Australian National University

Crawford School of Public Policy

TTPI

Tax and Transfer Policy Institute

Quantile treatment effects revisited: Uncovering the distributional consequences of a welfare experiment

TTPI - Working Paper 18/2023
November 2023

Mathias Sinning
Tax and Transfer Policy Institute
Crawford School of Public Policy, Australian National University

Abstract

Heterogeneous effects of welfare reforms on earnings, transfers, and income have been established theoretically and empirically. Evaluation studies often focus on quantile treatment effects (QTE), which rely on the marginal distributions of potential treatment and control outcomes. Parameters that depend on the joint distribution of potential outcomes, such as quantiles of the distribution of treatment effects (QDTE), receive less attention. We propose a strategy to identify these parameters. We leverage the property that, under random assignment, rank correlation coefficients between actual treatment and predicted control state outcomes are identical, irrespective of whether predictions are based on treatment or control units. To identify QDTE, we assume that all permutations of observation units satisfying this property are equally likely. Rearranging quantiles yields a generalized version of quantile treatment effects (GQTE). We employ a reweighting approach for identification under strong ignorability. We test the predictor strength and demonstrate that highly predictive covariates yield unbiased, consistent, and asymptotically normal estimators. Our analysis of Connecticut's Jobs First program reveals initial income increases for a larger fraction of participants than previously recognized. Long-term gains were at least twice as large as those derived from conventional QTE and concentrated at the lower end of the distribution.

JEL Codes: C13, C21, D04, I38

Keywords: Treatment effect heterogeneity, distribution of treatment effects, rank correlation, welfare experiments

** I thank Josh Angrist, Colin Cameron, Ashley Craig, Christian Gillitzer, Tue Gørgens, James Heckman, Kosuke Imai, Gary King, Myoung-Jae Lee, Seojeong Lee, Thomas Lemieux, Sangsoo Park, Jeffrey Smith and Xiang Zhou for helpful conversations and comments on earlier drafts. I also thank participants of the Applied Statistics Seminar at the Harvard University Institute of Quantitative Social Science, the Vancouver School of Economics Seminar at the University of British Columbia, and the 2019 Asia and China Meetings of the Econometric Society. This work was supported by the Australian Research Council under Grants DP150104247, LP160100810, and LP220100291. I report there are no competing interests to declare. All correspondence to Mathias Sinning, E-mail: mathias.sinning@anu.edu.au*

Tax and Transfer Policy Institute

Crawford School of Public Policy

College of Asia and the Pacific

+61 2 6125 9318

tax.policy@anu.edu.au

The Australian National University

Canberra ACT 0200 Australia

www.anu.edu.au

The Tax and Transfer Policy Institute (TTPI) is an independent policy institute that was established in 2013 with seed funding from the federal government. It is supported by the Crawford School of Public Policy of the Australian National University.

TTPI contributes to public policy by improving understanding, building the evidence base, and promoting the study, discussion and debate of the economic and social impacts of the tax and transfer system.

The Crawford School of Public Policy is the Australian National University's public policy school, serving and influencing Australia, Asia and the Pacific through advanced policy research, graduate and executive education, and policy impact.

1 Introduction

Understanding the distributional consequences of welfare reforms is important for the design of tax and transfer policies. Transfer programs, which aim to provide social security and enhance equity, can affect labor supply and earnings (Kline and Tartari, 2016). Some design features of these programs, such as a generous earnings disregard, can lead to increased earnings for certain segments of the distribution while simultaneously reducing earnings for other segments (Bitler et al., 2006). A key challenge when studying the effectiveness of transfer programs is that even ideal experiments can only be used to identify features of the distribution of treatment effects that depend on separate marginal distributions of treatment and control outcomes, such as the average treatment effect (ATE) or quantile treatment effects (QTE).[1]

Bitler et al. (2006) analyze data from Connecticut's Jobs First welfare experiment and find that QTE estimates display heterogeneous treatment responses consistent with predictions from labor supply theory. Using the same data source, Bitler et al. (2017) formally test and reject the hypothesis that subgroup-specific ATE can explain important features of the treatment effect heterogeneity revealed by QTE. However, the authors emphasize that QTE are different from quantiles of the distribution of treatment effects (QDTE) if the rank invariance assumption is violated. Bitler et al. (2005) provide evidence of rank reversals in a related study based on Canadian data, suggesting that their QTE estimates may not be interpreted as QDTE.[2]

This paper proposes an approach to identify QDTE. The approach is used to explore the distributional consequences of Connecticut's Jobs First welfare experiment on women's earnings, transfers, and income. We begin by considering rank correlation coefficients (Kendall, 1938) between actual treatment outcomes and predicted control

[1]Experiments are informative about ATE without any further identifying assumptions. In contrast, a rank invariance assumption is required to identify QTE as defined by Doksum (1974) and Lehmann (1974). Rank invariance implies that observation units maintain their relative position in the potential outcome distribution regardless of whether they are being assigned to the treatment group or the control group. Although rank invariance is a strong assumption, QTE estimation has become a standard approach in the treatment effects literature. Important methodological contributions include Chernozhukov and Hansen (2005), Firpo (2007) and Frölich and Melly (2013).

[2]QDTE are only identical to QTE if the rank invariance assumption holds and if QTE are monotonically increasing along the distribution. In general, the interpretation of QTE is quite different from that of QDTE (Lee, 2006). For example, consider a welfare reform that aims to increase individual income. In this context, a positive median treatment effect implies that the income of the median person has increased. In contrast, a positive median of the distribution of treatment effects indicates that the incomes of at least 50 percent of the study population have increased.

state outcomes in a scenario where treatment and control group sizes are equal. Under random assignment, we obtain identical rank correlation coefficients, regardless of whether we use treatment or control units to calculate predicted outcomes. We refer to this property as 'Rank Correlation Property' (RCP). We show that the rank correlation coefficient between potential treatment and control outcomes is identified under the assumption that all permutations of observation units satisfying the RCP are equally likely. We use the same assumption to identify QDTE as quantiles of average unit-level treatment effects derived from permutations of observation units satisfying the RCP. Rearranging QDTE yields a generalized version of quantile treatment treatment effects (GQTE), which do not require a rank invariance assumption but align with conventional QTE under rank invariance.[3]

Our identification strategy rests on modest assumptions because we focus on situations in which relevant control outcome predictors are observed. Many experimental studies collect pre-treatment outcomes along with other highly predictive covariates. We propose a simulation-based test to assess the strength of model predictors, measured by the rank correlation coefficient between actual and predicted control outcomes. We use simulated data to estimate specific target values of the rank correlation coefficient between potential treatment and control outcomes. We test whether the estimated rank correlation coefficients deviate significantly from these target values. We also employ a Kolmogorov-Smirnov test to ascertain whether deviations of estimated rank correlation coefficients from target values have an impact on our GQTE estimates. While we cannot test whether all permutations of observation units satisfying the RCP are equally likely, we demonstrate that the distribution of permutation-specific GQTE is approximately normal for each quantile, strongly supporting the use of averages associated with the assumption of equally likely permutations.

We extend our approach to identify our parameters of interest under strong ignorability. We employ a reweighting approach akin to Firpo (2007) to obtain reweighted versions of actual and predicted outcomes. We use these reweighted outcomes to specify a 'conditional RCP'. The identification of QDTE under strong ignorability is based

[3]The distribution of treatment effects also serves as a basis for deriving various other parameters (Heckman et al., 1997; Heckman and Vytlacil, 2007). In our empirical application, we will determine the proportion of women who benefit from Jobs First and calculate the total gains and losses resulting from the program.

on the assumption that all permutations of observation units satisfying the conditional RCP are equally likely. Finally, we extend our approach to situations in which treatment and control group sizes are unequal. Monte Carlo simulations confirm that our estimators are unbiased, consistent, and asymptotically normal if relevant predictors are available. We also find that estimated bootstrap standard errors are a good representation of the true sampling variation.

Following Bitler et al. (2006), we estimate separate effects for the periods preceding and following the implementation of time limits, which led to the exclusion of Jobs First women from welfare benefits. During the pre-time limit period (quarters 1-7 after the initiation of Jobs First), our estimated rank correlation coefficients between potential treatment and control outcomes range from approximately 0.7 to 0.8. During the post-time limit period (quarters 8-16), this range shifts to approximately 0.6 to 0.7 due to mobility over time. We test the relevance of our predictors and find that they are sufficiently strong to produce unbiased results.

Our findings confirm substantial heterogeneity in response to the Jobs First welfare experiment. Our GQTE estimates exhibit qualitative similarity to conventional QTE estimates during the pre-time limit period. In contrast, when we account for rank changes during the post-time limit period, we uncover both positive effects at the lower end and negative effects at the higher end of the earnings distribution. Negative effects at the higher end of the earnings distribution suggest that some women in the Jobs First group were either unable or unwilling to increase their earnings, even though incentives to reduce earnings to receive higher welfare payments disappeared with the implementation of time limits. Instead, our findings indicate that the initially observed negative effects of Jobs First on the earnings distribution were more persistent than previously recognized. Regarding income, we find that women with lower incomes are most responsive to reductions in transfers following the implementation of time limits. We also observe negative income effects at the higher end of the distribution, which align with the corresponding negative earnings effects.

We use QTE and GQTE estimates to assess the proportions of positive and negative treatment effects, revealing substantial disparities between the two methods. During the pre-time limit period, QTE estimates yield a lower proportion of positive income effects (85.9%) than GQTE estimates (99.0%). When focussing on effects that reach statistical

significance at a 0.1% level, the proportion of positive income effects based on QTE estimates drops to 52.5%, whereas the proportion based on GQTE estimates remains higher at 71.7%. During the post-time limit period, QTE and GQTE estimates exhibit similar proportions of statistically significant income effects, although the locations of these effects differ substantially between the two methods.

We also use our estimates to quantify the total gains and losses emanating from the Jobs First program. We find that the total gains and losses during the post-time limit period are preliminary driven by the impact of Jobs First on the earnings distribution. Moreover, our findings indicate that the increase in earnings successfully offset the majority of losses in transfers stemming from the implementation of time limits. We observe total gains in income of approximately $190,000 based on QTE estimates and around $383,000 based on GQTE estimates. When we apply a significance level of 0.1%, these gains decrease to approximately $43,000 based on QTE estimates, whereas the corresponding figure based on GQTE estimates is approximately $260,000, more than six times higher.

The remainder of this paper is structured as follows: Section 2 provides a brief overview of related literature. Section 3 introduces some basic ideas and defines our parameters of interest. Identification is covered in Section 4, followed by estimation in Section 5. The results of our empirical application are presented in Section 6. Section 7 concludes.

2 Related literature

The study of heterogeneous treatment effects has traditionally relied on estimating ATE within subgroups that share the same observed characteristics (Crump et al., 2008; Lee and Shaikh, 2013; List et al., 2019). In recent years, more adaptable approaches for estimating heterogeneous treatment effects have emerged, often leveraging machine learning techniques. Examples include Su et al. (2009), Hill (2011), Imai and Ratkovic (2013), Athey and Imbens (2016), Shalit et al. (2017), Chernozhukov et al. (2018), Wager and Athey (2018), Powers et al. (2018), Künzel et al. (2019), Hahn et al. (2020), and Nie and Wager (2021). While subgroup analysis can yield valuable insights, it may overlook critical aspects of treatment effect heterogeneity when essential information

is unobserved. Most importantly, subgroup analysis cannot be used to determine the distribution of treatment effects and the parameters derived from it. Our approach addresses this issue.

Our work builds on the growing literature on the distribution of treatment effects. A major strand of this literature focuses on the identification of features of the distribution of treatment effects without making any identifying assumptions. Because point identification of features beyond the average cannot be achieved without further assumptions, this literature uses partial identification and proposes methods to bound features of the distribution of treatment effects (Makarov, 1982; Rüschendorf, 1982; Frank et al., 1987; Williamson and Downs, 1990; Heckman et al., 1997; Fan and Wu, 2010; Fan and Park, 2010; Firpo and Ridder, 2019; Russel, 2021). Unfortunately, the bounds arising from this literature are often too wide to be informative.

Another strand imposes restrictions to tighten the bounds. Manski (1997) establishes bounds under the assumption of monotone treatment response. Heckman et al. (1997) explore bounds within the framework of a Roy model. Kim (2018) proposes a generalization of these restrictions by considering support restrictions on potential outcomes. Using the same data as in our analysis, Kline and Tartari (2016) derive bounds from restrictions on revealed preferences. Frandsen and Lefgren (2021) assume that potential outcomes are mutually stochastically increasing. Lee (2022) imposes stochastic dominance assumptions to bound the conditional distribution of treatment effects. While restrictions can substantially tighten the bounds on the distribution of treatment effects, there is no guarantee that these bounds will necessarily provide useful insights in empirical applications. Moreover, restrictions are often implausible in real-world settings.[4]

Our approach shares a conceptual link with Frandsen and Lefgren (2021), who consider the correlation between predicted treatment and control outcomes to draw inferences about the correlation between potential outcomes. We show that the rank correlation coefficient between potential outcomes is identified if highly predictive covariates are available and if all permutations of observation units satisfying the RCP are

[4]For example, Kline and Tartari (2016) refer to instances where decision makers have been found violating revealed preference restrictions. Frandsen and Lefgren (2021) advise against applying their approach if, for instance, unobserved characteristics are expected to be advantageous in the control state but detrimental in the treatment state.

equally likely. There is also a connection between our approach and the work of Fan and Park (2009), who show that knowledge of the specific value of Kendall's (1938) rank correlation coefficient between potential treatment and control outcomes does not improve the bounds on the distribution of treatment effects if the rank correlation coefficient is positive. We do not consider bounds on the distribution of treatment effects because we observe that the distribution of permutation-specific GQTE is approximately normal for each quantile. This finding strongly supports our focus on averages associated with equally likely permutations, rather than on bounds linked to permutations of observation units with near-zero probability of occurrence.

Our approach is more directly related to the work of Heckman et al. (1997), who derive features of the 'impact distribution' assuming knowledge of the rank correlation coefficient between potential outcomes.[5] Heckman et al. (1997) pair percentiles of the marginal distributions of treatment and control outcomes to estimate the mean of features of the impact distribution. They use random permutations of percentiles that produce a given rank correlation coefficient and consider each permutation as equally likely. We employ a similar approach by using random permutations of observation units satisfying the RCP. While Heckman et al. (1997) obtain average features of the impact distribution from permutations of percentiles, we obtain features of average unit-level treatment effects from permutations of observation units.[6]

3 Basic ideas and parameters of interest

Before defining our parameters of interest, we introduce some notation, drawing upon the potential outcomes framework introduced by Rubin (1974). Following Imbens and Rubin (2015), we consider a super-population of size N_{sp}, where N_{sp} is large but countable. Throughout the paper, we use the subscript 'sp' to refer to the super-population. For each unit i in the super-population, $i \in \{1, \ldots, N_{\text{sp}}\}$, we define the potential out-

[5]Several other studies impose restrictions on the dependence between potential outcomes and unobservables to identify the joint outcome distribution. Abbring and Heckman (2007) provide an overview of this literature.

[6]We illustrate in Section 6.2 that the impact distribution proposed by Heckman et al. (1997) captures the full amount of heterogeneity in the data by averaging over permutation-specific treatment effects that were arranged in ascending order, regardless of the location of control outcomes. In contrast, our identification strategy yields a distribution of treatment effects that is made up of average unit-level treatment effects associated with unit-level control outcomes.

comes $Y_i(1)$ and $Y_i(0)$, which are viewed as fixed. We define a treatment indicator, denoted as D_i, which takes on the value 1 if unit i is assigned to the treatment group, and 0 otherwise. For each unit i, only the potential outcome corresponding to their actual treatment received can be observed: $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$. We also consider a k-dimensional set of covariates X_i.

We present two commonly used assumptions regarding the assignment of units to treatment and control groups. One assumption asserts that units are assigned randomly:

ASSUMPTION 1 (Random Assignment): $Y_i(1), Y_i(0) \perp\!\!\!\perp D_i$.

Random assignment ensures that any differences between treatment and control outcomes are solely attributable to the treatment. Another commonly used assumption is the 'strong ignorability' assumption introduced by Rosenbaum and Rubin (1983). Under strong ignorability, the assignment of observation units to treatment and control groups is viewed as random (or as good as random), conditional on a set of covariates. The strong ignorability assumption can be stated as follows:

ASSUMPTION 2 (Strong Ignorability):
(a) (Unconfoundedness): $Y_i(1), Y_i(0) \perp\!\!\!\perp D_i | X_i$.
(b) (Common Support): $0 < Pr(D_i = 1|X_i = x) < 1$.

Assumption 2(a) is commonly referred to as 'unconfoundedness' (Rubin, 1990) because it implies that comparing units that share the same characteristics is sufficient to eliminate all confounding.[7,8] There are many situations in which this assumption is violated because selection is based on unobservables.[9] Assumption 2(b) requires the presence of both treated and untreated units for each value of $X = x$.

[7] Assumption 2(a) is also known as 'ignorability' (Rosenbaum and Rubin, 1983), 'selection on observables' (Heckman and Robb, 1985) and 'conditional independence assumption' (Lechner, 1999).

[8] Many studies assume unconfoundedness to identify average treatment effects although their identification only requires a conditional mean independence assumption (i.e., $\mathbb{E}[Y_i(d)|X_i, D_i] = \mathbb{E}[Y_i(d)|X_i]$, with $d = \{0, 1\}$), which is considerably weaker than uncounfoundedness (Heckman et al., 1998). In our case, unconfoundedness is relevant because we are not just interested in the mean but also in the quantiles of the distribution of treatment effects.

[9] Heckman and Vytlacil (2007) and Abbring and Heckman (2007) provide an overview of possible solutions associated with selection on unobservables.

Throughout the paper, we use $F_A(a)$ to denote the distribution function of a random variable A and $F_{A|B=b}(a|B=b)$ to denote the distribution function of A conditional on $B=b$. Moreover, for $u \in (0,1)$, we define $q_u(F_A(a)) = \inf\{q : \Pr(A \leq q) \geq u\}$ as the quantile function of the distribution function $F_A(a)$. We are particularly interested in the distribution functions of the potential outcomes, $F_{Y(1)}(y)$ and $F_{Y(0)}(y)$. Following Rubin (1974), we consider the difference between potential outcomes, $\Delta_i = Y_i(1) - Y_i(0)$, which allows us to specify the distribution function of treatment effects, $F_\Delta(\delta)$. Our parameters of interest are formally defined as follows.

Definition 1 (Population Quantiles of the Distribution of Treatment Effects): *For $u \in (0,1)$, the Population Quantiles of the Distribution of Treatment Effects are given by*

$$q_{\Delta,u} \quad = \quad q_u(F_\Delta(\delta)) = \inf\{\delta : \frac{1}{N_{\text{sp}}} \sum_{i=1}^{N_{\text{sp}}} \mathbf{1}\{\Delta_i \leq \delta\} \geq u\}. \tag{1}$$

While this paper focuses on the Population Quantiles of the Distribution of Treatment Effects (PQDTE), similar results can be obtained for the Population Quantiles of the Distribution of Treatment Effects on the Treated (PQDTT), $q_u(F_{\Delta|D=1}(\delta|D=1))$. We consider two different estimands: Quantiles of the Impact Distribution (QID) and Quantiles of the Distribution of Treatment Effects (QDTE). Our use of QID is inspired by Heckman et al. (1997), who present estimates of quantiles of the 'impact distribution'. Our QID estimator relies on quantiles of the distribution of unit-level treatment effects. While QID capture the full spectrum of heterogeneity in the data, they ignore that unit-level treatment effects associated with units having similar or identical control outcomes may be vastly different. We demonstrate that rearranging unit-level treatment effects by control outcome yields a generalized version of quantile treatment effects (GQTE), which do not require imposing a rank invariance assumption but align with conventional quantile treatment effects (QTE) under rank invariance. If the GQTE are monotonically increasing, they coincide with QDTE.

4 Identification

4.1 Measuring rank dependence

We use Kendall's (1938) rank correlation coefficient to specify the degree of rank dependence between potential outcomes in the treatment and control state. We consider the potential outcomes in the super-population. For any two pairs $(Y_i(1), Y_i(0))$ and $(Y_j(1), Y_j(0))$, $i, j \in \{1, \ldots, N_{\text{sp}}\}$, we define the product $\pi_{ij} = (Y_i(1) - Y_j(1))(Y_i(0) - Y_j(0))$ as concordant if $\pi_{ij} > 0$ and as discordant if $\pi_{ij} < 0$. The total number of comparable pairs is $N_{\text{sp}}(N_{\text{sp}} - 1)/2$. Given the number of concordant and discordant pairs, p_c and p_d, Kendall's (1938) rank correlation coefficient between potential treatment and control outcomes in the super-population may be written as

$$\widetilde{\tau}_{\text{sp}} = (p_c - p_d)/(N_{\text{sp}}(N_{\text{sp}} - 1)/2). \tag{2}$$

Throughout this paper, we consider a modified version of the rank correlation coefficient, which adjusts for ties,

$$\tau_{\text{sp}} = (p_c - p_d)/\sqrt{((N_{\text{sp}}(N_{\text{sp}} - 1)/2) - p_a)((N_{\text{sp}}(N_{\text{sp}} - 1)/2) - p_b)}, \tag{3}$$

where $p_a = \sum_{i=1}^{s_1} a_i(a_i - 1)/2$ and $p_b = \sum_{j=1}^{s_0} b_j(b_j - 1)/2$, and where a_i (b_j) is the number of tied $Y(1)$ ($Y(0)$) values in the ith (jth) set of ties, with s_1 (s_0) denoting the number of sets (Kendall and Gibbons, 1990).

In the following, we will use the notation '$\tau(\mathbf{V}_1, \mathbf{V}_0)$' to denote the rank correlation coefficient between two vectors $\mathbf{V}_1$ and $\mathbf{V}_0$ of equal length. Given the vectors $\mathbf{Y}_{\text{sp}}(1) = (Y_1(1), \ldots, Y_{N_{\text{sp}}}(1))'$ and $\mathbf{Y}_{\text{sp}}(0) = (Y_1(0), \ldots, Y_{N_{\text{sp}}}(0))'$, we refer to the case in which $\tau_{\text{sp}} = \tau(\mathbf{Y}_{\text{sp}}(1), \mathbf{Y}_{\text{sp}}(0)) = 1$ as perfect positive rank dependence. In this case, we can compare the highest ranked outcome in $\mathbf{Y}_{\text{sp}}(1)$ to the highest ranked outcome in $\mathbf{Y}_{\text{sp}}(0)$, the second-highest ranked outcome in $\mathbf{Y}_{\text{sp}}(1)$ to the second-highest ranked outcome in $\mathbf{Y}_{\text{sp}}(0)$, and so on. Similarly, we refer to the case in which $\tau_{\text{sp}} = \tau(\mathbf{Y}_{\text{sp}}(1), \mathbf{Y}_{\text{sp}}(0)) = -1$ as perfect negative rank dependence. This case involves comparing the lowest ranked outcome in $\mathbf{Y}_{\text{sp}}(1)$ to the highest ranked outcome in $\mathbf{Y}_{\text{sp}}(0)$, the second-lowest ranked outcome in $\mathbf{Y}_{\text{sp}}(1)$ to the second-highest ranked

outcome in $\mathbf{Y}_{\text{sp}}(0)$, and so on. Heckman et al. (1997) use perfect positive and negative rank dependence to obtain bounds on the distribution of treatment effects. They conclude that these bounds are too wide to provide meaningful insights.

4.2 Rank invariance

Rank invariance refers to the case of perfect positive rank dependence between potential treatment and control outcomes. The rank invariance assumption can be stated as follows.

ASSUMPTION 3 (Rank Invariance): $\tau_{\text{sp}} = 1$.

Under Assumption 3, the identification of our parameters of interest relies on knowledge of the separate marginal distributions of potential treatment and control outcomes. For $v \in (0,1)$, the QTE under rank invariance as defined by Doksum (1974) and Lehmann (1974) may be written as

$$QTE_v = q_v(F_{Y(1)}(y)) - q_v(F_{Y(0)}(y)).$$

For $u \in (0,1)$, identification of $q_{\Delta,u}$ under Assumptions 1 and 3 follows from identification of the marginal distributions $F_{Y(1)}(y)$ and $F_{Y(0)}(y)$ under random assignment. Fan and Park (2009) establish that the $q_{\Delta,u}$ are identified if the QTE_v are monotonically increasing in v. If the QTE_v are not monotonically increasing in v, then they have to be rearranged to coincide with the $q_{\Delta,u}$. Formally, if the QTE_v are monotonically increasing in v, then

$$\begin{aligned} q_{\Delta,u} &= q_u(F_\Delta(\delta)) \\ &= q_v(F_{Y(1)}(y)) - q_v(F_{Y(0)}(y)) \\ &= q_v(F_{Y(1)|D=1}(y|D=1)) - q_v(F_{Y(0)|D=0}(y|D=0)) \\ &= q_v(F_{Y|D=1}(y|D=1)) - q_v(F_{Y|D=0}(y|D=0)). \end{aligned} \tag{4}$$

The first equality in (4) follows from Definition 1. The second equality follows from Assumption 3 and from the proof of Lemma 4 in Fan and Park (2009). The third

equality follows from Assumption 1. The fourth equality follows from $Y_i = Y_i(1)D_i + Y_i(0)(1 - D_i)$.

