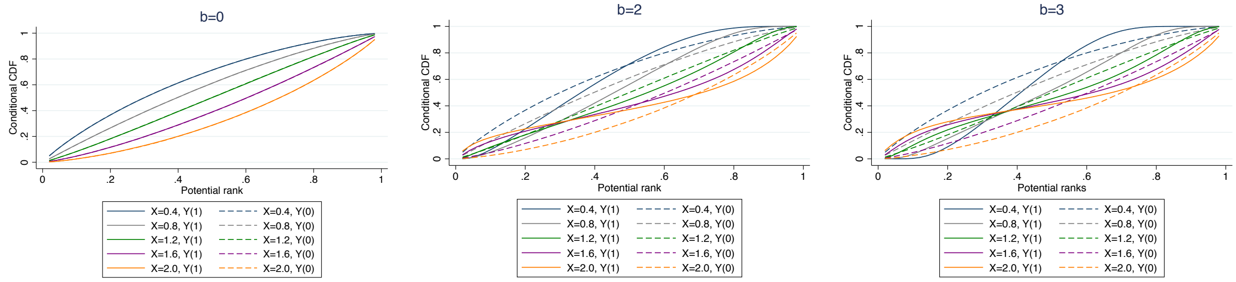
DGP 1: $\Pr(T = 0) = \Pr(T = 1) = 0.5$.

DGP 2: $\Pr(Z = 0) = \Pr(Z = 1) = 0.5$; $T = \mathbf{1}(0.15(Y_1 - Y_0) + Z - 0.5 > 0)$.

When treatment is endogenous, we have a generalized Roy model (Roy, 1951). Treatment is determined by an individual's idiosyncratic gain from the treatment $Y_1 - Y_0$ and additionally by an exogenous variable Z. Z can be taken as a random assignment indicator

*University of California, Irvine, email: yyd@uci.edu
†University of California, Davis, email: shushen@ucdavis.edu

Figure A.1: Conditional distributions of potential ranks



for a program, representing program incentives. Subtracting 0.5 normalizes $Z - 0.5$ to be mean zero.

Figure A.1 illustrates the design of DGP 1 for different values of b. In each graph, the solid and dotted lines represent $F_{U_1|X}(.|x)$ and $F_{U_0|X}(.|x)$, the conditional distributions of potential ranks under treatment and no treatment, respectively at $x = 0.4, 0.8, ..., 2$. When $b = 0$, rank is invariant to treatment, so the two conditional CDFs overlap at all values of X. When $b \neq 0$, rank similarity is violated. Figure A.1 shows that how the degree of violation increases with b. In addition, for $b = 2, 3$, rank similarity is violated more strongly at the lower quantiles.

The left half of Table A.1 presents simulation results under DGP 1. We conduct the mean test as well as the distributional test with four different sets of quantiles. For each $b = 0, 2, 3$, we draw samples of size 500, 1000, 1500, 2000 and 2500. All test statistics are constructed using bootstrapped variance-covariance matrices with 1,000 bootstrap repetitions. 1,000 simulations are performed for each test. Table A.1 reports the rejection rates at the 5% significance level.

Results for $b = 0$ under DGP 1 in Table A.1 show that both the distributional test and the mean test control size well. Results for $b = 2, 3$ show that the power of the proposed tests increases with the sample size. The rejection rate goes to one rapidly with the increase of the sample size. When $b = 2, 3$, as shown in Figure A.1, rank similarity is violated more seriously at the lower quantiles. Consequently, the distributional tests show greater power when these lower quantiles are included in the tests. In contrast, the mean test for rank

Table A.1: Small sample performance of the proposed tests

	DGP 1					DGP 2				
N	500	1000	1500	2000	2500	500	1000	1500	2000	2500
			$b = 0$					$b = 0$		
$\Omega = \{0.5\}$	0.034	0.039	0.051	0.040	0.053	0.025	0.036	0.041	0.038	0.057
$\Omega = \{0.2, 0.3, 0.4\}$	0.013	0.013	0.025	0.021	0.023	0.012	0.012	0.018	0.017	0.025
$\Omega = \{0.5, 0.6, 0.7, 0.8\}$	0.014	0.014	0.023	0.023	0.018	0.006	0.013	0.016	0.022	0.015
$\Omega = \{0.2, 0.3, ..., 0.8\}$	0.006	0.010	0.013	0.013	0.013	0.002	0.010	0.006	0.010	0.008
Mean Test	0.051	0.044	0.048	0.041	0.067	0.054	0.050	0.051	0.045	0.057
			$b = 2$					$b = 2$		
$\Omega = \{0.5\}$	0.074	0.150	0.232	0.303	0.388	0.084	0.242	0.379	0.522	0.615
$\Omega = \{0.2, 0.3, 0.4\}$	0.269	0.776	0.968	0.994	1.000	0.170	0.589	0.870	0.965	0.993
$\Omega = \{0.5, 0.6, 0.7, 0.8\}$	0.151	0.581	0.857	0.962	0.991	0.021	0.150	0.340	0.600	0.764
$\Omega = \{0.2, 0.3, ..., 0.8\}$	0.287	0.910	0.996	1.000	1.000	0.053	0.431	0.823	0.960	0.993
Mean Test	0.103	0.213	0.278	0.424	0.500	0.152	0.322	0.481	0.622	0.709
			$b = 3$					$b = 3$		
$\Omega = \{0.5\}$	0.143	0.335	0.512	0.640	0.800	0.113	0.293	0.441	0.617	0.700
$\Omega = \{0.2, 0.3, 0.4\}$	0.817	0.999	1.000	1.000	1.000	0.284	0.783	0.975	1.000	1.000
$\Omega = \{0.5, 0.6, 0.7, 0.8\}$	0.306	0.880	0.992	1.000	1.000	0.020	0.198	0.450	0.704	0.865
$\Omega = \{0.2, 0.3, ..., 0.8\}$	0.836	0.999	1.000	1.000	1.000	0.093	0.634	0.949	1.000	1.000
Mean Test	0.340	0.659	0.853	0.941	0.971	0.191	0.441	0.602	0.772	0.843

similarity has much lower power due to the fact that rank similarity is violated only at part of the distribution.

Figure A.2 demonstrates the power performance of the proposed tests visually. The left graph presents evidence when the sample size is fixed at 1,000 and the value of b varies; the right graph presents evidence when b is fixed at $b = 2$ and the sample size varies. As is clear from these graphs, the distributional tests at a wide range of quantile values in general have better small sample performance given the DGP under study.

Figure A.2: Small sample performance of the proposed tests: exogenous treatment

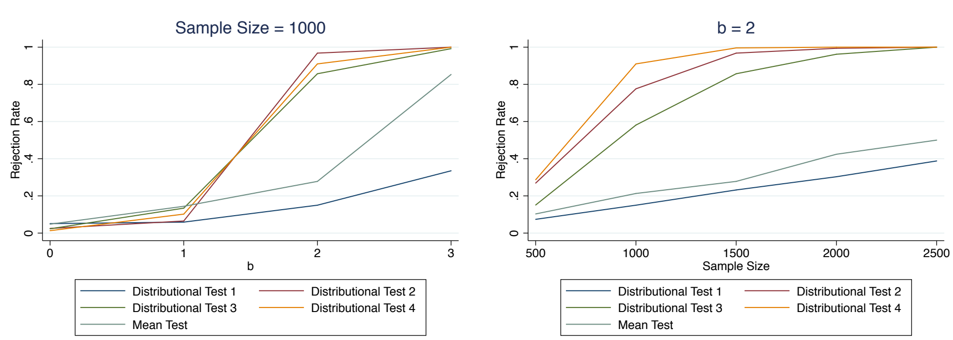


Figure A.3: Conditional distributions of potential ranks among compliers

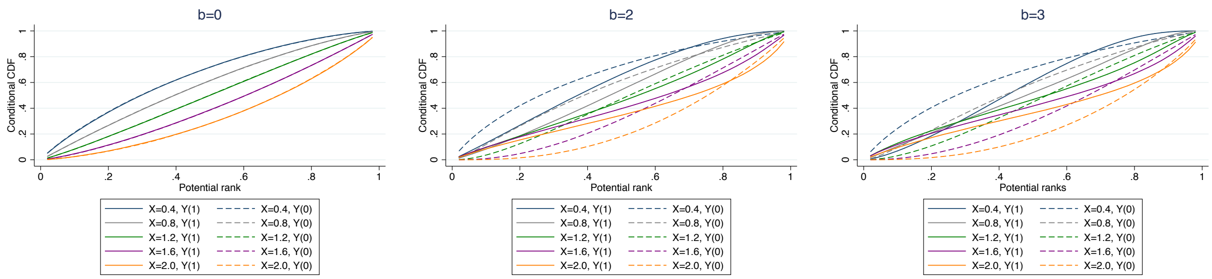
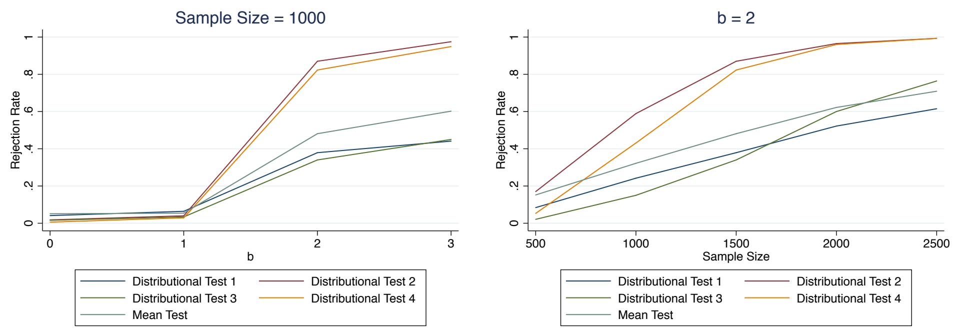


