Partial Identification of the Distribution of Treatment Effects with an Application to the Knowledge Is Power Program (KIPP)

Brigham R. Frandsen* Lars J. Lefgren*

July 13, 2020

Abstract

We bound the distribution of treatment effects under plausible and testable assumptions on the joint distribution of potential outcomes, namely that potential outcomes are mutually stochastically increasing. We show how to test the empirical restrictions implied by those assumptions. The resulting bounds substantially sharpen bounds based on classical inequalities. We apply our method to estimate bounds on the distribution of effects of attending a Knowledge is Power Program (KIPP) charter school on student achievement, and find that a substantial majority of students' math achievement benefitted from attendance, especially those who would have fared poorly in a traditional classroom.

1 Introduction

What fraction of patients benefit from a certain medical procedure? What is the median response of subjects to a treatment? How does the effect of a novel education intervention vary by the outcome that would have been realized in a traditional classroom? Questions such as these are often of great interest to researchers, policy makers, and individuals.

*Department of Economics, Brigham Young University. The authors thank Brigham Young University College of Family, Home, and Social Sciences for financial support. This research was approved by the Brigham Young University Institutional Review Board. We thank Ryan Hill for excellent research assistance.

For example, parents and policy makers may be rightly concerned about a program that harms a substantial fraction of participants despite a positive average effect, particularly given that a positive average effect often depends on how outcomes are scaled. Additionally, parents considering enrolling a child in an education intervention may have private information regarding their child's likely outcome in a traditional classroom. Knowledge about how the effects of an intervention vary across students who would do well or do poorly in a conventional setting can then improve the efficiency of enrollment decisions and minimize the probability that the student is harmed by placement in an ill-suited program.

Yet even ideal experimental data, under standard assumptions, cannot identify the answers to questions such as these concerning the distribution of treatment effects. The reason for this is that experimental data identify the separate marginal distributions of potential outcomes under treatment and control, not the joint distribution. Consequently, researchers can identify parameters that are functions of the marginal distributions, such as quantile treatment effects, which compare the distributions of potential outcomes at different quantiles, or average treatment effects. Researchers cannot, however, identify parameters that depend on the joint distribution of potential outcomes such as the fraction of subjects harmed by the treatment, the median treatment effect, or the expected treatment effect given a subject's outcome in the control state.

While the distribution of treatment effects is not point identified, it may be bounded. The marginal distributions of potential outcomes themselves imply bounds on the joint distribution via the classical Fréchet-Hoeffding limits. Bounds based on these limits, however, are typically very wide, precluding meaningful economic inferences. We develop a method that provides much tighter bounds. Our method relies on an assumption about the joint distribution of potential outcomes that is both plausible in many economic contexts and testable. In particular, we assume that potential outcomes are mutually stochastically increasing: the distribution of outcomes under treatment among individuals who would have realized a higher outcome in the control state (weakly) stochastically dominates that among individuals who would have realized a lower outcome in the control state, and vice versa. In an education setting, this means that if student A performs better than student B in the control state, student A would likely have outperformed student B in the treated state, and vice versa. The assumption should be plausible in many economic settings, and indeed is even implied by some models of how individuals select into programs (Heckman et al., 1997).

This assumption substantially sharpens the classical bounds on the overall distribution of treatment effects, and also implies bounds on the conditional distribution of treatment effects at each point of the control and treated distribution. Thus we can place informative bounds on parameters such as the overall fraction of individuals harmed by treatment and the median treatment effect. We also can bound the average treatment effect, median treatment effect, and probability of being harmed for a student who performed, say, one standard deviation below the control mean in the absence of treatment. We can create similar bounds for a treated individual who performed one standard deviation below the treated mean.

What can be learned from our approach beyond what can be learned from traditional analyses of treatment effect heterogeneity? The traditional approach to analyzing treatment effect heterogeneity focuses on how average effects vary by subgroup (see List et al., 2019; Lee and Shaikh, 2013; Crump et al., 2008; Djebbari and Smith, 2008; and Bitler et al., 2017 for discussion). Examining average treatment effects by subgroup can be useful for targeting policy interventions and exploring mechanisms, but it cannot answer questions regarding the fraction of individuals helped or harmed by a treatment without strong assumptions like rank invariance or constant treatment effects within subgroups. Similarly, individuals often have information regarding their potential outcome in the control or treated state not available to researchers. In this case bounds on the distribution of treatment effects given one's potential outcome in the treated or control state *conditional on observed variables* provides relevant information beyond average treatment effects by observed subgroup.