Our parameters of interest are also identified when we replace Assumptions 1 and 3 by Assumptions 2 and 3. Under Assumption 2, the marginal distributions $F_{Y(d)}(y)$, with $d = \{0, 1\}$, are identified through

$$F_{Y(d)}(y) \quad = \quad \int F_{Y(d)|X=x,D=d}(y|X = x, D = d)dF_X(x). \tag{5}$$

Identification of our parameters of interest follows from the identification of $QTE_v = q_v(F_{Y(1)}(y)) - q_v(F_{Y(0)}(y))$. If the QTE_v are monotonically increasing in v, then $q_{\Delta,u} = q_v(F_{Y(1)}(y)) - q_v(F_{Y(0)}(y))$. Similar to (4), if the QTE_v are not monotonically increasing in v, they have to be rearranged to coincide with our parameters of interest.

4.3 Imperfect rank dependence

In this section, we consider identification under Assumption 1 in situations where $0 \leq \tau_{\text{sp}} \leq 1$. We ignore situations where $\tau_{\text{sp}} < 0$ because they are expected to be uncommon in real-world applications. Extending our approach to these situations is straightforward. Moreover, we focus on the special case in which treatment and control group sizes are equal. Specifically, in the super-population of size N_{sp}, N^1_{sp} units are randomly assigned to the treatment group and $N^0_{\text{sp}} = N_{\text{sp}} - N^1_{\text{sp}}$ units are randomly assigned to the control group, with $N^1_{\text{sp}} = N^0_{\text{sp}} = N_{\text{sp}}/2$. We discuss more general cases in Section 4.4 below.

4.3.1 Rank correlation coefficient

We define $Y_{id} = Y_i$ if $D_i = d$, $d \in \{0, 1\}$, and consider the $((N_{\text{sp}}/2) \times 1)$-vectors of observable treatment and control outcomes $\mathbf{Y}_{\text{sp},d} = (Y_{1d}, \ldots, Y_{(N_{\text{sp}}/2)d})'$. Because the ranks of $\mathbf{Y}_{\text{sp},1}$ relative to the ranks of $\mathbf{Y}_{\text{sp},0}$ are unknown, we consider all possible permutations of observable treatment outcomes and define the $((N_{\text{sp}}/2) \times (N_{\text{sp}}/2))$-permutation matrix $\mathbf{\Pi}_{\text{sp},p}$, $p \in \mathcal{P}_{\text{sp}} = \{1, \ldots, (N_{\text{sp}}/2)!\}$.[10]

[10]Instead of keeping the ranks of $\mathbf{Y}_{\text{sp},0}$ fixed and considering permutations of $\mathbf{Y}_{\text{sp},1}$, an alternative approach involves keeping the ranks of $\mathbf{Y}_{\text{sp},1}$ fixed and considering permutations of $\mathbf{Y}_{\text{sp},0}$. We have chosen the former approach because it enables us to determine the average unit-level treatment effect on control group members if they had received the treatment. Moreover, this approach allows us to sort the average unit-level treatment effects by control outcome to obtain the GQTE. See Section 6.2 for details.

Under Assumption 1, the rank correlation coefficient between potential treatment and control outcomes may be written as

$$\tau_{\text{sp}} = \sum_{p \in \mathcal{P}_{\text{sp}}} \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}) = \tau_{\text{sp}}]\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}). \tag{6}$$

Equation (6) illustrates that knowledge of the probabilities $\Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}) = \tau_{\text{sp}}], p \in \mathcal{P}_{\text{sp}}$, would solve the identification problem. To make progress towards this goal, we use the set of covariates to obtain predicted outcomes. We define the $(N_{\text{sp}} \times k)$-covariate matrix $\mathbf{X}_{\text{sp}}$ and the corresponding $((N_{\text{sp}}/2) \times k)$-covariate matrix $\mathbf{X}_{\text{sp},0}$ for members of the control group. We obtain the rank correlation coefficients between actual and predicted outcomes in three steps:

1. We calculate the predicted outcomes $\widehat{\mathbf{Y}}_{\text{sp}} = \mathbf{X}_{\text{sp}}(\mathbf{X}'_{\text{sp},0}\mathbf{X}_{\text{sp},0})^{-1}\mathbf{X}'_{\text{sp},0}\mathbf{Y}_{\text{sp},0}$.[11] We use the elements of the $(N_{\text{sp}} \times 1)$-vector $\widehat{\mathbf{Y}}_{\text{sp}} = (\widehat{Y}_1, \ldots, \widehat{Y}_{N_{\text{sp}}})'$ to define $\widehat{Y}_{id} = \widehat{Y}_i$ if $D_i = d$, $d \in \{0, 1\}$, and consider the $((N_{\text{sp}}/2) \times 1)$-vectors of predicted treatment and control outcomes $\widehat{\mathbf{Y}}_{\text{sp},d} = (\widehat{Y}_{1d}, \ldots, \widehat{Y}_{(N_{\text{sp}}/2)d})'$.

2. We obtain the rank correlation coefficient between actual and predicted outcomes for members of the control group, $\tau(\mathbf{Y}_{\text{sp},0}, \widehat{\mathbf{Y}}_{\text{sp},0})$.

3. We obtain the rank correlation coefficient between actual and predicted outcomes for members of the treatment group, $\tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})$.

In the following, we focus on situations where $|\tau(\mathbf{Y}_{\text{sp},0}, \widehat{\mathbf{Y}}_{\text{sp},0})| \geq |\tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})|$. Focusing on these situations greatly facilitates the presentation of our approach. We ignore situations where $|\tau(\mathbf{Y}_{\text{sp},0}, \widehat{\mathbf{Y}}_{\text{sp},0})| < |\tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})|$ because they are rare in empirical applications. We leave the study of these situations for future work.[12]

[11] Alternative methods, including machine learning algorithms, can be used to predict the values or ranks of the observed outcomes. We obtain predicted outcomes in a linear regression framework and consider continuous outcome variables because we find that this approach is sufficient to fit the data used in our empirical application. To mitigate potential problems associated with overfitting, we use the same set of covariates as in Bitler et al. (2006).

[12] The experimental study of Gillitzer and Sinning (2020) provides an example in which the out-of-sample prediction of ranks is stronger than the in-sample prediction. The authors investigate the effectiveness of a reminder from the tax office on tax payments. Instead of observing pre-intervention outcomes, they observe the initial amount of tax owed, which is a strong predictor of tax payments of the treatment group (who receive a reminder) but a relatively weak predictor of tax payments of the control group (who do not receive a reminder).

We make the following assumption about the relevance of predictors:

ASSUMPTION 4 (Relevance of predictors): $\tau(\mathbf{Y}_{\text{sp},0}, \widehat{\mathbf{Y}}_{\text{sp},0})$ *is positive and sufficiently large.*

Assumption 4 requires the availability of relevant control outcome predictors. The set of covariates used in our analysis contains pre-intervention outcomes, which constitute strong predictors of post-intervention control outcomes. We are able to formally test the strength of predictors, as discussed below.

It is useful to consider the rank correlation coefficient $\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},0})$ associated with the rank correlation coefficient $\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0})$. Under Assumptions 1 and 4, permutations of $\mathbf{Y}_{\text{sp},1}$ must satisfy the condition $\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},0}) = \tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})$ because the rank correlation coefficient between actual and predicted outcomes must be the same, regardless of whether its calculation is based on predicted treatment or control outcomes. We refer to this property as 'Rank Correlation Property' (RCP). Permutations of $\mathbf{Y}_{\text{sp},1}$ that do not satisfy the RCP occur with a probability of zero.[13]

Therefore,

$$\tau_{\text{sp}} = \sum_{p \in \mathcal{S}_{\text{sp}}} \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}) = \tau_{\text{sp}}]\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}), \tag{7}$$

where $\mathcal{S}_{\text{sp}} = \{p \in \mathcal{P}_{\text{sp}} \mid \tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},0}) = \tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})\}$. Equation (7) follows from equation (6), Assumptions 1 and 4, and from

$$\begin{aligned}
&\textstyle\sum_{p \in \mathcal{P}_{\text{sp}}} \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}) = \tau_{\text{sp}}] \\
&= \sum_{p \in \mathcal{S}_{\text{sp}}} \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}) = \tau_{\text{sp}} \mid \tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},0}) = \tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})] \\
&\times \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},0}) = \tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})] \\
&+ \sum_{p \in \mathcal{S}'_{\text{sp}}} \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}) = \tau_{\text{sp}} \mid \tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},0}) \neq \tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})] \\
&\times \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},0}) \neq \tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})],
\end{aligned}$$

[13] It is possible to impose a more stringent version of the RCP by considering permutations that satisfy both $\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},0}) = \tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})$ and $\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},0}, \widehat{\mathbf{Y}}_{\text{sp},1}) = \tau(\mathbf{Y}_{\text{sp},0}, \widehat{\mathbf{Y}}_{\text{sp},0})$. While implementing this stricter version increases complexity, sensitivity checks indicate that it yields very similar results. We leave the use of the stricter RCP for future research.

where $\mathcal{S}'_{\text{sp}} = \{p \in \mathcal{P}_{\text{sp}} \mid \tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},0}) \neq \tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})\}$. Permutations that do not satisfy the RCP occur with a probability of zero due to random assignment under Assumption 1, $\Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},0}) \neq \tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})] = 0$ for all $p \in \mathcal{P}_{\text{sp}}$. Using Bayes' law, $\Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}) = \tau_{\text{sp}}] = \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}) = \tau_{\text{sp}} \mid \tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},0}) = \tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})] \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},0}) = \tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})]$ for all $p \in \mathcal{S}_{\text{sp}}$.

We make the following assumption about permutations with a positive probability of occurrence:

Assumption 5 (Equally likely permutations): *All permutations of treatment group outcomes with a positive probability of occurrence are equally likely to occur.*

Assumption 5 implies that

$$\Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}) = \tau_{\text{sp}}] = \frac{\mathbf{1}\{\Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}) = \tau_{\text{sp}}] > 0\}}{\sum_{p \in \mathcal{P}_{\text{sp}}} \mathbf{1}\{\Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}) = \tau_{\text{sp}}] > 0\}}$$

for all $p \in \mathcal{P}_{\text{sp}}$. In the absence of further information, all permutations of $\mathbf{Y}_{\text{sp},1}$ that satisfy the RCP have a positive probability of occurrence under Assumptions 1 and 4. Therefore, under Assumptions 1, 4, and 5, the rank correlation coefficient between potential treatment and control outcomes is identified through

$$\tau_{\text{sp}} = \frac{1}{n_{\text{sp},p}} \sum_{p \in \mathcal{S}_{\text{sp}}} \tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}), \tag{8}$$

with $n_{\text{sp},p} = \sum_{p \in \mathcal{P}_{\text{sp}}} \mathbf{1}\{\Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \mathbf{Y}_{\text{sp},0}) = \tau_{\text{sp}}] > 0\} = \sum_{p \in \mathcal{S}_{\text{sp}}} \mathbf{1}\{\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},0}) = \tau(\mathbf{Y}_{\text{sp},1}, \widehat{\mathbf{Y}}_{\text{sp},1})\}$.

4.3.2 Quantiles of the distribution of treatment effects

In this section, we apply the same reasoning as in the previous section to identify our parameters of interest under Assumptions 1, 4 and 5. We define $\Delta_{\text{sp},p} = \mathbf{\Pi}_{\text{sp},p}\mathbf{Y}_{\text{sp},1} - \mathbf{Y}_{\text{sp},0}$. Under Assumption 1, the distribution of treatment effects may be written as

$$F_{\Delta}(\delta) = F_{\Delta'}(\delta'), \tag{9}$$

where

$$\Delta^{'} = \sum_{p \epsilon \mathcal{P}_{\text{sp}}} \Pr[F_{\Delta_{\text{sp},p}}(\delta_{\text{sp},p}) = F_{\Delta}(\delta)]\Delta_{\text{sp},p}.$$

Knowledge of the probabilities $\Pr[F_{\Delta_{\text{sp},p}}(\delta_{\text{sp},p}) = F_{\Delta}(\delta)]$, $p \in \mathcal{P}_{\text{sp}}$, would solve the identification problem. Permutations of $\mathbf{Y}_{\text{sp},1}$ that do not satisfy the RCP occur with a probability of zero under Assumption 1. Therefore,

$$\Delta^{'} = \sum_{p \epsilon \mathcal{S}_{\text{sp}}} \Pr[F_{\Delta_{\text{sp},p}}(\delta_{\text{sp},p}) = F_{\Delta}(\delta)]\Delta_{\text{sp},p}.$$

In the absence of further information, all permutations of $\mathbf{Y}_{\text{sp},1}$ that satisfy the RCP have a positive probability of occurrence under Assumptions 1 and 4. Therefore, under Assumptions 1, 4, and 5, the distribution of treatment effects is identified through

$$F_{\Delta}(\delta) = F_{\Delta'}(\delta^{'}), \tag{10}$$

where $\Delta' = \frac{1}{n_{\text{sp},p}} \sum_{p \epsilon \mathcal{S}_{\text{sp}}} \Delta_{\text{sp},p}$. Identification of our parameters of interest follows from

$$q_{\Delta,u} = q_u(F_{\Delta}(\delta)) = q_u(F_{\Delta'}(\delta^{'})).$$

It is important to note that $\Delta' = \frac{1}{n_{\text{sp},p}} \sum_{p \epsilon \mathcal{S}_{\text{sp}}} \mathbf{\Pi}_{\text{sp},p} \mathbf{Y}_{\text{sp},1} - \mathbf{Y}_{\text{sp},0}$ involves averaging over permutations of $\mathbf{Y}_{\text{sp},1}$ while keeping the ranks of $\mathbf{Y}_{\text{sp},0}$ fixed. Computing quantiles of the distribution of permutation-specific treatment effects before averaging over permutations would produce a different result. Under Assumptions 1, 4, and 5, we may write

$$\begin{aligned} q_{\Delta,u}^{QID} &= \sum_{p \epsilon \mathcal{P}_{\text{sp}}} \Pr[q_u(F_{\Delta_{\text{sp},p}}(\delta_{\text{sp},p})) = q_u(F_{\Delta}(\delta))] q_u(F_{\Delta_{\text{sp},p}}(\delta_{\text{sp},p})) \\ &= \frac{1}{n_{\text{sp},p}} \sum_{p \epsilon \mathcal{S}_{\text{sp}}} q_u(F_{\Delta_{\text{sp},p}}(\delta_{\text{sp},p})). \end{aligned} \tag{11}$$

The parameters specified by (11) bear resemblance to those of Heckman et al. (1997) who estimate quantiles and other parameters of the 'impact distribution' based on assumed values of τ_{sp}. Averaging quantiles of permutation-specific distributions of treatment effects ignores that the location of control outcomes changes each time the quan-

tiles are computed. Consequently, this approach does not yield QDTE. We provide a detailed discussion of this issue in Section 6.2.

4.4 Controlling for covariates

In this section, we extend our approach in two important ways. Firstly, we consider identification of our parameters of interest under strong ignorability. Secondly, we proof identification of our parameters of interest for both equal and unequal group sizes.

4.4.1 Equal group sizes

We first consider the case in which $N_{\text{sp}}^1 = N_{\text{sp}}^0$ to identify our parameters of interest under Assumptions 2, 4, and 5. Using results presented in Firpo (2007), we may write the marginal distributions $F_{Y(d)}(y)$, $d = \{0,1\}$, in terms of inverse propensity score weighted averages,

$$F_{Y(d)}(y) \quad = \quad \frac{1}{N_{\text{sp}}} \sum_{i=1}^{N_{\text{sp}}} w_{id} \mathbf{1}\{Y_i \leq y\}, \tag{12}$$

where $w_{i1} = \frac{D_i}{Pr(D_i=1|X_i)}$ and $w_{i0} = \frac{1-D_i}{1-Pr(D_i=1|X_i)}$. Following Parzen (1979), we consider the order statistics $Z_{(1)d} \leq \ldots \leq Z_{(N_{\text{sp}}^d)d}$, where

$$Z_{(i)d} \quad = \quad F_{Y(d)}^{-1}[\theta_d] = \inf\{y : \frac{1}{N_{\text{sp}}} \sum_{i=1}^{N_{\text{sp}}} w_{id} \mathbf{1}\{Y_i \leq y\} \geq \theta_d\}, \tag{13}$$

with $(i-1)/N_{\text{sp}}^d < \theta_d \leq i/N_{\text{sp}}^d$, $i \in \{1, \ldots, N_{\text{sp}}^d\}$. We define the $(N_{\text{sp}}^d \times 1)$-vectors $\mathbf{Z}_{\text{sp},d} = (Z_{1d}, \ldots, Z_{N_{\text{sp}}^d d})'$. Reweighting can impact the within-group ranks of both observed and predicted outcomes. The identification of our parameters of interest under Assumptions 2, 4, and 5 relies on using reweighted outcomes and adopting the approach outlined in Section 4.3.2. Theorem 1 formally states this result.

THEOREM 1 (Identification of $q_{\Delta,u}$ with Equal Group Sizes): *Let $N_{\text{sp}}^1 = N_{\text{sp}}^0$. Then, under Assumptions 2, 4, and 5, the Population Quantiles of the Distribution of Treatment Effects, $q_{\Delta,u}$, are identified from data on (Y, D, X).*

All proofs are provided in Appendix C. Theorem 1 states that our parameters of interest are identified from data on (Y, D, X) under Assumptions 2, 4, and 5 if $N^1_{\text{sp}} = N^0_{\text{sp}}$.

4.4.2 Unequal group sizes

Since Assumption 2 is applicable to N^d_{sp} order statistics, as demonstrated in the previous section, it also extends to multiples of N^d_{sp} order statistics. We leverage this property to describe the quantile functions of treatment and control outcomes using $N^1_{\text{sp}}N^0_{\text{sp}}$ values.[14] Formally, we consider the order statistics $Z_{(1)d} \leq \ldots \leq Z_{(N^1_{\text{sp}}N^0_{\text{sp}})d}$, where $Z_{(i')d} = \inf\{y : \frac{1}{N_{\text{sp}}}\sum_{i=1}^{N_{\text{sp}}} w_{id}\mathbf{1}\{Y_i \leq y\} \geq \theta_d\}$ with $(i'-1)/(N^1_{\text{sp}}N^0_{\text{sp}}) < \theta_d \leq i'/(N^1_{\text{sp}}N^0_{\text{sp}})$, $i' \in \{1, \ldots, N^1_{\text{sp}}N^0_{\text{sp}}\}$. Moreover, we define the $(N^1_{\text{sp}}N^0_{\text{sp}} \times n_d)$-transformation matrix $\mathbf{M}_{\text{sp},d}$ that transforms the $(N^d_{\text{sp}} \times 1)$-vectors $\mathbf{Z}_{\text{sp},d}$ into the $(N^1_{\text{sp}}N^0_{\text{sp}} \times 1)$-vectors

$$\mathbf{M}_{\text{sp},d}\mathbf{Z}_{\text{sp},d} = (\underbrace{Z_{1d}, \ldots, Z_{1d}}_{N^{1-d}_{\text{sp}}}, \ldots, \underbrace{Z_{N^d_{\text{sp}}d}, \ldots, Z_{N^d_{\text{sp}}d}}_{N^{1-d}_{\text{sp}}})'.$$

Considering $N^1_{\text{sp}}N^0_{\text{sp}}$ values of the order statistics allows us to obtain a generalized version of Theorem 1, which involves using the transformed outcome variables $\mathbf{M}_{\text{sp},d}\mathbf{Z}_{\text{sp},d}$ instead of $\mathbf{Z}_{\text{sp},d}$. While using $N^1_{\text{sp}}N^0_{\text{sp}}$ order statistics addresses the problem of unequal group sizes, the challenges resulting from uncertainty about the ranking of treatment outcomes relative to control outcomes remain unchanged. The identification of our parameters of interest continues to be based on permutations of $\mathbf{Z}_{\text{sp},1}$. Theorem 2 states our main result.

THEOREM 2 (Identification of $q_{\Delta,u}$): *Under Assumptions 2, 4, and 5, the Population Quantiles of the Distribution of Treatment Effects, $q_{\Delta,u}$, are identified from data on (Y, D, X).*

Theorem 2 states that our parameters of interest are identified from data on (Y, D, X) under Assumptions 2, 4 and 5. Because Theorem 2 holds for any $N^1_{\text{sp}}, N^0_{\text{sp}} \in \mathbb{N}_+$, it includes Theorem 1 as a special case. In the case of rank invariance, our parameters of

[14] Applying this approach to large samples may pose computational challenges. However, the use of frequency weights can substantially alleviate this issue by reducing the necessary number of rows in a transformed dataset.

interest simplify to $q_{\Delta,u} = q_v(F_{Z_1}(z_1)) - q_v(F_{Z_0}(z_0))$, where $Z_1 = \mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}$ and $Z_0 = \mathbf{M}_{\text{sp},0}\mathbf{Z}_{\text{sp},0}$.[15] As noted earlier, the differences between quantiles have to be monotonically increasing in v to obtain this result.

5 Estimation

Using the framework introduced by Imbens and Rubin (2015), we consider a finite Simple Random Sample (SRS) of fixed size N drawn from the super-population and define R_i as a sampling indicator, which takes on the value 1 if unit i is sampled, and 0 otherwise. We define the treatment indicator W_i, which takes on the value 1 if unit i is sampled and assigned to the treatment group, and 0 if unit i is assigned to the control group. We simplify the presentation by assigning all unsampled units to the control group, implying $W_i = 0$ if $R_i = 0$.

We observe N_1 units randomly assigned to the treatment group and $N_0 = N - N_1$ units randomly assigned to the control group. We define the $(N \times 1)$-vectors of potential outcomes in the finite sample, $\mathbf{Y}(d) = (Y_1(d), \ldots, Y_N(d))'$, $d = \{0, 1\}$. We also define the $(N_d \times 1)$-vectors of observed treatment and control outcomes, $\mathbf{Y}_d = (Y_{1d}, \ldots, Y_{N_d d})'$. We consider the $(N \times k)$-covariate matrix $\mathbf{X}$ and the corresponding $(N_0 \times k)$-covariate matrix $\mathbf{X}_0$ of the control group sample. Using the control group sample, we estimate a linear regression model to obtain the parameter vector $\widehat{\gamma}_0 = (\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{Y}_0$. The vector of predicted values is given by $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\gamma}_0$. Using the elements of $\widehat{\mathbf{Y}}$, we define $\widehat{Y}_{id} = \widehat{Y}_i$ if $W_i = d$ and consider the $(N_d \times 1)$-vectors of predicted values $\widehat{\mathbf{Y}}_d = (\widehat{Y}_{id}, \ldots, \widehat{Y}_{N_d d})$. Moreover, we define the $(N_1 \times N_1)$- permutation matrix $\mathbf{\Pi}_p, p \in \mathcal{P} = \{1, \ldots, N_1!\}$.

Given the sample $Y_1, \ldots, Y_N$, the empirical distribution functions of $F_{Y(d)}(y)$ may be written as $\widehat{F}_{Y(d)}(y) = \frac{1}{N}\sum_{i=1}^{N} w_{id}\mathbf{1}\{Y_i \leq y\}$, where $w_{i1} = \frac{W_i}{\Pr(W_i|X_i)}$ and $w_{i0} = \frac{1-W_i}{1-\Pr(W_i=1|X_i)}$. For

$$i' \in \{\{j_{11}, \ldots, j_{1N_{(1-d)}}\}, \ldots, \{j_{N_d 1}, \ldots, j_{N_d N_{(1-d)}}\}\} = \{1, \ldots, N_1 N_0\},$$

we define the order statistics $Z_{(1)d} \leq \ldots \leq Z_{(N_1 N_0)d}$ as values of the empirical quantile function of continuous and monotonically increasing empirical distribution functions

[15] Figure A3d demonstrates that our approach produces very similar results to the approach of Firpo (2007) under rank invariance.

$\widehat{F}_{Y(d)}(y)$,

$$Z_{(i')d} = \widehat{F}^{-1}_{Y(d)}[\theta_d] = \inf\{y : \frac{1}{N}\sum_{i=1}^{N} w_{id}\mathbf{1}\{Y_i \leq y\} \geq \theta_d\},$$

with $(i'-1)/(N_0N_1) < \theta_d \leq i'/(N_0N_1), i' \in \{1,\ldots,N_0N_1\}$. We define the $(N_1N_0 \times 1)$-vectors

$$\begin{aligned}\mathbf{V}_d &= (Z_{(j_{11})d},\ldots,Z_{(j_{1N_{(1-d)}})d},\ldots,Z_{(j_{N_d1})d},\ldots,Z_{(j_{N_dN_{(1-d)}})d})' \\ &= (\underbrace{Z_{(1)d},\ldots,Z_{(1)d}}_{N_{(1-d)}},\ldots,\underbrace{Z_{(N_d)d},\ldots,Z_{(N_d)d}}_{N_{(1-d)}})',\end{aligned}$$

where

$$Z_{(i)d} = \widehat{F}^{-1}_{Y(d)}[\theta_d] = \inf\{y : \frac{1}{N}\sum_{i=1}^{N} w_{id}\mathbf{1}\{Y_i \leq y\} \geq \theta_d\},$$

with $(i-1)/N_d < \theta_d \leq i/N_d$, $i \in \{1,\ldots,N_d\}$. We define the $(N_d \times 1)$-vectors $\mathbf{Z}_d = (Z_{1d},\ldots,Z_{N_d}d)'$. Similarly, we define

$$\widehat{Z}_{(i)d} = \inf\{\widehat{y} : \frac{1}{N}\sum_{i=1}^{N} w_{id}\mathbf{1}\{\widehat{Y}_i \leq \widehat{y}\} \geq \theta_d\}$$

and the $(N_d \times 1)$-vectors $\widehat{\mathbf{Z}}_d = (\widehat{Z}_{1d},\ldots,\widehat{Z}_{N_dd})'$. We consider the permutations $\mathbf{\Omega}_d\mathbf{V}_d = \mathbf{M}_d\mathbf{Z}_d$, where $\mathbf{\Omega}_d$ are $(N_1N_0 \times N_1N_0)$-permutation matrices that transform the $(N_d \times 1)$-vectors $\mathbf{Z}_d$ into the vectors $(N_1N_0 \times 1)$-vectors

$$\mathbf{M}_d\mathbf{Z}_d = (\underbrace{Z_{1d},\ldots,Z_{1d}}_{N_{(1-d)}},\ldots,\underbrace{Z_{N_dd},\ldots,Z_{N_dd}}_{N_{(1-d)}})'.$$

5.1 Rank correlation coefficient

Our estimator of the rank correlation coefficient between potential treatment and control outcomes τ_{sp} is given by

$$\widehat{\tau} = \frac{1}{n_p}\sum_{p\in\mathcal{S}} \tau(\mathbf{M}_1\mathbf{\Pi}_p\mathbf{Z}_1, \mathbf{M}_0\mathbf{Z}_0),$$

where $\mathcal{S} = \{p \in \mathcal{P} | \tau(\mathbf{M}_1\mathbf{\Pi}_p\mathbf{Z}_1, \mathbf{M}_0\widehat{\mathbf{Z}}_0) = \tau(\mathbf{Z}_1, \widehat{\mathbf{Z}}_1)\}$ and $n_p = \sum_{p\in\mathcal{S}} \mathbf{1}\{\tau(\mathbf{M}_1\mathbf{\Pi}_p\mathbf{Z}_1, \mathbf{M}_0\widehat{\mathbf{Z}}_0) = \tau(\mathbf{Z}_1, \widehat{\mathbf{Z}}_1)\}$.