Figure A.4: Small sample performance of the proposed tests: endogenous treatment



Next we study the performance of the proposed tests under DGP 2 where treatment is endogenous. Figure A.3 illustrates the design of DGP 2. In each graph, the solid and dotted lines represent $F_{U_1|C,X}(.|x)$ and $F_{U_0|C,X}(.|x)$, the conditional distributions of compliers' ranks under treatment and no treatment, respectively, at $x = 0.4, 0.8, ..., 2$. Again the violation of rank similarity is greater when the value of b is larger, especially at the lower quantiles.

The right half of Table A.1 presents the simulation results under DGP 2. The proposed tests again control size well. The power of the tests increases quickly with the sample size. The distributional tests at a range of quantile values again outperform the mean and median tests in power. Note that at any given sample size, tests for DGP 2 with endogenous treatment has substantially lower power than tests for DGP 1 where the treatment is exogenous. This is not surprising, since the ranks of always takers and never takers do not change by construction and the effective sample size is determined by the fraction of compliers. Finally, Figure A.4 visually illustrates the small sample performance of the proposed tests with endogenous treatment.

B Proofs

Proof of Lemma 1

For simplicity, here we assume that both $\mathbf{X}$ and V are continuous. When $\mathbf{X}$ and V are discrete, one can simply replace the probability density with probability mass functions and integration with summation below.

Part 1: By Bayes' Rule,

$$f_{\mathbf{X},V|U_t}(\mathbf{x},v|\tau) = \frac{f_{U_t|\mathbf{X},V}(\tau|\mathbf{x},v)f_{\mathbf{X},V}(\mathbf{x},v)}{\iint f_{U_t|\mathbf{X},V}(\tau|\mathbf{x},v)f_{\mathbf{X},V}(\mathbf{x},v)d\mathbf{x}dv} \text{ for } t = 0, 1.$$

By the definition of rank similarity, $f_{U_0|\mathbf{X},V}(\tau|\mathbf{x},v) = f_{U_1|\mathbf{X},V}(\tau|\mathbf{x},v)$. It follows immediately from Bayes' Rule that $f_{\mathbf{X},V|U_1}(\mathbf{x},v|\tau) = f_{\mathbf{X},V|U_0}(\mathbf{x},v|\tau)$, which in turn implies that $F_{\mathbf{X},V|U_0}(\mathbf{x},v|\tau) \equiv F_{\mathbf{X},V|U_1}(\mathbf{x},v|\tau)$.

Part 2: By the definition of rank similarity, $f_{U_0|\mathbf{X},V}(\tau|\mathbf{x},v) = f_{U_1|\mathbf{X},V}(\tau|\mathbf{x},v)$. Since $f_{U_t|\mathbf{X}}(\tau|\mathbf{x}) = \int f_{U_0|\mathbf{X},V}(\tau|\mathbf{x},v)dF_{V|\mathbf{X}}(v|\mathbf{x})$, we know $f_{U_0|\mathbf{X}}(\tau|\mathbf{x}) = f_{U_1|\mathbf{X}}(\tau|\mathbf{x})$, which implies that $F_{U_0|\mathbf{X}}(\tau|\mathbf{x}) = F_{U_1|\mathbf{X}}(\tau|\mathbf{x})$.

Proof of Lemma 2

Like in the proof of Lemma 2, we assume without loss of generality that both $\mathbf{X}$ and V are continuous.

Lemma 1 shows that rank similarity implies that $f_{U_0|\mathbf{X}}(\tau|\mathbf{x}) = f_{U_1|\mathbf{X}}(\tau|\mathbf{x})$. The following proves that $f_{U_0|\mathbf{X}}(\tau|\mathbf{x}) = f_{U_1|\mathbf{X}}(\tau|\mathbf{x})$ implies rank similarity given the condition stated in the lemma.

By Bayes' rule

$$f_{U_t|\mathbf{X},V}(\tau|\mathbf{x},v) = \frac{f_{\mathbf{X},V,U_t}(\mathbf{x},v,\tau)}{f_{\mathbf{X},V}(\mathbf{x},v)} = \frac{f_{V|\mathbf{X},U_t}(v|\mathbf{x},\tau)f_{U_t|\mathbf{X}}(u_t|\mathbf{x})f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{X},V}(\mathbf{x},v)}.$$

Therefore, $f_{U_1|\mathbf{X},V}(\tau|\mathbf{x},v) = f_{U_0|\mathbf{X},V}(\tau|\mathbf{x},v)$ and rank similarity holds.

Proof of Theorem 1

Identification of $F_{U_t|C,\mathbf{X}}(\tau|\mathbf{x})$ for $t=0,1$, $\tau\in(0,1)$ and $\mathbf{x}\in\mathcal{X}_C$:

Given Assumption 1, and for $\mathbf{x}\in\mathcal{X}_C$ $E(T|Z=1,\mathbf{X}=\mathbf{x})-E(T|Z=0,\mathbf{X}=\mathbf{x})\neq 0$, following the standard LATE arguments (see, e.g. Abadie 2003), $F_{U_t|C,\mathbf{X}}(\tau|\mathbf{x})$ for $t=0,1$ are identified as follows

$$
\begin{aligned}
&F_{U_1|C,\mathbf{X}}(\tau|\mathbf{x}) = E\left[\mathbf{1}(U_1\leq\tau)|C,\mathbf{X}=\mathbf{x}\right] = E\left[\mathbf{1}(Y_1\leq q_{1|C}(\tau))|C,\mathbf{X}=\mathbf{x}\right]\\
=\ &\frac{E\left[\mathbf{1}(Y\leq q_{1|C}(\tau))T|Z=1,\mathbf{X}=\mathbf{x}\right]-E\left[\mathbf{1}(Y\leq q_{1|C}(\tau))T|Z=0,\mathbf{X}=\mathbf{x}\right]}{E\left[T|Z=1,\mathbf{X}=\mathbf{x}\right]-E\left[T|Z=0,\mathbf{X}=\mathbf{x}\right]};\\
&F_{U_0|C,\mathbf{X}}(\tau|\mathbf{x}) = E\left[\mathbf{1}(U_0\leq\tau)|C,\mathbf{X}=\mathbf{x}\right] = E\left[\mathbf{1}(Y_0\leq q_{0|C}(\tau))|C,\mathbf{X}=\mathbf{x}\right]\\
=\ &\frac{E\left[\mathbf{1}(Y\leq q_{0|C}(\tau))(1-T)|Z=1,\mathbf{X}=\mathbf{x}\right]-E\left[\mathbf{1}(Y\leq q_{0|C}(\tau))(1-T)|Z=0,\mathbf{X}=\mathbf{x}\right]}{E\left[1-T|Z=1,\mathbf{X}=\mathbf{x}\right]-E\left[1-T|Z=0,\mathbf{X}=\mathbf{x}\right]}.
\end{aligned}
$$

Recall $I(\tau)\equiv\mathbf{1}\left(Y\leq Tq_{1|C}(\tau)+(1-T)q_{0|C}(\tau)\right)$. The above equations can be re-written as

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

for t=0,1, which is equation (2) in Theorem 1.

