Assessing Heterogeneity of Treatment Effects*

Tetsuya Kaji[1] and Jianfei Cao[2]

[1]University of Chicago
[2]Northeastern University

June 28, 2023

Abstract

Treatment effect heterogeneity is of major interest in economics, but its assessment is often hindered by the fundamental lack of identification of the individual treatment effects. For example, we may want to assess the effect of insurance on the health of otherwise unhealthy individuals, but it is infeasible to insure only the unhealthy, and thus the causal effects for those are not identified. Or, we may be interested in the shares of winners from a minimum wage increase, while without observing the counterfactual, the winners are not identified. Such heterogeneity is often assessed by quantile treatment effects, which do not come with clear interpretation and the takeaway can sometimes be equivocal. We show that, with the quantiles of the treated and control outcomes, the ranges of these quantities are identified and can be informative even when the average treatment effects are not significant. Two applications illustrate how these ranges can inform us about heterogeneity of the treatment effects.

*This work is supported by the Liew Family Junior Faculty Fellowship and the Richard N. Rosett Faculty Fellowship at the University of Chicago Booth School of Business and the support for visiting faculty at the Department of Economics at the Massachusetts Institute of Technology. We thank Yijing Zhang for excellent research assistance. We thank Rachael Meager, Toru Kitagawa, Kosuke Imai, Alberto Abadie, Whitney Newey, Ismael Mourifié, Eric Auerbach, Ivan Canay, Federico Bugni, Filip Obradovic, Muyang Ren, Josh Angrist, and seminar participants at various places for helpful comments.

1 Introduction

Heterogeneity of the treatment effects is a major concern in various places of economics. In causal inference with instrumental variables, heterogeneous treatment effects play a key role in the correct interpretation of the estimates (Imbens and Angrist, 1994). In optimal policy targeting and ethical intervention design, assessment of heterogeneity is vital in improving welfare and promoting equity (Kitagawa and Tetenov, 2018, 2021). In the empirical examination of economic theory, it is essential to account for heterogeneous impacts that are consistent with the theoretical prediction (Bitler et al., 2006).

The assessment of heterogeneity is often hindered by the lack of identification of the individual treatment effects. Because of the fundamental impossibility of assigning more than one treatment status to one subject, the joint distribution of the treated outcome and the control outcome is not identified. When the average treatment effect (ATE) is of concern, this lack of identification is not an issue since the ATE only depends on the marginal distributions of the potential outcomes. There are situations, however, in which delicate assessment of heterogeneity is needed.

Consider an investigation of moral hazard in health care (e.g., Aron-Dine et al., 2013). Ideally, we want to know how health insurance affects medical spending across different levels of untreated medical spending. However, since the joint distribution of the treated outcome and the control outcome is not identified, such treatment effects are not identified. In another example, consider the evaluation of a welfare program for low-income and unemployed households (e.g., Bitler et al., 2006). The theoretical prediction is often ambiguous even on who benefits and who loses from the program, so the average effects do not draw the complete picture. However, again, due to the lack of identification of the joint distribution, the shares of winners or losers are not identified.

In these settings where heterogeneity is of primary interest, economists often use the quantile treatment effects (QTEs) to grasp a sense of heterogeneity. While QTEs can suggest the presence of heterogeneity, they can be hard to assess or interpret if we are not willing to assume a strong condition of rank invariance.

In this paper, we focus on two quantities that summarize heterogeneity and complement QTEs, and show that—although they are not point-identified—we can calculate the ranges in which they reside. Using two empirical applications, we show that these ranges can be informative, even when the ATE estimates are not significant.

The first quantity is what we call the *subgroup treatment effect*, which is the ATE for the subpopulation whose control outcome is in some range. For example, when the lower bound for the subgroup treatment effect is positive, we may conclude that the subgroup of interest receives at least some positive treatment effects. We illustrate the usefulness of this concept in Section 3 drawing an example from Tarozzi et al. (2015). The second quantity is what we call the *subgroup proportion of winners* (or *of losers*), which is the share of those whose treated outcome exceeds (or falls under, respectively) the control outcome in the subpopulation whose control outcome is in some range. For example, if the proportion of winners is above one fourth, we may conclude that there are at least a quarter of subjects whose outcomes improved thanks to the treatment. Section 4 illustrates the usefulness of this concept in an example from Bitler et al. (2006).

There is enormous literature on (partial) identification of various heterogeneity in program evaluation. Close to our paper are Heckman et al. (1997) and Tetenov (2012). Heckman et al. (1997) discussed bounds on the proportion of winners and the distribution of treatment effects and investigated how the assumption of rational choice by the participants helps sharpening the bounds. Tetenov (2012) derived bounds on positive and absolute treatment effects and the proportion of winners on the entire population. A key departure of our analysis is that we consider subgroups that are defined by ranges of the control outcome, which gives a finer picture of who receives how much treatment effects.

The rest of the paper is organized as follows. Section 2 lays down the notation used in the paper and clarifies the extent to which the proposed methods can be applied. Section 3 introduces the bounds on the subgroup treatment effects and illustrates their use drawing examples from Tarozzi et al. (2015). Section 4 presents the bounds on the subgroup proportions of winners and applies them to revisit Bitler et al. (2006). Section 5 concludes. Appendix A collects the proofs.

2 Notation and the Scope of Applicability

We employ the standard notation of the potential outcome framework; Y_0 denotes the potential outcome without the treatment and Y_1 the potential outcome with the treatment. The marginal cumulative distribution functions (cdfs) of the potential outcomes are denoted by F_0 for the control and F_1 for the treated. We define the quantile functions of the potential outcomes, Q_0 and Q_1, by the left-continuous generalized inverses of F_0 and F_1. We denote the normalized rank of Y_0 by $U \in [0, 1]$. Mathematically, this is equivalent to having a uniformly distributed random variable $U \sim U(0, 1)$ that satisfies $Y_0 = Q_0(U)$. We define a *winner (from the treatment)* to be a subject with $Y_1 > Y_0$ and a *loser (from the treatment)* a subject with $Y_1 < Y_0$. We use the notation $(x)_+ := \max\{x, 0\}$ and $(x)_- := (-x)_+$. We denote the left limit of a cdf by $F(a-) := \lim_{x \nearrow a} F(x)$.

The key assumption in this paper is that the marginal distributions of the potential outcomes are identified. This applies to settings beyond the standard randomized controlled trials (RCTs) with perfect compliance. In RCTs with monotone compliance, Abadie (2002) discussed derivation of the marginal distributions of potential outcomes for compliers. In standard difference-in-differences (change-in-changes) models, Athey and Imbens (2006) provided the marginal distributions for the treated, which was extended to synthetic control settings by Gunsilius (2023). In models with selection on observables, Firpo et al. (2009) developed unconditional quantile regression to obtain the marginal distributions, which was generalized to models with instrumental variables by Frölich and Melly (2013) and to models with multiple and continuous treatments by Powell (2020). Our method applies to all of these settings.