Under Assumption 1, our estimator simplifies to $\widehat{\tau} = \frac{1}{n_p}\sum_{p\in\mathcal{S}} \tau(\mathbf{M}_1\mathbf{\Pi}_p\mathbf{Y}_1, \mathbf{M}_0\mathbf{Y}_0)$. When treatment and control group sample sizes are equal, $N_1 = N_0 = N/2$, it further simplifies to $\widehat{\tau} = \frac{1}{n_p}\sum_{p\in\mathcal{S}} \tau(\mathbf{\Pi}_p\mathbf{Y}_1, \mathbf{Y}_0)$.

Due to the large number of possible permutations, we use the following steps to estimate τ_{sp} in our empirical application:

1. We search for a random permutation of $\mathbf{Z}_1$ that satisfies $\tau(\mathbf{M}_1\mathbf{\Pi}_p\mathbf{Z}_1, \mathbf{M}_0\widehat{\mathbf{Z}}_0) = \tau(\mathbf{Z}_1, \widehat{\mathbf{Z}}_1)$ and calculate $\tau(\mathbf{\Pi}_p\mathbf{Z}_1, \mathbf{Z}_0)$.

2. We repeat the previous step P times.

3. We calculate $\widehat{\tau}_P = \frac{1}{P}\sum_{p=1}^{P} \tau(\mathbf{M}_1\mathbf{\Pi}_p\mathbf{Z}_1, \mathbf{M}_0\mathbf{Z}_0)$.

Table 1 presents estimates of τ_{sp} for alternative values of $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ and $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1)$ based on simulation data. Each estimate involves 100 random permutations of $2 \times 16{,}772$ observation units ($N_1 = N_0 = N/2$=16,772), consistent with the restricted analysis sample using in our application below. For simplicity, we consider a model with a single continuous covariate and exclude the possibility of ties. We obtain bootstrap standard errors for the estimated parameters by drawing a random sample with replacement, stratified by treatment group to ensure that $N_1 = N_0$, for each permutation.[16]

The results in Table 1 illustrate that it is possible to obtain precise estimates of τ_{sp}, even if the strength of the model predictors, measured by $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$, is only moderate. For $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0) \geq 0.3$, our estimates are identical to the target values if we round to one decimal place. For $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0) = 0.2$, we observe reasonably precise estimates of τ_{sp} if $\tau_{\text{sp}} < 0.9$. However, we obtain an estimate of 0.83 if $\tau_{\text{sp}} = 0.9$, indicating that it is more challenging to estimate τ_{sp} when the ranks of the potential treatment and control outcomes are highly correlated.[17]

<Table 1 about here.>

[16]The estimation of bootstrap standard errors involves the use of ties due to random sampling with replacement.

[17]The reason for this result is that the underlying values of $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1)$, which are presented in Table A1 of Appendix A, are very close to each other when τ_{sp} is close to 1 and $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ is small.

The use of simulated data to estimate target values of τ_{sp} for given values of $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ allows us to develop a statistical test for assessing the strength of model predictors, measured by $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$. We test whether our estimated rank correlation coefficients deviate significantly from these target values. We also employ a Kolmogorov-Smirnov test to ascertain whether deviations in estimated rank correlation coefficients from target values impact our QDTE estimates. Our test results indicate that our predictors are sufficiently strong to yield unbiased QDTE estimates. A more detailed discussion of our test results is provided in Appendix A.

Results of Monte Carlo simulations (presented in Tables B1-B4 of Appendix B) are in line with the findings presented in Table 1. In particular, we find that our estimator of τ_{sp} is fairly precise even for relatively small sample sizes and unbiased if $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ is sufficiently large. Biases observed for low values of $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ and large values of τ_{sp} decline as the sample size increases. Our simulation results also confirm the validity of estimated bootstrap standard errors.

5.2 Quantiles of the Distribution of Treatment Effects

For $u \in (0, 1)$, the PQDTE estimator is given by

$$\widehat{q}_{\Delta,u} \quad = \quad q_u(F_{\Delta^{\text{obs}}}(\delta^{\text{obs}})),$$

where $\Delta^{\text{obs}} = \frac{1}{n_p} \sum_{p \in \mathcal{S}} \mathbf{M}_1 \mathbf{\Pi}_p \mathbf{Z}_1 - \mathbf{M}_0 \mathbf{Z}_0$. We impose Assumption 1 and consider the case in which $N_1 = N_0 = N/2$ to show that $\widehat{q}_{\Delta,u}$ is an unbiased estimator of $q_{\Delta,u}$. In this case, our estimator may be written as

$$\widehat{q}_{\Delta,u} \quad = \quad \inf\{\delta^{\text{obs}} : \frac{1}{N_1} \sum_{i=1}^{N_{\text{sp}}} R_i(1 - W_i) \mathbf{1}\{\Delta_i^{obs} \leq \delta^{\text{obs}}\} \geq u\},$$

where Δ_i^{obs} is the ith element of $\Delta^{\text{obs}} = \frac{1}{n_p} \sum_{p \in \mathcal{S}} \mathbf{\Pi}_p \mathbf{Y}_1 - \mathbf{Y}_0$. Following Imbens and Rubin (2015), we can take the expectation of our estimator over the randomization distribution

by conditioning on the $(N_{\text{sp}} \times 1)$-vector $\mathbf{R} = (R_1, \ldots, R_{N_{\text{sp}}})'$, and on $\mathbf{Y}_{\text{sp}}(1)$ and $\mathbf{Y}_{\text{sp}}(0)$:

$$\begin{aligned}
\mathbb{E}_W[\widehat{q}_{\Delta,u}|\mathbf{R}, \mathbf{Y}_{\text{sp}}(1), \mathbf{Y}_{\text{sp}}(0)] &= \inf\{\delta^{\text{obs}} : \frac{1}{N_0}\sum_{i=1}^{N_{\text{sp}}} R_i \mathbb{E}_W[1 - W_i]\mathbf{1}\{\Delta_i^{\text{obs}} \leq \delta^{\text{obs}}\} \geq u\} \\
&= \inf\{\delta^{\text{obs}} : \frac{1}{N_0}\sum_{i=1}^{N_{\text{sp}}} R_i \frac{N_0}{N}\mathbf{1}\{\Delta_i^{\text{obs}} \leq \delta^{\text{obs}}\} \geq u\} \qquad (14) \\
&= \inf\{\delta : \frac{1}{N}\sum_{i=1}^{N_{\text{sp}}} R_i \mathbf{1}\{\Delta_i \leq \delta\} \geq u\} \\
&= q^{\text{fs}}_{\Delta,u},
\end{aligned}$$

where $q^{\text{fs}}_{\Delta,u}$ are the finite-population quantiles of the distribution of treatment effects. Taking expectations over the distribution generated by the random sampling yields

$$\begin{aligned}
\mathbb{E}_{\text{sp}}[q^{\text{fs}}_{\Delta,u}|\mathbf{Y}_{\text{sp}}(1), \mathbf{Y}_{\text{sp}}(0)] &= \inf\{\delta : \frac{1}{N}\sum_{i=1}^{N_{\text{sp}}} \mathbb{E}_{\text{sp}}[R_i]\mathbf{1}\{\Delta_i \leq \delta\} \geq u\} \\
&= \inf\{\delta : \frac{1}{N}\sum_{i=1}^{N_{\text{sp}}} \frac{N}{N_{\text{sp}}}\mathbf{1}\{\Delta_i \leq \delta\} \geq u\} \qquad (15) \\
&= \inf\{\delta : \frac{1}{N_{\text{sp}}}\sum_{i=1}^{N_{\text{sp}}} \mathbf{1}\{\Delta_i \leq \delta\} \geq u\} \\
&= q_{\Delta,u}.
\end{aligned}$$

We may use a similar approach to study the QID estimator

$$\widehat{q}^{\,QID}_{\Delta,u} = \frac{1}{n_p}\sum_{p \in \mathcal{S}} \widehat{q}^{\,QID}_{\Delta_p,u} = \frac{1}{n_p}\sum_{p \in \mathcal{S}} q_u(F_{\Delta_p^{\text{obs}}}(\delta_p^{\text{obs}})), \qquad (16)$$

where $\Delta_p^{\text{obs}} = \mathbf{\Pi}_p \mathbf{Y}_1 - \mathbf{Y}_0$. It is possible to take expectations over the randomization distribution by conditioning on $\mathbf{R}$, $\mathbf{Y}_{\text{sp}}(1)$, and $\mathbf{Y}_{\text{sp}}(0)$ to show that

$$\mathbb{E}_W[\widehat{q}^{\,QID}_{\Delta_p,u}|\mathbf{R}, \mathbf{Y}_{\text{sp}}(1), \mathbf{Y}_{\text{sp}}(0)] = q^{\text{fs}}_{\Delta_p,u},$$

where $q^{\text{fs}}_{\Delta_p,u}$ are the permutation-specific finite-population quantiles of the distribution of treatment effects. Taking expectations over the distribution generated by the random sampling yields

$$\mathbb{E}_{\text{sp}}[q^{\text{fs}}_{\Delta_p,u}|\mathbf{Y}_{\text{sp}}(1), \mathbf{Y}_{\text{sp}}(0)] = q_{\Delta_p,u},$$

where $q_{\Delta_p,u}$ are the permutation-specific quantiles of the distribution of treatment effects in the super-population. While the permutation-specific elements of the QID estimator are unbiased, averaging over these elements does not yield $q_{\Delta,u}$. We illustrate this issue in our empirical application. Due to the large number of possible permutations, we use Steps 1 and 2 presented in the previous section to obtain QDTE and QID estimates.

6 Application

6.1 Background and data

Connecticut introduced the Jobs First program in response to the nationwide mandate for welfare reform initiated by the Personal Responsibility and Work Opportunity Reconciliation Act in 1996. To assess the program's effectiveness, the state conducted a randomized experiment in collaboration with the Manpower Development Research Corporation (MDRC). The experiment compared the Jobs First program with the previous Aid to Families with Dependent Children (AFDC) welfare program for low-income single parents with children. Jobs First incorporates key elements commonly found in US welfare programs, including earnings disregards, work requirements, and financial sanctions for non-compliance with these requirements. The program introduced a combination of both positive and negative work incentives. Detailed descriptions of the key elements of Jobs First and AFDC can be found in Bitler et al. (2006) and Kline and Tartari (2016).

Our analysis relies on public use files from MDRC's experimental evaluation of Connecticut's Jobs First waiver from AFDC rules. The files contain rounded data on quarterly earnings, as well as monthly welfare and food stamps income, covering a period of two years leading up to program assignment and extending for a minimum of four years thereafter. Following Bitler et al. (2006), our analysis focuses on three outcome variables: earnings, transfers (cash welfare plus food stamps), and income (the sum of earnings and transfers). MDRC's dataset contains a sample of 4,803 cases, with 2,396 assigned to Jobs First and 2,407 to AFDC. To obtain an analysis sample with equal treatment and control group sizes, we randomly exclude 11 observations from the AFDC sample to create a restricted AFDC sample of 2,396 observations. This sample

restriction enables us to illustrate our approach for equal group sizes. Subsequently, we expand our approach to accommodate scenarios in which treatment and control group sizes differ.

Table 2 replicates the summary statistics of pre-treatment characteristics presented in Bitler et al. (2006) and provides additional results for the restricted AFDC sample. We observe that the exclusion of 11 randomly chosen observations from the AFDC sample has no significant impact on the sample means, with all p-values associated with the comparison of group means of unrestricted and restricted AFDC samples exceeding 0.9.

<Table 2 about here.>

Because our approach relies on predicting values or ranks of observed outcomes, we employ the same set of covariates as used in Bitler et al. (2006) to mitigate potential problems associated with overfitting. A complete list of covariates can be found in Appendix B of Bitler et al. (2006). We use the same set of covariates to estimate our parameters of interest under strong ignorability.

6.2 QID, QDTE and GQTE

Heckman et al. (1997) pair percentiles of marginal distributions of treatment and control outcomes to calculate the mean of parameters of the impact distribution. They use random permutations of percentiles to achieve a given value of τ_{sp}.[18] We modify their approach in two ways that have minimal impact on our results. Firstly, we use restricted analysis samples with equal treatment and control group sizes. This change enables us to consider random permutations of unit-level outcomes instead of relying on percentiles of marginal distributions of treatment and control outcomes. This modification produces slightly more accurate results. The impact of imposing sample restrictions is negligible in our application, as demonstrated below (see Figure A3 in Appendix A).

Secondly, instead of assuming a specific value of τ_{sp}, we focus on random permutations that satisfy the RCP. Figures 1a and 1b depict the distributions of rank correlation coefficients resulting from 100 random permutations of $\mathbf{Y}_1$ that satisfy the RCP. We observe that the rank correlation distributions are relatively narrow, suggesting that QID

[18]See the discussion related to Table 5A in Heckman et al. (1997) for details.

estimates based on the RCP will closely resemble QID estimates based on a fixed value of τ_{sp} that coincides with the support of the corresponding rank correlation distribution.

Given these modifications, our QID estimator may be written as

$$QID \quad = \quad \frac{1}{P}\sum_{p=1}^{P} q_u(F_{\Delta_p^{\text{obs}}}(\delta_p^{\text{obs}})), \tag{17}$$

where $\Delta_p^{\text{obs}} = \mathbf{\Pi}_p\mathbf{Y}_1 - \mathbf{Y}_0$. Figures 1c-1h present QID estimates for earnings, transfers, and income in pre- and post-time limit periods based on our restricted analysis samples. The figures confirm that the discrepancy between the percentile approach, which relies on permutations consistent with a fixed value $\widehat{\tau}$, and the use of unit-level treatment effects in conjunction with the RCP is negligible.

<Figure 1 about here.>

Even though our data is entirely different from that used by Heckman et al. (1997), we observe similar overall patterns in our findings. Specifically, we notice substantially larger extremes (both positive and negative) in the tails of the impact distributions for lower values of $\widehat{\tau}$. For example, in the case of earnings during the pre-time limit period (Figure 1c) based on $\widehat{\tau} = 0.71$, the values of the impact distribution range from approximately −$1,300 to $1,400. In the case of earnings during the post-time limit period (Figure 1d) based on $\widehat{\tau} = 0.62$, the corresponding values of the impact distribution range from approximately −$3,100 to $3,100. We observe similar patterns for transfers (Figures 1e-1f) and income (Figures 1g-1h).[19]

Heckman et al. (1997) conclude that impact distributions require a high degree of positive rank dependence because low values of τ_{sp} produce negative values of the impact distribution that are too large to be plausible.[20] In the following, we offer an alternative explanation. Using our restricted analysis sample for the pre-time limit period, we obtain permutation-specific QID estimates for selected random permutations of income. We present three permutation-specific impact distributions in Figures 2a-2c.

[19]We find that the rank correlation coefficients during the post-time limit period are lower than those observed during the pre-time limit period due to greater mobility with time after the intervention.

[20]Using data on earnings of women aged 22 and above from an experimental evaluation of training programmes financed under Title IIA of the US Job Training Partnership Act (JTPA), they argue that their results are only reasonable for $\tau_{\text{sp}} \geq 0.8$.

We also select a control group member whose income matches the median control group income and pinpoint their location on each permutation-specific impact distribution.

We observe that the selected control group member is located near the lower end of the first permutation-specific impact distribution (Figure 2a), near the median of the second distribution (Figure 2b), and near the upper end of the third distribution (Figure 2c). This finding illustrates that our parameters of interest cannot be obtained by averaging over quantiles of permutation-specific impact distributions because this approach fails to capture the underlying unit-level treatment effect we would expect to observe for any given member of the control group.

<Figure 2 about here.>

Figure 2d presents both average unit-level treatment effects and QDTE estimates. Our findings indicate that the Jobs First program led to income increases across the entire income distribution during the pre-time limit period. This finding is broadly consistent with the results of Bitler et al. (2006), who report non-negative QTE estimates across the income distribution during the pre-time limit period. However, while Bitler et al. (2006) find zero QTE estimates for the bottom 10 quantiles due to zero income in both groups, our QDTE estimates are strictly positive. This difference arises because we compare each control group member to a set of treatment group members associated with random permutations that satisfy the RCP. In the absence of rank invariance, we compare control group members with zero income to treatment group members whose income is either zero or positive. Consequently, averaging over the distribution of permutation-specific treatment effects produces positive QDTE estimates.[21]

Rearranging the average unit-level treatment effects shown in Figure 2d yields GQTE estimates, which are presented in Figure 2e. GQTE estimates are obtained by averaging over average unit-level treatment effects within quantile groups of the control outcome. This approach provides us with a generalized summary measure of quantile treatment effects that does not rely on a rank invariance assumption. The GQTE provides an estimate of the treatment effect that would have been observed if control group members at a given point of the control outcome distribution, such as the

[21] An important challenge in estimating QDTE in real-world applications is the handling of ties. We address this challenge by maintaining a random but fixed ranking of ties in control group outcomes, while allowing for varying ranks of treatment group outcomes through alternative permutations.

median, had received the treatment. Calculating the average within quantile groups is essential to obtain a measure that captures the heterogeneity around the ATE.[22]

Another perspective on GQTE estimates is to view them as averages of permutation-specific GQTE estimates. Permutation-specific GQTE estimates are obtained by averaging unit-level treatment effects within quantile groups of the control outcome. They differ from permutation-specific QID estimates in terms of how the unit-level treatment effects are aggregated. While permutation-specific QIDs capture the full extent of treatment effect heterogeneity in the data, permutation-specific GQTEs aggregate both positive and negative unit-level treatment effects within quantile groups. In contrast to permutation-specific QIDs, permutation-specific GQTEs have the property that control group members at the median of their outcome distribution are always located at the median, regardless of the chosen permutation.

We use permutation-specific GQTEs for income during the pre-time limit period to illustrate the implications of imposing Assumption 5. Figure A2 in Appendix A presents the density functions of permutation-specific GQTEs at selected quantiles, along with the GQTE and associated density functions across quantiles. We observe that the density functions approximate normal distributions as the number of permutations increases (Figures A2a-A2e). We also find that 50 percent of the probability mass of the density functions concentrates near the GQTE (Figure A2f), providing strong support for calculating averages in accordance with Assumption 5. Additionally, our findings illustrate the potential for developing tests to compare alternative identifying assumptions, such as comparing GQTE based on Assumption 5 to GQTE based on alternative weighting schemes.

Figure 2f demonstrates that our approach yields nearly identical results to the conventional QTE approach under rank invariance. While the conventional QTE approach involves calculating quantiles of separate marginal distributions of treatment and control outcomes, our approach is based on averaging unit-level treatment effects within quantile groups of control outcomes. In cases where the rank invariance assumption holds, we have no basis for favoring one approach over another. However, in contrast to our approach, the conventional QTE approach can only be applied if the rank invariance

[22]Using the median instead of the average within quantile groups, for example, could potentially yield non-positive treatment effects across the entire distribution, even though the ATE is positive.

assumption holds.

6.3 Quantiles of the distribution of treatment effects

Figure 3 presents QDTE estimates for earnings, transfers, and income, both before and after time limits take effect. In line with the results of Bitler et al. (2006) presented in Figure A1, we find substantial heterogeneity in responses to the welfare reform. During the pre-time limit period, we observe a range of effects on earnings, spanning from approximately −$470 to $370 (Figure 3a). This range broadly aligns with the range of −$300 to $500 depicted in Figure A1a. However, after time limits take effect, we discover a much wider range of effects on earnings, varying from approximately −$1,020 to $900 (Figure 3b), which differs considerably from the corresponding range of −$100 to $800 presented in Figure A1b.

<Figure 3 about here.>

The effects of Jobs First on transfers during the pre-time limit period span a range of approximately −$30 to $540 (Figure 3c), compared to the range of $0 to $700 observed in Figure A1c. The effects on transfers during the post-time limit period range from approximately −$470 to $200 (Figure 3d), which broadly aligns with the range of −$550 to $100 presented in Figure A1d. The effects on income are largely driven by the patterns observed for earnings. The effects on income during the pre-time limit period (Figure 3e), ranging from approximately $40 to $570, are similar to the range of $0 to $800 observed in Figure A1e. In contrast, the effects on income, ranging from approximately −$940 to $880 (Figure 3f), deviate substantially from the range of −$300 to $300 presented in Figure A1f.

In contrast to the results of Bitler et al. (2006), which rely on rank invariance, our findings explicitly account for rank changes. We obtain rank correlation coefficients between potential treatment and control outcomes ranging from approximately 0.7 to 0.8 during the pre-time limit period. While Bitler et al. (2006) maintain the rank invariance assumption across time periods, our analysis reveals a considerable decline in the estimated rank correlation between potential treatment and control outcomes over time, reaching approximately 0.6 to 0.7 during the post-time limit period. The dramatic

differences in the effects on earnings and income during the post-time limit period can be attributed to mobility over time following the intervention.

6.4 Generalized quantile treatment effects

Figure 4 depicts GQTE estimates for earnings, transfers, and income during pre- and post-time limit periods. Bitler et al. (2006) find that Jobs First had no impact at the bottom of the earnings distribution, increased earnings in the middle, and reduced earnings at the top before time limits took effect (Figure A1a).[23] Our GQTE estimates largely support this overall conclusion (Figure 4a). However, when accounting for the possibility of rank changes, we observe small positive effects at the bottom of the earnings distribution, with a gradual increase from $55 at quantile 1 to $250 at quantile 45, and a further rise to $355 at quantile 54. These effects can be attributed to underlying positive average unit-level treatment effects. Specifically, we compare women in the AFDC group with zero earnings to women in the Jobs First group who have either zero or positive earnings.[24] In contrast, Bitler et al. (2006) report effects that are exactly zero for the first 48 quantiles of the earnings distribution because earnings are equal to zero for 48 percent of person quarters in the Jobs First sample and 55 percent of person quarters in the AFDC sample.

After time limits take effect, Bitler et al. (2006) observe relatively large positive effects in the middle of the earnings distribution but find no impact at the bottom and top segments (Figure A1b). This finding suggests that the implementation of time limits, which render Jobs First women ineligible for welfare benefits (with possible extensions for some), reinforces positive earnings effects while eliminating negative ones. In contrast, our GQTE estimates reveal increases in both positive effects at the lower

[23]Zero effects observed at the bottom of the earnings distribution can be attributed to the comparison of women without earnings in both groups. According to static labor supply theory, assignment to Jobs First should lead to increased earnings, transfers, and income in the middle of the distribution if substitution effects dominate income effects. Negative effects of Jobs First on earnings at the top of the distribution may be attributed to the behavioral-induced eligibility effect (Ashenfelter, 1983). This effect arises from individual decisions to reduce labor supply (and earnings) in order to increase welfare payments.

[24]This approach is consistent with static labor supply theory. Bitler et al. (2006) note that if a woman with zero earnings in the AFDC group was assigned to the Jobs First group, she could either continue to have zero earnings or enter the labor market and increase her income while still receiving the same amount of transfers. Therefore, the expected unit-level treatment effect for this woman is positive.

end of the earnings distribution, ranging from \$153 at quantile 1 to \$840 at quantile 46, and negative effects at the higher end, ranging from −\$18 at quantile 60 to −\$1,076 at quantile 98 (Figure 4b). Although incentives to reduce earnings to receive higher welfare payments disappeared with the implementation of time limits, our findings suggest that women at the higher end of the earnings distribution in the Jobs First group were either unable or unwilling to increase their earnings. Instead, we find that negative effects become more pronounced over time when we account for rank changes, indicating that the initially observed negative effects of Jobs First on the earnings distribution were more persistent than previously recognized.

We observe positive effects of Jobs First on transfers for a substantial fraction of the distribution (Figure 4c), reflecting that the Jobs First program's earnings disregard up to the poverty line was more generous than that of AFDC. Bitler et al. (2006) report zero effects for the bottom 20 quantiles of the transfer distribution (Figure A1c) because women with no earnings under AFDC and those with earnings below the poverty line under Jobs First receive maximum benefit payments. However, when accounting for rank changes, we observe small positive effects at the lower end of the distribution, ranging from \$119 at quantile 1 to \$338 at quantile 18. Moreover, we find substantial positive effects (exceeding \$350) for all quantiles ranging from 19 to 50, while observing small or statistically insignificant effects towards the higher end of the distribution. Overall, these findings are similar to those of Bitler et al. (2006) and align with the predictions of static labor supply theory.