Next we derive equation (3) in Theorem 1:

Case 1: for any $\mathbf{x}\in\mathcal{X}_C$, $E\left[T|Z=1,\mathbf{X}=\mathbf{x}\right]-E\left[T|Z=0,\mathbf{X}=\mathbf{x}\right]\neq 0$. Based on the results in equation (2), we know that equation (1) holds if and only if

$$
F_{U_1|C,\mathbf{X}}(\tau|\mathbf{x})-F_{U_0|C,\mathbf{X}}(\tau|\mathbf{x}) = \frac{E\left[I(\tau)|Z=1,\mathbf{X}=\mathbf{x}\right]-E\left[I(\tau)|Z=0,\mathbf{X}=\mathbf{x}\right]}{E\left[T|Z=1,\mathbf{X}=\mathbf{x}\right]-E\left[T|Z=0,\mathbf{X}=\mathbf{x}\right]} = 0.
$$

Case 2: for any $\mathbf{x}\in\mathcal{X}/\mathcal{X}_C$, $T_0=T_1$ by Assumption 1.3. Further

$$
\begin{aligned}
E\left[I(\tau)|Z=z,\mathbf{X}=\mathbf{x}\right] &= E\left[\mathbf{1}(Y\leq q_1(\tau))T+\mathbf{1}(Y\leq q_0(\tau))(1-T)|Z=z,\mathbf{X}=\mathbf{x}\right]\\
&= E\left[\mathbf{1}(Y_1\leq q_1(\tau))T_z+\mathbf{1}(Y_0\leq q_0(\tau))(1-T_z)|Z=z,\mathbf{X}=\mathbf{x}\right]\\
&= E\left[\mathbf{1}(Y_1\leq q_1(\tau))T_z+\mathbf{1}(Y_0\leq q_0(\tau))(1-T_z)|\mathbf{X}=\mathbf{x}\right],
\end{aligned}
$$

where the second equality holds by the definition of Y and T_0, T_1 and the third equality holds by Assumption 1.1. $T_0 = T_1$ then means that Equation (3) holds trivially for $\mathbf{x} \in \mathcal{X}/\mathcal{X}_C$.

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

$$E\left[I(\tau)|Z=1,\mathbf{X}=\mathbf{x}\right] - E\left[I(\tau)|Z=0,\mathbf{X}=\mathbf{x}\right] = 0 \quad \forall \mathbf{x} \in \mathcal{X}.$$

Proof for the fact that $\sum_{j=1}^{J} m_j^1(\tau)P\left[\mathbf{X}=\mathbf{x}_j\right] = \sum_{j=1}^{J} m_j^0(\tau)P\left[\mathbf{X}=\mathbf{x}_j\right]$

Recall that we write down the null hypothesis for the rank similarity test $H_0 : m_j^0(\tau_k) = m_j^1(\tau_k)$, we only included $j = 1, ..., J-1$. This is because for all $\tau \in (0,1)$,

$$\sum_{j=1}^{J} m_j^1(\tau)\Pr\left(\mathbf{X}=\mathbf{x}_j\right) = \sum_{j=1}^{J} m_j^0(\tau)\Pr\left(\mathbf{X}=\mathbf{x}_j\right).$$

Next, we prove the above equation. Notice that for all $\tau \in (0,1)$,

$$\begin{aligned}
& m_j^1(\tau) - m_j^0(\tau) = E\left[I(\tau)|\mathbf{X}=\mathbf{x}_j, Z=1\right] - E\left[I(\tau)|\mathbf{X}=\mathbf{x}_j, Z=0\right] \\
=& E\left[1\left(Y \le q_{1|C}(\tau)T + q_{0|C}(\tau)(1-T)\right)|\mathbf{X}=\mathbf{x}_j, Z=1\right] \\
& - E\left[1\left(Y \le q_{1|C}(\tau)T + q_{0|C}(\tau)(1-T)\right)|\mathbf{X}=\mathbf{x}_j, Z=0\right] \\
=& E\left[1\left(Y_1 \le q_{1|C}(\tau)\right)|\mathbf{X}=\mathbf{x}_j, Z=1, C\right]\Pr\left(C|\mathbf{X}=\mathbf{x}_j, Z=1\right) \\
& + E\left[1\left(Y_1 \le q_{1|C}(\tau)\right)|\mathbf{X}=\mathbf{x}_j, Z=1, A\right]\Pr\left(A|\mathbf{X}=\mathbf{x}_j, Z=1\right) \\
& + E\left[1\left(Y_0 \le q_{0|C}(\tau)\right)|\mathbf{X}=\mathbf{x}_j, Z=1, N\right]\Pr\left(N|\mathbf{X}=\mathbf{x}_j, Z=1\right) \\
& - E\left[1\left(Y_0 \le q_{0|C}(\tau)\right)|\mathbf{X}=\mathbf{x}_j, Z=0, C\right]\Pr\left(C|\mathbf{X}=\mathbf{x}_j, Z=0\right) \\
& - E\left[1\left(Y_1 \le q_{1|C}(\tau)\right)|\mathbf{X}=\mathbf{x}_j, Z=0, A\right]\Pr\left(A|\mathbf{X}=\mathbf{x}_j, Z=0\right) \\
& - E\left[1\left(Y_0 \le q_{0|C}(\tau)\right)|\mathbf{X}=\mathbf{x}_j, Z=0, N\right]\Pr\left(N|\mathbf{X}=\mathbf{x}_j, Z=0\right) \\
=& E\left[1\left(Y_1 \le q_{1|C}(\tau)\right) - 1\left(Y_0 \le q_{0|C}(\tau)\right)|\mathbf{X}=\mathbf{x}_j, C\right]\Pr\left(C|\mathbf{X}=\mathbf{x}_j\right) \\
=& \Pr\left[Y_1 \le q_{1|C}(\tau), C|\mathbf{X}=\mathbf{x}_j\right] - \Pr\left[Y_0 \le q_{0|C}(\tau), C|\mathbf{X}=\mathbf{x}_j\right].
\end{aligned}$$

The fourth equality follows from the conditional independence assumption that $(Y_1, Y_0, T_1, T_0) \perp Z|\mathbf{X}$.

Therefore,

$$\sum_{j=1}^{J} \left(m_j^1(\tau) - m_j^0(\tau)\right) \Pr(\mathbf{X} = \mathbf{x}_j) = \Pr\left[Y_1 \le q_{1|C}(\tau), C\right] - \Pr\left[Y_0 \le q_{0|C}(\tau), C\right]$$

$$= \Pr\left[Y_1 \le q_{1|C}(\tau)|C\right]\Pr(C) - \Pr\left[Y_0 \le q_{0|C}(\tau)|C\right]\Pr(C) = (\tau - \tau)\Pr(C) = 0.$$

Proof of Theorem 2

Proof. Let $\hat{p}_j^z = \frac{n_j^z}{n}$ be the nonparametric estimator of $p_j^z = P_{Z,X}(z, \mathbf{x}_j)$. The nonparametric estimator $\hat{m}_j^z(\tau_k)$ and its population counterpart $m_j^z(\tau_k)$ can be rewritten as

$$\hat{m}_j^z(\tau_k) = \sum_{t=0,1} \frac{1}{n} \sum_{i}^{n} \mathbf{1}\left(Y_i \le \hat{q}_{t|C}(\tau_k), T_i = t, Z_i = z, \mathbf{X}_i = \mathbf{x}_j\right)/\hat{p}_j^z,$$

$$m_j^z(\tau_k) = \sum_{t=0,1} F_{Y|T,Z,\mathbf{X}}(q_{t|C}(\tau_k)|t, z, \mathbf{x}_j) P_{T|Z,\mathbf{X}}(t|z, \mathbf{x}_j).$$

Define intermediate statistics

$$\check{m}_j^z(\tau_k) = \sum_{t=0,1} \frac{1}{n} \sum_{i}^{n} \mathbf{1}\left(Y_i \le \hat{q}_{t|C}(\tau_k), T_i = t, Z_i = z, \mathbf{X}_i = \mathbf{x}_j\right)/p_j^z,$$

$$\tilde{m}_j^z(\tau_k) = \sum_{t=0,1} F_{Y|T,Z,\mathbf{X}}(\hat{q}_{t|C}(\tau_k)|t, z, \mathbf{x}_j) P_{T|Z,\mathbf{X}}(t|z, \mathbf{x}_j)..$$