The bounds on the treatment effect distribution conditional on potential outcomes can be calculated using standard nonparametric or semiparametric regression techniques. The bounds on the overall treatment effect distribution can be calculated using either of two methods. The first, and computationally faster, method simply integrates over the conditional bounds. This yields somewhat conservative bounds, since it implicitly imposes worst-case assumptions pointwise across the distribution, while the uniformly worst-case assumption would not correspond to the worst case at every point. The second method is computationally more intensive but yields tighter bounds, and involves searching across the set of joint distributions of control and treated outcomes that satisfy stochastic increasingness but yield the observed marginal distributions.

We show how incorporating covariates that are predictive of outcomes in the control and treatment states can substantially tighten the bounds. Consequently, our bounds will be most informative when highly predictive covariates exist. Our bounds will also tend to be more informative regarding the fraction of students benefitted by treatment when the average treatment effect is large in absolute value.

We propose a test of stochastic increasingness that probes the implication that potential outcomes are positively correlated. Although the correlation between potential outcomes cannot be computed directly, we can compute the correlation of predicted outcomes in the control and treament states. If the covariates are sufficiently predictive of outcomes in the treatment and control states and if the predicted outcomes are sufficiently highly correlated themselves, we demonstrate that potential outcomes in the treated and control states are positively correlated—a necessary condition for our assumption of stochastic increasingess.

We apply our method to calculate bounds on the distribution of effects on student achievement of the Knowledge Is Power Program (KIPP) charter school in Lynn, Massachusetts. Our bounds imply that the substantial majority of students who attended the charter school benefitted in terms of mathematics achievement. Furthermore, our bounds allow us to say definitively that students who would have performed poorly in the control state experienced a large positive average treatment effect and were very likely to benefit from KIPP attendance. These results are novel relative to prior findings on KIPP attendance, which focus on the average treatment effect. Our bounds are also much tighter than bounds that do not impose stochastic increasingness.

The next section describes our paper's contribution relative to prior work on the distribution of treatment effects. Section 3 develops our econometric framework, defines the restrictions we propose, shows how to test these assumptions, derives the implied bounds on the distribution of treatment effects, and shows how they are identified in the data. Section 4 applies the bounds and the testing procedure in the KIPP setting. Section 5 concludes.

2 Relationship to Previous Literature on the Distribution of Treatment Effects

Prior researchers have developed methods to bound the distribution of treatment effects. Williamson and Downs (1990), Heckman et al. (1997), Fan and Wu (2010) and Russell (2019) derive bounds on features of the joint distribution of potential outcomes using only information contained in the marginal distributions. Fan and Park (2010) show how to perform inference on these bounds. Firpo and Ridder (2010) develop uniform versions of the bounds, resulting in tighter bounds on functionals of the treatment effect distribution. These papers rely on the fact that the marginal distributions of control and treatment outcomes themselves restrict the joint distribution via the well-known Fréchet-Hoeffding bounds. Unfortunately, these bounds, which place no additional restrictions on the joint distribution of outcomes, tend to be quite wide. Often, one cannot rule out harm to a substantial majority of subjects, even in the presence of a positive average effect. Furthermore, bounds on the conditional distribution of treatment effects given the outcome in the non-treated state tend to be extremely wide since any outcome in the support of the control distribution can correspond to any outcome in the support of the treated distribution. For these reasons, such bounds tend to preclude meaningful economic inferences.

Additional restrictions are therefore required to meaningfully bound the distribution of treatment effects. Manski (1997) proposes the restriction that treatment responses are monotone, and derives the resulting bounds. Kim (2014) generalizes this approach to other deterministic restrictions on the support of treatment effects. These restrictions allow the bounds to be substantially tightened, but may be too strong to be plausible in many empirical settings.

Economic models sometimes imply restrictions that can meaningfully bound the distribution of treatment effects. Heckman et al. (1997) employ Roy-type assumptions on treatment selection and Kline and Tartari (2016) use a more general framework in which treatments alter utilities in known directions to tighten bounds on the distribution of treatment effects. Our method complements these approaches in that it applies in settings where the utility consequences of treatment alternatives may not be known by the researcher.