In the post-time limit period, Bitler et al. (2006) find that the effects of Jobs First on transfers are precisely zero for the first 47 quantiles, and negative for quantiles 48 to 96 (Figure A1d). They highlight that reductions in the rate of cash assistance are the main contributing factor to the negative effects on transfers. Interestingly, they also report a positive cash assistance gap conditional on eligibility, which can be attributed to the generous earnings disregard provided by Jobs First. When considering rank changes, we uncover small positive effects of Jobs First on transfers. These effects gradually increase from \$13 at quantile 1 to \$200 at quantile 48, and then decrease to \$4 at quantile 54 (Figure 4d). From quantile 55 to 98, we observe negative effects, with a minimum of −\$452 at quantile 72. Overall, our findings regarding transfers during the post-time limit period are similar to those of Bitler et al. (2006), indicating that rank

changes had a limited impact on the observed effects.

<Figure 4 about here.>

Bitler et al. (2006) report that Jobs First had no impact on the first 10 quantiles of the income distribution. From quantile 11 to quantile 37, the effects were modest or not statistically significant. However, they observed substantial increases in incomes in the middle and at the top of the distribution, reaching up to $800 at quantile 82 (Figure A1e). Our findings are largely consistent with these results, indicating that the contribution of rank changes to the observed effects was moderate (Figure 4e). Specifically, our analysis reveals positive (although not always statistically significant) effects at the bottom of the distribution. The effects range from $66 at quantile 1 to $379 at quantile 10, gradually declining to $117 at quantile 24. For the remainder of the distribution, we observe positive effects ranging from $147 at quantile 25 to $513 at quantile 89.

Bitler et al. (2006) find that the effects of Jobs First on income during the post-time limit period exhibit a distinct pattern. Specifically, they report that the effects are zero for the first 18 quantiles, negative for the subsequent 22 quantiles, and then turn positive from quantile 46 to quantile 89, before eventually returning to zero or becoming negative (Figure A1f). Bitler et al. (2006) predict that while some women in the Jobs First group may experience an increase in earnings that exceeds the loss in transfers due to the implementation of time limits, this is not the case for others. They also predict that income increases will occur at higher income levels than income decreases if offered wages remain constant. Our estimates validate the first prediction but do not support the second (Figure 4f). When considering rank changes, our analysis reveals substantial positive effects that gradually increase from $150 at quantile 1 to $863 at quantile 18, and then gradually decrease to $60 at quantile 31. Between quantile 32 and quantile 67, we observe small negative effects of approximately −$100 or less (in absolute terms), followed by a decline from −$113 at quantile 68 to −$950 at quantile 98. These findings suggest that women with lower incomes are most responsive to reductions in transfers following the implementation of time limits. Moreover, the negative income effects observed at the higher end of the distribution align with the corresponding negative effects of Jobs First on earnings.

In sum, the discrepancies between our findings and those presented in Bitler et al. (2006) can be attributed entirely to the explicit inclusion of rank changes in our analysis. While static labor supply theory would not have predicted certain effects stemming from rank changes, our results align well with theoretical predictions, particularly during the pre-time limit period. Deviations from static labor supply theory during the post-time limit period can be attributed to persistent effects observed during the pre-time limit period and increased mobility over time.

6.5 Additional parameters of interest

The treatment effect distributions presented in Figures 3 and 4 permit inferences about the proportion of women who experience improvements, setbacks, or remain unaffected by the Jobs First program. Table 3 presents the proportion of positive and negative treatment effects on earnings, transfers, and income during the pre- and post-time limit periods, as derived from QTE estimates and GQTE estimates. The table includes the overall proportion of treatment effects, regardless of their significance levels, as well as the proportions of treatment effects that are statistically significant at the 5%, 1% and 0.1% level, respectively.[25]

When examining earnings during the pre-time limit period, our findings reveal substantial disparities in the proportions of positive and negative effects derived from QTE and GQTE estimates. Based on QTE estimates, 34.3% of the effects of Jobs First on earnings are positive, while 12.1% are negative. In contrast, the proportion of positive effects based on GQTE estimates is 68.7%, whereas the proportion of negative effects is 31.3%. A considerable gap in the proportion of positive effects remains when we focus on effects that are statistically significant at the 0.1% level (18.2% based on QTE versus 58.6% based on GQTE). However, the gap in the proportion of negative effects disappears when we consider a 0.1% significance level. A similar pattern can be observed for the post-time limit period. These findings highlight substantial disparities between QTE and GQTE estimates in determining the proportions of 'winners' and 'losers' of the Jobs First program.

[25]We present alternative significance levels to address the issue of inflated type I error rates associated with conducting the same hypothesis test at each percentile of the distribution. We defer the application of alternative tests, such as the bootstrap test of distributional treatment effects proposed in Abadie (2002), to future work.

<Table 3 about here.>

According to QTE estimates, the overall proportion of positive effects on transfers during the pre-time limit period is 77.8%, while GQTE estimates indicate a higher proportion of 97.0%. At a 0.1% significance level, the proportions of positive effects are 55.6% based on QTE estimates and 78.8% based on GQTE estimates. Both QTE and GQTE reveal no significantly negative effects of Jobs First on transfers before the implementation of time limits. During the post-time limit period, we observe notable differences in the proportions of positive effects. While QTE estimates indicate no significantly positive effects, GQTE estimates show that 49.5% of effects are positive at a 5% significance level, and 10.1% are positive at a 0.1% significance level. In contrast, the proportions of negative effects do not differ much between QTE and GQTE estimates. Overall, we find considerable disparities in the proportion of positive effects obtained from QTE and GQTE estimates. Nonetheless, in addition to assessing the proportions of treatment effects, it is equally important to examine their magnitude. We will discuss this issue below.

Our analysis uncovers notable discrepancies in the proportion of positive treatment effects on income between QTE and GQTE estimates during the pre-time limit period. Specifically, the overall proportion of positive effects based on QTE estimates is considerably lower at 85.9% compared to the corresponding proportion of 99.0% based on GQTE estimates. When focusing on effects that reach statistical significance at a 0.1% level, the proportion of positive effects based on QTE estimates decreases to 52.5%, while the proportion based on GQTE estimates remains higher at 71.7%. We observe no significantly negative effects on income, regardless of whether we consider QTE or GQTE estimates.

Interestingly, QTE and GQTE estimates yield similar proportions of statistically significant effects on income during the post-time limit period. Specifically, we find that the proportion of positive effects at a 5% significance level is 23.2% based on QTE estimates and 24.2% based on GQTE estimates. When considering a 0.1% significance level, these proportions decrease to 6.1% based on QTE estimates and 15.2% based on GQTE estimates. Moreover, we observe that the proportion of negative effects at a 5% significance level is 7.1% based on QTE estimates and 6.1% based on GQTE estimates. We do not find any negative effects on income when applying a 0.1% signifi-

cance level. Despite the observed similarities in the proportions of positive and negative treatment effects on income, the locations of these effects differ substantially between QTE estimates (Figure A1f) and GQTE estimates (Figure 4f).

Table 4 provides an overview of the total gains and losses, and presents the standard deviation of treatment effects for earnings, transfers, and income during the pre- and post-time limit periods. In the following, we focus on the most pronounced disparities observed between QTE estimates and GQTE estimates, which occur for earnings and income, particularly in the post-time limit period. Specifically, when considering the overall gains in earnings during quarters 8-16, QTE estimates yield total gains of approximately \$382,000, while GQTE estimates reveal higher gains of approximately \$649,000. When applying a 0.1% significance level, these gains decrease slightly to approximately \$301,000 based on QTE estimates and to approximately \$616,000 based on GQTE estimates. These findings indicate that the total gains resulting from the Jobs First program were more than twice as large as previously recognized. Figure 4b reveals that a considerable fraction of these gains occurred at the lower end of the earnings distribution, contrasting with the results presented in Figure A1b, which suggests a concentration of gains in the middle of the distribution.

<Table 4 about here.>

In addition to higher total gains, we also observe higher total losses in earnings when accounting for rank changes. However, total losses decrease substantially when we apply a stringent significance level of 0.1%. In this case, we observe total losses of approximately \$31,000 based on GQTE estimates, and no losses based on QTE estimates. Figure 4b illustrates that the observed losses occur at the higher end of the earnings distribution. Accounting for rank changes more than doubles the standard deviation of treatment effects on earnings in quarters 8-16, from 272.6 based on QTE estimates to 578.2 based on GQTE estimates.

Turning to income during the post-time limit period, we observe total gains in income of approximately \$190,000 based on QTE estimates when statistical significance is ignored. Based on GQTE estimates, total gains in income are more than twice as large, reaching around \$383,000. When considering a 0.1% significance level, total gains in income decrease to approximately \$43,000 based on QTE estimates. The

corresponding figure based on GQTE estimates is more than six times higher, totaling approximately $260,000. We also observe total losses in income during the post-time limit period, which amount to approximately $51,000 based on QTE estimates when a 5% significance level is applied. The corresponding figure based on GQTE estimates is approximately $107,000. However, total losses disappear when a 0.1% significance level is applied.

Taken together, our findings suggest that the total gains and losses observed during the post-time limit period are primarily driven by the impact of Jobs First on the earnings distribution. Gains in earnings were successful in mitigating the majority of losses in transfers stemming from the implementation of time limits, which led to the exclusion of Jobs First women from welfare benefits. While the QTE estimates (Figure A1f) indicate that losses occur below the median of the income distribution and gains occur above it, our GQTE estimates (Figure 4f) reveal that gains occur at the lower end of the distribution, while losses, to the extent that they are statistically significant, occur at the higher end.

7 Conclusions

This paper introduces an approach to identify quantiles of the distribution of treatment effects (QDTE) and examines the distributional impacts of Connecticut's Jobs First welfare experiment on women's earnings, transfers, and income. Our focus on situations in which relevant outcome predictors are available enables us to identify our parameters of interest under modest assumptions. We leverage a property we term the 'Rank Correlation Property' (RCP), which establishes that, under random assignment, rank correlation coefficients between actual treatment outcomes and predicted control state outcomes are identical, regardless of whether predictions are based on treatment or control units. We show that the rank correlation coefficient between potential outcomes is identified under the assumption of equally likely permutations of observation units satisfying the RCP. We use the same assumption to identify QDTE as quantiles of average unit-level treatment effects. Rearranging quantiles yields generalized quantile treatment effects (GQTE), which do not rely on a rank invariance assumption but align with conventional QTE under rank invariance. We employ reweighting methods to

identify our parameters of interest under strong ignorability. Monte Carlo simulations validate the unbiasedness, consistency, and asymptotic normality of our estimators when relevant outcome predictors are available.

We find substantial heterogeneity in the effects of Jobs First on earnings, transfers, and income. We estimate the effects separately for the periods before and after the implementation of time limits. During the pre-time limit period, our GQTE estimates align qualitatively with conventional QTE estimates. In contrast, accounting for rank changes during the post-time limit period reveals both positive effects at the lower end and negative effects at the higher end of the earnings distribution, indicating that the initially observed negative effects on earnings were more persistent than previously recognized. Our findings also suggest that women with lower incomes are most responsive to reductions in transfers following the implementation of time limits. Negative income effects at the higher end of the distribution are consistent with the corresponding negative earnings effects.

We use QTE and GQTE estimates to determine the proportions of positive and negative treatment effects. During the pre-time limit period, GQTE estimates show a higher proportion of positive income effects compared to QTE estimates. During the post-time limit period, QTE and GQTE estimates exhibit similar proportions of income effects, although the locations of these effects differ substantially between the two methods. Total gains and losses resulting from Jobs First are mainly driven by changes in the earnings distribution, with GQTE-based estimates showing considerably higher gains compared to QTE-based estimates. These findings highlight the importance of accounting for rank changes when estimating distributional effects.

Tables and Figures

TABLE 1: ESTIMATION OF τ_{sp} FOR ALTERNATIVE VALUES OF $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ AND $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1)$

	τ_{sp}								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Target	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
$\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$									
0.9	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
0.8	0.91	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10
	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
0.7	0.92	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
0.6	0.89	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
0.5	0.88	0.79	0.70	0.60	0.50	0.40	0.30	0.20	0.10
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
0.4	0.87	0.78	0.70	0.61	0.50	0.40	0.30	0.21	0.11
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
0.3	0.85	0.77	0.69	0.62	0.51	0.41	0.31	0.21	0.11
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
0.2	0.83	0.75	0.69	0.63	0.52	0.42	0.32	0.22	0.12
	(0.01)	(0.02)	(0.01)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)
0.1	0.76	0.71	0.69	0.66	0.57	0.47	0.37	0.26	0.17
	(0.06)	(0.02)	(0.01)	(0.03)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)

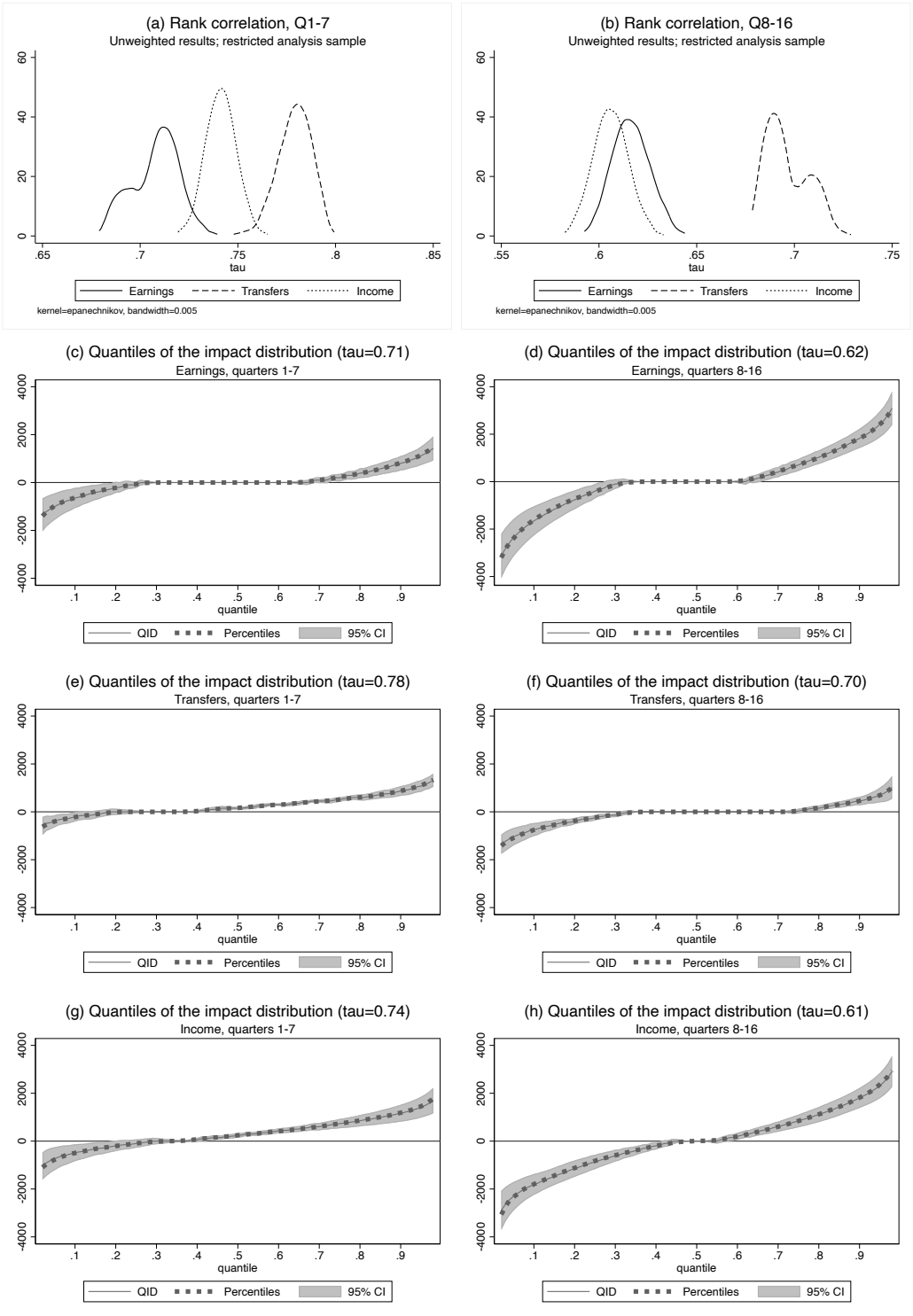
Note: This table presents estimated values of τ_{sp} using simulation data for alternative values of $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ and $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1)$. Each estimate is based on 100 random permutations of 2 × 16,772 observation units. Bootstrap standard errors (100 replications), which are presented in parentheses, were estimated by drawing a random sample with replacement stratified by group membership for each permutation.

Table 2: Pre-Treatment Characteristics

	Means			p-values		
		AFDC				
	Jobs First	U	R	p_U	p_R	p_{UR}
White	0.362	0.348	0.348	0.307	0.305	0.996
Black	0.368	0.371	0.370	0.836	0.857	0.978
Hispanic	0.207	0.216	0.217	0.423	0.416	0.989
Never married	0.654	0.661	0.660	0.624	0.663	0.957
Div/wid/sep/living apart	0.332	0.327	0.328	0.715	0.753	0.960
HS dropout	0.350	0.334	0.335	0.242	0.279	0.933
HS diploma/GED	0.583	0.604	0.603	0.155	0.168	0.967
More than HS diploma	0.066	0.062	0.061	0.557	0.504	0.935
More than two children	0.235	0.214	0.213	0.089	0.079	0.953
Mother younger than 25	0.289	0.297	0.297	0.573	0.547	0.968
Mother age 25-34	0.410	0.418	0.417	0.609	0.639	0.967
Mother older than 34	0.301	0.286	0.286	0.264	0.267	0.996
Recipient (stock) sample	0.624	0.593	0.592	0.029	0.024	0.942
Average quarterly values						
Earnings	678.91	785.90	785.97	0.009	0.010	0.999
AFDC payments	890.82	835.11	832.47	0.015	0.011	0.907
Food stamps	352.12	339.35	338.45	0.156	0.129	0.918
Fraction of quarters with						
Any earnings	0.322	0.351	0.351	0.006	0.006	0.977
Any AFDC payments	0.573	0.544	0.543	0.026	0.021	0.934
Any food stamps	0.607	0.598	0.597	0.485	0.443	0.943
N	2,396	2,407	2,396			

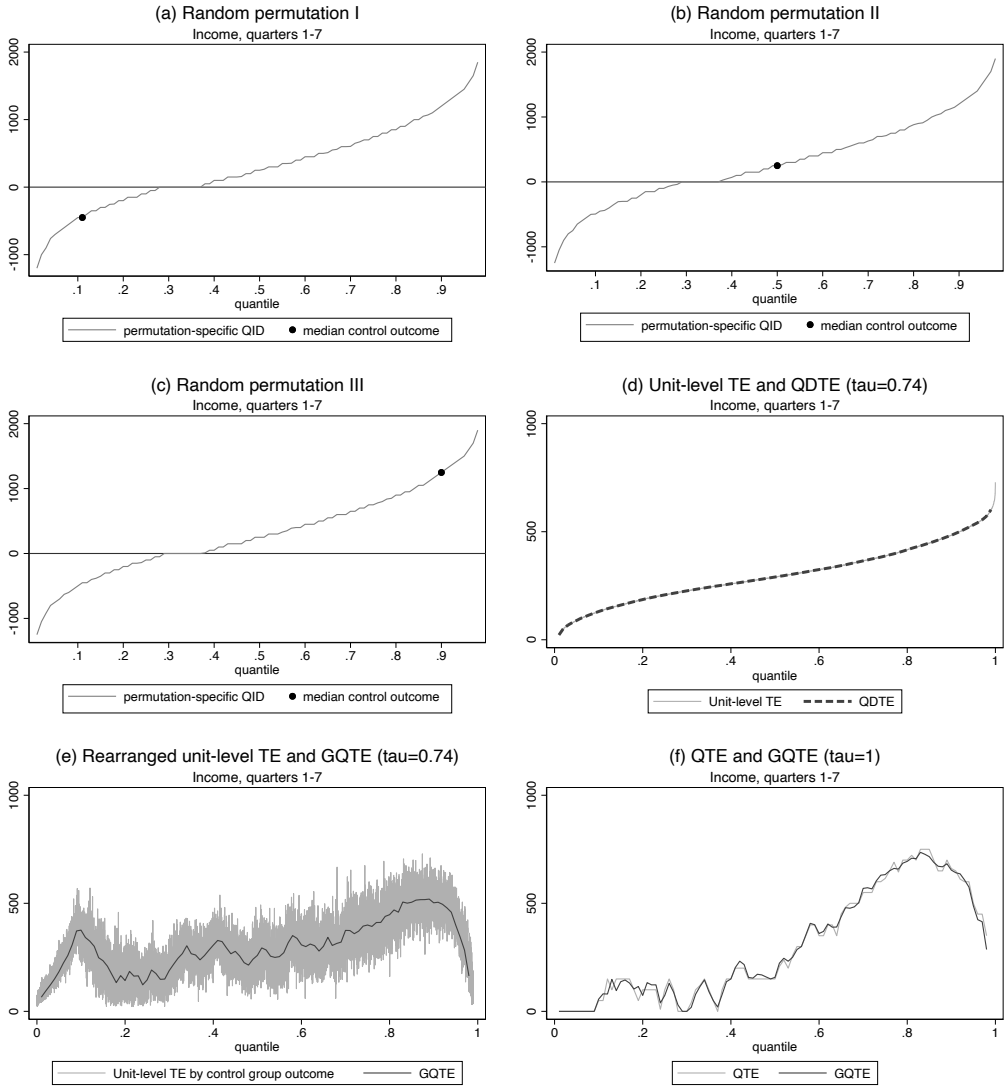
Note: AFDC-U: unrestricted AFDC sample; AFDC-R: restricted AFDC sample. p-values refer to the comparison of means between groups. p_U: Jobs First sample vs. AFDC-U sample. p_R: Jobs First sample vs. AFDC-R sample. p_{UR}: AFDC-U sample vs. AFDC-R sample. Several variables have missing values for a small number of observations (marital status: 220 observations, educational attainment: 288 observations, number of children: 161 observations).

FIGURE 1: RANK CORRELATION DISTRIBUTIONS AND QUANTILES OF THE IMPACT DISTRIBUTION (QID)



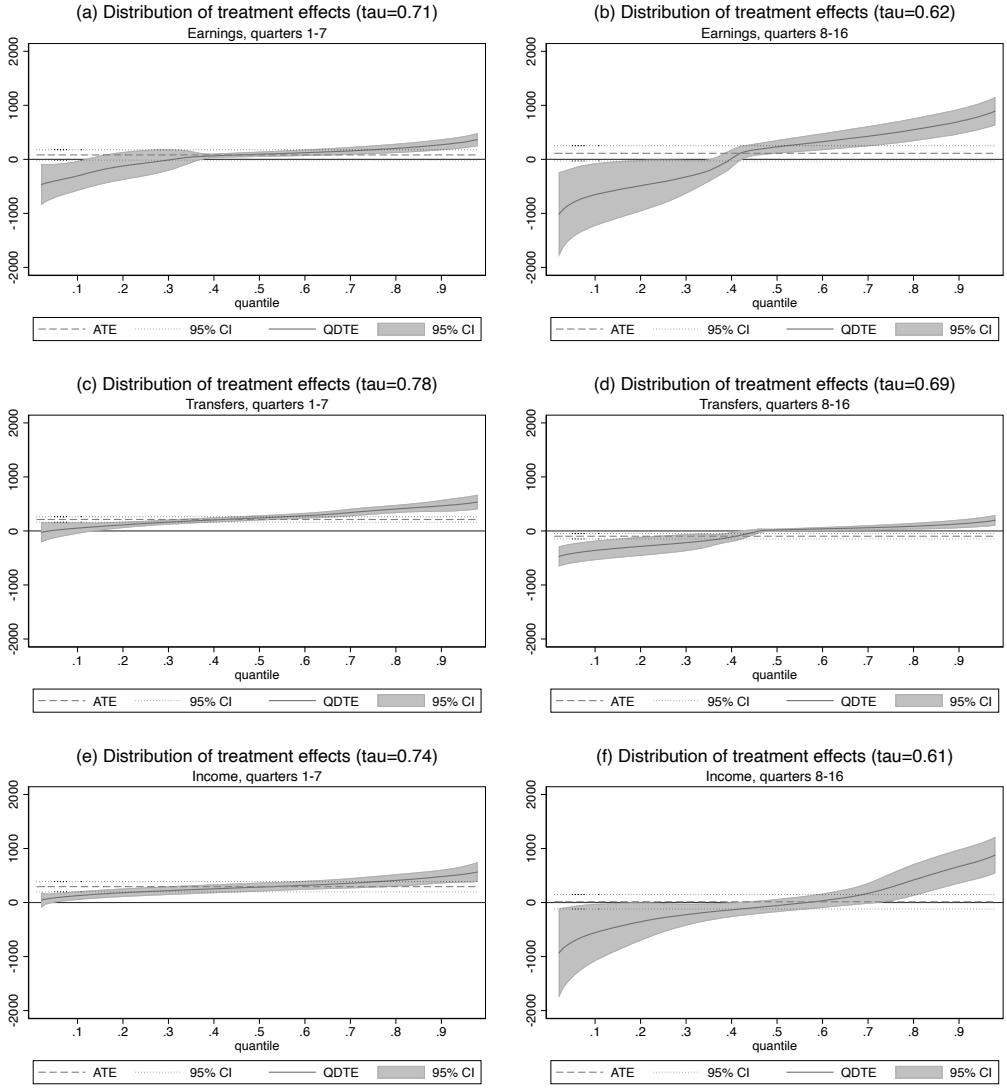
Note: Unweighted results; restricted analysis sample. Figures 1a and 1b: rank correlation distributions based on 100 permutations of treatment outcomes that satisfy the RCP. Kernel density estimates were plotted over a range that exceeds the minimum and maximum values observed by ±0.0075. Figures 1c-1h: Confidence intervals are based on clustered bootstrap standard errors (100 replications) to account for repeated observations. Top and bottom percentiles not included due to high sampling variability.