We can decompose $\sqrt{n}\left(\hat{m}_j^z(\tau_k) - m_j^z(\tau_k)\right)$ such that

$$\sqrt{n}\left(\hat{m}_j^z(\tau_k) - m_j^z(\tau_k)\right) = \sqrt{n}\left(\hat{m}_j^z(\tau_k) - \check{m}_j^z(\tau_k)\right) + \sqrt{n}\left(\check{m}_j^z(\tau_k) - \tilde{m}_j^z(\tau_k)\right) + \sqrt{n}\left(\tilde{m}_j^z(\tau_k) - m_j^z(\tau_k)\right)$$

$$= I + II + III.$$

For all $z = 0, 1$, $j = 1, ..., J-1$, and $k = 1, ..., K$,

$$
\begin{aligned}
I &= \sum_{t=0,1} \frac{1}{n} \sum_{i}^{n} \mathbf{1}\left(Y_i \leq \hat{q}_{t|C}(\tau_k), T_i = t, Z_i = z, \mathbf{X}_i = \mathbf{x}_j\right) \frac{-\sqrt{n}(\hat{p}_j^z - p_j^z)}{\hat{p}_j^z p_j^z} \\
&= -\frac{\sum_{t=0,1} E[Y_i \leq \hat{q}_{t|C}(\tau_k), T_i = t, Z_i = z, \mathbf{X}_i = \mathbf{x}_j]}{(p_j^z)^2} \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} \mathbf{1}(Z_i = z, \mathbf{X}_i = \mathbf{x}_j) - p_j^z \right) + o_p(1), \\
&= -\frac{m_j^z(\tau_k)}{p_j^z} \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} \mathbf{1}(Z_i = z, \mathbf{X}_i = \mathbf{x}_j) - p_j^z \right) + o_p(1).
\end{aligned}
$$

The second equality follows from the fact that the concerned indicator functions are P-Glivenko-Cantelli, and that $\hat{p}_j^z$ is consistent. The third equality follows from the consistency of $\hat{q}_{t|C}(\tau_k)$ and the fact that $F_{Y,T,Z,\mathbf{X}}(., t, z, \mathbf{x}_j)$ is continuous in a neighborhood of $q_{t|C}(\tau_k)$.

Next, notice that for all $z = 0, 1$, $j = 1, ..., J-1$, $k = 1, ..., K$, and $t = 0, 1$,

$$
\begin{aligned}
&\int \left[\mathbf{1}(Y \leq \hat{q}_{t|C}(\tau_k), T = t, Z = z, \mathbf{X} = \mathbf{x}_j) - \mathbf{1}(Y \leq q_{t|C}(\tau_k), T = t, Z = z, \mathbf{X} = \mathbf{x}_j)\right]^2 dF_{Y,T,Z,\mathbf{X}}(Y, T, Z, \mathbf{X}) \\
&= \left[F_{Y,T,Z,\mathbf{X}}(\hat{q}_{t|C}(\tau_k), t, z, \mathbf{x}_j) + F_{Y,T,Z,\mathbf{X}}(q_{t|C}(\tau_k), t, z, \mathbf{x}_j) - 2F_{Y,T,Z,\mathbf{X}}(\min(\hat{q}_{t|C}(\tau_k), q_{t|C}(\tau_k)), t, z, \mathbf{x}_j)\right] \\
&\xrightarrow{p} 0.
\end{aligned}
$$

Then by Lemma 19.24 of Van der Vaart (1998),

$$
II = \sqrt{n}\left(\check{m}_j^z(\tau_k) - \tilde{m}_j^z(\tau_k)\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\frac{I_i(\tau_k)\mathbf{1}(Z_i = z, \mathbf{X}_i = \mathbf{X}_j)}{p_j^z} - m_j^z(\tau_k) \right] + o_p(1).
$$

Last, by continuity of $f_{Y|T,Z,\mathbf{X}}(y|t, z, \mathbf{x})$ in a neighborhood of $q_{t|C}(\tau_k)$, we have that

$$
\begin{aligned}
III =& \sqrt{n}\left(\tilde{m}_j^z(\tau_k) - m_j^z(\tau_k)\right) \\
=& \sum_{t=0,1} p_{T|Z,\mathbf{X}}(t|z, \mathbf{x}_j) f_{Y|T,Z,\mathbf{X}}(\tilde{q}_{t|C}(\tau_k)|t, z, \mathbf{x}_j)\sqrt{n}\left(\hat{q}_{t|C}(\tau_k) - q_{t|C}(\tau_k)\right) \\
=& \sum_{t=0,1} p_{T|Z,\mathbf{X}}(t|z, \mathbf{x}_j) f_{Y|T,Z,\mathbf{X}}(q_{t|C}(\tau_k)|t, z, \mathbf{x}_j)\sqrt{n}\left(\hat{q}_{t|C}(\tau_k) - q_{t|C}(\tau_k)\right) + o_p(1),
\end{aligned}
$$

where $\tilde{q}_{t|C}(\tau_k)$ is a value between $q_{t|C}(\tau_k)$ and $\hat{q}_{t|C}(\tau_k)$.

By Frolich and Melly (2013), we have

$$\sqrt{n}\left(\hat{q}_{1|C}(\tau_k) - q_{1|C}(\tau_k)\right) = -\frac{\sqrt{n}\Upsilon_n(q_{1|C}(\tau_k), \hat{w})}{P_c f_{1|C}(q_{1|C}(\tau_k))} + o_p(1)$$

where $\Upsilon_n(q_{1|C}(\tau_k), \hat{w})$ as is defined in Frolich and Melly (2013), is equal to

$$\begin{aligned}
\Upsilon_n(q_{1|C}(\tau_k), \hat{w}) =& \frac{1}{n}\sum_{i=1}\left(1(Y_i \leq q_{1|C}(\tau_k)) - \tau_k\right)\hat{w}_i T_i \\
=&\frac{1}{n}\sum_{i=1}\left(\frac{Z_iT_i}{\hat{\pi}(\mathbf{X}_i)} - \frac{(1-Z_i)T_i}{1-\hat{\pi}(\mathbf{X}_i)}\right)\left(1(Y_i \leq q_{1|C}(\tau_k)) - \tau_k\right) \\
=&\frac{1}{n}\sum_{i=1}\left(\frac{Z_iT_i}{\pi(\mathbf{X}_i)} - \frac{(1-Z_i)T_i}{1-\pi(\mathbf{X}_i)}\right)\left(1(Y_i \leq q_{1|C}(\tau_k)) - \tau_k\right) \\
&- \frac{1}{n}\sum_{i=1}\left(\frac{Z_iT_i}{\pi^2(\mathbf{X}_i)} + \frac{(1-Z_i)T_i}{(1-\pi(\mathbf{X}_i))^2}\right)1(Y_i \leq q_{1|C}(\tau_k)) - \tau_k)\left(\hat{\pi}(\mathbf{X}_i) - \pi(\mathbf{X}_i)\right) + o_p(1/\sqrt{n}) \\
=&\frac{1}{n}\sum_{i=1}\left(\frac{Z_iT_i}{\pi(\mathbf{X}_i)} - \frac{(1-Z_i)T_i}{1-\pi(\mathbf{X}_i)}\right)1(Y_i \leq q_{1|C}(\tau_k)) - \tau_k) \\
&- \frac{1}{n}\sum_{i=1}\sum_{z=0,1}\frac{P_{T|Z,\mathbf{X}}(1|z,\mathbf{X}_i)\left(F_{Y|T,Z,\mathbf{X}}(q_{1|C}(\tau_k)|1,z,\mathbf{X}_i) - \tau_k\right)}{p_{Z|\mathbf{X}}(z|\mathbf{X}_i)}\left(Z_i - \pi(\mathbf{X}_i)\right) + o_p(1/\sqrt{n}),
\end{aligned}$$

where $p_{Z|\mathbf{X}}(z|\mathbf{X}_i)$ is used to denote $\pi(\mathbf{X}_i)$ and $1 - \pi(\mathbf{X}_i)$ with $z = 1, 0$, respectively. The fourth equality holds because the V-statistic could be projected to a U-statistic as is shown in Frolich and Melly (2013). Therefore,

$$\begin{aligned}
\sqrt{n}\Upsilon_n(q_{1|C}(\tau_k), \hat{w}) =& \frac{1}{\sqrt{n}}\sum_{i=1}\psi_1(Y_i, T_i, Z_i, \mathbf{X}_i), \\
\psi_1(Y_i, T_i, Z_i, \mathbf{X}_i) =& \frac{Z_iT_i}{\pi(\mathbf{X}_i)}\left[\mathbf{1}\left(Y_i \leq q_{1|C}(\tau_k)\right) - \tau_k - \left(F_{Y|T,Z,\mathbf{X}}(q_{1|C}(\tau_k)|1,1,\mathbf{X}_i) - \tau_k\right)\right] \\
&- \frac{(1-Z_i)T_i}{1-\pi(\mathbf{X}_i)}\left[\mathbf{1}\left(Y_i \leq q_{1|C}(\tau_k)\right) - \tau_k - \left(F_{Y|T,Z,\mathbf{X}}(q_{1|C}(\tau_k)|1,0,\mathbf{X}_i) - \tau_k\right)\right] \\
&+ \frac{Z_iT_i - P_{T|Z,\mathbf{X}}(1|1,\mathbf{X}_i)(Z_i - \pi(\mathbf{X}_i))}{\pi(\mathbf{X}_i)}\left(F_{Y|T,Z,\mathbf{X}}(q_{1|C}(\tau_k)|1,1,\mathbf{X}_i) - \tau_k\right) \\
&- \frac{(1-Z_i)T_i + P_{T|Z,\mathbf{X}}(1|0,\mathbf{X}_i)(Z_i - \pi(\mathbf{X}_i))}{1-\pi(\mathbf{X}_i)}\left(F_{Y|T,Z,\mathbf{X}}(q_{1|C}(\tau_k)|1,0,\mathbf{X}_i) - \tau_k\right).
\end{aligned}$$