Moreover, the marginal distributions can be replaced with conditional distributions conditional on covariates, so our framework applies to cases with conditional randomization, for which conditional quantile regression estimates the conditional distributions of the potential outcomes (Koenker and Bassett, 1978; Chernozhukov and Hansen, 2005; Abadie et al., 2003).

3 Subgroup Treatment Effects

There is literature that considers access to microfinance as a way to reduce poverty in developing countries. Tarozzi et al. (2015) used data from an RCT conducted in rural Ethiopia from 2003 to 2006 to assess the impact of microfinance on various outcomes.

The brief outline of the experiment is as follows. Randomization was carried out at the Peasant Association (PA) level, which is a local administrative unit. Out of 133 PAs, 34 PAs were randomly assigned to microcredit and 33 PAs to the control. The remaining 66 PAs were assigned to a different combination of treatments, which we will not use. The compliance rate was 78%, and our analysis below focuses on the intent-to-treat (ITT) effects as did Tarozzi et al. (2015). We leave further details of the experiment to Tarozzi et al. (2015).

Tarozzi et al. (2015) noted that "most loans were initiated to fund crop cultivation or animal husbandry, with 80 percent of the 1,388 loans used for working capital or investment in these sectors . . ." In light of this, we look at two outcome variables: (1) the net cash revenues from crops and (2) the total value of livestock owned.

3.1 Who Benefits from Microfinance?

For both of these outcomes, we are interested in knowing how much impact microfinance have had on those who were most in need, i.e., those who would have attained low outcome in the absence of the treatment (Abadie et al., 2018). For example, we may want to see whether a small amount of money helps increase the output of those who are poor, which is approximated by low Y_0.

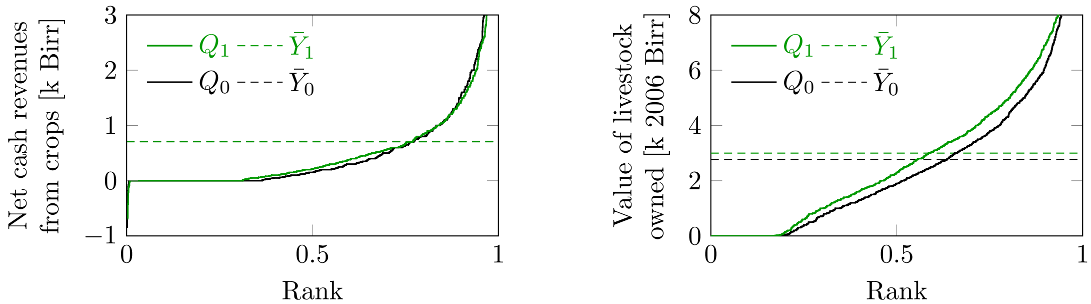
Precisely, we are interested in the ATE conditional on $U < b$ for some level b, that is,

$$\mathbb{E}[Y_1 - Y_0 \mid U < b].$$

Note here that when Y_0 is continuously distributed, conditioning on $U < b$ is equivalent to conditioning on $Y_0 < Q_0(b)$. We call this quantity the *subgroup treatment effect* (STE). The STE is not point-identified since it depends on the joint distribution of Y_0 and Y_1, but using their quantile functions, we can infer (Theorem 3)

$$\frac{1}{b}\int_0^b [Q_1(u) - Q_0(u)]du \leq \mathbb{E}[Y_1 - Y_0 \mid U < b] \leq \frac{1}{b}\int_0^b [Q_1(1-u) - Q_0(u)]du. \quad (1)$$

While QTEs in general cannot be interpreted as individual treatment effects without the rank invariance assumption, (1) states that the *integral* of the QTEs can be interpreted as the lower bound for the STE *without any assumption.* The lower bound is exact when rank invariance holds, but it is also expected to be close when rank invariance approximately holds. This idea that individuals do not change ranks too drastically by the treatment would be a reasonable presumption in many applications. (Meanwhile the upper bound may not be very informative in many cases.) A possible takeaway from (1) is that, if the lower bound is greater than zero, we know that the poor individuals receive some positive treatment effects even though we do not know



(a) Quantiles for the net cash revenues from crops. The KS p-value = 0.014.

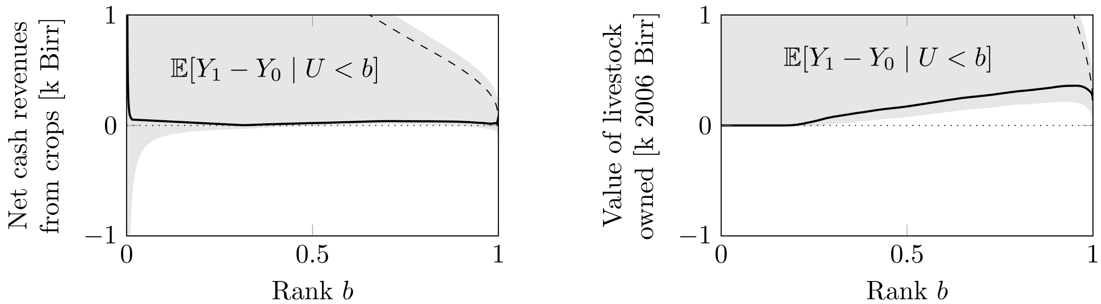
(b) Quantiles for the total value of livestock owned. The KS p-value = 0.0002.

Figure 1: Empirical quantile functions for the two outcomes in the treated and control groups in Tarozzi et al. (2015). The difference of the green dashed line and the black dashed line is the estimated ATE, which is insignificant in either case. However, the KS test indicates non-null effects in both outcomes.

exactly how much.

Now we turn to estimates from the data. We first note that the ATE estimates for both outcomes are not significant. Tarozzi et al. (2015) wrote "We document that despite substantial increases in borrowing in areas assigned to treatment the null of no impact cannot be rejected for a large majority of outcomes." However, this does not mean that there is no evidence of heterogeneity in the data. In fact, the weighted Kolmogorov–Smirnov (KS) test of equal distributions (Monahan, 2011, p. 358) suggests that the marginal distributions of Y_0 and Y_1 are significantly different from each other for both outcome variables (Figure 1).