Figure 2: Relation between QID, QDTE and GQTE, Income, Quarters 1-7



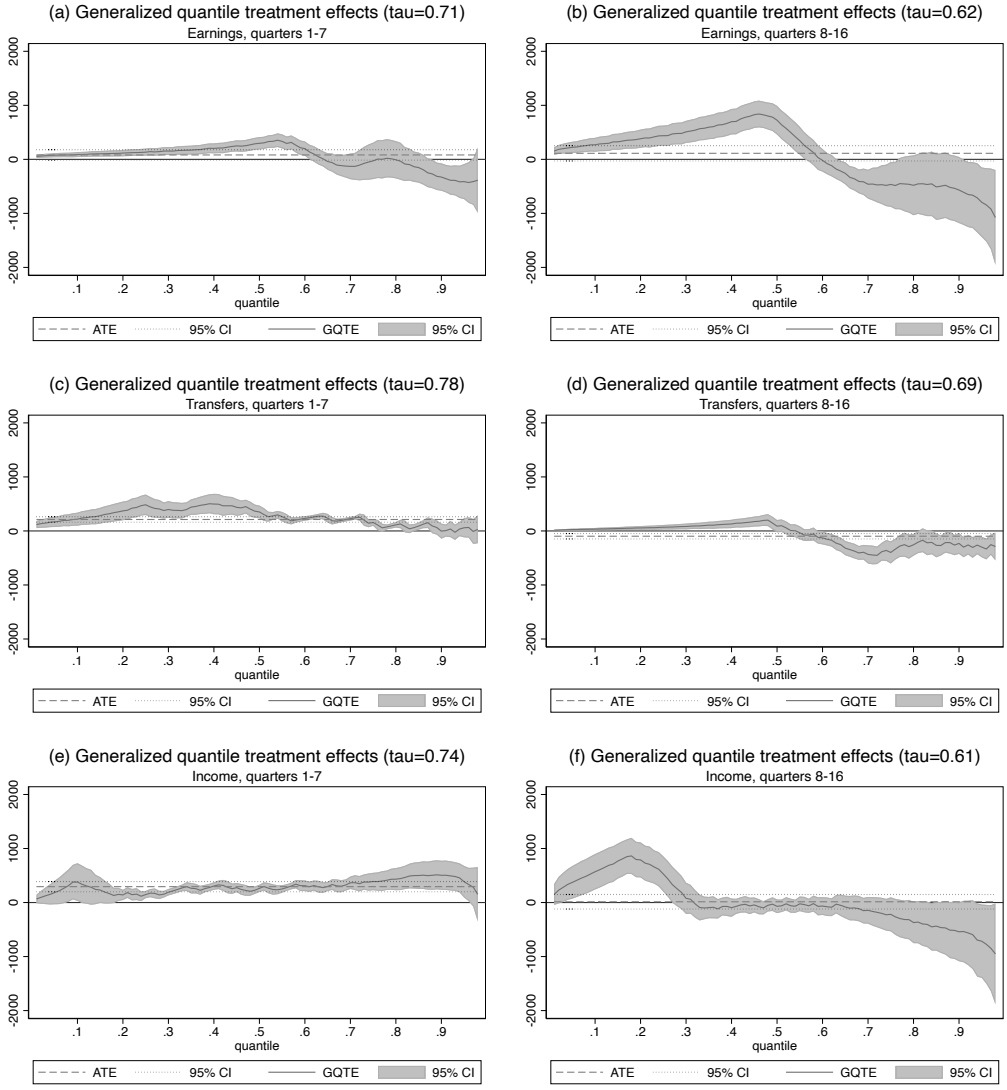
Note: Unweighted results; restricted analysis sample. Figures 2a-2c: QID for selected random permutations of treatment outcomes. Figures 2d-2e: Average unit-level treatment effects based on 100 random permutations of treatment outcomes. Figure 2f: Results based on rank invariance assumption. Figures 2a-2c and 2e-2f: Top percentile not included due to high sampling variability.

FIGURE 3: QUANTILES OF THE DISTRIBUTION OF TREATMENT EFFECTS (QDTE)



Note: Weighted results; unrestricted analysis sample. Confidence intervals based on clustered bootstrap standard errors (100 replications) to account for repeated observations. Top and bottom percentiles not included due to high sampling variability.

Figure 4: Generalized Quantile Treatment Effects (GQTE)



Weighted results; unrestricted analysis sample. Confidence intervals based on clustered bootstrap standard errors (100 replications) to account for repeated observations. Top percentile not included due to high sampling variability.

Table 3: Winners and Losers: Proportion of Positive and Negative Treatment Effects

	Earnings				Transfers				Income			
	Q1-7		Q8-16		Q1-7		Q8-16		Q1-7		Q8-16	
	(1) QTE	(2) GQTE	(3) QTE	(4) GQTE	(5) QTE	(6) GQTE	(7) QTE	(8) GQTE	(9) QTE	(10) GQTE	(11) QTE	(12) GQTE
$\widehat{\tau}$	1	0.71	1	0.62	1	0.78	1	0.69	1	0.74	1	0.61
% > 0												
Total	34.3	68.7	36.4	59.6	77.8	97.0	1.0	54.5	85.9	99.0	43.4	31.3
$p < 0.05$	31.3	61.6	27.3	56.6	64.6	83.8	0.0	49.5	66.7	81.8	23.2	24.2
$p < 0.01$	22.2	60.6	22.2	55.6	60.6	82.8	0.0	46.5	59.6	75.8	16.2	21.2
$p < 0.001$	18.2	58.6	20.2	53.5	55.6	78.8	0.0	10.1	52.5	71.7	6.1	15.2
% < 0												
Total	12.1	31.3	15.2	40.4	0.0	3.0	47.5	45.5	0.0	1.0	26.3	68.7
$p < 0.05$	1.0	9.1	0.0	22.2	0.0	0.0	33.3	37.4	0.0	0.0	7.1	6.1
$p < 0.01$	0.0	4.0	0.0	7.1	0.0	0.0	26.3	26.3	0.0	0.0	2.0	0.0
$p < 0.001$	0.0	0.0	0.0	3.0	0.0	0.0	17.2	15.2	0.0	0.0	0.0	0.0

Note: Weighted results; unrestricted analysis sample. $\widehat{\tau}$ represents the assumed or estimated rank correlation coefficient between potential treatment and control outcomes. Proportion of positive and negative treatment effects: 'Total' refers to the overall proportion of treatment effects, irrespective of their significance levels. Proportions of treatment effects denoted by $p < 0.05$, $p < 0.01$ and $p < 0.001$ are statistically significant at the 5%, 1% and 0.1% levels, respectively. Bootstrap standard errors (100 replications) were clustered to account for repeated observations.

Table 4: Total Gains and Losses, and Standard Deviation of Treatment Effects

	Earnings				Transfers				Income			
	Q1-7		Q8-16		Q1-7		Q8-16		Q1-7		Q8-16	
	(1) QTE	(2) GQTE	(3) QTE	(4) GQTE	(5) QTE	(6) GQTE	(7) QTE	(8) GQTE	(9) QTE	(10) GQTE	(11) QTE	(12) GQTE
$\widehat{\tau}$	1	0.71	1	0.62	1	0.78	1	0.69	1	0.74	1	0.61
Gains (USD '000s)												
Total	276	259	382	649	517	604	2	102	742	704	190	383
$p < 0.05$	269	254	354	640	493	589	0	96	704	621	137	352
$p < 0.01$	203	251	317	634	480	586	0	92	668	580	105	328
$p < 0.001$	184	241	301	616	464	575	0	37	633	560	43	260
Losses (USD '000s)												
Total	53	179	43	516	0	1	246	287	0	24	108	439
$p < 0.05$	7	80	0	344	0	0	215	258	0	0	51	107
$p < 0.01$	0	37	0	70	0	0	187	199	0	0	14	0
$p < 0.001$	0	0	0	31	0	0	146	131	0	0	0	0
σ	201.7	244.3	272.6	578.2	212.9	146.9	147.4	188.9	272.6	171.0	166.1	498.2

Note: Weighted results; unrestricted analysis sample. $\widehat{\tau}$ represents the assumed or estimated rank correlation coefficient between potential treatment and control outcomes. Gains and losses: 'Total' refers to cumulative gains and losses, irrespective of their significance levels. Gains and losses denoted by $p < 0.05$, $p < 0.01$ and $p < 0.001$ are statistically significant at the 5%, 1% and 0.1% levels, respectively. σ denotes the standard deviation of (generalized) quantile treatment effects. Bootstrap standard errors (100 replications) were clustered to account for repeated observations.

References

Abadie, A. (2002): "Bootstrap Tests of Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 97, 284–292.

Abbring, J. and J. J. Heckman (2007): "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, Amsterdam, vol. 6B, chap. 72, 5145–5303.

Ashenfelter, O. (1983): "Determining Participation in Income-Tested Social Programs," *Journal of the American Statistical Association*, 78, 517–525.

Athey, S. and G. Imbens (2016): "Recursive Partitioning for Heterogeneous Causal Effects," *Proceedings of the National Academy of Sciences (PNAS)*, 113, 7353–7360.

Bitler, M., J. Gelbach, and H. Hoynes (2005): "Distributional impacts of the self-sufficiency project," Working Paper 11626, NBER, Cambridge, MA.

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

——— (2017): "Can Variation in Subgroups' Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment," *Review of Economics and Statistics*, 99, 683–697.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernández-Val (2018): "Generic Machine Learning Inference on Heterogenous Treatment Effects in Randomized Experiments," NBER Working Paper Series 24678, National Bureau of Economic Research (NBER), Cambridge, MA.

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

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2008): "Nonparametric Tests for Treatment Effect Heterogeneity," *Review of Economics and Statistics*, 90, 389–405.

Doksum, K. (1974): "Empirical Probability Plots and Statistical Inference for Nonlinear Models," *The Annals of Statistics*, 2, 267–277.

Fan, Y. and S. Park (2010): "Sharp Bounds on the Distribution of Treatment Effects and their Statistical Inference," *Econometric Theory*, 26, 931–951.

Fan, Y. and S. S. Park (2009): "Partial Identification of the Distribution of Treatment Effects and its Confidence Sets," *Advances inEconometrics*, 25, 3–70.

Fan, Y. and J. Wu (2010): "Partial Identification of the Distribution of Treatment Effects in Switching Regime Models and Its Confidence Sets," *The Review of Economic Studies*, 77, 1002–1041.

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

Firpo, S. and G. Ridder (2019): "Partial Identification of the Treatment Effect Distribution and its Functionals," *Journal of Econometrics*, 213, 210–234.

Frandsen, B. R. and L. J. Lefgren (2021): "Partial Identification of the Distribution of Treatment Effects with an Application to the Knowledge is Power Program (KIPP)," *Quantitative Economics*, 12.

Frank, M. J., R. B. Nelsen, and B. Schweizer (1987): "Best-possible Bounds for the Distribution of a Sum - A Problem of Kolmogorov," *Probability Theory and Related Fields*, 74, 199–211.

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

Gibbons, J. D. and S. Chakraborti (2011): *Nonparametric Statistical Inference*, Boca Raton, FL: Chapman and Hall/CRC, 5th ed.

Gillitzer, C. and M. Sinning (2020): "Nudging Businesses to Pay their Taxes: Does Timing Matter?" *Journal of Economic Behavior and Organization*, 169, 284–300.

Hahn, P. R., J. S. Murray, and C. Carvalho (2020): "Bayesian Regression Tree Models for Causal Inference: Regularization, Confounding, and Heterogeneous Effects," *Bayesian Analysis*, 15, 965–1056.

Heckman, J. J., H. Ichimura, and P. Todd (1998): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261–294.

Heckman, J. J. and R. Robb (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman and B. Singer, Cambridge University Press, vol. 10, 156–245.

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

Heckman, J. J. and E. J. Vytlacil (2007): "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Economic Estimators to Evaluate Social Programs and to Forecast Their Effects in New Environments," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, Elsevier, Amsterdam, vol. 6B, chap. 71, 4875–5143.

Hill, J. (2011): "Bayesian Nonparametric Modeling for Causal Inference," *Journal of Computational and Graphical Statistics*, 20, 217–240.

Imai, K. and M. Ratkovic (2013): "Estimating Treatment Effect Heterogeneity in Randomized Program Evaluation," *The Annals of Applied Statistics*, 7, 443–470.

Imbens, G. W. and D. B. Rubin (2015): *Causal Inference for Statistics, Social, and Biomedical Sciences*, Cambridge Books, Cambridge University Press.

Kendall, M. G. (1938): "A New Measure of Rank Correlation," *Biometrika*, 30, 81–93.

Kendall, M. G. and J. D. Gibbons (1990): *Rank Correlation Methods*, New York: Oxford University Press, 5th ed.

Kim, J. H. (2018): "Identifying the Distribution of Treatment Effects under Support Restrictions," arXiv, e-print 1410.5885.

Kline, P. and M. Tartari (2016): "Bounding the Labor Supply Response to a Randomized Welfare Experiment: A Revealed Preference Approach," *American Economic Review*, 106, 972–1014.

Künzel, S. R., J. S. Sekhon, P. J. Bickel, and B. Yu (2019): "Metalearners for Estimating Heterogeneous Treatment Effects using Machine Learning," *Proceedings of the National Academy of Sciences (PNAS)*, 116, 4156–4165.

Lechner, M. (1999): "Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification," *Journal of Business and Economic Statistics*, 17, 74–90.

Lee, M. J. (2006): *Micro-Econometrics for Policy, Program and Treatment Effects*, Oxford University Press.

Lee, S. (2022): "Partial Identification and Inference for Conditional Distributions of Treatment Effects," arXiv, e-print 2108.00723.

Lee, S. and A. M. Shaikh (2013): "Multiple Testing and Heterogeneous Treatment Effects: Re-evaluating the Effects of Progresa on School Enrollment," *Journal of Applied Econometrics*, 29, 612–626.

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

List, J. A., A. M. Shaikh, and Y. Xu (2019): "Multiple Hypothesis Testing in Experimental Economics," *Experimental Economics*, 22, 773–793.

Makarov, G. D. (1982): "Estimates for the Distribution Function of a Sum of Two Random Variables When the Marginal Distributions are Fixed," *Theory of Probability and its Applications*, 26, 803–806.

Manski, C. (1997): "The Mixing Problem in Programme Evaluation," *Review of Economic Studies*, 64, 537–553.

Nie, X. and S. Wager (2021): "Quasi-Oracle Estimation of Heterogeneous Treatment Effects," *Biometrika*, 108, 299–319.

Parzen, E. (1979): "Nonparametric Statistical Data Modeling," *Journal of the American Statistical Association*, 74, 105–121.

Powers, S., J. Qian, K. Jung, A. S. N. H. Shah, T. Hastie, and R. Tibshirani (2018): "Some Methods for Heterogeneous Treatment Effect Estimation in High Dimensions," *Statistics in Medicine*, 37, 1767–1787.

Rosenbaum, P. and D. B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41–55.

Rubin, D. (1990): "Formal Modes of Statistical Inference for Causal Effects," *Journal of Statistical Planning and Inference*, 25, 279–292.

Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Non-Randomized Studies," *Journal of Educational Psychology*, 66, 688–701.

Rüschendorf, L. (1982): "Random Variables with Maximum Sums," *Advances in Applied Probability*, 14, 623–632.

Russel, T. M. (2021): "Sharp Bounds on Functionals of the Joint Distribution in the Analysis of Treatment Effects," *Journal of Business and Economic Statistics*, 39, 532–546.

Shalit, U., F. D. Johansson, and D. Sontag (2017): "Estimating Individual Treatment Effect: Generalization Bounds and Algorithms," *Proceedings of Machine Learning Research*, 3076–3085.

Su, X., C.-L. Tsai, H. Wang, D. M. Nickerson, and B. Li (2009): "Subgroup Analysis via Recursive Partitioning," *Journal of Machine Learning Research*, 10, 141–158.

Wager, S. and S. Athey (2018): "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, 113, 1228–1242.

Williamson, R. C. and T. Downs (1990): "Probabilistic Arithmetic I. Numerical Methods for Calculating Convolutions and Dependency Bounds," *International Journal of Approximate Reasoning*, 4, 89–158.

Appendix A – Additional results

1. QTE results under rank invariance

Figure A1 displays the replication results of conventional QTE estimates originally presented in Bitler et al. (2006), using the approach of Firpo (2007). In contrast to Bitler et al. (2006), who report 90 percent confidence intervals, we present 95 percent confidence intervals throughout the paper. While we were unable to replicate the exact confidence intervals of Bitler et al. (2006) due to their reliance on bootstrap standard errors, we adopt their approach of estimating clustered bootstrap standard errors to account for the presence of repeated observations.

2. Intuition behind the identification of τ_{sp}

To provide context for the identification of τ_{sp}, it is useful to examine the values of $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ and $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1)$ we would expect to observe if τ_{sp} was known. Table A1 presents simulated values of $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1)$ corresponding to various values of $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ and τ_{sp} based on averaging over 100 random permutations of $2 \times 16{,}772$ observation units ($N_1 = N_0 = N/2 = 16{,}772$), which matches our restricted analysis sample. For simplicity, we assume no ties and consider a model with a single continuous covariate vector $\mathbf{X}_1$ of length $N \times 1$. We choose 100 random permutations of $\mathbf{X}_1$ for different values of $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0) = \tau(\mathbf{Y}_0, \mathbf{X}_1)$. We also choose 100 random permutations of $\mathbf{Y}_1$ for various values of τ_{sp}. We can use the permutations of $\mathbf{Y}_1$ to obtain simulated values of $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1) = \tau(\mathbf{Y}_1, \mathbf{X}_1)$ because the ranks of $\mathbf{X}_1$ in the treatment group sample are the same as those in the control group sample.

We initially consider the case in which the rank invariance assumption holds, $\tau_{\text{sp}} = 1$. In this case, we expect to observe an identical rank correlation coefficient between actual and predicted outcomes in both groups, $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1) = \tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$, irrespective of how well we predict the ranks of the control outcomes. The values in Column (1) of Table A1 confirm this expectation: $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1)$ equals $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ if $\tau_{\text{sp}} = 1$. Similarly, for cases in which the set of covariates predicts the ranks of the control outcomes perfectly, $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0) = 1$, any deviation of $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1)$ from 1 is expected to emanate from an imperfect rank correlation between potential treatment and control outcomes.

The first row of Table A1 validates this point by showing that $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1)$ equals τ_{sp} if $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0) = 1$.

We observe that for the remaining cases in which $0 \leq \tau_{sp} < 1$ and $0 \leq \tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0) < 1$, $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1)$ deviates from τ_{sp} because our predictions of the ranks of control outcomes are not perfect. At the same time, $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1)$ deviates from $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ because the rank invariance assumption does not hold. The simulated values presented in Table A1 form a symmetric matrix because deviations of τ_{sp} from 1 have the same impact on $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1)$ as deviations of $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ from 1. Table A1 focuses on cases in which τ_{sp} and $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ are non-negative. Extending the analysis to cases in which $\tau_{sp} < 0$ and $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0) < 0$ produces a mirror image of the positive values presented in Table A1. Similarly, analyzing cases in which $\tau_{sp} < 0$ and $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0) \geq 0$ or $\tau_{sp} \geq 0$ and $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0) < 0$ produces a negative mirror image of the positive values in Table A1.

3. Testing the predictive strength of covariates

We test the predictive strength of covariates by creating 100 simulated datasets. Each simulated dataset s, $s = \{1, \ldots, 100\}$, contains observed outcomes and predicted values, with permutations of observation units tailored to a specific target value τ_{sp} and a specific rank correlation coefficient $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$. We compare the target values to estimated rank correlation coefficients. For simplicity, we use our restricted analysis sample to create simulated datasets with equal treatment and control group sizes.[26] Within each dataset, we find a random permutation of $\mathbf{Y}_1$ that satisfies the RCP. We use the permutations derived from the 100 datasets to estimate $\widehat{\tau} = (1/100) \sum_{s=1}^{100} \tau(\mathbf{\Pi}_s \mathbf{Y}_1, \mathbf{Y}_0)$, where $\mathbf{\Pi}_s$, $s = \{1, \ldots, 100\}$, is a $(N_1 \times N_1)$-permutation matrix.

To ascertain whether our estimate deviates significantly from the target value, we conduct a hypothesis test with the null hypothesis $H_0 : \tau_{sp} - \widehat{\tau} = 0$ against the alternative hypothesis $H_1 : \tau_{sp} - \widehat{\tau} \neq 0$. Our test results, which are presented in Panel A of Table A2, confirm that our predictors generally possess sufficient strength to yield unbiased estimates of the rank correlation between potential treatment and control outcomes. However, we do observe a small but statistically significant bias for transfers during the pre-time limit period.

[26] It is possible to accommodate different treatment and control group sizes by using an approach akin to that employed in Section 4.4.2.

To explore how deviations in estimated rank correlation coefficients from target values affect our QDTE estimates, we obtain QDTE estimates corresponding to different values of τ_{sp}. We use a two-sided Kolmogorov-Smirnov test to determine whether differences between alternative rank correlation coefficients translate into significant differences between QDTE functions. The test results, which are presented in Panel B of Table A2, indicate that the differences between estimated rank correlation coefficients and target values have no significant impact on our QDTE estimates.

4. Implications of assuming equally likely permutations

The primary identifying assumption in this paper is that all permutations satisfying the RCP are equally likely. While it is not possible to test this assumption, we can explore its implications for GQTE estimates, which are derived from averages of permutation-specific GQTE estimates. In Figure A2, we present permutation-specific GQTEs for income during the pre-time limit period. Figures A2a-A2e illustrate the density functions of permutation-specific GQTEs at selected quantiles. Our findings indicate that these density functions converge toward approximately normal distributions as the number of permutations increases. Moreover, we observe that even a relatively modest number of permutations ($P = 100$) yields a reasonably accurate approximation.

Figure A2f depicts the GQTE and the underlying density functions of permutation-specific GQTEs across percentiles. We observe that 50 percent of the probability mass of the underlying density functions is concentrated in close proximity to the GQTE. This observation provides strong evidence in support of calculating averages under the assumption of equally likely permutations. Our findings also emphasize the limitations on knowledge generation when deriving bounds from permutations with extremely low probability of occurrence. Moreover, our results highlight the potential for developing tests to compare alternative identifying assumptions. For instance, one could compare GQTE based on equally likely permutations to GQTE based on an alternative weighting scheme. The exploration of this issue remains a subject for future research.

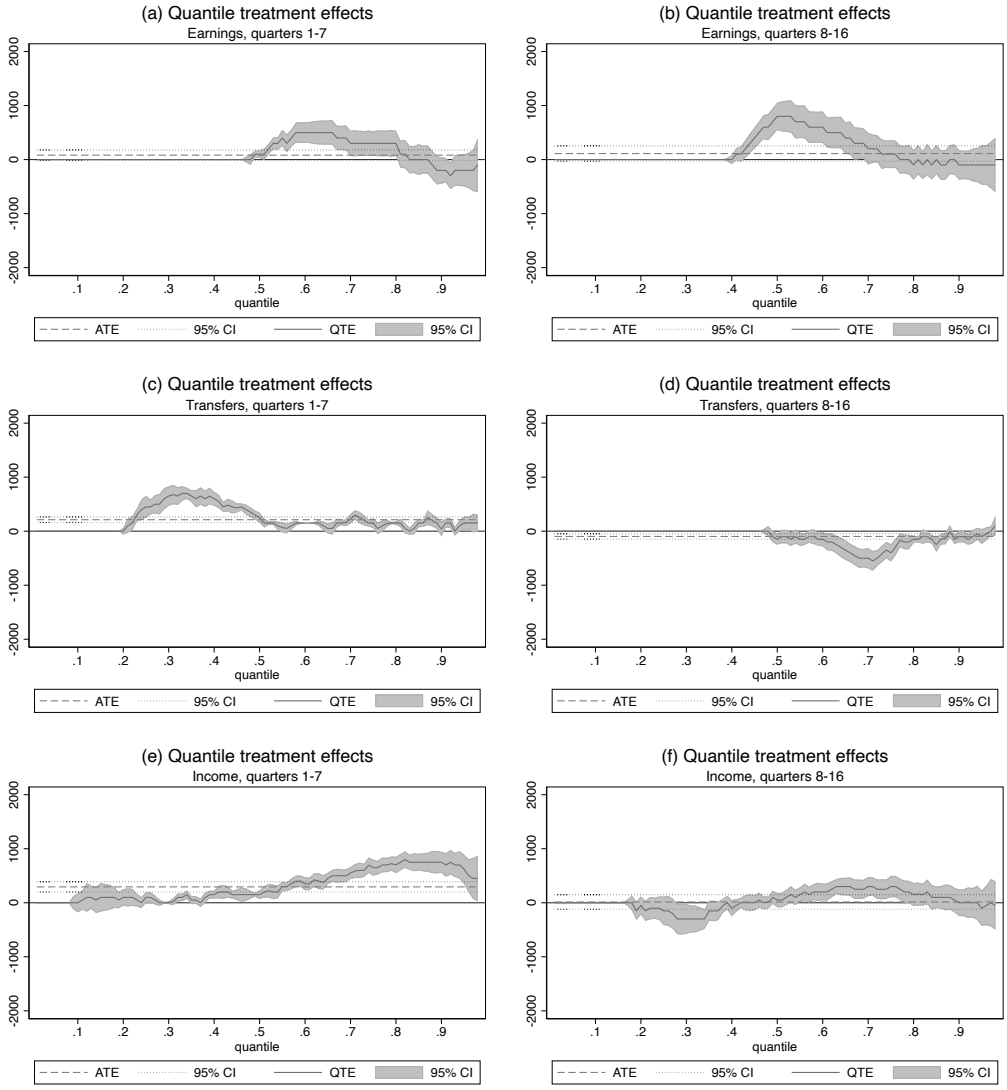
5. Impact of sample restriction and reweighting

Figure A3 presents QDTE, GQTE, and QTE estimates for income during the pre-time limit period, illustrating the consequences of two factors: imposing a sample restriction on the AFDC sample and employing a reweighting approach to control for covariates. We find that excluding 11 randomly selected observations from the AFDC sample to obtain a restricted analysis sample with balanced treatment and control group sizes has minimal influence on our results. We also observe that, due to random assignment of observation units to treatment and control groups, the use of a reweighting approach does not change our results qualitatively.