Likewise, one can show that

$$\begin{aligned}\psi_0(Y_i, T_i, Z_i, \mathbf{X}_i) = &-\frac{Z_i(1-T_i)}{\pi(\mathbf{X}_i)}\left[\mathbf{1}\left(Y_i \le q_{0|C}(\tau_k)\right) - \tau_k - \left(F_{Y|T,Z,\mathbf{X}}(q_{0|C}(\tau_k)|0,1,\mathbf{X}_i) - \tau_k\right)\right] \\ &+ \frac{(1-Z_i)(1-T_i)}{1-\pi(\mathbf{X}_i)}\left[\mathbf{1}\left(Y_i \le q_{0|C}(\tau_k) - \tau_k\right) - \left(F_{Y|T,Z,\mathbf{X}}(q_{0|C}(\tau_k)|0,0,\mathbf{X}_i) - \tau_k\right)\right] \\ &- \frac{Z_i(1-T_i) - P_{T|Z,\mathbf{X}}(0|1,\mathbf{X}_i)(Z_i - \pi(\mathbf{X}_i))}{\pi(\mathbf{X}_i)}\left(F_{Y|T,Z,\mathbf{X}}(q_{0|C}(\tau_k)|0,1,\mathbf{X}_i) - \tau_k\right) \\ &+ \frac{(1-Z_i)(1-T_i) + P_{T|Z,\mathbf{X}}(0|0,\mathbf{X}_i)(Z_i - \pi(\mathbf{X}_i))}{1-\pi(\mathbf{X}_i)}\left(F_{Y|T,Z,\mathbf{X}}(q_{0|C}(\tau_k)|0,0,\mathbf{X}_i) - \tau_k\right).\end{aligned}$$

Combining all the results yields

$$\begin{aligned}\sqrt{n}\left(\hat{m}_j^z(\tau_k) - m_j^z(\tau_k)\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\left(I_i(\tau_k) - m_j^z(\tau_k)\right)\mathbf{1}(Z_i = z, \mathbf{X}_i = \mathbf{x}_j)}{p_j^z}\right. \\ &\quad \left. - \sum_{t=0,1} f_{Y|T,Z,\mathbf{X}}(q_{t|C}(\tau_k)|t,z,\mathbf{x}_j)p_{T|Z,\mathbf{X}}(t|z,\mathbf{x}_j)\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_t(Y_i,T_i,Z_i,\mathbf{X}_i)}{P_c f_{t|C}(q_{t|C}(\tau_k))}\right] + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_j^z(\tau_k, Y_i, T_i, Z_i, \mathbf{X}_i) + o_p(1)\end{aligned}$$

The theorem is then proven by applying the Central Limit Theorem. ∎

Proof of Corollary 1

Proof. Under H_0, $\mathbf{m}^1 = \mathbf{m}^0$. So the test statistic

$$\begin{aligned}W &= n\left((\hat{\mathbf{m}}^1 - \hat{\mathbf{m}}^0) - (\mathbf{m}^1 - \mathbf{m}^0)\right)'\hat{\mathbf{V}}^{-1}\left((\hat{\mathbf{m}}^1 - \hat{\mathbf{m}}^0) - (\mathbf{m}^1 - \mathbf{m}^0)\right) \\ &\Rightarrow \chi^2(K(J-1))\end{aligned}$$

as $n \to \infty$. The convergence result follows from Theorem 2 and the fact the $\hat{\mathbf{V}}$ is a consistent estimator of $\mathbf{V}$. Therefore when the null is true,

$$P(\textit{reject the null}) = P(W > c_\alpha) \to \alpha.$$

Under the alternative, $\mathbf{m}^1 - \mathbf{m}^0 = A$, which is a $K(J-1) \times 1$ vector of constants that are not all zero. Then

$$W = n\left((\hat{\mathbf{m}}^1 - \hat{\mathbf{m}}^0) - (\mathbf{m}^1 - \mathbf{m}^0) + A\right)' \hat{\mathbf{V}}^{-1}\left((\hat{\mathbf{m}}^1 - \hat{\mathbf{m}}^0) - (\mathbf{m}^1 - \mathbf{m}^0) + A\right) \to \infty,$$

as $n \to \infty$. Therefore when the null is not true,

$$P(\textit{reject the null}) = P(W > c_\alpha) \to 1.$$

∎

Proof of Corollary 2

Proof. First consider the case with $t = 1$. For all $\tau \in (0,1)$,

$$\begin{aligned}
&\hat{\mathsf{R}}(y,1) = \frac{1}{S}\sum_{s=1}^{S} \mathbf{1}\left(\left(\hat{q}_{1|C}(\tau^s)\right) \leq y\right) \\
=&\frac{1}{S}\sum_{s=1}^{S}\left[\mathbf{1}\left(\left(q_{1|C}(\tau^s)\right) \leq y\right) + \mathbf{1}\left(\left(\hat{q}_{1|C}(\tau^s)\right) \leq y\right) - \mathbf{1}\left(\left(q_{1|C}(\tau^s)\right) \leq y\right)\right] \\
\leq&\frac{1}{S}\sum_{s=1}^{S}\mathbf{1}\left(\left(q_{1|C}(\tau^s)\right) \leq y\right) + \frac{1}{S}\sum_{s=1}^{S}\mathbf{1}\left(-\left|\hat{q}_{1|C}(\tau^s) - q_{1|C}(\tau^s)\right| \leq y \leq \left|\hat{q}_{1|C}(\tau^s) - q_{1|C}(\tau^s)\right|\right) \\
\leq&\frac{1}{S}\sum_{s=1}^{S}\mathbf{1}\left(\left(q_{1|C}(\tau^s)\right) \leq y\right) + \frac{1}{S}\sum_{s=1}^{S}\mathbf{1}\left(-|\epsilon| \leq y \leq |\epsilon|\right) \\
=&\int_0^1 \mathbf{1}\left(\left(q_{1|C}(\tau)\right) \leq y\right) d\tau + \int_0^1 \mathbf{1}\left(-|\epsilon| \leq y \leq |\epsilon|\right) d\tau + o_p(1), \\
=&R(y,1) + o_p(1),
\end{aligned}$$

where we define $\epsilon = \sup_{\tau\in(0,1)}\left|\hat{q}_{1|C}(\tau) - q_{1|C}(\tau)\right|$. The last equality holds as $\epsilon = o_p(1)$, following results in Hsu et al. (2016).

Similarly, we can show the convergence result for t=0. Therefore, $\hat{\mathsf{R}}(y,t) \xrightarrow{p} \mathsf{R}(y,t)$ as $S, n \to \infty$ for both $t = 0, 1$.