Figure 2 shows the bounds on the STE given by (1). The black solid lines indicate the theoretical lower bounds of $\mathbb{E}[Y_1 - Y_0 \mid U < b]$ across different values of b. The black dashed lines are the upper bounds, but we do not scale the figures to cover the whole of upper bounds since they are uninformatively large in most values of b. The gray areas are the pointwise 95% confidence intervals for the STE at each point of b constructed by the method of Imbens and Manski (2004). While we cannot exclude the possibility that microfinance has no effect on the crop revenues across various levels of crop revenues (Figure 2a), the effect on the value of livestock is significantly positive for the mid-quantile levels of the livestock value (Figure 2b). The flat zero region on the left is likely those who do not own any livestock, so Figure 2b suggests that the effect of microcredit on those who has a modest to low level of livestock is



(a) Bounds for the STEs on the net cash revenues from crops.

(b) Bounds for the STEs on the total value of livestock owned.

Figure 2: Lower bounds for the STEs in Tarozzi et al. (2015). The black lines indicate the estimated lower bounds for $\mathbb{E}[Y_1 - Y_0 \mid U < b]$ for each b. The gray areas are the pointwise 95% confidence intervals.

expected to be at least as big as a few hundred Birr.[1]

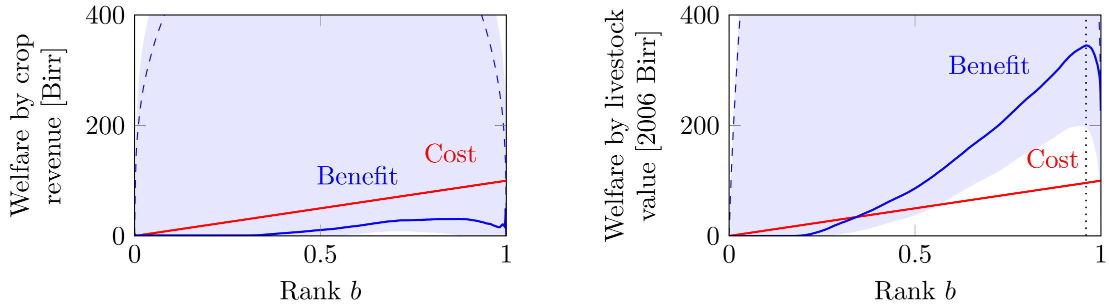
3.2 Who Should Be Treated?

These bounds can be used to determine who to assign treatment to for attaining higher welfare. The literature on the welfare-maximizing assignment of the treatment was initiated by Kitagawa and Tetenov (2018). They discussed a mechanism to decide the treatment assignment based on covariates that maximizes the empirical welfare.

Meanwhile, policymakers may want to assign treatment to those who are most in need, that is, those whose outcome are low without treatment. When the dataset has a comonly observed baseline Y_0 in the pre-treatment period, we may estimate the treatment effect conditional on the predicted Y_0 and consider optional assignment based on it (Abadie et al., 2018). However, even when we only have a one-time cross-sectional data, we can compute the bounds on the treatment effects conditional on Y_0 and discuss assignment mechanisms based on Y_0. In Tarozzi et al. (2015), there were two waves of surveys, one for pre-treatment and the other for post-treatment, but each wave surveyed different sets of households, so we cannot use the approach by Abadie et al. (2018) to estimate the treatment effects conditional on predicted Y_0. We apply our method to the post-treatment data to discuss optimal assignment of treatment based on Y_0.

To illustrate the point, consider assigning microfinance to those whose Y_0 is less

[1] In January 2006, 100 Birr was worth about 11.4 USD.



(a) Welfare measured by the net cash revenues from crops in Birr. The welfare lower bound is maximized at $b = 0$, i.e., by treating nobody.

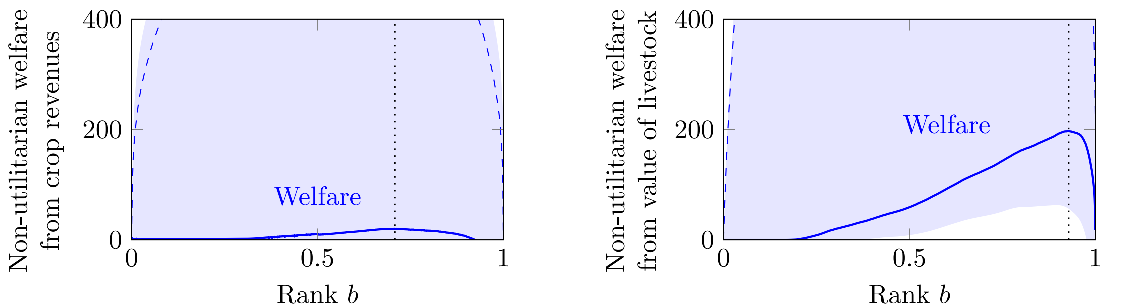
(b) Welfare measured by the total value of livestock owned in 2006 Birr. The dotted line ($b = 0.96$) maximizes the welfare lower bound.

Figure 3: Welfare bounds when the treatment is assigned to individuals whose Y_0 is below rank b. The solid blue lines give the lower bounds for the benefit, $\mathbb{E}[(Y_1 - Y_0)\mathbb{1}\{U < b\}]$, while the thin dashed blue lines give the upper bounds. The red lines are the social cost, which is taken to be 100 Birr per treated individual, $\mathbb{E}[100 \cdot \mathbb{1}\{U < b\}]$. The blue areas are the pointwise 95% confidence intervals for the social benefit.

than some value for both outcomes. The social benefit is the aggregate treatment effect, which equals $\mathbb{E}[(Y_1 - Y_0)\mathbb{1}\{U < b\}]$ in the per-capita basis. Since the average amount of outstanding loans from the microfinance for the treated individuals was 299 Birr, suppose conservatively that about one third of the balance goes sour, that is, the social cost is 100 Birr per capita. In sum, the social welfare from the assignment mechanism $U < b$ is

$$\mathbb{E}[(Y_1 - Y_0)\mathbb{1}\{U < b\}] - \mathbb{E}[100 \cdot \mathbb{1}\{U < b\}].$$

Figure 3 shows the graphs of the social benefits and costs for the two outcomes. The solid blue lines are the lower bounds for the social benefit per capita, and the dashed blue lines are the upper bounds. The blue shaded areas indicate the pointwise 95% confidence intervals for the social benefits using Imbens and Manski (2004). For the net cash revenues from crops, the lower bounds of the social benefits never exceed the social costs, so it may not be justifiable to assign microfinance in order to maximize the crop revenues (Figure 3a). On the other hand, the benefits measured by the value of livestock exceed the costs in the middle range of b (Figure 3b). As the subgroup welfare is not identified, we may go with the maximin approach: assign treatment so as to maximize the worst-case welfare. The lower bound for the social welfare



(a) Non-utilitarian welfare measured by the net cash revenues from crops. The dotted line ($b = 0.7081$) maximizes the welfare lower bound.

(b) Non-utilitarian welfare measured by the total value of livestock owned. The dotted line ($b = 0.928$) maximizes the welfare lower bound.