Figures A3c and A3d depict unweighted and weighted QTE and GQTE estimates under rank invariance for our unrestricted analysis sample.[27] These results confirm that our approach yields very similar results to the approach of Firpo (2007) under rank invariance. While the QTE estimates presented in Figure A3d involve calculating reweighted quantiles of unweighted outcomes, our GQTE estimates are based on calculating unweighted quantiles of reweighted outcomes. We have no clear preference for one approach over another if the rank invariance assumption holds. However, in contrast to the approach of Firpo (2007), our approach remains applicable in situations where the rank invariance assumption is violated.

[27] Unweighted QTE and GQTE estimates under rank invariance for our restricted analysis sample are presented in Figure 2f.

Figure A1: Quantile Treatment Effects under Rank Invariance



Note: Weighted results; unrestricted analysis sample. ATE: Confidence intervals are based on clustered standard errors to account for repeated observations. QTE: Confidence intervals are based on clustered bootstrap standard errors (100 replications) to account for repeated observations. Top percentile not included due to high sampling variability.

TABLE A1: SIMULATED VALUES OF $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1)$ FOR GIVEN VALUES OF $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ AND τ_{sp}

	τ_{sp}										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
$\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0.0
1.0	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00
0.9	0.90	0.86	0.78	0.69	0.59	0.49	0.39	0.30	0.20	0.10	0.00
0.8	0.80	0.78	0.72	0.64	0.56	0.47	0.38	0.28	0.19	0.10	0.00
0.7	0.70	0.69	0.64	0.59	0.51	0.43	0.35	0.26	0.18	0.09	0.00
0.6	0.60	0.59	0.56	0.51	0.46	0.39	0.32	0.24	0.16	0.08	0.00
0.5	0.50	0.49	0.47	0.43	0.39	0.33	0.27	0.21	0.14	0.07	0.00
0.4	0.40	0.39	0.38	0.35	0.32	0.27	0.22	0.17	0.12	0.06	0.00
0.3	0.30	0.30	0.28	0.26	0.24	0.21	0.17	0.13	0.09	0.05	0.00
0.2	0.20	0.20	0.19	0.18	0.16	0.14	0.12	0.09	0.06	0.03	0.00
0.1	0.10	0.10	0.09	0.09	0.08	0.07	0.06	0.04	0.03	0.02	-0.00
0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	0.00

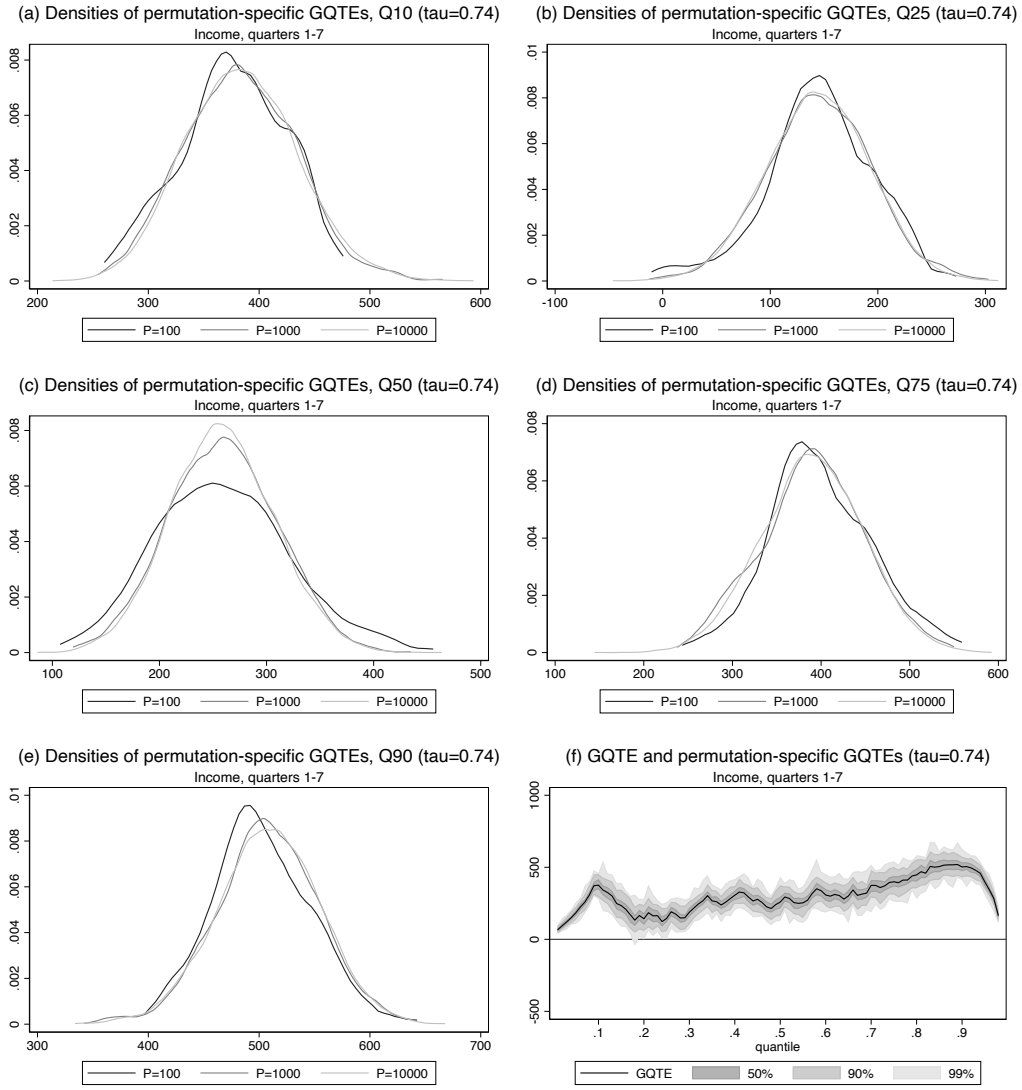
Note: This table presents simulated values of $\tau(\mathbf{Y}_1, \widehat{\mathbf{Y}}_1)$ using simulation data for alternative values of $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$ and τ_{sp}. Each simulated value is based on 100 random permutations of 2 × 16,772 observation units.

Table A2: Simulation-based test for strength of model predictors and Kolmogorov-Smirnov test for equality of distribution functions

	Earnings		Transfers		Income	
	(1)	(2)	(3)	(4)	(5)	(6)
	Q1-7	Q8-16	Q1-7	Q8-16	Q1-7	Q8-16
Panel A.						
$\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$	0.35	0.31	0.35	0.30	0.32	0.27
τ_{sp}	Estimate					
0.80			0.78 [0.00]			
0.76					0.74 [0.06]	
0.72	0.71 [0.46]					
0.70				0.69 [0.17]		
0.61		0.62 [0.29]				
0.60						0.61 [0.20]
Panel B.						
τ_1, τ_2	Kolmogorov-Smirnov test (p-values)					
0.80, 0.78			0.994			
0.76, 0.74					0.696	
0.72, 0.71	1.000					
0.70, 0.69				1.000		
0.61, 0.62		0.577				
0.60, 0.61						1.000

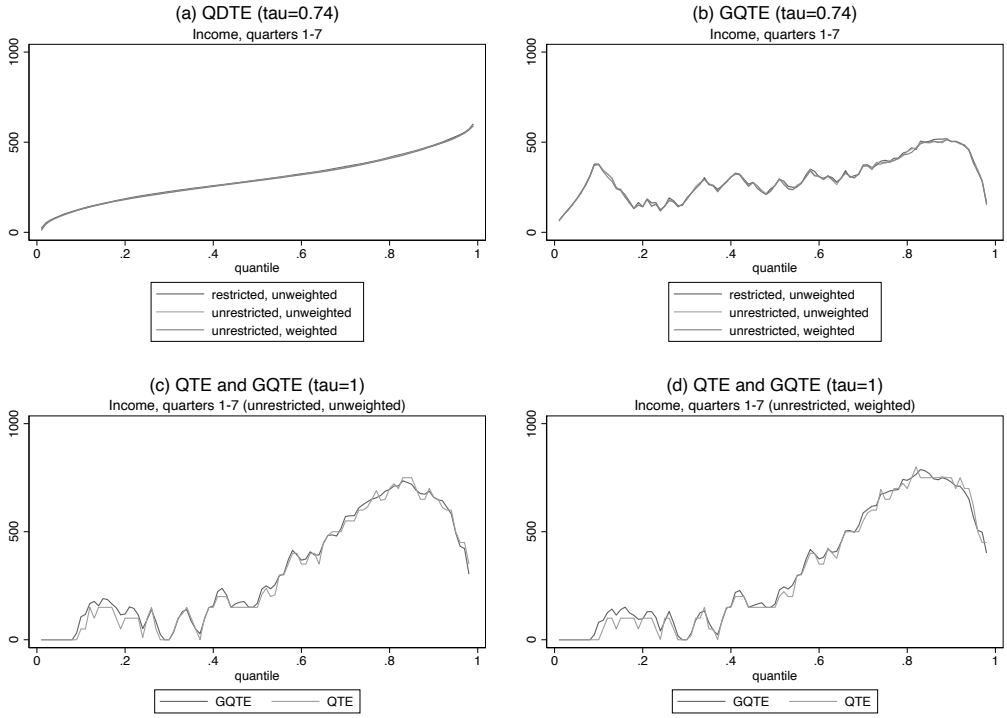
Note: Panel A: We generate 100 simulation datasets based on our restricted analysis sample. Each dataset contains random permutations of observation units consistent with a specific target value τ_{sp} and a given rank correlation between actual and predicted control outcomes $\tau(\mathbf{Y}_0, \widehat{\mathbf{Y}}_0)$. We use these datasets to estimate $\widehat{\tau}$ while varying τ_{sp} over the range from 0.6 to 0.8. Panel A presents the target values and their corresponding estimates, which conform with the estimates shown in Figures 1a and 1b. We test whether the target values deviate significantly from the corresponding estimates. The associated p-values for the two-sided test comparing target values and estimates are reported in brackets. Panel B: Two-sided Kolmogorov-Smirnov test. We compare QDTE estimates under two different assumed values of the rank correlation coefficient between potential treatment and control outcomes, τ_1 and τ_2. Panel B presents the associated p-values, which were calculated using a counting algorithm as described in Gibbons and Chakraborti (2011).

Figure A2: GQTE and permutation-specific GQTEs, Income, quarters 1-7



Note: GQTE estimates are derived from averaging permutation-specific GQTE estimates. Figures A2a-A2e: density functions of permutation-specific GQTEs at selected quantiles. Figure A2f: GQTE and underlying density functions of permutation-specific GQTEs across percentiles. Top percentile not included due to high sampling variability.

FIGURE A3: IMPACT OF SAMPLE RESTRICTION AND REWEIGHTING, INCOME, QUARTERS 1-7



Note: Figures A3b-A3d: Top percentile not included due to high sampling variability.

Appendix B – Monte Carlo simulation

We use Monte Carlo simulations to study the finite sample behavior of our estimators. We consider two normally distributed $(N \times 1)$-vectors, $\mathbf{Y}_1 \sim N(\mu_1, \sigma_1^2)$ and $\mathbf{Y}_0 \sim N(\mu_0, \sigma_0^2)$. Moreover, we consider a single $(N \times 1)$-covariate vector $\mathbf{X}_1$, which contains the ranks of a random variable without ties.

1. Rank correlation coefficients

We use random permutations of $\mathbf{Y}_1$ to generate simulation datasets in which the rank correlation coefficient $\tau_{\text{sp}} = \tau(\mathbf{Y}_1, \mathbf{Y}_0)$ ranges from 0 to 0.9. We also use random permutations to vary the predictive strength of $\mathbf{X}_1$ by adjusting the rank correlation coefficient $\tau_c = \tau(\mathbf{Y}_0, \mathbf{X}_1)$ over the range from 0.1 to 0.9. For each combination of τ_{sp} and τ_c, we create 100 simulation datasets with a sample size of 500 (250 observation units in each group) and an additional 100 simulation datasets with a sample size of 5,000 (2,500 observation units in each group). Within each simulation dataset, we estimate τ_{sp} using 100 random permutations of $\mathbf{Y}_1$ satisfying the RCP. We obtain bootstrap standard errors for each estimate using 100 replications. We repeat this process 100 times for each sample and each combination of τ_{sp} and τ_c to obtain our Monte Carlo simulation results.

Tables B1 and B2 present the results for datasets with 500 and 5,000 observation units, respectively. The results confirm the precision of our estimator of τ_{sp}, even when the sample size is relatively small. They also confirm that the estimator is unbiased when τ_c exceeds a certain threshold. Specifically, for $\tau_c \geq 0.4$, our estimator performs well in terms of bias and root mean square error (RMSE). For $\tau_c \geq 0.3$, our estimator maintains its accuracy as long as τ_{sp} remains below or equal to 0.8. We also find that the biases diminish notably as the sample size increases. Tables B3 and B4 report the estimated bootstrap standard errors, their lower and upper 5th percentile, and the 90 percent coverage rate. The results confirm that our estimated bootstrap standard errors are a good representation of the true sampling variation.

2. QID, QDTE and GQTE

We study the finite sample behavior of QID, QDTE, and GQTE estimators. We derive target values for the case in which $\mathbf{Y}_1$ and $\mathbf{Y}_0$ are independent to establish their statistical properties, including unbiasedness, consistency, and asymptotic normality. We also assess the validity of estimated bootstrap standard errors for the QDTE and GQTE estimators.

For $u \in (0,1)$, the quantile function of the QID estimator under independence is given by

$$(\mu_1 - \mu_0) + \sqrt{\sigma_1^2 + \sigma_0^2}\Phi^{-1}(u),$$

where $\Phi(\cdot)$ represents the CDF of the standard normal distribution. Following Fan and Park (2009), we consider a scenario where $\mathbf{Y}_1 \sim N(2,2)$ and $\mathbf{Y}_0 \sim N(1,1)$. We estimate the QID at selected quantiles, including Q10, Q25, Q50, Q75, and Q90, and subsequently compare our estimates to the target values of the quantile function at these quantiles. The respective target values are −1.22, −0.17, 1.00, 2.17, and 3.22.

The quantile function of the QID estimator is based on arranging treatment effects in ascending order, capturing the full amount of heterogeneity in the data, without considering the location of control outcomes. Our objective is to compare each control outcome to a set of treatment outcomes resulting from permutations that have a positive probability of occurrence. Under independence, this means we compare each control outcome to all possible treatment outcomes. Assuming equally likely permutations, we calculate the average over all permutations to obtain the GQTE. Therefore, the quantile function of the GQTE estimator under independence is obtained by comparing the expected value of the treatment outcome, μ_1, to the quantile function of $\mathbf{Y}_0$. The quantile function of $\mathbf{Y}_0$ is given by $\mu_0 + \sigma_0\Phi^{-1}(u)$. Consequently, the quantile function of the GQTE estimator under independence is

$$\mu_1 - (\mu_0 + \sigma_0\Phi^{-1}(u)).$$

In the scenario where $\mathbf{Y}_1 \sim N(2,2)$ and $\mathbf{Y}_0 \sim N(1,1)$, the target values at Q10, Q25, Q50, Q75, and Q90 are 2.28, 1.67, 1.00, 0.33, and −0.28, respectively. To derive

the quantile function of the QDTE estimator, the values of the quantile function of the GQTE estimator have to be rearranged to be monotonically increasing. Therefore, the corresponding target values of the QDTE are −0.28, 0.33, 1.00, 1.67, and 2.28.

Figures B1 and B2 provide evidence of the unbiasedness, consistency, and asymptotic normality of our estimators. We focus on a scenario in which $\tau_c = 0.4$. Estimation is based on 100 random permutations of observation units. Figure B1 illustrates the convergence of our estimators towards the relevant target values at sample sizes of 500, 5,000, and 50,000. Figure B2 illustrates that the distributions of a recentered and rescaled version of our estimators become increasingly indistinguishable from a normal distribution as the sample size increases. This finding underscores the asymptotic normality property of our estimators. Figures B3 and B5 present bootstrap standard errors for selected quantiles of QDTE and GQTE estimators. These findings confirm that increased sample size and the availability of highly predictive covariates contribute to enhanced precision. In Figures B4 and B6, we report the 90 percent coverage rates of QDTE and GQTE estimators. These results validate the accuracy of our estimated bootstrap standard errors.

Table B1: Monte Carlo simulation (500 observations; 100 replications): Point estimates of τ_{sp} resulting from 100 random permutations

τ_{sp}	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00
τ_c =.9										
Mean	0.899	0.800	0.699	0.599	0.500	0.400	0.301	0.200	0.099	0.010
Standard deviation	0.008	0.008	0.008	0.007	0.007	0.008	0.009	0.008	0.008	0.001
Bias	-0.001	0.000	-0.001	-0.001	0.000	0.000	0.001	-0.000	-0.001	0.010
RMSE	0.008	0.008	0.008	0.007	0.007	0.008	0.009	0.008	0.008	0.010
τ_c =.8										
Mean	0.909	0.801	0.699	0.602	0.498	0.402	0.301	0.199	0.101	0.013
Standard deviation	0.016	0.016	0.014	0.014	0.014	0.014	0.016	0.015	0.015	0.006
Bias	0.009	0.001	-0.001	0.002	-0.002	0.002	0.001	-0.001	0.001	0.013
RMSE	0.018	0.016	0.014	0.014	0.014	0.015	0.016	0.015	0.015	0.015
τ_c =.7										
Mean	0.912	0.801	0.699	0.604	0.501	0.403	0.297	0.203	0.101	0.018
Standard deviation	0.020	0.027	0.023	0.022	0.019	0.022	0.021	0.023	0.025	0.011
Bias	0.012	0.001	-0.001	0.004	0.001	0.003	-0.003	0.003	0.001	0.018
RMSE	0.023	0.027	0.023	0.023	0.019	0.022	0.022	0.023	0.025	0.021
τ_c =.6										
Mean	0.892	0.801	0.697	0.604	0.498	0.401	0.297	0.200	0.105	0.021
Standard deviation	0.023	0.036	0.031	0.034	0.031	0.029	0.033	0.030	0.036	0.014
Bias	-0.008	0.001	-0.003	0.004	-0.002	0.001	-0.003	-0.000	0.005	0.021
RMSE	0.024	0.036	0.031	0.035	0.031	0.029	0.033	0.030	0.036	0.026
τ_c =.5										
Mean	0.873	0.800	0.699	0.612	0.495	0.405	0.310	0.206	0.109	0.031
Standard deviation	0.023	0.037	0.047	0.045	0.047	0.046	0.045	0.048	0.040	0.025
Bias	-0.027	0.000	-0.001	0.012	-0.005	0.005	0.010	0.006	0.009	0.031
RMSE	0.036	0.037	0.047	0.046	0.048	0.047	0.046	0.048	0.041	0.040
τ_c =.4										
Mean	0.843	0.777	0.698	0.627	0.505	0.413	0.306	0.196	0.106	0.035
Standard deviation	0.023	0.044	0.043	0.069	0.065	0.058	0.061	0.062	0.057	0.033
Bias	-0.057	-0.023	-0.002	0.027	0.005	0.013	0.006	-0.004	0.006	0.035
RMSE	0.062	0.049	0.043	0.075	0.065	0.060	0.061	0.062	0.058	0.048
τ_c =.3										
Mean	0.813	0.758	0.704	0.616	0.509	0.410	0.320	0.217	0.116	0.058
Standard deviation	0.027	0.046	0.061	0.097	0.082	0.094	0.095	0.086	0.070	0.045
Bias	-0.087	-0.042	0.004	0.016	0.009	0.010	0.020	0.017	0.016	0.058
RMSE	0.091	0.062	0.061	0.098	0.083	0.094	0.097	0.088	0.072	0.073
τ_c =.2										
Mean	0.772	0.728	0.671	0.615	0.524	0.438	0.341	0.235	0.155	0.104
Standard deviation	0.024	0.050	0.076	0.092	0.119	0.116	0.129	0.125	0.122	0.094
Bias	-0.128	-0.072	-0.029	0.015	0.024	0.038	0.041	0.035	0.055	0.104
RMSE	0.130	0.087	0.081	0.093	0.122	0.122	0.136	0.129	0.134	0.140
τ_c =.1										
Mean	0.725	0.678	0.655	0.589	0.551	0.481	0.371	0.351	0.282	0.173
Standard deviation	0.024	0.056	0.070	0.109	0.130	0.170	0.210	0.197	0.214	0.168
Bias	-0.175	-0.122	-0.045	-0.011	0.051	0.081	0.071	0.151	0.182	0.173
RMSE	0.177	0.134	0.083	0.110	0.140	0.189	0.221	0.248	0.281	0.241

TABLE B2: MONTE CARLO SIMULATION (5,000 OBSERVATIONS; 100 REPLICATIONS): POINT ESTIMATES OF τ_{sp} RESULTING FROM 100 RANDOM PERMUTATIONS

τ_{sp}	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00
τ_c =.9										
Mean	0.901	0.799	0.699	0.599	0.500	0.401	0.301	0.200	0.100	0.007
Standard deviation	0.003	0.003	0.004	0.003	0.003	0.003	0.003	0.003	0.003	0.001
Bias	0.001	-0.001	-0.001	-0.001	-0.000	0.001	0.001	0.000	-0.000	0.007
RMSE	0.003	0.003	0.004	0.003	0.003	0.003	0.003	0.003	0.003	0.007
τ_c =.8										
Mean	0.907	0.800	0.698	0.599	0.499	0.400	0.301	0.200	0.100	0.009
Standard deviation	0.004	0.006	0.005	0.005	0.004	0.006	0.005	0.005	0.005	0.001
Bias	0.007	0.000	-0.002	-0.001	-0.001	-0.000	0.001	0.000	0.000	0.009
RMSE	0.009	0.006	0.006	0.005	0.005	0.006	0.005	0.005	0.005	0.009
τ_c =.7										
Mean	0.915	0.803	0.698	0.601	0.501	0.399	0.300	0.199	0.100	0.010
Standard deviation	0.009	0.008	0.007	0.008	0.007	0.007	0.008	0.008	0.007	0.002
Bias	0.015	0.003	-0.002	0.001	0.001	-0.001	0.000	-0.001	0.000	0.010
RMSE	0.018	0.009	0.007	0.008	0.007	0.007	0.008	0.008	0.007	0.010
τ_c =.6										
Mean	0.898	0.804	0.698	0.602	0.503	0.400	0.300	0.200	0.103	0.012
Standard deviation	0.006	0.011	0.010	0.010	0.010	0.012	0.010	0.010	0.008	0.004
Bias	-0.002	0.004	-0.002	0.002	0.003	0.000	-0.000	-0.000	0.003	0.012
RMSE	0.007	0.012	0.011	0.010	0.011	0.012	0.010	0.010	0.009	0.013
τ_c =.5										
Mean	0.874	0.795	0.696	0.599	0.501	0.400	0.301	0.203	0.102	0.016
Standard deviation	0.010	0.013	0.011	0.013	0.012	0.013	0.014	0.013	0.012	0.008
Bias	-0.026	-0.005	-0.004	-0.001	0.001	0.000	0.001	0.003	0.002	0.016
RMSE	0.027	0.013	0.012	0.013	0.012	0.013	0.014	0.013	0.012	0.018
τ_c =.4										
Mean	0.858	0.782	0.697	0.605	0.500	0.401	0.304	0.205	0.108	0.019
Standard deviation	0.013	0.015	0.013	0.018	0.020	0.018	0.018	0.018	0.017	0.011
Bias	-0.042	-0.018	-0.003	0.005	0.000	0.001	0.004	0.005	0.008	0.019
RMSE	0.044	0.023	0.014	0.019	0.020	0.018	0.019	0.019	0.019	0.022
τ_c =.3										
Mean	0.834	0.774	0.694	0.616	0.511	0.412	0.311	0.212	0.114	0.026
Standard deviation	0.013	0.019	0.018	0.029	0.024	0.030	0.026	0.025	0.027	0.016
Bias	-0.066	-0.026	-0.006	0.016	0.011	0.012	0.011	0.012	0.014	0.026
RMSE	0.067	0.032	0.019	0.033	0.027	0.032	0.029	0.028	0.030	0.030
τ_c =.2										
Mean	0.793	0.747	0.693	0.628	0.533	0.438	0.331	0.226	0.131	0.041
Standard deviation	0.014	0.023	0.020	0.033	0.042	0.042	0.047	0.042	0.039	0.030
Bias	-0.107	-0.053	-0.007	0.028	0.033	0.038	0.031	0.026	0.031	0.041
RMSE	0.107	0.058	0.021	0.043	0.053	0.056	0.057	0.049	0.050	0.051
τ_c =.1										
Mean	0.729	0.710	0.693	0.658	0.591	0.509	0.400	0.300	0.210	0.116
Standard deviation	0.011	0.016	0.019	0.036	0.064	0.074	0.084	0.083	0.082	0.070
Bias	-0.171	-0.090	-0.007	0.058	0.091	0.109	0.100	0.100	0.110	0.116
RMSE	0.171	0.092	0.020	0.068	0.111	0.132	0.131	0.130	0.137	0.136