To show the weak convergence result stated in the corollary, notice that

$$
\begin{aligned}
&\bar{m}_j^z \equiv E[U|Z=z,\mathbf{X}=\mathbf{x}] = E\left[\int_0^1 \mathbf{1}\left(\left(Tq_{1|C}(\tau)+(1-T)q_{0|C}(\tau)\right)\leq Y\right)\Bigg|Z=z,\mathbf{X}=\mathbf{x}\right]\\
=&\int_0^1 \left(1-m_j^z(\tau)\right)d\tau,\\
&\ddot{m}_j^z = \frac{1}{n_j^z}\sum_{Z_i=z,\mathbf{X}_i=\mathbf{x}_j}\frac{1}{S}\sum_{s=1}^{S}\mathbf{1}\left(\left(T_i\hat{q}_{1|C}\left(\tau^s\right)+(1-T_i)\hat{q}_{0|C}\left(\tau^s\right)\right)\leq Y_i\right) = \frac{1}{S}\sum_{s=1}^{S}\left(1-\hat{m}_j^z(\tau^s)\right),\\
&\sqrt{n}\left(\ddot{m}_j^1-\ddot{m}_j^0\right) = \sqrt{n}\left(\frac{1}{S}\sum_{s=1}^{S}\left(1-\hat{m}_j^1(\tau^s)\right)-\left(1-\hat{m}_j^0(\tau^s)\right)\right) = \frac{1}{S}\sum_{s=1}^{S}\sqrt{n}\left(\hat{m}_j^0(\tau^s)-\hat{m}_j^1(\tau^s)\right)\\
=&\int_0^1 \sqrt{n}\left(\hat{m}_j^0(\tau)-\hat{m}_j^1(\tau)\right)d\tau + \frac{1}{S}\sum_{s=1}^{S}\sqrt{n}\left(\hat{m}_j^0(\tau^s)-\hat{m}_j^1(\tau^s)\right) - \int_0^1 \sqrt{n}\left(\hat{m}_j^0(\tau)-\hat{m}_j^1(\tau)\right)d\tau\\
=&I+II.
\end{aligned}
$$

Since under the null hypothesis, $\bar{m}_j^1 - \bar{m}_j^0 = 0$ for all $j = 1, ..., J-1$,

$$
\begin{aligned}
I &= \int_0^1 \sqrt{n}\left(\hat{m}_j^0(\tau)-\bar{m}_j^0(\tau)-\left(\hat{m}_j^1(\tau)-\bar{m}_j^1(\tau)\right)\right)d\tau\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^1 \left(\phi_j^0(\tau,Y_i,T_i,Z_i,\mathbf{X}_i)-\phi_j^1(\tau,Y_i,T_i,Z_i,\mathbf{X}_i)\right)d\tau + o_p(1).
\end{aligned}
$$

The second equality follows from the influence function representation in the proof of Theorem 2 and the dominanted convergence theorem.

Meanwhile, $II = o_p(1)$ because as $S, N \to \infty$,

$$
E(II) = E[I] \xrightarrow{p} 0, \quad Var(II) = \frac{1}{S}Var(I) \xrightarrow{p} 0.
$$

Applying the central limit theorem to I then proves the result of this Corollary. ∎

C Extensions

C.1 Many Discrete Covariates

This appendix discusses the case where J, the number of the unique values of $\mathbf{X}$, to go to infinity as the sample size n goes to infinity. This case is practically relevant when researchers are interested in adding more controls to $\mathbf{X}$ as the sample size increases. Notice that when $J \to \infty$ as $n \to \infty$, the asymptotic distribution in Theorem 2 no longer holds because the density $p_{Z,X}$ in the denominator goes to zero in the limit. This section provides a new test for that situation.

We make the following assumptions on the data.

Assumption A.1. *1) The data* $\{Y_i, T_i, Z_i, \mathbf{X}_i\}_{i=1}^n$ *is a random sample from* $(Y, T, Z, \mathbf{X})$*. 2) For all* $\tau \in \Omega = \{\tau_1, \tau_2, ..., \tau_K\}$ *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* $q_{t|C}(\tau)$*. 3) Let* $n_j = \sum_{i=1}^n \mathbf{1}(\mathbf{X} = \mathbf{x_j})$*.* $n_j \asymp n/J$ *uniformly over* j*, i.e. there exist* $0 < c \leq C < \infty$ *such that* $cn/J \leq n_j \leq Cn/J$ *for all* $j = 1, ..., J$*.* $\lim_{n\to\infty} J/n = 0$*. 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* $n, J \to \infty$*. 5) Let* $f_{Y|\mathbf{X}}(y|\mathbf{x}_j)$ *be the conditional density of* Y *given* $\mathbf{X}$*. For* $j = 1, ..., J$*,* $f_{Y|\mathbf{X}}(y|\mathbf{x}_j)$ *is positive and continuouly differentiable in a neighborhood of* $q_{t|C}(\tau)$*, for all* $t = 0, 1$ *and* $\tau \in \Omega$*.*

Recall that $n_j^z = \sum_{i=1}^n \mathbf{1}(Z_i = z, \mathbf{X}_i = \mathbf{x}_j)$ for $z = 0, 1$. The following corollary provides the asymptotic distribution of the estimator $\sqrt{n}\left(\hat{\mathsf{m}}_j^1 - \hat{\mathsf{m}}_j^0\right)$ for $j = 1, .., J-1$ when n, $J \to \infty$.

Corollary A.1. *Given Assumptions 1 and A.1 we have*

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

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

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

The corollary shows that when $J \to \infty$ as $n \to \infty$, the estimation error for the unconditional quantiles in the first stage, which is of order $\sqrt{n}$, is small enough relative to the estimation error in the second stage and hence can be ignored.

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

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

where $\hat{\mathbf{V}}_j$ is a consistent estimator of $\mathbf{V}_j$. Let $\hat{V}_{j;k,k'}$ be the (k,k')-th element of $\mathbf{V}_j$,

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

The test statistic under the null hypothesis of rank similarity can be constructed as

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

Given the rate condition stated in Assumption A.1.3 and results in de Jong and Bierens (1994), $W_{largeJ} \Rightarrow N(0,1)$ as $J \to \infty$ under the null. Let c_α be the $(1-\alpha) \times 100$-*th* percentile of the $N(0,1)$ distribution. The *one-sided* decision rule of the test is to "reject the null hypothesis H_0 if $W_{largeJ} > c_\alpha$". Note that the asymptotic theory of this test relies on having both J and the sample size at each value $\mathbf{X} = \mathbf{x}_j$ go to infinity, which might be a reasonable assumption when one has big data at hand.

Proof of Corollary A.1

Proof. Let $\tilde{m}_j^z(\tau_k) = E\left[\mathbf{1}\left(Y_i \le T_i \hat{q}_{1|C}(\tau_k) + (1-T_i)\,\hat{q}_{0|C}(\tau_k)\right)|Z_i = z, \mathbf{X}_i = \mathbf{x}_j\right]$, we have

$$\begin{aligned}\sqrt{n_j^z}\left(\hat{m}_j^z(\tau_k) - m_j^z(\tau_k)\right) &= \sqrt{n_j^z}\left(\hat{m}_j^z(\tau_k) - \tilde{m}_j^z(\tau_k)\right) + \sqrt{n_j^z}\left(\tilde{m}_j^z(\tau_k) - m_j^z(\tau_k)\right) \\ &= I + II.\end{aligned}$$

Similar to the arguments in the proof of Theorem 2, by the consistency of $\hat{q}_{1|C}(\tau_k)$ and $\hat{q}_{0|C}(\tau_k)$ and by Lemma 19.24 of Van der Vaart (1998), we have that for all $z = 0, 1$, $k = 1, ..., K$, and $j = 1, ..., J-1$,

$$I = \frac{1}{\sqrt{n_j^z}} \sum_{Z_i = z, \mathbf{X}_i = \mathbf{x}_j} \left(I_i(\tau_k) - m_j^z(\tau_k)\right) + o_p(1).$$

Meanwhile, for all $z = 0, 1$, $k = 1, ..., K$, and $j = 1, ..., J-1$,

$$\begin{aligned}II =& \sqrt{n_j^z} E\left[\mathbf{1}\left(Y_i \le T_i \hat{q}_{1|C}(\tau_k) + (1-T_i)\,\hat{q}_{0|C}(\tau_k)\right)|Z_i = z, \mathbf{X}_i = \mathbf{x}_j)\right] \\ &- \sqrt{n_j^z} E\left[\mathbf{1}\left(Y_i \le T_i q_{1|C}(\tau_k) + (1-T_i)\,q_{0|C}(\tau_k)\right)|Z_i = z, \mathbf{X}_i = \mathbf{x}_j)\right] \\ =& \sqrt{n_j^z/n} f_{Y|T,Z,\mathbf{X}}(\bar{q}_1^k|1, z, \mathbf{x}_j) P_{T|Z,\mathbf{X}_j}(z, \mathbf{x}_j) \sqrt{n}\left(\hat{q}_{1|C}(\tau_k) - q_{1|C}(\tau_k)\right) \\ &+ \sqrt{n_j^z/n} f_{Y|T,Z,\mathbf{X}}(\bar{q}_0^k|0, z, \mathbf{x}_j)\left(1 - P_{T|Z,\mathbf{X}_j}(z, \mathbf{x}_j)\right)\sqrt{n}\left(\hat{q}_{0|C}(\tau_k) - q_{0|C}(\tau_k)\right) \\ =& o_p(1),\end{aligned}$$

where $\bar{q}_t^k$ is a value between $\hat{q}_{t|C}(\tau_k)$ and $q_{t|C}(\tau_k)$ for both $t = 0, 1$. The second equality follows from the Mean Value Theorem, and the last equality follows from the boundedness of $f_{Y|T,Z,\mathbf{X}}(.|t, z, \mathbf{x}_j)$ in a neighborhood of $q_{t|C}(\tau_k)$ and the weak convergence result of $\hat{q}_{0|C}(\tau_k)$ shown in Frolich and Melly (2013).