Figure 4: Non-utilitarian welfare bounds when the treatment is assigned to individuals whose Y_0 is below rank b. The welfare function weighs the loss 10% more than the gain. The solid blue lines give the lower bounds for the welfare, $\mathbb{E}[h(Y_1 - Y_0)\mathbb{1}\{U < b\}]$, while the thin dashed blue lines give the upper bounds. The blue areas are the pointwise 95% confidence intervals for the social welfare.

(the difference between the benefits and the costs) is maximized at $b = 0.96$, which corresponds to providing microfinance to individuals with the value of livestock below 9,200 Birr.

Our framework also allows for non-utilitarian welfare. Let h be a welfare function and suppose the policymaker wants to maximize the sum of $h(Y_1 - Y_0)$. If h cannot be written in the form of $g(Y_1) - g(Y_0)$ for some g, h is called *non-utilitarian welfare*. If h is nondecreasing and concave, Theorem 1 implies

$$\frac{1}{b}\int_0^b h(Q_1(b-u) - Q_0(u))du \leq \mathbb{E}[h(Y_1 - Y_0) \mid U < b] \leq \frac{1}{b}\int_0^b h(Q_1(1-b+u) - Q_0(u))du. \quad (2)$$

As an example, consider a loss-averse non-utilitarian welfare function

$$h(x) = \begin{cases} x & x \geq 0, \\ 1.1x & x < 0, \end{cases}$$

which overweighs the loss 10% more than the gain from the treatment. Such loss aversion may arise naturally when the provider of the treatment can be held responsible for negative treatment effects, such as doctors (Bordley, 2009) or policymakers

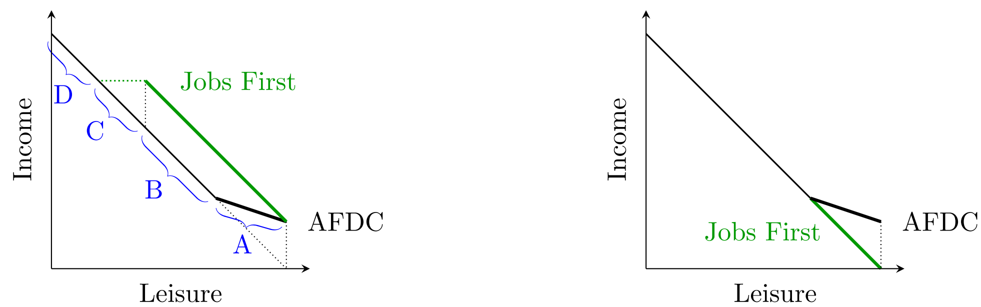
(Nicholson and Hellman, 2020).[2] Figure 4 shows the range of social welfare in (2) when the treatment eligibility is capped by a variable threshold. The vertical dotted black lines indicate the points where the lower bounds of the welfare are maximized. Again, taking the maximin approach to partially-identified social welfare, the maximizer for the lower bound in Figure 4b is found to be $b = 0.928$, which is when the access to microfinance is granted to those with the baseline total value of livestock owned below 7,299 Birr. The maximum of the lower bound in Figure 4a is attained at $b = 0.708$, which corresponds to the baseline net cash revenues from crops below 520 Birr, and is statistically significant, but whether the welfare improvement in this case is economically significant is a judgment call.

4 Subgroup Proportions of Winners

In 1996, the welfare program for low-income women went over a reform in a way that provided more generous benefits but with time limits on program participation. Bitler et al. (2006) argued that labor supply theory predicted heterogeneous effects of this reform on earnings, transfers, and income, and this called for more than the mean-impact analysis to evaluate the consequences of the reform. In particular, Bitler et al. (2006) used QTEs on RCT data to demonstrate that the effects were heterogeneous across different levels of outcome variables. In this section, we aim to push forward the boundary of quantitative knowledge we can infer from the data and provide complementary evidence to their findings.

Figure 5 illustrates the stylized budget constraints faced by women supported by Aid to Families with Dependent Children (AFDC) and by the Jobs First program in Connecticut. The horizontal axis is the time for leisure, which is a complement of working hours, and the vertical axis is the income. Assuming that leisure and consumption are normal goods, each program participant enjoys the income and leisure pair on the budget constraint. Under AFDC, the budget constraint is given by the two segments of black solid lines in Figure 5. The segment A corresponds to receiving the benefits from AFDC. The Jobs First program replaces the AFDC benefits with the green solid line, raising the budget constraint for individuals in A, B, and C. The Jobs First program also comes with a time limit; after 21 months, the

[2] On a related note, Heckman et al. (1997) questions utilitarian welfare in the context where individuals can choose the treatment.



(a) Budget constraints before time limit. (b) Budget constraints after time limit.

Figure 5: Income-leisure budget constraints under AFDC and under Jobs First (Bitler et al., 2006, Figure 1). Food stamp benefits are not included in the figures.

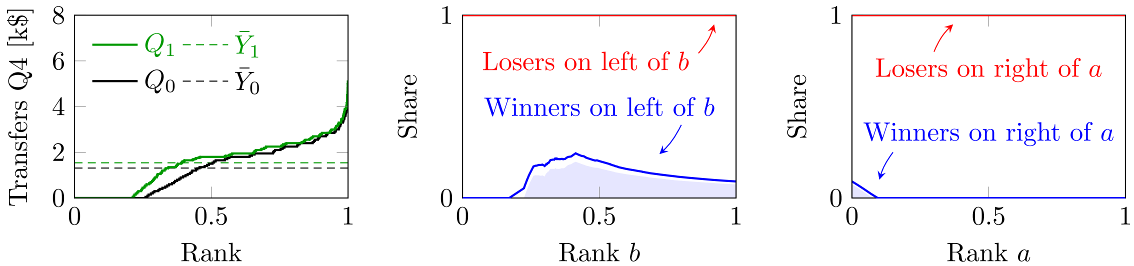
budget constraint is pushed down to just the thin black solid and dotted line with no benefits.[3] Before the time limit (Figure 5a), people in segments A, B, and C will move to somewhere on the green line; people in segment D may or may not move to a point on the green line. After the time limit (Figure 5b), individuals receive zero cash transfers and become worse off than AFDC (they may still receive food stamps).

As we will see below, this observation makes diverse predictions on how people react. Our outcome variables are transfers, earnings, and income, and individuals in different locations of the budget constraint may have responded in opposite directions. Bitler et al. (2006) used QTEs to reveal this heterogeneity, while also facing limitations that QTEs were not individual treatment effects. Our goal here is to provide more specific, quantitative analysis on the proportion and location of winners and losers. In this regard, Theorem 4 tells that the proportion of winners in an arbitrary interval $a < U < b$ is bounded by

$$\frac{1}{b-a} \sup_{x \in (a,b)} [x - a - F_1(Q_0(x))]_+ \leq P(Y_0 < Y_1 \mid a < U < b)$$
$$\leq P(Y_0 \leq Y_1 \mid a < U < b) \leq 1 - \frac{1}{b-a} \sup_{x \in (a,b)} [b - x - 1 + F_1(Q_0(x)-)]_+. \quad (3)$$

The lower bound can be interpreted as the proportion of "identified" winners, and one minus the upper bound as the proportion of "identified" losers. The difference between these bounds is the proportion of individuals for whom the data alone cannot conclude the direction of the treatment effect without making further assumptions.