Table B3: Bootstrap standard errors from Monte Carlo simulation
(500 observations; 100 replications)

τ_{sp}	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00
τ_c =.9										
Mean	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.002
Upper 5th percentile	0.009	0.009	0.009	0.009	0.008	0.009	0.009	0.009	0.009	0.003
Lower 5th percentile	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.007	0.002
90% coverage rate (%)	94	88	92	91	95	91	88	85	90	97
τ_c =.8										
Mean	0.016	0.016	0.014	0.015	0.015	0.016	0.015	0.015	0.015	0.006
Upper 5th percentile	0.018	0.018	0.016	0.017	0.017	0.017	0.017	0.017	0.017	0.008
Lower 5th percentile	0.014	0.014	0.012	0.013	0.013	0.014	0.014	0.013	0.013	0.005
90% coverage rate (%)	89	88	90	91	94	93	87	90	92	94
τ_c =.7										
Mean	0.020	0.026	0.022	0.023	0.024	0.023	0.022	0.024	0.023	0.010
Upper 5th percentile	0.022	0.029	0.025	0.026	0.026	0.026	0.025	0.026	0.026	0.013
Lower 5th percentile	0.018	0.022	0.020	0.021	0.022	0.021	0.019	0.021	0.021	0.008
90% coverage rate (%)	93	89	89	92	97	91	93	91	87	90
τ_c =.6										
Mean	0.021	0.036	0.032	0.034	0.033	0.032	0.032	0.033	0.031	0.016
Upper 5th percentile	0.023	0.040	0.037	0.038	0.037	0.036	0.036	0.036	0.035	0.019
Lower 5th percentile	0.019	0.032	0.027	0.029	0.028	0.028	0.028	0.029	0.028	0.013
90% coverage rate (%)	86	91	87	90	93	94	89	93	89	95
τ_c =.5										
Mean	0.022	0.042	0.043	0.047	0.044	0.043	0.044	0.044	0.042	0.022
Upper 5th percentile	0.025	0.047	0.048	0.052	0.049	0.049	0.049	0.048	0.047	0.026
Lower 5th percentile	0.020	0.037	0.037	0.041	0.038	0.038	0.039	0.039	0.037	0.018
90% coverage rate (%)	91	96	89	92	88	84	90	86	90	88
τ_c =.4										
Mean	0.024	0.044	0.055	0.062	0.062	0.059	0.060	0.059	0.055	0.034
Upper 5th percentile	0.027	0.049	0.063	0.070	0.069	0.068	0.067	0.066	0.060	0.041
Lower 5th percentile	0.021	0.040	0.048	0.055	0.054	0.051	0.054	0.052	0.049	0.027
90% coverage rate (%)	91	88	96	86	91	90	92	88	90	92
τ_c =.3										
Mean	0.025	0.044	0.063	0.080	0.088	0.087	0.085	0.084	0.074	0.050
Upper 5th percentile	0.028	0.049	0.070	0.090	0.099	0.098	0.095	0.093	0.084	0.059
Lower 5th percentile	0.023	0.039	0.055	0.071	0.080	0.075	0.075	0.075	0.065	0.042
90% coverage rate (%)	85	88	88	84	91	89	88	88	96	94
τ_c =.2										
Mean	0.025	0.045	0.070	0.096	0.117	0.131	0.133	0.128	0.111	0.083
Upper 5th percentile	0.028	0.052	0.077	0.105	0.130	0.143	0.146	0.146	0.129	0.100
Lower 5th percentile	0.022	0.039	0.062	0.085	0.104	0.118	0.119	0.115	0.097	0.070
90% coverage rate (%)	95	87	92	89	89	95	90	94	89	90
τ_c =.1										
Mean	0.026	0.051	0.082	0.117	0.150	0.177	0.192	0.201	0.196	0.182
Upper 5th percentile	0.029	0.059	0.095	0.133	0.166	0.193	0.210	0.216	0.211	0.202
Lower 5th percentile	0.023	0.044	0.069	0.104	0.131	0.160	0.174	0.187	0.180	0.163
90% coverage rate (%)	93	92	95	95	94	96	87	93	90	92

Table B4: Bootstrap standard errors from Monte Carlo simulation
(5,000 observations; 100 replications)

τ_{sp}	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10	0.00
τ_c =.9										
Mean	0.003	0.003	0.004	0.003	0.003	0.003	0.003	0.003	0.003	0.001
Upper 5th percentile	0.003	0.003	0.004	0.004	0.004	0.004	0.004	0.003	0.004	0.001
Lower 5th percentile	0.002	0.003	0.004	0.003	0.003	0.003	0.003	0.003	0.003	0.001
90% coverage rate (%)	88	90	91	90	95	92	92	94	92	94
τ_c =.8										
Mean	0.005	0.006	0.006	0.005	0.005	0.005	0.005	0.005	0.005	0.001
Upper 5th percentile	0.006	0.006	0.006	0.006	0.006	0.005	0.006	0.006	0.006	0.002
Lower 5th percentile	0.004	0.005	0.005	0.005	0.004	0.004	0.005	0.004	0.004	0.001
90% coverage rate (%)	90	92	92	91	96	84	90	91	93	95
τ_c =.7										
Mean	0.009	0.008	0.008	0.008	0.007	0.007	0.008	0.008	0.007	0.002
Upper 5th percentile	0.010	0.009	0.008	0.008	0.008	0.008	0.009	0.008	0.008	0.003
Lower 5th percentile	0.008	0.007	0.007	0.007	0.006	0.007	0.007	0.007	0.006	0.002
90% coverage rate (%)	88	88	93	87	92	91	87	91	88	94
τ_c =.6										
Mean	0.007	0.011	0.010	0.010	0.010	0.011	0.011	0.010	0.010	0.004
Upper 5th percentile	0.008	0.012	0.011	0.011	0.012	0.012	0.012	0.012	0.011	0.005
Lower 5th percentile	0.006	0.010	0.009	0.009	0.009	0.010	0.009	0.009	0.009	0.003
90% coverage rate (%)	91	93	88	90	92	86	89	90	93	94
τ_c =.5										
Mean	0.010	0.012	0.012	0.014	0.014	0.015	0.014	0.014	0.014	0.006
Upper 5th percentile	0.011	0.014	0.014	0.016	0.015	0.016	0.015	0.015	0.015	0.008
Lower 5th percentile	0.009	0.011	0.011	0.012	0.012	0.014	0.012	0.012	0.012	0.005
90% coverage rate (%)	92	91	95	91	93	91	89	90	93	87
τ_c =.4										
Mean	0.012	0.016	0.015	0.020	0.020	0.019	0.018	0.018	0.018	0.010
Upper 5th percentile	0.013	0.018	0.018	0.022	0.022	0.022	0.021	0.020	0.021	0.012
Lower 5th percentile	0.011	0.014	0.013	0.017	0.018	0.017	0.016	0.016	0.016	0.008
90% coverage rate (%)	85	89	91	93	88	90	90	91	95	90
τ_c =.3										
Mean	0.014	0.020	0.018	0.027	0.028	0.027	0.026	0.026	0.026	0.016
Upper 5th percentile	0.015	0.023	0.021	0.030	0.031	0.030	0.029	0.029	0.029	0.019
Lower 5th percentile	0.012	0.018	0.015	0.023	0.024	0.024	0.023	0.023	0.024	0.014
90% coverage rate (%)	90	93	88	83	92	90	90	92	93	90
τ_c =.2										
Mean	0.015	0.025	0.021	0.036	0.042	0.042	0.042	0.041	0.040	0.031
Upper 5th percentile	0.016	0.027	0.025	0.040	0.047	0.048	0.047	0.045	0.045	0.035
Lower 5th percentile	0.013	0.022	0.018	0.032	0.036	0.038	0.038	0.036	0.036	0.027
90% coverage rate (%)	91	97	96	96	92	90	86	89	88	90
τ_c =.1										
Mean	0.011	0.016	0.023	0.042	0.066	0.080	0.083	0.083	0.082	0.072
Upper 5th percentile	0.012	0.018	0.026	0.047	0.072	0.089	0.094	0.094	0.092	0.080
Lower 5th percentile	0.010	0.014	0.019	0.036	0.060	0.070	0.073	0.074	0.074	0.065
90% coverage rate (%)	82	95	95	93	91	92	88	92	89	93