Now, let $\lambda_{n,j} = n_j^0 / \left(n_j^1 + n_j^0\right)$. Then

$$\begin{aligned}
&\sqrt{\frac{n_j^1 n_j^0}{n_j^1 + n_j^0}} \left\{\hat{m}_j^1(\tau_k) - \hat{m}_j^0(\tau_k) - \left(m_j^1(\tau_k) - m_j^0(\tau_k)\right)\right\} \\
=&\sqrt{\lambda_{n,j}} \frac{1}{\sqrt{n_j^1}} \sum_{Z_i=1,\mathbf{X}_i=\mathbf{x}_j} \left(I_i(\tau_k) - m_j^1(\tau_k)\right) - \sqrt{1-\lambda_{n,j}} \frac{1}{\sqrt{n_j^0}} \sum_{Z_i=0,\mathbf{X}_i=\mathbf{x}_j} \left(I_i(\tau_k) - m_j^0(\tau_k)\right) \\
&+ o_p(1).
\end{aligned}$$

For any j, compiling all K quantiles together gives

$$\begin{aligned}
&\sqrt{\frac{n_j^1 n_j^0}{n_j^1 + n_j^0}} \left\{\hat{\mathsf{m}}_j^1 - \hat{\mathsf{m}}_j^0 - \left(\mathsf{m}_j^1 - \mathsf{m}_j^0\right)\right\} \\
&\Rightarrow \sqrt{1-\pi(\mathbf{x}_j)} N_1(0, \mathbf{V}_1) + \sqrt{\pi(\mathbf{x}_j)} N_0(0, \mathbf{V}_0),
\end{aligned}$$

where N_1 and N_0 are two independent normally distributed random variable because the data are i.i.d. and the variance-covariance matrix $\mathbf{V}_1$ and $\mathbf{V}_2$ have (k, k')-th element equal to $m_j^1(\tau_k) - m_j^1(\tau_k) m_j^1(\tau_{k'})$ and $m_j^0(\tau_k) - m_j^0(\tau_k) m_j^0(\tau_{k'})$, respectively. Then we know that

$$\begin{aligned}
&\sqrt{\frac{n_j^1 n_j^0}{n_j^1 + n_j^0}} \left\{\hat{\mathsf{m}}_j^1 - \hat{\mathsf{m}}_j^0 - \left(\mathsf{m}_j^1 - \mathsf{m}_j^0\right)\right\} \\
&\Rightarrow N(0, \mathbf{V}_j),
\end{aligned}$$

where the (k, k')-th element of $\mathbf{V}_j$ equal to

$$(1-\pi(\mathbf{x}_j)) \left(m_j^1(\tau_k \wedge \tau_{k'}) - m_j^1(\tau_k) m_j^1(\tau_{k'})\right) + \pi(\mathbf{x}_j) \left(m_j^0(\tau_k \wedge \tau_{k'}) - m_j^0(\tau_k) m_j^0(\tau_{k'})\right).$$

∎

C.2 Continuous Covariates

This appendix considers the case where covariates $\mathbf{X}$ are continuous. Let $\mathbf{X} = (X_1 ... X_L)$ be a L dimensional vector of continuous variables. For any $\mathbf{x} \in \mathcal{X}$ and $z = 0, 1$, define $m_k^z(\mathbf{x}) = E[I(\tau_k)|Z = z, \mathbf{X} = \mathbf{x}]$. Let $m_k(.) = m_k^1(.) - m_k^0(.)$, we are interested in testing the following null hypothesis.

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

The following discusses estimation of $m_k(\mathbf{x})$ as well as a Kolmogorov-Smirnov type test statistic. Similar as before let $\widehat{q}_{0|C}(\tau_k)$ and $\widehat{q}_{1|C}(\tau_k)$ be $\sqrt{N}$-consistent estimators of $q_{0|C}(\tau_k)$ and $q_{1|C}(\tau_k)$, respectively. With product kernel functions $\mathcal{K}_{h_z}(\mathbf{X}_i - \mathbf{x}) = \frac{1}{h_z}\mathcal{K}\left(\frac{\mathbf{X}_i - \mathbf{x}}{h_z}\right)$ for $z = 0, 1$, and bandwidths $h_0, h_1 \to 0$, we can define the local linear estimators for $\hat{m}_k^0(\mathbf{x})$ and $\hat{m}_k^1(\mathbf{x})$ as the intercepts a_0 and a_1 in the following minimization problems

$$\min_{a_0, b_{01}, ... b_{0L}} \sum_{Z_i=0} \left[\mathbf{1}\left(Y_i \le \hat{q}_{1|C}(\tau_k)T_i + \hat{q}_{0|C}(\tau_k)(1 - T_i)\right) - a_0 - \sum_{l=1}^{L} b_{0l}(X_{i,l} - x_l)\right]^2 \mathcal{K}_{h_0}(\mathbf{X}_i - \mathbf{x}),$$
$$\min_{a_1, b_{11}, ... b_{1L}} \sum_{Z_i=1} \left[\mathbf{1}\left(Y_i \le \hat{q}_{1|C}(\tau_k)T_i + \hat{q}_{0|C}(\tau_k)(1 - T_i)\right) - a_1 - \sum_{l=1}^{L} b_{1l}(X_{i,l} - x_l)\right]^2 \mathcal{K}_{h_1}(\mathbf{X}_i - \mathbf{x}).$$

We can then estimate $m_k(\mathbf{x})$ by $\hat{m}_k(\mathbf{x}) = \hat{m}_k^0(\mathbf{x}) - \hat{m}_k^1(\mathbf{x})$. Let $s_k(\mathbf{x})$ be the standard error of $\hat{m}_k(\mathbf{x})$ for all $k = 1, ..., K$, which can be estimated using the asymptotic formula (Fan and Gijbels, 1996) or bootstrap. The test statistic can then be defined as

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

With a significance level α, the null hypothesis is rejected if $KS > c_\alpha$, where c_α is the critical value that satisfies

$$\lim_{n\to\infty} P\left(\sup_{k,\mathbf{x}} \left|\frac{\hat{m}_k(\mathbf{x}) - m_k(\mathbf{x})}{s_k(\mathbf{x})}\right| > c_\alpha\right) \le \alpha. \tag{A.1}$$

Assumption A.2. *1) The data* $\{Y_i, T_i, Z_i, \mathbf{X}_i\}_{i=1}^n$ *is a random sample 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)* $\mathbf{X}|Z = z$ *has a conditional density that is bounded away from zero on convex* $\mathcal{X}$ *for both* $z = 0, 1$. $\hat{\pi}(\mathbf{x})$ *is uniformly consistent, i.e.,* $\sup_{\mathbf{x}\in\mathcal{X}} |\hat{\pi}(\mathbf{x}) - \pi(\mathbf{x})| \xrightarrow{p} 0$*; 4)* $I_i(\tau_k) - m_k^z(\mathbf{x})|Z = z, \mathbf{X} = \mathbf{x}$ *has a conditional density that is bounded from above and from below away from zero uniformly over* $\mathbf{x} \in \mathcal{X}$*,* $z \in \{0, 1\}$ *and* $\tau_k \in \Omega$*; 5) for all* $k = 1, ..., K$*,* $m_k(\mathbf{x})$ *is twice continuously differentiable. Its first derivative is bounded uniformly over* $\mathbf{x} \in \mathcal{X}$ *and* $\tau_k \in \Omega$*; the kernel* K *has compact support and two continuous derivatives, and satisfies* $\int uK(u)\,du = 0$ *and* $\int K(u)\,du = 1$*; 6) the bandwidths,* h_0 *and* h_1*, satisfy that* $nh_0^{L+2} \to \infty$*,* $nh_0^{L+2} \to \infty$*,* $nh_0^{L+4} \to 0$*, and* $nh_0^{L+4} \to 0$ *as* $n \to \infty$*.*

Let $\hat{\epsilon}_{k,i} = I_i(\tau_k) - \hat{m}_k^1(\mathbf{x}_i)Z_i - \hat{m}_k^0(\mathbf{x}_i)(1 - Z_i)$ and $\hat{m}_k^*(\mathbf{x})$ be a multiplier process such that

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

with $\{\eta_i\}_{i=1}^n$ simulated from i.i.d. $N(0, 1)$, independent of data. Let c_α as the $(1 - \alpha) \times 100$-*th* percentile of the simulated process $\sup_{k,\mathbf{x}} \left|\frac{\hat{m}_k^*(\mathbf{x})}{s_k(\mathbf{x})}\right|$. Given Assumptions 1 and A.2, the multiplier bootstrap critical value c_α defined above satisfies the condition required in equation (A.1). See the arguments in Example 7 of Chernozhukov et al. (2011). Details are omitted for the interest of space. Consistency of this test then follows.