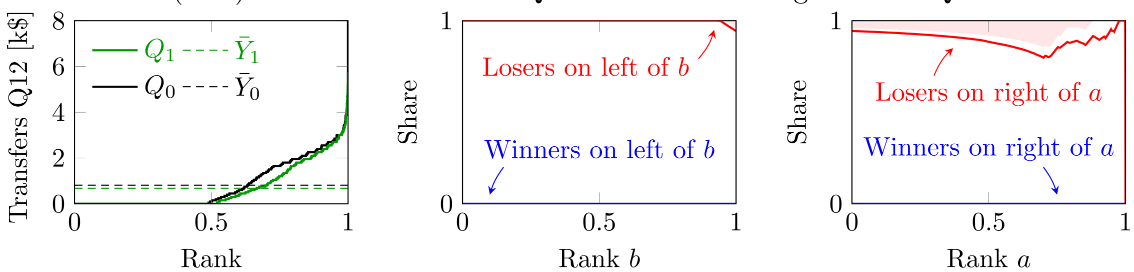
[3]There was 6-month extension under some conditions.



(a) Quantile functions in Q4. ATE = 230.1 (30.9).

(b) Winner bounds on the left tail in Q4.

(c) Winner bounds on the right tail in Q4.



(d) Quantile functions in Q12. ATE = −133.5 (29.9).

(e) Winner bounds on the left tail in Q12.

(f) Winner bounds on the right tail in Q12.

Figure 6: Estimated bounds on the proportions of winners in terms of transfers in Q4 and Q12. The proportion of winners is estimated to lie between the blue and red lines. The areas above the pointwise 95% confidence intervals (viz. significant losers) are shaded in red; the areas below (viz. significant winners) shaded in blue. Significant winners are found on the left tail in Q4 and significant losers on the right tail in Q12.

4.1 Heterogeneous Effects on Transfers

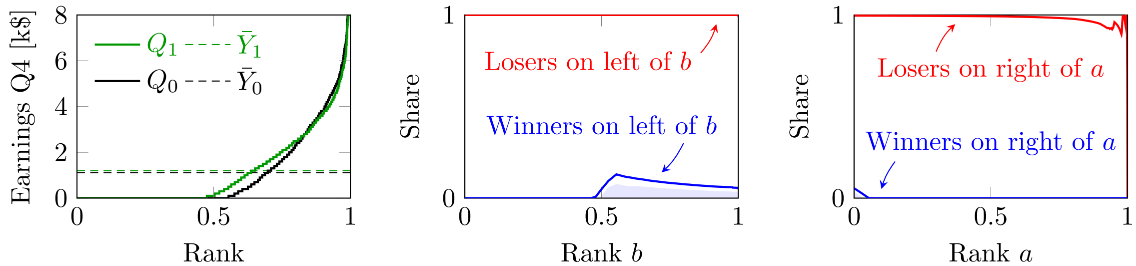
The first variable of interest is the transfers. In Figure 5, this corresponds to the difference of the actual income and the corresponding income on the budget constraint without benefits. The observed transfers also include the food stamps. Individuals in segments A, B, and C as well as some of D will receive more transfers in Jobs First before the time limit (Figure 5a), and individuals in A will receive less (may still receive nonzero transfers from food stamps) after the time limit (Figure 5b).

Figure 6 shows the quantile functions and the bounds on the proportions of winners for transfers in Quarter 4 (Q4) and Quarter 12 (Q12). Figures 6a and 6d give the quantile functions of Y_1 and Y_0 for transfers in Q4 and Q12. The mean effect in Q4 is 230.1 with standard error 30.9, and the mean effect in Q12 is -133.5 with standard error 29.9. Figures 6b and 6e plot the lower bounds for $P(Y_1 > Y_0 \mid U < b)$ as the blue lines and the upper bounds for $P(Y_1 \geq Y_0 \mid U < b)$ as the red lines. These bounds are half-median-unbiased as developed in Chernozhukov et al. (2013). For example, at $b = 0.414$, the lower bound gives 0.244, which means that among individuals who receive transfers less than \$1,000 under AFDC (that is, Y_0 lower than the 41.4th percentile), at least 24.4% of them receive transfers under Jobs First larger than they would under AFDC in Q4. The blue shaded areas on the bottom and the red shaded areas on the top indicate the region outside the pointwise 95% confindence intervals for the proportions of winners constructed by the same method as Chernozhukov et al. (2013). In other words, the blue shaded area indicates the "significant" proportion of winners and the red area the "significant" proportion of losers. Figures 6c and 6f plot the same when the conditioning is replaced with the right tail, $U > a$. Among individuals who receive transfers greater than \$1,350 ($a = 0.701$) under AFDC, 20.2% or more of them receive less transfers under Jobs First (Figure 6f).

4.2 Heterogeneous Effects on Earnings

The second variable of interest is the earnings from work, which are a decreasing function of leisure. In Figure 5a, individuals in segment A might adjust the working hours in either direction, and hence their earnings can go in either direction; individuals in segments B and C work less and receive less earnings; individuals in segment D might as well work less for less earnings if they move to the green line, or otherwise do not change working hours. After the time limit in Figure 5b, individuals in segment A will increase the working hours if leisure is a normal good.

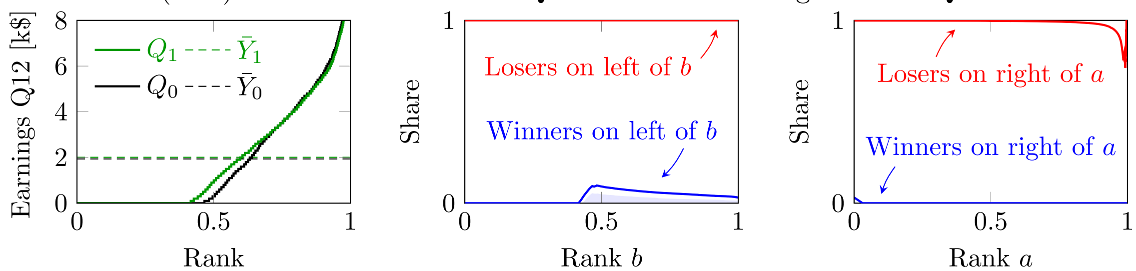
Figure 7 shows the corresponding quantile functions and bounds on the proportions of winners for earnings in Q4 and in Q12. The ATE of Jobs First on earnings is not significant in either Q4 or Q12; in particular, it is estimated to be 86.0 with standard error 53.9 in Q4 (Figure 7a) and 71.5 with standard error 75.5 in Q12 (Figure 7d). Figure 7b shows the proportions of winners and losers in the subgroups defined by $U < b$. In the very left tail, there is neither significant proportion of winners or of losers. Unlike the theory predicts, however, as we include the median and above, there appears a significant proportion of winners in earnings but not of losers. For example, one possibility is that at low working hours, a small monetary support from Jobs First can help individuals free their time from daily chores and they use the free time to work more, that is, leisure may be an inferior good when much of



(a) Quantile functions in Q4. ATE = 86.0 (53.9).