FIGURE B1: MONTE CARLO SIMULATION (100 REPLICATIONS): CONSISTENCY

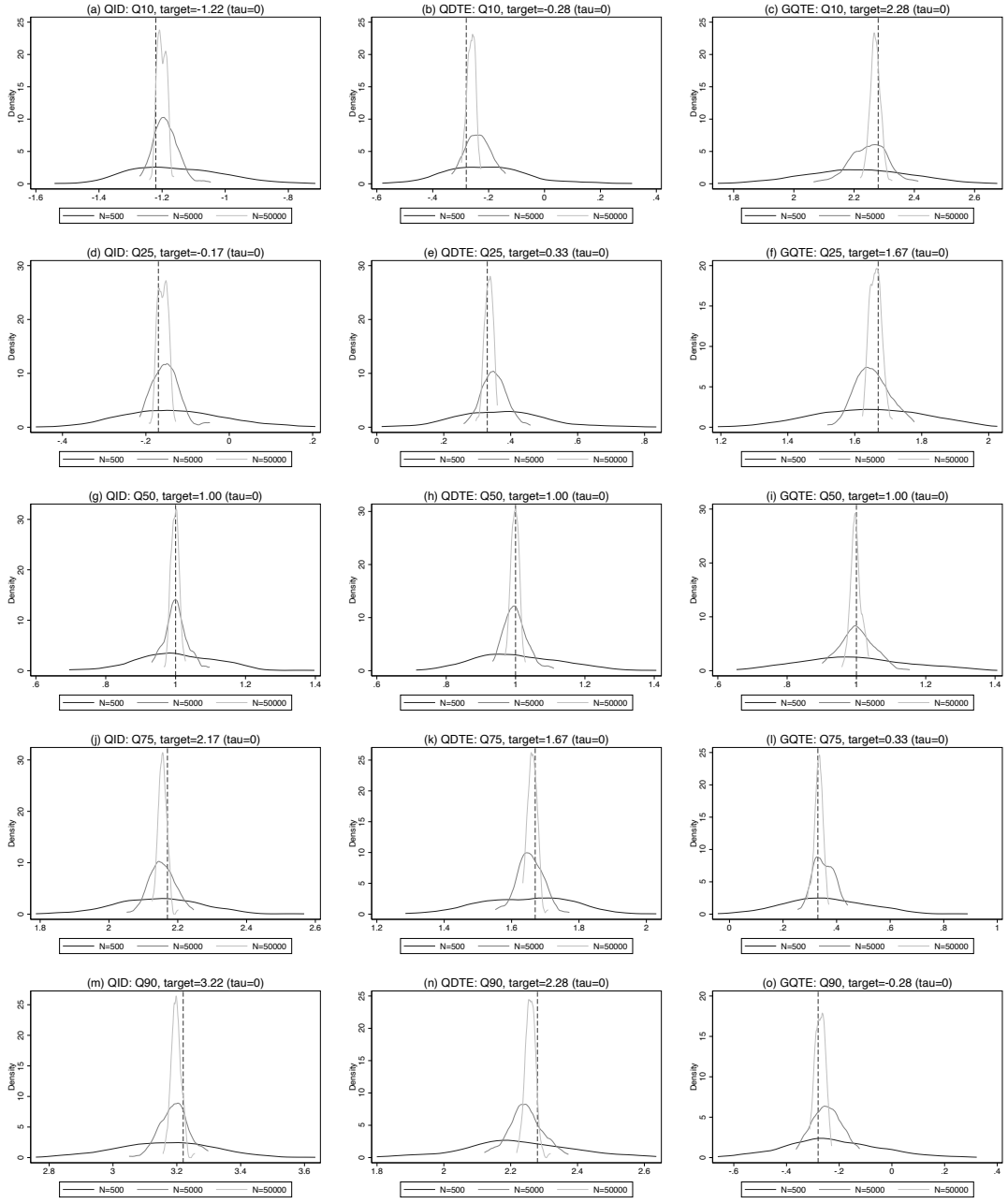


FIGURE B2: MONTE CARLO SIMULATION (100 REPLICATIONS): ASYMPTOTIC NORMALITY

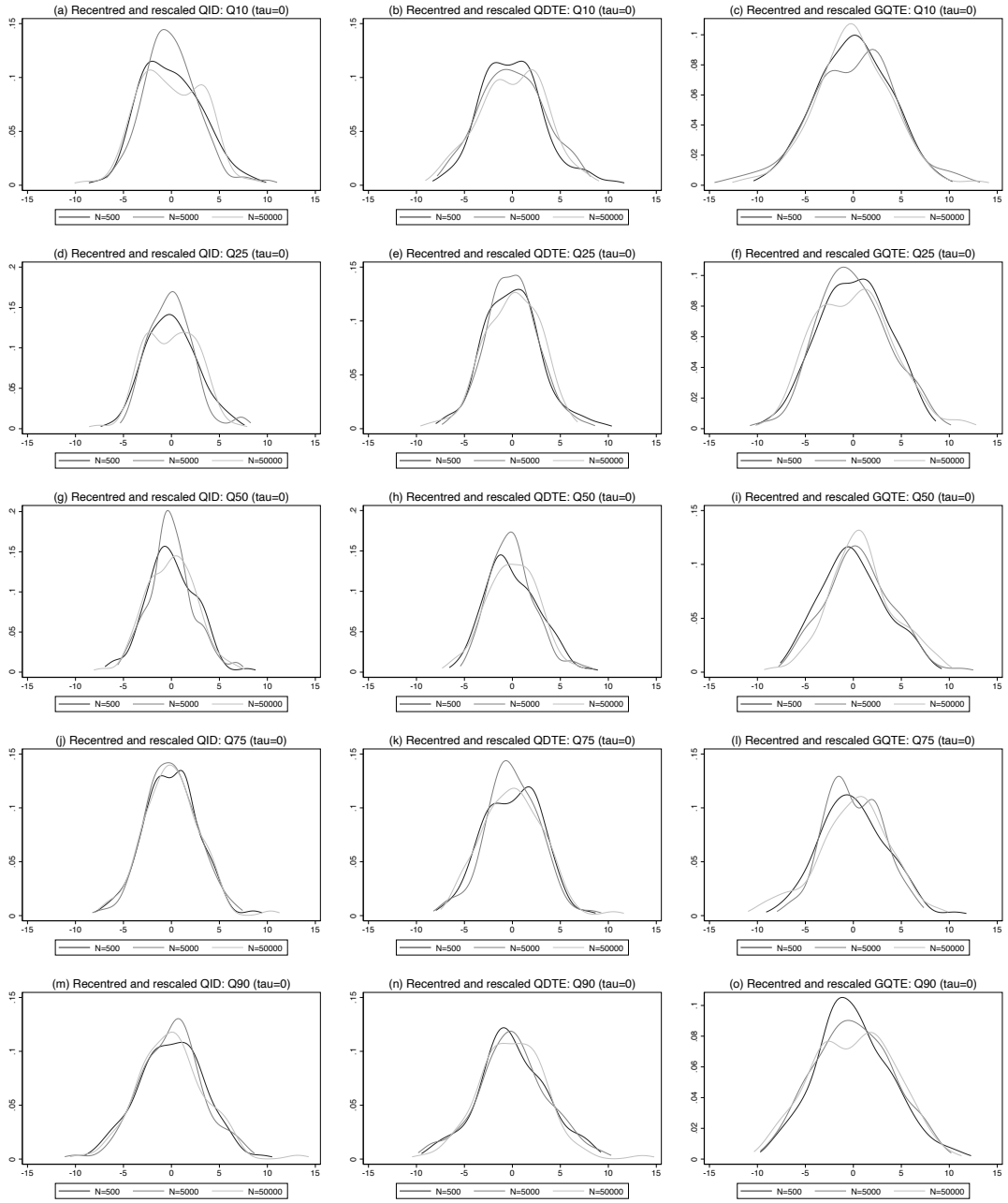


FIGURE B3: MONTE CARLO SIMULATION (100 REPS.): BOOTSTRAP STANDARD ERRORS FOR SELECTED QUANTILES OF QDTE

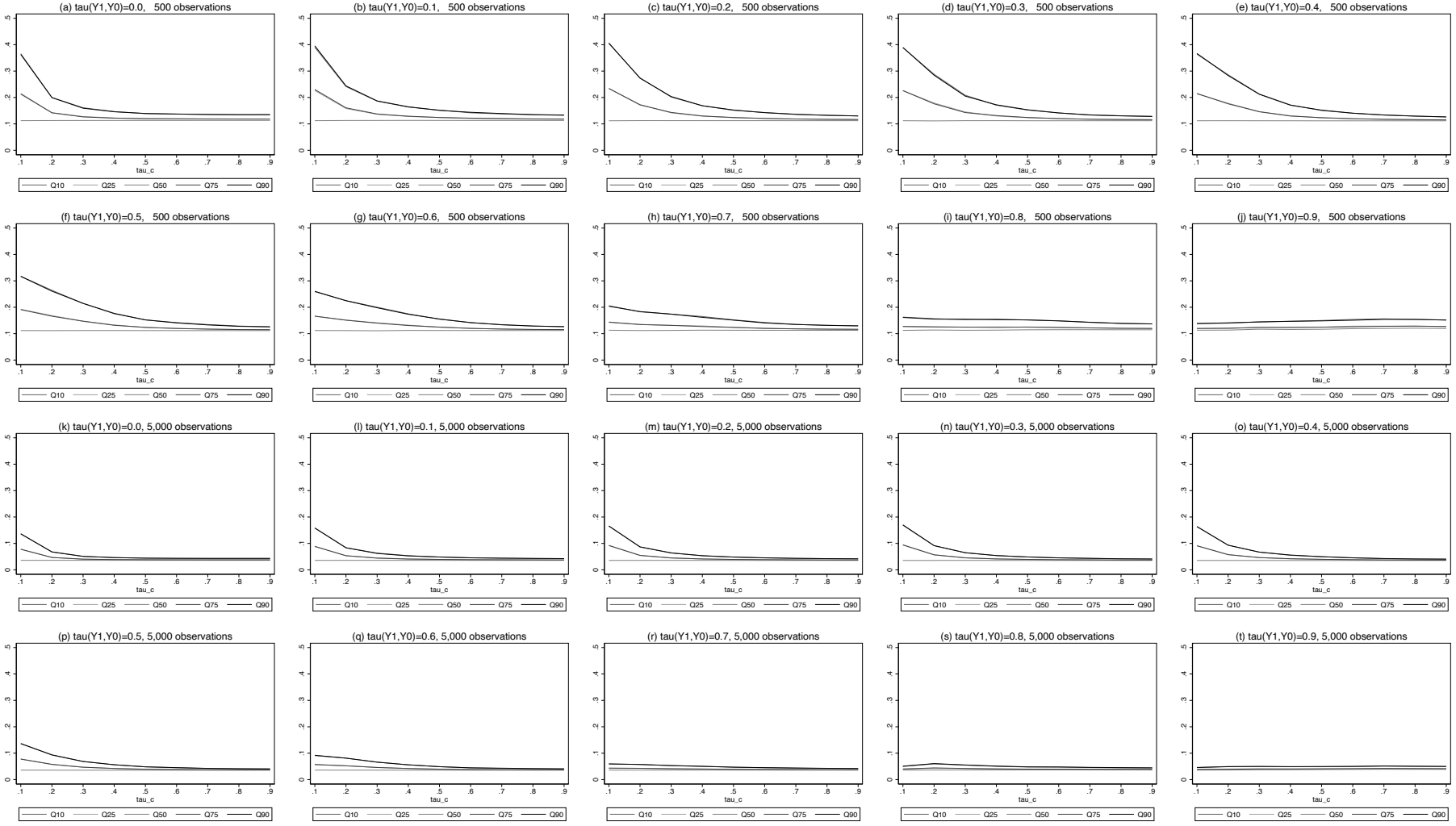


FIGURE B4: MONTE CARLO SIMULATION (100 REPLICATIONS): 90% COVERAGE RATE FOR SELECTED QUANTILES OF QDTE

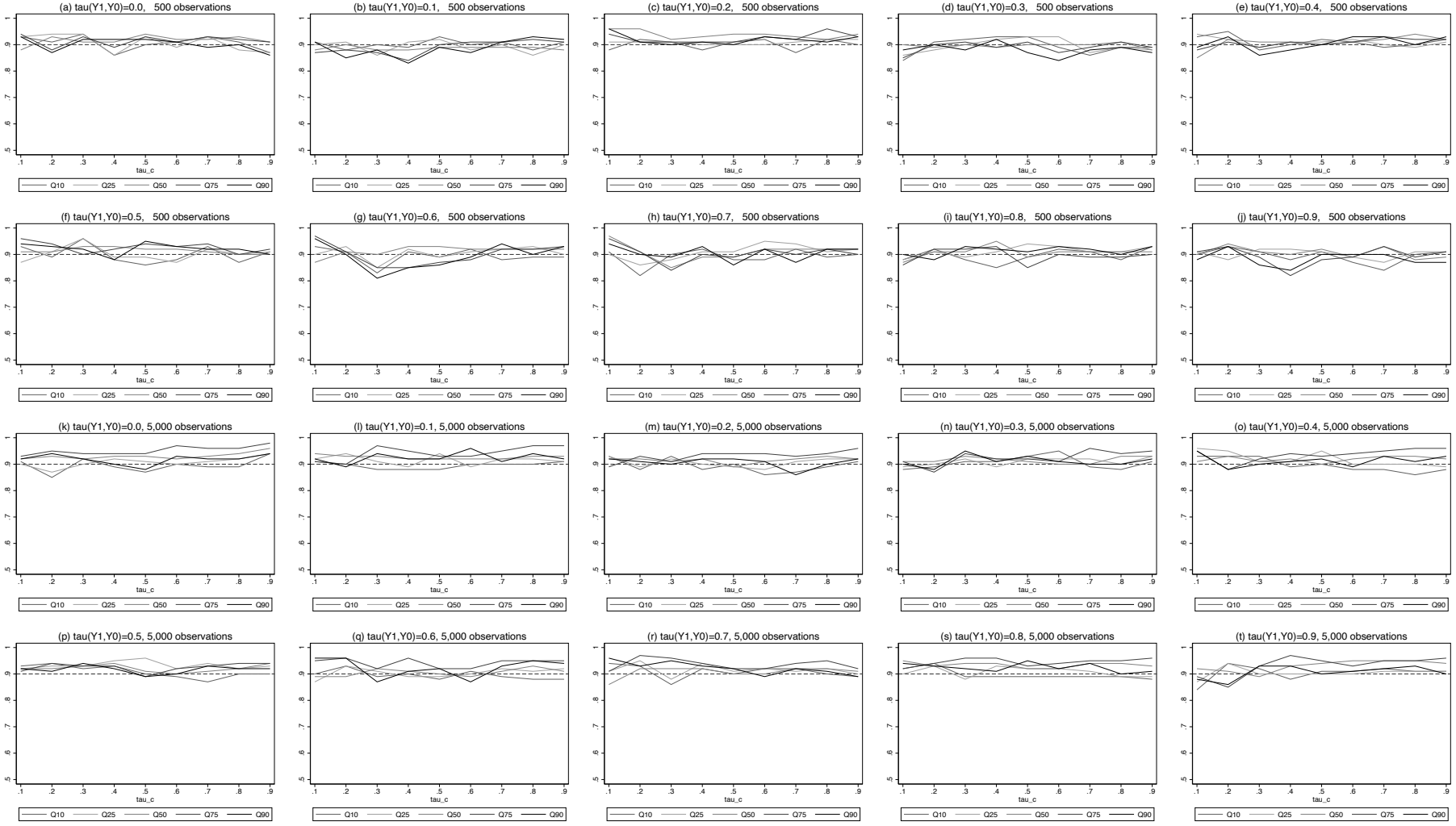


FIGURE B5: MONTE CARLO SIMULATION (100 REPS.): BOOTSTRAP STANDARD ERRORS FOR SELECTED QUANTILES OF GQTE

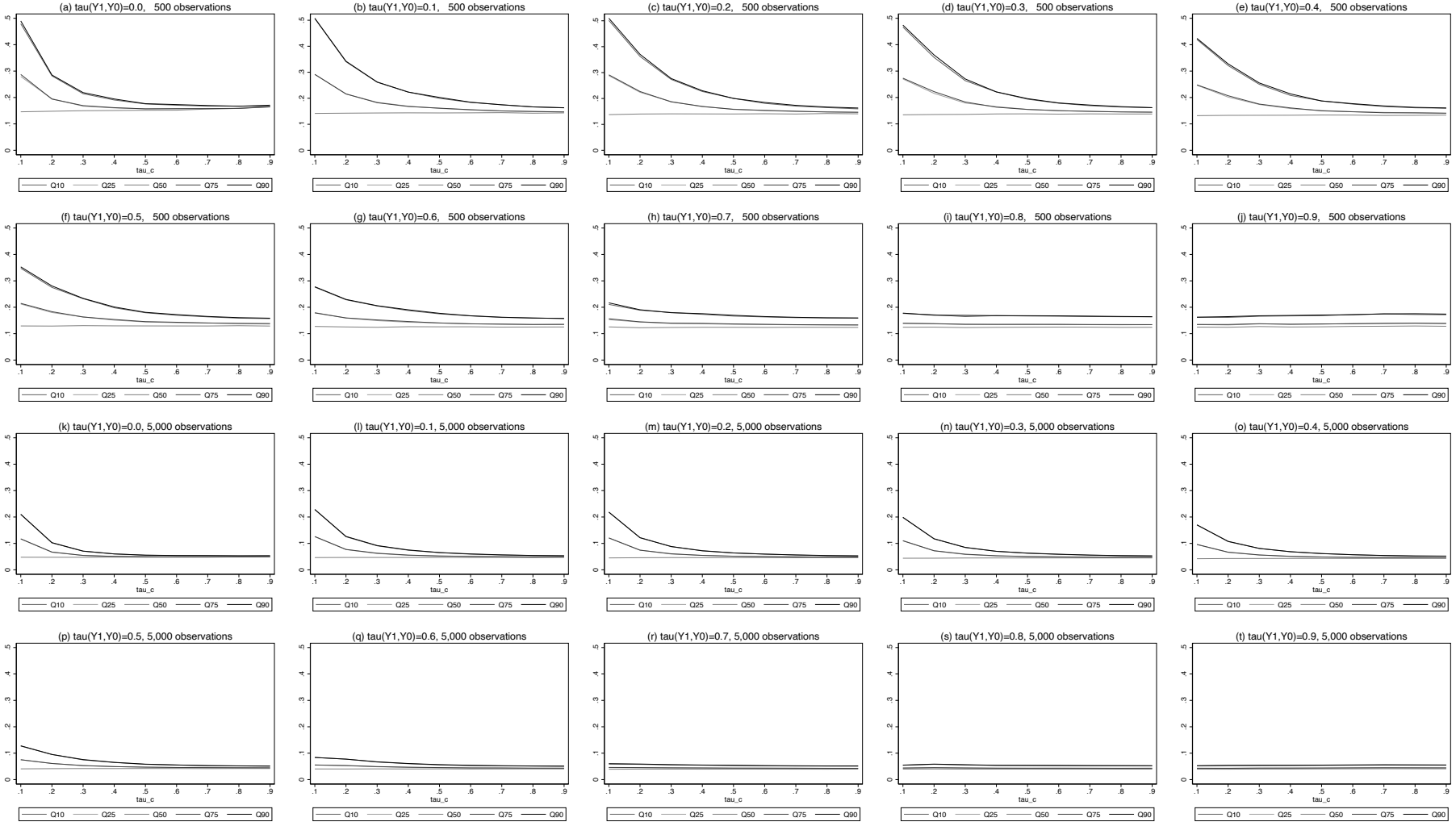
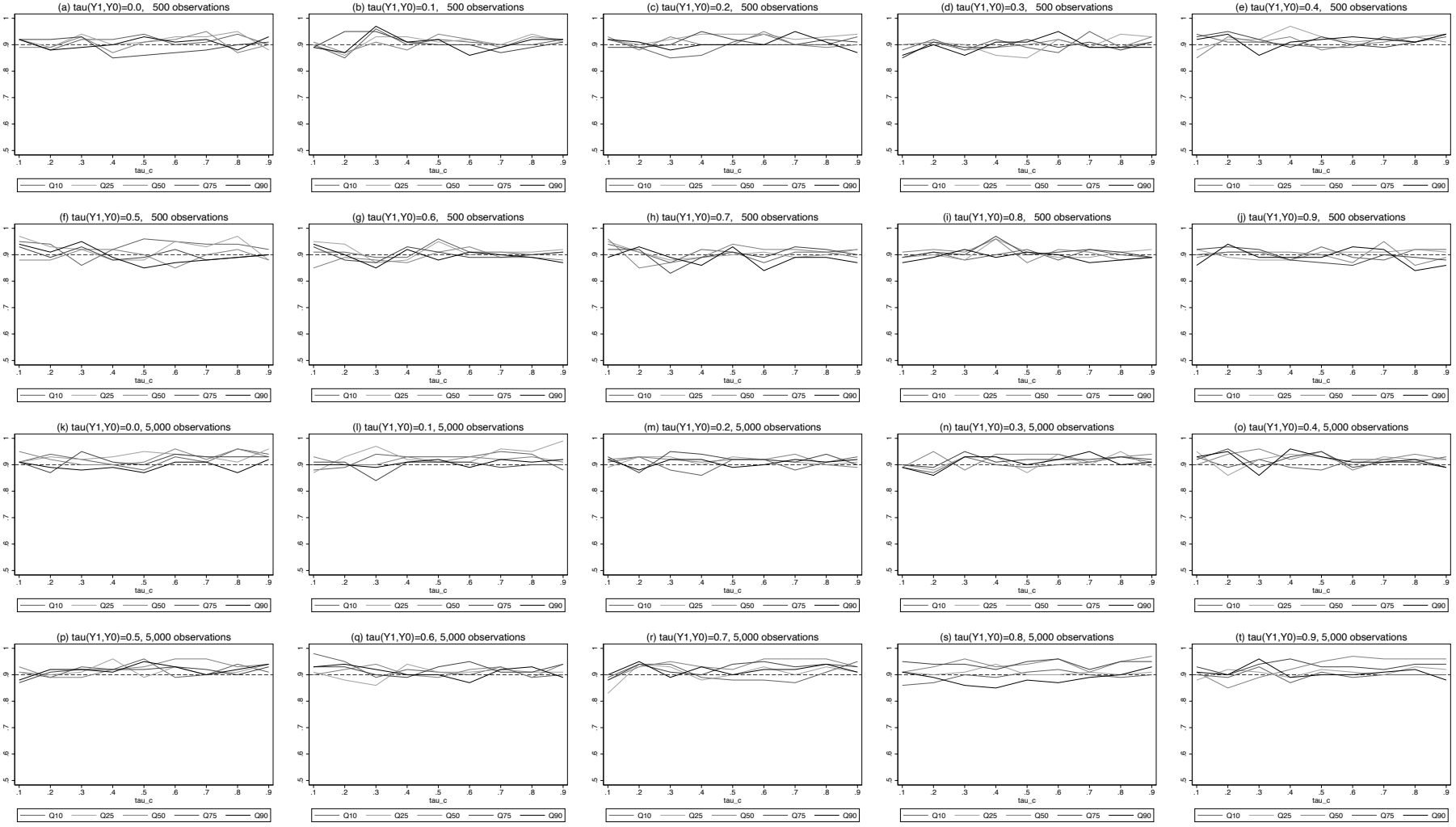


FIGURE B6: MONTE CARLO SIMULATION (100 REPLICATIONS): 90% COVERAGE RATE FOR SELECTED QUANTILES OF GQTE



Appendix C – Proofs

PROOF OF THEOREM 1: Using Lemma 1 of Firpo (2007), the distribution functions $F_{Y(d)}(y)$, $d = \{0,1\}$, can be expressed in terms of weighted averages: $F_{Y(d)}(y) = \frac{1}{N_{\text{sp}}} \sum_{i=1}^{N_{\text{sp}}} w_{id} \mathbf{1}\{Y_i \leq y\}$, with $w_{i1} = \frac{D_i}{Pr(D_i=1|X_i)}$ and $w_{i0} = \frac{1-D_i}{1-Pr(D_i=1|X_i)}$. The order statistics $Z_{(1)d} \leq \ldots \leq Z_{(N_{\text{sp}}^d)d}$ are the values of the quantile functions derived from continuous and monotonically increasing distribution functions $F_{Y(d)}(y)$,

$$Z_{(i)d} \quad = \quad F_{Y(d)}^{-1}[\theta_d] = \inf\{y : \frac{1}{N_{\text{sp}}} \sum_{i=1}^{N_{\text{sp}}} w_{id} \mathbf{1}\{Y_i \leq y\} \geq \theta_d\}, \tag{18}$$

with $(i-1)/N_{\text{sp}}^d < \theta_d \leq i/N_{\text{sp}}^d$, $i \in \{1, \ldots, N_{\text{sp}}^d\}$. Define the $(N_{\text{sp}}^d \times 1)$-vectors $\mathbf{Z}_{\text{sp},d} = (Z_{1d}, \ldots, Z_{N_{\text{sp}}^d d})'$. Consider the $(N_{\text{sp}} \times k)$-covariate matrix $\mathbf{X}_{\text{sp}}$, and the corresponding $((N_{\text{sp}}/2) \times k)$-covariate matrix $\mathbf{X}_{\text{sp},0}$ for members of the control group. The predicted outcomes are given by $\widehat{\mathbf{Y}}_{\text{sp}} = \mathbf{X}_{\text{sp}}(\mathbf{X}'_{\text{sp},0}\mathbf{X}_{\text{sp},0})^{-1}\mathbf{X}'_{\text{sp},0}\mathbf{Y}_{\text{sp},0}$. The elements of the $(N_{\text{sp}} \times 1)$-vector $\widehat{\mathbf{Y}}_{\text{sp}} = (\widehat{Y}_1, \ldots, \widehat{Y}_{N_{\text{sp}}})'$ are used to obtain the order statistics $\widehat{Z}_{(i)d} = \inf\{\widehat{y} : \frac{1}{N_{\text{sp}}} \sum_{i=1}^{N_{\text{sp}}} w_{id} \mathbf{1}\{\widehat{Y}_i \leq \widehat{y}\} \geq \theta_d\}$ and to define the $(N_{\text{sp}}^d \times 1)$-vectors $\widehat{\mathbf{Z}}_{\text{sp},d} = (\widehat{Z}_{1d}, \ldots, \widehat{Z}_{N_{\text{sp}}^d d})'$.

Let $\Delta_{\text{sp},p}^Z = \mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1} - \mathbf{Z}_{\text{sp},0}$. Under Assumption 2, the distribution of treatment effects may be written as

$$F_\Delta(\delta) = F_{\Delta^Z}(\delta^Z), \tag{19}$$

where

$$\Delta^Z \quad = \quad \sum_{p \in \mathcal{P}_{\text{sp}}} \Pr[F_{\Delta_{\text{sp},p}^Z}(\delta_{\text{sp},p}^Z) = F_\Delta(\delta)] \Delta_{\text{sp},p}^Z. \tag{20}$$

After controlling for covariates, permutations that do not satisfy the condition $\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},1})$ occur with a probability of zero under Assumption 2. Therefore,

$$\Delta^Z \quad = \quad \sum_{p \in \mathcal{S}_{\text{sp}|X}} \Pr[F_{\Delta_{\text{sp},p}}^Z(\delta_{\text{sp},p}^Z) = F_\Delta(\delta)] \Delta_{\text{sp},p}^Z, \tag{21}$$

where $\mathcal{S}_{\text{sp}|X} = \{p \in \mathcal{P}_{\text{sp}} \mid \tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},1})\}$. Equation (21) follows from

equation (20), Assumptions 2 and 4, and from

$$
\begin{aligned}
\textstyle\sum_{p\in\mathcal{P}_{\text{sp}}} \quad & \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\mathbf{Z}_{\text{sp},0}) = \tau_{\text{sp}}] \\
= \quad & \sum_{p\in\mathcal{S}_{\text{sp}|X}} \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\mathbf{Z}_{\text{sp},0}) = \tau_{\text{sp}} \mid \tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},1})] \\
\times \quad & \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},1})] \\
+ \quad & \sum_{p\in\mathcal{S}'_{\text{sp}|X}} \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\mathbf{Z}_{\text{sp},0}) = \tau_{\text{sp}} \mid \tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},0}) \neq \tau(\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},1})] \\
\times \quad & \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},0}) \neq \tau(\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},1})],
\end{aligned}
$$

where $\mathcal{S}'_{\text{sp}|X} = \{p \in \mathcal{P}_{\text{sp}} \mid \tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},0}) \neq \tau(\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},1})\}$. Under Assumption 2, $\Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},0}) \neq \tau(\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},1})] = 0$ for all $p \in \mathcal{P}_{\text{sp}}$. Using Bayes' law,

$$
\begin{aligned}
& \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\mathbf{Z}_{\text{sp},0}) = \tau_{\text{sp}}] \\
= \quad & \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\mathbf{Z}_{\text{sp},0}) = \tau_{\text{sp}} \mid \tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},1})] \\
\times \quad & \Pr[\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},1})]
\end{aligned}
$$

for all $p \in \mathcal{S}_{\text{sp}|X}$. Under Assumption 5,

$$
\Pr[F_{\Delta^Z_{\text{sp},p}}(\delta^Z_{\text{sp},p}) = F_\Delta(\delta)] = \frac{\mathbf{1}\{\Pr[F_{\Delta^Z_{\text{sp},p}}(\delta^Z_{\text{sp},p}) = F_\Delta(\delta)] > 0\}}{\sum_{\mathcal{P}_{\text{sp}}} \mathbf{1}\{\Pr[F_{\Delta^Z_{\text{sp},p}}(\delta^Z_{\text{sp},p}) = F_\Delta(\delta)] > 0\}} \tag{22}
$$

for all $p \in \mathcal{P}_{\text{sp}}$. All permutations of $\mathbf{Z}_{\text{sp},1}$ that satisfy $\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},1})$ have a positive probability of occurrence under Assumptions 2 and 4. Therefore, under Assumptions 2, 4, and 5, the distribution of treatment effects is identified through

$$
F_\Delta(\delta) \quad = \quad F_{\Delta^Z}(\delta^Z), \tag{23}
$$

where $\Delta^Z = \frac{1}{n_{p|X}} \sum_{p\in\mathcal{S}_{\text{sp}|X}} \Delta^Z_{\text{sp},p}$ with $n_{p|X} = \sum_{p\in\mathcal{P}_{\text{sp}}} \mathbf{1}\{\Pr[F_{\Delta^Z_{\text{sp},p}}(\delta^Z_{\text{sp},p}) = F_\Delta(\delta)] > 0\} = \sum_{p\in\mathcal{S}_{\text{sp}|X}} \mathbf{1}\{\tau(\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1},\widehat{\mathbf{Z}}_{\text{sp},1})\}$. Identification of the Population Quantiles of the Distribution of Treatment Effects follows from

$$
q_{\Delta,u} \quad = \quad q_u(F_\Delta(\delta)) = q_u(F_{\Delta^Z}(\delta^Z)). \tag{24}
$$

Q.E.D.

PROOF OF THEOREM 2: Using Lemma 1 of Firpo (2007), the distribution functions $F_{Y(d)}(y)$, $d = \{0,1\}$, can be expressed in terms of weighted averages: $F_{Y(d)}(y) = \frac{1}{N_{\text{sp}}} \sum_{i=1}^{N_{\text{sp}}} w_{id} \mathbf{1}\{Y_i \leq y\}$, with $w_{i1} = \frac{D_i}{Pr(D_i=1|X_i)}$ and $w_{i0} = \frac{1-D_i}{1-Pr(D_i=1|X_i)}$. Let

$$i' \in \{\{j_{11}, \ldots, j_{1N_{\text{sp}}^{(1-d)}}\}, \ldots, \{j_{N_{\text{sp}}^d 1}, \ldots, j_{N_{\text{sp}}^d N_{\text{sp}}^{(1-d)}}\}\} = \{1, \ldots, N_{\text{sp}}^1 N_{\text{sp}}^0\},$$

and define the order statistics $Z_{(1)d} \leq \ldots \leq Z_{(N_{\text{sp}}^1 N_{\text{sp}}^0)d}$ as values of the quantile functions derived from continuous and monotonically increasing distribution functions $F_{Y(d)}(y)$,

$$Z_{(i')d} = F_{Y(d)}^{-1}[\theta_d] = \inf\{y : \frac{1}{N_{\text{sp}}} \sum_{i=1}^{N_{\text{sp}}} w_{id} \mathbf{1}\{Y_i \leq y\} \geq \theta_d\}, \tag{25}$$

with $(i'-1)/(N_{\text{sp}}^1 N_{\text{sp}}^0) < \theta_d \leq i'/(N_{\text{sp}}^1 N_{\text{sp}}^0)$, $i' \in \{1, \ldots, N_{\text{sp}}^1 N_{\text{sp}}^0\}$.

Define the $(N_{\text{sp}}^1 N_{\text{sp}}^0 \times 1)$-vectors

$$\begin{aligned} \mathbf{V}_{\text{sp},d} &= (Z_{(j_{11})d}, \ldots, Z_{(j_{1N_{\text{sp}}^{(1-d)}})d}, \ldots, Z_{(j_{N_{\text{sp}}^d 1})d}, \ldots, Z_{(j_{N_{\text{sp}}^d N_{\text{sp}}^{(1-d)}})d})' \\ &= (\underbrace{Z_{(1)d}, \ldots, Z_{(1)d}}_{N_{\text{sp}}^{(1-d)}}, \ldots, \underbrace{Z_{(N_{\text{sp}}^d)d}, \ldots, Z_{(N_{\text{sp}}^d)d}}_{N_{\text{sp}}^{(1-d)}})', \end{aligned}$$

where

$$Z_{(i)d} = F_{Y(d)}^{-1}[\theta_d] = \inf\{y : \frac{1}{N_{\text{sp}}} \sum_{i=1}^{N_{\text{sp}}} w_{id} \mathbf{1}\{Y_i \leq y\} \geq \theta_d\}, \tag{26}$$

with $(i-1)/N_{\text{sp}}^d < \theta_d \leq i/N_{\text{sp}}^d$, $i \in \{1, \ldots, N_{\text{sp}}^d\}$. Define the $(N_{\text{sp}}^d \times 1)$-vectors $\mathbf{Z}_{\text{sp},d} = (Z_{1d}, \ldots, Z_{N_{\text{sp}}^d d})'$. Consider the $(N_{\text{sp}} \times k)$-covariate matrix $\mathbf{X}_{\text{sp}}$, and the corresponding $((N_{\text{sp}}/2) \times k)$-covariate matrix $\mathbf{X}_{\text{sp},0}$ for members of the control group. The predicted outcomes are given by $\widehat{\mathbf{Y}}_{\text{sp}} = \mathbf{X}_{\text{sp}}(\mathbf{X}'_{\text{sp},0}\mathbf{X}_{\text{sp},0})^{-1}\mathbf{X}'_{\text{sp},0}\mathbf{Y}_{\text{sp},0}$. The elements of the $(N_{\text{sp}} \times 1)$-vector $\widehat{\mathbf{Y}}_{\text{sp}} = (\widehat{Y}_1, \ldots, \widehat{Y}_{N_{\text{sp}}})'$ are used to obtain the order statistics $\widehat{Z}_{(i)d} = \inf\{\widehat{y} : \frac{1}{N_{\text{sp}}} \sum_{i=1}^{N_{\text{sp}}} w_{id} \mathbf{1}\{\widehat{Y}_i \leq \widehat{y}\} \geq \theta_d\}$ and to define the $(N_{\text{sp}}^d \times 1)$-vectors $\widehat{\mathbf{Z}}_{\text{sp},d} = (\widehat{Z}_{1d}, \ldots, \widehat{Z}_{N_{\text{sp}}^d d})'$.

Consider the permutations $\mathbf{\Omega}_{\text{sp},d}\mathbf{V}_{\text{sp},d} = \mathbf{M}_{\text{sp},d}\mathbf{Z}_{\text{sp},d}$, where $\mathbf{\Omega}_{\text{sp},d}$ are $(N_{\text{sp}}^1 N_{\text{sp}}^0 \times N_{\text{sp}}^1 N_{\text{sp}}^0)$-permutation matrices, and where $\mathbf{M}_{\text{sp},d}$ are $(N_{\text{sp}}^1 N_{\text{sp}}^0 \times N_{\text{sp}}^d)$-transformation matrices that

transform the $(N_{\text{sp}}^{d} \times 1)$-vectors $\mathbf{Z}_{\text{sp},d}$ into the $(N_{\text{sp}}^{1} N_{\text{sp}}^{0} \times 1)$-vectors

$$\mathbf{M}_{\text{sp},d}\mathbf{Z}_{\text{sp},d} \;=\; (\underbrace{Z_{1d},\ldots,Z_{1d}}_{N_{\text{sp}}^{(1-d)}},\ldots,\underbrace{Z_{N_{\text{sp}}^{d}d},\ldots,Z_{N_{\text{sp}}^{d}d}}_{N_{\text{sp}}^{(1-d)}})'.$$

Let $\Delta_{\text{sp},p}^{Z} = \mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1} - \mathbf{M}_{\text{sp},0}\mathbf{Z}_{\text{sp},0}$. Under Assumption 2, the distribution of treatment effects may be written as

$$F_{\Delta}(\delta) = F_{\Delta^{Z}}(\delta^{Z}), \tag{27}$$

where

$$\Delta^{Z} \;=\; \sum_{p\in\mathcal{P}_{\text{sp}}} \Pr[F_{\Delta_{\text{sp},p}^{Z}}(\delta_{\text{sp},p}^{Z}) = F_{\Delta}(\delta)]\Delta_{\text{sp},p}^{Z}. \tag{28}$$

After controlling for covariates, permutations that do not satisfy the condition $\tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},1})$ occur with a probability of zero under Assumption 2. Therefore,

$$\Delta^{Z} \;=\; \sum_{p\in\mathcal{S}_{\text{sp}|X}} \Pr[F_{\Delta_{\text{sp},p}}^{Z}(\delta_{\text{sp},p}^{Z}) = F_{\Delta}(\delta)]\Delta_{\text{sp},p}^{Z}, \tag{29}$$

where $\mathcal{S}_{\text{sp}|X} = \{p \in \mathcal{P}_{\text{sp}} \mid \tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},1})\}$. Equation (29) follows from equation (28), Assumptions 2 and 4, and from

$$\begin{aligned}
&\textstyle\sum_{p\in\mathcal{P}_{\text{sp}}} \quad \Pr[\tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\mathbf{Z}_{\text{sp},0}) = \tau_{\text{sp}}] \\
&= \quad \sum_{p\in\mathcal{S}_{\text{sp}|X}} \Pr\big[\tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\mathbf{Z}_{\text{sp},0}) = \tau_{\text{sp}} \mid \\
&\qquad\quad \tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},1})\big] \\
&\times \quad \Pr[\tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},1})] \\
&+ \quad \sum_{p\in\mathcal{S}'_{\text{sp}|X}} \Pr\big[\tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\mathbf{Z}_{\text{sp},0}) = \tau_{\text{sp}} \mid \\
&\qquad\quad \tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\widehat{\mathbf{Z}}_{\text{sp},0}) \neq \tau(\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},1})\big] \\
&\times \quad \Pr[\tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\widehat{\mathbf{Z}}_{\text{sp},0}) \neq \tau(\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},1})],
\end{aligned}$$

where $\mathcal{S}'_{\text{sp}|X} = \{p \in \mathcal{P}_{\text{sp}} \mid \tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\widehat{\mathbf{Z}}_{\text{sp},0}) \neq \tau(\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},1})\}$. Under Assump-

tion 2, $\Pr[\tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\widehat{\mathbf{Z}}_{\text{sp},0}) \neq \tau(\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},1})] = 0$ for all $p \in \mathcal{P}_{\text{sp}}$. Using Bayes' law,

$$\begin{aligned} & \Pr[\tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\mathbf{Z}_{\text{sp},0}) = \tau_{\text{sp}}] \\ = \; & \Pr[\tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\mathbf{Z}_{\text{sp},0}) = \tau_{\text{sp}} \mid \\ & \tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},1})] \\ \times \; & \Pr[\tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},1})] \end{aligned}$$

for all $p \in \mathcal{S}_{\text{sp}|X}$. Under Assumption 5,

$$\Pr[F_{\Delta^Z_{\text{sp},p}}(\delta^Z_{\text{sp},p}) = F_\Delta(\delta)] = \frac{\mathbf{1}\{\Pr[F_{\Delta^Z_{\text{sp},p}}(\delta^Z_{\text{sp},p}) = F_\Delta(\delta)] > 0\}}{\sum_{\mathcal{P}_{\text{sp}}} \mathbf{1}\{\Pr[F_{\Delta^Z_{\text{sp},p}}(\delta^Z_{\text{sp},p}) = F_\Delta(\delta)] > 0\}} \tag{30}$$

for all $p \in \mathcal{P}_{\text{sp}}$. All permutations of $\mathbf{Z}_{\text{sp},1}$ that satisfy $\tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},1})$ have a positive probability of occurrence under Assumptions 2 and 4. Therefore, under Assumptions 2, 4, and 5, the distribution of treatment effects is identified through

$$F_\Delta(\delta) \;\; = \;\; F_{\Delta^Z}(\delta^Z), \tag{31}$$

where $\Delta^Z = \frac{1}{n_{p|X}} \sum_{p \in \mathcal{S}_{\text{sp}|X}} \Delta^Z_{\text{sp},p}$ with $n_{p|X} = \sum_{p \in \mathcal{P}_{\text{sp}}} \mathbf{1}\{\Pr[F_{\Delta^Z_{\text{sp},p}}(\delta^Z_{\text{sp},p}) = F_\Delta(\delta)] > 0\} = \sum_{p \in \mathcal{S}_{\text{sp}|X}} \mathbf{1}\{\tau(\mathbf{M}_{\text{sp},1}\mathbf{\Pi}_{\text{sp},p}\mathbf{Z}_{\text{sp},1}, \mathbf{M}_{\text{sp},0}\widehat{\mathbf{Z}}_{\text{sp},0}) = \tau(\mathbf{Z}_{\text{sp},1}, \widehat{\mathbf{Z}}_{\text{sp},1})\}$. Identification of the Population Quantiles of the Distribution of Treatment Effects follows from

$$q_{\Delta,u} \;\; = \;\; q_u(F_\Delta(\delta)) = q_u(F_{\Delta^Z}(\delta^Z)). \tag{32}$$

Q.E.D.