Heckman et al. (1997) and Fan and Park (2009) show the bounds on the distribution of treatment effects can sometimes be tightened if one assumes that a dependence measure (e.g., Kendall's τ) between potential outcomes is known. Our results complement this approach by relaxing the need to specify a known measure, and instead assumes only the direction of dependence is known. The restriction we propose, stochastic increasingness of potential outcomes (defined in the following section), implies substantially tighter bounds than those of Williamson and Downs (1990), should be plausible in many applied settings, and is testable.

Stochastic increasingness or related assumptions have been used in other settings. In a sample selection setting, Blundell et al. (2007), Lechner and Melly (2010), and Blanco et al. (2013) employed stochastic dominance assumptions across working and nonworking individuals to estimate features of the distribution of wages. Imai (2008) adopted a similar approach when the source of sample selection is attrition. To estimate measures of intergenerational mobility, Chetty et al. (2016) imposed stochastic increasingness of fathers' and sons' earnings. Heckman and Smith (1998) examine whether treated potential outcomes second-order stochastically dominate untreated potential outcomes to test the ex ante rationality of program participation.

3 Econometric Framework

Consider a binary indicator, D, for a treatment that possibly affects a continuously distributed outcome Y with support $\mathcal{Y}$. Let $Y(1)$ and $Y(0)$ be potential outcomes with and without treatment, with marginal cdfs F_1 and F_0. Observed variables are the outcome, $Y = Y(D)$, and the treatment indicator D. For clarity, we first consider the case without covariates, and where treatment D is independent of potential outcomes.

We show in the appendix how covariates may be incorporated to tighten the bounds, and how instrumental variables methods can be incorporated, if necessary, to aid in identification.

The parameters of interest in this paper are features of the distribution of treatment effects $\Delta := Y(1) - Y(0)$ (with support $\mathcal{D}$), including the cdf, F_Δ; the conditional cdf given $Y(d)$, $F_{\Delta|Y(d)}, d \in \{0,1\}$; and the expectation conditional on $Y(d)$, $E[\Delta|Y(d)], d \in \{0,1\}$. These parameters are typically of policy and economic importance, but, unlike the marginal distributions of potential outcomes, are not directly identified by experimental data. The parameters are not identified because they depend on the joint distribution of $Y(0)$ and $Y(1)$, which are never jointly observed. The marginal distributions F_1 and F_0 themselves impose some restrictions on the joint distribution via the Fréchet-Hoeffding bounds, but these are rarely tight enough to imply economically meaningful restrictions. As discussed above, economically meaningful bounds in the current literature require strong, typically implausible assumptions. The bounds we construct here substantially improve upon the Fréchet-Hoeffiding bounds and the related bounds on the distribution of treatment effects derived by Williamson and Downs (1990) and discussed by Fan and Park (2010) and Fan et al. (2014) by imposing natural—and testable—restrictions on the joint distribution of potential outcomes.

3.1 Bounding the Distribution of Treatment Effects

The separate distributions of $Y(0)$ and $Y(1)$ themselves imply bounds on the joint distribution of $(Y(1), Y(0))$ and also the distribution of $Y(1) - Y(0)$. The well-known Fréchet-Hoeffding bounds provide upper and lower bounds on the joint distribution of $(Y(1), Y(0))$, while the following expressions due to Williamson and Downs

(1990) provide upper and lower bounds on the distribution of their difference—that is, the distribution of treatment effects:

$$F_{\Delta}^{L}(t) = \sup_{y} \max\{F_1(y) - F_0(y-t), 0\}, \quad (1)$$

$$F_{\Delta}^{U}(t) = 1 + \inf_{y} \min\{F_1(y) - F_0(y-t), 0\}. \quad (2)$$

These bounds, while attractive in that they impose no restrictions on the joint distribution of $(Y(1), Y(0))$, are often uninformative. Further restrictions are required to provide more informative bounds.