(b) Winner bounds on the left tail in Q4.

(c) Winner bounds on the right tail in Q4.



(d) Quantile functions in Q12. ATE = 71.5 (75.5).

(e) Winner bounds on the left tail in Q12.

(f) Winner bounds on the right tail in Q12.

Figure 7: Estimated bounds on the proportions of winners in earnings (proportional to working hours) in Q4 and Q12. Significant winners are found on the left in Q4, but no significant losers are found.

leisure time is spent for chores. Interestingly, this significant proportion of winners does not seem to disappear after the time limit (Figure 7e). In the upper tail, there is no significant proportion of winners or losers (Figures 7c and 7f).

5 Conclusion

In this paper, we proposed two new bounds related to the heterogeneity of the treatment effects that complement the commonly used QTE.

The first bounds were on the STE, which is the ATE in a subgroup whose Y_0 resides in an interval specified by the researcher. In Section 3, we went over an empirical example in which we are interested in assessing the effect of microfinance on poor individuals, and found that individuals whose total value of livestock owned was not too high received some positive ATE. We also applied our bounds to the problem of policy targeting in the context of social welfare. We considered two measures of welfare: one was linear in the treatment effect with a per-capita fixed cost and the other was non-utilitarian with a loss-aversive preference. We illustrated how our bounds can suggest the treatment assignment mechanisms that maximize the lower bounds of the social welfare.

The second bounds were on the proportions of winners and of losers in the same subgroup. In Section 4, we investigated the heterogeneous effects of welfare reform on transfers and earnings predicted by labor supply theory. For transfers, our bounds were able to confirm the presence of winners before the time limit and of losers after the time limit as the stylized theory predicted. For earnings, our bounds detected a significant proportion of winners in the mid- to left tail of the earnings distribution, which was not entirely in line with the theoretical prediction.

In both examples, we saw cases where our bounds could detect significant STE or proportion of winners even when the ATE estimates were not significant.

Appendix

A Theoretical Results

In this section, we formally state and prove our bounds.

The following theorem yields the second-order stochastic dominance bounds on the subgroup distribution of treatment effects. This extends the existing results on the entire distribution of treatment effects (Fan and Park, 2010, Lemma 2.2).

Theorem 1 (Second-order stochastic dominance bounds on subgroup distribution of treatment effects)**.** *Let $0 \leq a < b \leq 1$. Conditional on $a < U < b$, we have*

$$Q_1(b-U) - Q_0(U) \preceq_2 Y_1 - Y_0 \preceq_2 Q_1(1-b+U) - Q_0(U),$$

where $A \preceq_2 B$ means that the distribution of A is second-order stochastically dominated by that of B.

Proof. Let $h : \mathbb{R} \to \mathbb{R}$ be a nondecreasing concave function. Consider the optimal transport from $y \in (a, b)$ to $x \in (0, b-a)$ with cost $h(Q_1(x) - Q_0(y))$. Then, Rachev and Rüschendorf (1998, Theorem 3.1.2) imply

$$\mathbb{E}[h(Q_1(b-V_0) - Q_0(V_0))] \leq \mathbb{E}[h(Q_1(V_1) - Q_0(V_0))]$$

where $V_0 \sim U(a,b)$ and $V_1 \sim U(0,b-a)$. Then, Beiglböck et al. (2009, Theorem 3) imply that there exist functions ϕ and ψ such that for $x \in (0,b-a)$ and $y \in (a,b)$, $\phi(x)+\psi(y) \leq h(Q_1(x)-Q_0(y))$ with equality when $x = b-y$.[4] Here, ϕ is nondecreasing since for every $y' \geq y$,

$$\phi(b-y)+\psi(y) = h(Q_1(b-y)-Q_0(y)) \geq h(Q_1(b-y')-Q_0(y)) \geq \phi(b-y')+\psi(y).$$

Thus, we see that for $x \in (0,1)$ and $y \in (0,1)$,

$$\phi(x \wedge (b-a))+\psi(y)\mathbb{1}\{y \in (a,b)\}-\phi(b-a)\mathbb{1}\{y \notin (a,b)\} \leq h(Q_1(x)-Q_0(y))\mathbb{1}\{y \in (a,b)\}$$

with equality when $x = b-y$ and $y \in (a,b)$ or $x \geq b-a$ and $y \notin (a,b)$. Thus, Beiglböck et al. (2009, Theorem 3) imply

$$\mathbb{E}[h(Q_1(b-U)-Q_0(U))\mathbb{1}\{a<U<b\}] \leq \mathbb{E}[h(Y_1-Y_0)\mathbb{1}\{a<U<b\}],$$

that is, $Q_1(b-U)-Q_0(U) \preceq_2 Y_1-Y_0$ conditional on $a<U<b$.

Similarly, Rachev and Rüschendorf (1998, Theorem 3.1.2) imply

$$\mathbb{E}[h(Q_1(V_1)-Q_0(V_0))] \leq \mathbb{E}[h(Q_1(1-b+V_0)-Q_0(V_0))]$$

where $V_0 \sim U(a,b)$ and $V_1 \sim U(1-b+a,1)$. Beiglböck et al. (2009, Theorem 3) imply that there exist functions ϕ and ψ such that for $x \in (1-b+a,1)$ and $y \in (a,b)$, $h(Q_1(x)-Q_0(y)) \leq \phi(x)+\psi(y)$ with equality when $x = 1-b+y$. Here, ϕ is nondecreasing since for every $y' \geq y$,

$$\begin{aligned}\phi(1-b+y)+\psi(y) &= h(Q_1(1-b+y)-Q_0(y)) \\ &\leq h(Q_1(1-b+y')-Q_0(y)) \leq \phi(1-b+y')+\psi(y).\end{aligned}$$

Thus, we see that for $x \in (0,1)$ and $y \in (0,1)$,

$$\begin{aligned}&h(Q_1(x)-Q_0(y))\mathbb{1}\{y \in (a,b)\} \\ &\qquad \leq \phi(x \vee (1-b+a))+\psi(y)\mathbb{1}\{y \in (a,b)\}-\phi(1-b+a)\mathbb{1}\{y \notin (a,b)\}\end{aligned}$$

with equality when $x = 1-b+y$ and $y \in (a,b)$ or $x \leq 1-b+a$ and $y \notin (a,b)$. Thus,

[4] Beiglböck et al. (2009, Theorem 3) assume nonnegative costs, but the theorem trivially extends to costs bounded from below. Then optimality follows for cost $h(Q_1(\cdot)-Q_0(\cdot)) \vee A$ for every $A \in \mathbb{R}$, so let $A \to -\infty$.