C.3 Testing Conditional Ranks

This appendix briefly discusses testing ranks in the conditional distribution of potential outcomes. In this case, additional covariates are needed for testing, other than those included in the conditioning set of conditional distribution of potential outcomes. Therefore, this test is feasible only when this conditioning set is small or one has a large set of covariates.

Let $\mathbf{X}_1$ be the covariates in the conditioning set of the conditional distribution of potential outcomes, and $\mathbf{X}_2$ be the additional covariates one use for testing. Let $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$. Denote the support of $\mathbf{X}_1$ as $\mathcal{X}_1$ and that of $\mathbf{X}_2$ as $\mathcal{X}_2$. For $t = 0, 1$, and $\mathbf{x}_1 \in \mathcal{X}_1$, further let the conditional distribution of potential outcome be $F_{t|C,\mathbf{X}_1}(y|\mathbf{x}_1) = \Pr(Y_t \leq y | T_1 > T_0, \mathbf{X}_1 = \mathbf{x}_1)$. Then define the conditional potential rank conditional on $\mathbf{X}_1$ among compliers as $U_{t|C,\mathbf{X}_1=\mathbf{x}_1} = F_{t|C,\mathbf{X}_1}(Y_t|\mathbf{x}_1)$, $t = 0, 1$. Rank similarity given $\mathbf{X}_1 = \mathbf{x}_1$ in this case implies

$$F_{U_{0|C,\mathbf{X}_1=\mathbf{x}_1}|C,\mathbf{X}_1=\mathbf{x}_1,\mathbf{X}_2=\mathbf{x}_2}(\tau|\mathbf{x}_1,\mathbf{x}_2) = F_{U_{1|C,\mathbf{X}_1=\mathbf{x}_1}|C,\mathbf{X}_1=\mathbf{x}_1,\mathbf{X}_2=\mathbf{x}_2}(\tau|\mathbf{x}_1,\mathbf{x}_2) \quad \text{(A.2)}$$

for all $\tau \in (0, 1)$ and $\mathbf{x}_2 \in \widetilde{\mathcal{X}}_2$, where $\widetilde{\mathcal{X}}_2$ is the support of $\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1$.

Let the conditional quantile function be $q_{t|C,\mathbf{X_1}}(\tau|\mathbf{x}_1) = F^{-1}_{t|C,\mathbf{X}_1}(\tau|\mathbf{x}_1)$ for $t = 0, 1$ and $\tau \in (0, 1)$. Assume that Parts 2), 3) and 4) of Assumption 1 hold, and that Part 1) of Assumption 1 holds conditioning on $\mathbf{X}_1$ or conditioning on $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$. Following Frolich and Melly (2013), one can identify $q_{t|C,\mathbf{X_1}}(\tau|\mathbf{x}_1)$ for $t = 0, 1$ by

$$\left(q_{0|C,\mathbf{X_1}}(\tau|\mathbf{x}_1), q_{1|C,\mathbf{X_1}}(\tau|\mathbf{x}_1)\right) = \arg\min_{q_0,q_1} E\left[\rho_\tau(Y - q_0(1-T) - q_1 T)\omega^{FM}|\mathbf{X}_1 = \mathbf{x}_1\right], \quad \text{(A.3)}$$

where $\omega^{FM} = \left(\frac{Z}{\pi(\mathbf{X})} - \frac{1-Z}{1-\pi(\mathbf{X})}\right)(2T-1)$ and $\pi(\mathbf{X}) \equiv E[Z|\mathbf{X}]$ is the instrument probability.

Define the conditional rank indicator as

$$\tilde{I}(\tau|\mathbf{x}_1) \equiv \mathbf{1}\left(Y \leq \left(T q_{0|C,\mathbf{X_1}}(\tau|\mathbf{x}_1) + (1-T)\, q_{1|C,\mathbf{X_1}}(\tau|\mathbf{x}_1)\right)\right).$$

Analogous to Theorem 1, equation (A.2) holds if and only if for all $\tau \in (0, 1)$ and $\mathbf{x_2} \in \widetilde{\mathcal{X}}_2$,

$$E\left[\tilde{I}(\tau|\mathbf{x}_1)|Z = 1, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2\right] = E\left[\tilde{I}(\tau, \mathbf{x}_1)|Z = 0, \mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2\right]. \quad \text{(A.4)}$$

One can then test rank similarity in this case as follows. First estimate $q_{t|C,\mathbf{X_1}}(\tau|\mathbf{x}_1)$ using equation (A.3), and then test whether equation (A.4) holds for all $\tau \in (0, 1)$ and $\mathbf{x_2} \in \widetilde{\mathcal{X}}_2$, replacing $q_{t|C,\mathbf{X_1}}(\tau|\mathbf{x}_1)$ $t = 0, 1$ with their estimates.

Alternatively, assume a linear model for conditional quantiles $\widetilde{q}_{t|C}(\tau)$, $t = 0, 1$, conditional on $\mathbf{X}_1$. Let Part 1) of Assumption 1 hold conditioning on $\mathbf{X}_1$. Following Abadie et al. (2002), one can identify the conditional quantiles as follows

$$\left(\widetilde{q}_{0|C}\left(\tau\right), \widetilde{q}_{1|C}\left(\tau\right)\right) = \arg\min_{q_0, q_1} E\left[\rho_\tau(Y_i - q_0(1-T) - q_1 T - \mathbf{X}_1'\gamma)\omega^{AAI}\right], \tag{A.5}$$

where $\omega^{AAI} = 1 - \frac{T(1-Z)}{\pi(\mathbf{X}_1)} - \frac{(1-T)Z}{1-\pi(\mathbf{X}_1)}$ and $\pi\left(\mathbf{X}_1\right) = E\left[Z|\mathbf{X}_1\right]$.[1]

The conditional rank indicator can then be defined as

$$\tilde{I}(\tau) \equiv \mathbf{1}\left(Y \leq \left(T\widetilde{q}_{0|C}\left(\tau\right) + \left(1-T\right)\widetilde{q}_{1|C}\left(\tau\right)\right)\right).$$

Analogous to Theorem 1, rank similarity implies that for all $\tau \in (0,1)$ and $\mathbf{x}_2 \in \mathcal{X}_2$,

$$E\left[\tilde{I}(\tau)|Z=1, \mathbf{X}_2 = \mathbf{x}_2\right] = E\left[\tilde{I}\left(\tau\right)|Z=0, \mathbf{X}_2 = \mathbf{x}_2\right]. \tag{A.6}$$

To test rank similarity for the conditional distribution of potential outcomes conditional on $\mathbf{X}_1$, one can first estimate conditional quantiles $\widetilde{q}_{t|C}\left(\tau\right)$, $t = 0, 1$ using equation (A.5), and then test whether equation (A.6) holds for all $\tau \in (0,1)$ and $\mathbf{x_2} \in \mathcal{X}_2$, replacing $\widetilde{q}_{t|C}\left(\tau\right)$ $t = 0, 1$ with their estimates.

[1]To ensure nonnegativity, ω^{FM} and ω^{AAI} can be replaced with their projections onto Y, T and $\mathbf{X}_1$.

References

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

Chernozhukov, V., S. Lee, and A. Rosen (2011): "Intersection Bounds: Estimation and Inference," *Working Paper*.

de Jong, R. M. and H. J. Bierens (1994): "On the Limit Behavior of a Chi-Square Type Test If the Number of Conditional Moments Tested Approaches Infinity," *Econometric Theory*, 10(1), 70–90.

Frolich, M. and B. Melly (2013): "Unconditional Quantile Treatment Effects Under Endogeneity," *Journal of Business & Economic Statistics*, 31, 346–357.

Hsu, Y.-C., R. P. Lieli, and T.-C. Lai (2016): "Estimation and Inference for Distribution Functions and Quantile Functions in Endogenous Treatment Effect Models," *Working paper*.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers (New Series)*, 3, 135–146.

Van der Vaart, A. W. (1998): *Asymptotic Statistics*, Cambridge University Press.