Beiglböck et al. (2009, Theorem 3) imply

$$\mathbb{E}[h(Y_1 - Y_0)\mathbb{1}\{a < U < b\}] \leq \mathbb{E}[h(Q_1(1 - b + U) - Q_0(U))\mathbb{1}\{a < U < b\}],$$

that is, $Y_1 - Y_0 \preceq_2 Q_1(1 - b + U) - Q_0(U)$ conditional on $a < U < b$. ■

Theorem 1 is equivalent to the following lemma.

Lemma 2. *Let $f : \mathbb{R} \to \mathbb{R}$ be a nondecreasing convex function and $g : \mathbb{R} \to \mathbb{R}$ a nonincreasing convex function. For every $0 \leq a < b \leq 1$,*

$$\int_a^b f(Q_1(u-a)-Q_0(u))du \leq \mathbb{E}[f(Y_1-Y_0)\mathbb{1}\{a<U<b\}] \leq \int_a^b f(Q_1(1-u+a)-Q_0(u))du,$$
$$\int_a^b g(Q_1(1-b+u)-Q_0(u))du \leq \mathbb{E}[g(Y_1-Y_0)\mathbb{1}\{a<U<b\}] \leq \int_a^b g(Q_1(b-u)-Q_0(u))du.$$

Proof. Let $\tilde{f}(x) = -f(-x)$, $\tilde{U} = 1 - U$, and $\tilde{Y}_j = -Y_j$ so the quantile function of $\tilde{Y}_j$ is $\tilde{Q}_j(u) = -Q_j(1 - u)$. Since $\tilde{f}$ is nondecreasing and concave, Theorem 1 implies

$$\begin{aligned}\mathbb{E}[\tilde{f}(\tilde{Q}_1(1-a-\tilde{U})-\tilde{Q}_0(\tilde{U}))\mathbb{1}\{1-b < \tilde{U} < 1-a\}] &\leq \mathbb{E}[\tilde{f}(\tilde{Y}_1-\tilde{Y}_0)\mathbb{1}\{1-b < \tilde{U} < 1-a\}] \\ &\leq \mathbb{E}[\tilde{f}(\tilde{Q}_1(a + \tilde{U}) - \tilde{Q}_0(\tilde{U}))\mathbb{1}\{1 - b < \tilde{U} < 1 - a\}],\end{aligned}$$

which reduces to the first claim. Since $-g$ is nondecreasing and concave, the second claim follows straightforwardly from Theorem 1. ■

With this, we can prove the bounds on the STE.

Theorem 3 (Bounds on subgroup treatment effects)**.** *For every $0 \leq a < b \leq 1$,*

$$\int_a^b [Q_1(u-a)-Q_0(u)]du \leq \mathbb{E}[(Y_1-Y_0)\mathbb{1}\{a < U < b\}] \leq \int_a^b [Q_1(1+a-u)-Q_0(u)]du,$$
$$\int_a^b [Q_1(u-a)-Q_0(u)]_+du \leq \mathbb{E}[(Y_1-Y_0)_+\mathbb{1}\{a<U<b\}] \leq \int_a^b [Q_1(1+a-u)-Q_0(u)]_+du,$$
$$\int_a^b [Q_1(1-b+u)-Q_0(u)]_-du \leq \mathbb{E}[(Y_1-Y_0)_-\mathbb{1}\{a<U<b\}] \leq \int_a^b [Q_1(b-u)-Q_0(u)]_-du.$$

Proof. All results follow directly from Lemma 2. ■

Remark. We may use bootstrap to calculate the standard errors for the bounds in Theorems 1 and 3 and Lemma 2 (Kaji, 2018). Then, we can construct confidence intervals using Imbens and Manski (2004) or Stoye (2009).

The next theorem bounds the subgroup proportions of winners and losers.

Theorem 4 (Bounds on subgroup proportions of winners and losers). *For every* $0 \leq a < b \leq 1$,

$$\sup_{x \in (a,b)} [x - a - F_1(Q_0(x))]_+ \leq P(a < U < b, Y_0 < Y_1)$$
$$\leq b - a - \sup_{x \in (a,b)} [b - x - 1 + F_1(Q_0(x))]_+,$$
$$\sup_{x \in (a,b)} [b - x - 1 + F_1(Q_0(x)-)]_+ \leq P(a < U < b, Y_0 > Y_1)$$
$$\leq b - a - \sup_{x \in (a,b)} [x - a - F_1(Q_0(x)-)]_+.$$

There exists a joint distribution of (Y_0, Y_1) *for which a lower bound is attained; there exists a joint distribution of* (Y_0, Y_1) *for which an upper bound is arbitrarily tight. All bounds are nonincreasing and Lipschitz in* a *and nondecreasing and Lipschitz in* b.

Proof. For the lower bound, observe that

$$P(a < U < b, Q_0(U) < Y_1) \geq \sup_{x \in (a,b)} P(a < U < x, Q_0(x) < Y_1)$$
$$\geq \sup_{x \in (a,b)} [P(a < U < x) + P(Q_0(x) < Y_1) - 1]_+ = \sup_{x \in (a,b)} [x - a - F_1(Q_0(x))]_+,$$

where the second inequality uses the lower Fréchet inequality.[5] To show tightness, let $\Delta = \sup_{x \in (a,b)} [x - F_1(Q_0(x))]_+$ and consider $Y_1 = Q_1(U_1)$ such that $U_1 = U - a - \Delta$ if $a + \Delta < U < b$. The joint distribution of (U_1, U) when $U \notin (a + \Delta, b)$ is arbitrary. Then $U = u \in (a + \Delta, b)$ are not winners since

$$Q_1(u - a - \Delta) - Q_0(u) \leq Q_1(u - a - \Delta) - Q_1(F_1(Q_0(u))) \leq 0,$$

where the last inequality follows from the monotonicity of Q_1 and

$$(u - a - \Delta) - F_1(Q_0(u)) = [u - a - F_1(Q_0(u))] - \Delta \leq 0.$$

This means that the lower bound is binding for this joint distribution.

[5]We thank Ismael Mourifié for suggesting the use of the Fréchet inequality.

To derive the upper bound, observe that

$$\begin{aligned} P(a < U < b, Q_0(U) < Y_1) &= P(a < U < b) - P(a < U < b, Y_1 \leq Q_0(U)) \\ &\leq b - a - \sup_{x \in (a,b)} [b - x - 1 + F_1(Q_0(x))]_+, \end{aligned}$$

where the inequality follows as the lower bound. For tightness, let $\Delta = b - a - \sup_{x \in (a,b)}[b - x - 1 + F_1(Q_0(x))]_+$ and, for arbitrary $\varepsilon > 0$, consider $Y_1 = Q_1(U_1)$ such that $U_1 = U - a + 1 - \Delta + \varepsilon$ if $a < U < a + \Delta - \varepsilon$. The joint distribution of (U_1, U) for $U \notin (a, a + \Delta - \varepsilon)$ is arbitrary. Then $U = u \in (a, a + \Delta - \varepsilon)$ are winners since

$$\begin{aligned} F_1(Q_1(u - a + 1 - \Delta + \varepsilon)) - F_1(Q_0(u)) &\geq u - a + 1 - \Delta + \varepsilon - F_1(Q_0(u)) \\ &= b - a - [b - u - 1 + F_1(Q_0(u))] - \Delta + \varepsilon \geq \varepsilon > 0, \end{aligned}$$

which implies $Q_1(u - a + 1 - \Delta + \varepsilon) - Q_0(u) > 0$. Since this holds for arbitrarily small $\varepsilon > 0$, the upper bound can be arbitrarily tight.

The lower bound is nondecreasing and Lipschitz in b since for every $\varepsilon > 0$,

$$\begin{aligned} \sup_{x \in (a,b)} [x - a - F_1(Q_0(x))]_+ &\leq \sup_{x \in (a,b+\varepsilon)} [x - a - F_1(Q_0(x))]_+ \\ &\leq \sup_{x \in (a,b+\varepsilon)} [x - a - F_1(Q_0(x \wedge b))]_+ \leq \sup_{x \in (a,b)} [x - a - F_1(Q_0(x))]_+ + \varepsilon. \end{aligned}$$

The lower bound is nonincreasing in a since for every $\varepsilon > 0$,

$$\sup_{x \in (a-\varepsilon,b)} [x - a + \varepsilon - F_1(Q_0(x))]_+ \geq \sup_{x \in (a-\varepsilon,b)} [x - a - F_1(Q_0(x))]_+ \geq \sup_{x \in (a,b)} [x - a - F_1(Q_0(x))]_+$$

and is Lipschitz in a since for every $\varepsilon > 0$,

$$\begin{aligned} \sup_{x \in (a-\varepsilon,b)} [x - a + \varepsilon - F_1(Q_0(x))]_+ &\leq \sup_{x \in (a-\varepsilon,a]} (x - a + \varepsilon)_+ \vee \sup_{x \in (a,b)} [x - a + \varepsilon - F_1(Q_0(x))]_+ \\ &\leq \varepsilon \vee \left(\sup_{x \in (a,b)} [x - a - F_1(Q_0(x))]_+ + \varepsilon \right) = \sup_{x \in (a,b)} [x - a - F_1(Q_0(x))]_+ + \varepsilon. \end{aligned}$$

The same properties follow analogously for the upper bound.

The loser bounds can be likewise derived. ■

Remark. Theorem 4 can be trivially extended to the bounds on $P(a < U < b, Y_1 - Y_0 < c)$ by shifting the distribution of either Y_0 or Y_1. Then, setting $a = 0$ and $b = 1$ makes them reduce to the classical Makarov bounds (Makarov, 1981).

Remark. Theorem 4 is in the form of intersection bounds for which Chernozhukov et al. (2013) provided a method for estimation and inference.

References

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

ABADIE, A., J. ANGRIST, AND G. IMBENS (2003): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117.

ABADIE, A., M. M. CHINGOS, AND M. R. WEST (2018): "Endogenous Stratification in Randomized Experiments," *Review of Economics and Statistics*, 100, 567–580.

ARON-DINE, A., L. EINAV, AND A. FINKELSTEIN (2013): "The RAND Health Insurance Experiment, Three Decades Later," *Journal of Economic Perspectives*, 27, 197–222.

ATHEY, S. AND G. W. IMBENS (2006): "Identification and Inference in Nonlinear Difference-In-Differences Models," *Econometrica*, 74, 431–497.

BEIGLBÖCK, M., M. GOLDSTERN, G. MARESCH, AND W. SCHACHERMAYER (2009): "Optimal and Better Transport Plans," *Journal of Functional Analysis*, 256, 1907–1927.

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.

BORDLEY, R. F. (2009): "The Hippocratic Oath, Effect Size, and Utility Theory," *Medical Decision Making*, 29, 377–379.

CHERNOZHUKOV, V. AND C. HANSEN (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81, 667–737.

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

Firpo, S., N. M. Fortin, and T. Lemieux (2009): "Unconditional Quantile Regressions," *Econometrica*, 77, 953–973.

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

Gunsilius, F. F. (2023): "Distributional Synthetic Controls," *Econometrica*, 91, 1105–1117.

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.

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Imbens, G. W. and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

Kaji, T. (2018): "Essays on Asymptotic Methods in Econometrics," Ph.D. thesis, Massachusetts Institute of Technology.

Kitagawa, T. and A. Tetenov (2018): "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 86, 591–616.

——— (2021): "Equality-Minded Treatment Choice," *Journal of Business & Economic Statistics*, 39, 561–574.

Koenker, R. and G. Bassett, Jr (1978): "Regression Quantiles," *Econometrica*, 46, 33–50.

Makarov, G. D. (1981): "Estimates for the Distribution Function of a Sum of Two Random Variables When the Marginal Distributions are Fixed," *Theory of Probability & Its Applications*, 26, 803–806.

Monahan, J. F. (2011): *Numerical Methods of Statistics*, Cambridge: Cambridge University Press, second ed.

Nicholson, K. M. and D. Hellman (2020): "Opioid Prescribing and the Ethical Duty to Do No Harm," *American Journal of Law & Medicine*, 46, 297–310.

Powell, D. (2020): "Quantile Treatment Effects in the Presence of Covariates," *Review of Economics and Statistics*, 102, 994–1005.

Rachev, S. T. and L. Rüschendorf (1998): *Mass Transportation Problems*, vol. I: Theory, New York: Springer-Verlag.

Stoye, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 1299–1315.

Tarozzi, A., J. Desai, and K. Johnson (2015): "The Impacts of Microcredit: Evidence from Ethiopia," *American Economic Journal: Applied Economics*, 7, 54–89.

Tetenov, A. (2012): "Identification of Positive Treatment Effects in Randomized Experiments with Non-Compliance," Working paper.