The restriction we propose assumes that potential outcomes are mutually stochastically increasing:

Definition 1 *Potential outcomes* $Y(0)$ *and* $Y(1)$ *are* ***mutually stochastically increasing*** *if* $\Pr(Y(1) \leq t | Y(0) = y)$ *and* $\Pr(Y(0) \leq t | Y(1) = y)$ *are each non-increasing in* y *almost everywhere.*

Lehmann (1966) described the property of stochastic increasingness, referring to it as positive regression dependence. It means that individuals with higher potential outcomes in one treatment state draw from a more favorable—in the first-order stochastic dominant sense—conditional distribution of outcomes in the other state. It is a generalization of constant treatment effects restrictions and the rank invariance assumption discussed in Heckman et al. (1997) and Chernozhukov and Hansen (2005). Under rank invariance, an individual's position in the control distribution is implied by her position in the treated distribution. Hence, rank invariance implies that treatment effects are constant conditional on $Y(0)$. Stochastic increasingness, however, makes no restrictions on treatment effect heterogeneity other than what is implied by the support of potential outcomes, even conditional on $Y(0)$. Stochastic increasingness is satisfied whenever $Y(1)$ and $Y(0)$ are positively likelihood ratio dependent, and it implies that $Y(1)$ and $Y(0)$ are positively correlated (in levels and rank). Stochastic increasingness implies positive quadrant dependence, a condition assumed in Bhattacharya et al. (2012). They imposed positive quandrant dependence between unobserved determinants of the outcome and treatment selection in order to construct bounds on treatment effects. We impose stochastic increasingness not between the outcome and treatment, but between potential outcomes. Stochastic increasingness rules out negative dependence between potential outcomes and can be tested, as we discuss below in Section 3.4.

Stochastically increasing potential outcomes should be a plausible assumption in many economic settings. For example, students with strong family backgrounds and high levels of prior knowledge are likely to outperform students without such advantages in most settings, including treatment and control situations. In a clinical setting, the pre-treatment level of morbidity would tend to cause those who do well in the control group to also do well in the treatment group. Unemployed workers with strong literacy and numeracy skills are likely to do better than workers without such skills both in a control setting as well as a treatment setting in which they've been randomized into a job training program. All of these situations would satisfy stochastic increasingness and seem very plausible. Note that stochastic increasingness does not restrict the possible support of individual-level treatment effects, as other approaches do (Manski, 1997; Kim, 2014); the support of treatment effects is governed only by the supports of $Y(0)$ and $Y(1)$.

Situations in which stochastic increasingness would be violated, in contrast, often seem unusual. This would happen if, on average, unobserved characteristics that were beneficial in the control state were harmful in the treatment state. This would also occur if two latent skills had different relative skill prices in the treatment and control state and were strongly negatively correlated. Similarly, interventions designed with the goal of reversing relative positions of subjects could be problematic. In such cases, one ought to be cautious about using our methodology.

3.1.1 Bounds on the treatment effect distribution given $Y(0)$ or $Y(1)$

Under the stochastically increasing property, the conditional distribution of the treatment effect given $Y(0)$ (or $Y(1)$) can be sharply bounded by a function of the separate marginal distributions of $Y(0)$ and $Y(1)$, as the following theorem establishes.

Theorem 2 *Suppose $Y(1)$ and $Y(0)$ are continuously distributed and mutually stochastically increasing. Then $F_{\Delta|Y(0)}(t|Y(0)) := \Pr(\Delta \leq t|Y(0))$ is sharply bounded from below pointwise in t and $Y(0)$ by*

$$F^{L}_{\Delta|Y(0)}(t|Y(0)) := \begin{cases} 0 & , \quad Y(0) > F_0^{-1}(F_1(Y(0)+t)) \\ \frac{F_1(Y(0)+t)-F_0(Y(0))}{1-F_0(Y(0))} & , \quad Y(0) \leq F_0^{-1}(F_1(Y(0)+t)) \end{cases} \tag{3}$$

and from above by

$$F^{U}_{\Delta|Y(0)}(t|Y(0)) := \begin{cases} \frac{F_1(Y(0)+t)}{F_0(Y(0))} & , \quad Y(0) \geq F_0^{-1}(F_1(Y(0)+t)) \\ 1 & , \quad Y(0) < F_0^{-1}(F_1(Y(0)+t)) \end{cases} . \tag{4}$$

Proof. See the appendix. ■

Theorem 2 gives bounds on the conditional distribution of treatment effects—which in general depends on the unidentified joint distribution of $(Y(0), Y(1))$—as a function of the separate marginal distributions of potential outcomes, which are identified. The bounds themselves are proper probability distributions. Mutual stochastic increasingness also implies analogous bounds on the conditional distribution of treatment effects given $Y(1)$, denoted $F^{L}_{\Delta|Y(1)}$ and $F^{U}_{\Delta|Y(1)}$, expressions for which are provided in the proof. The result assumes outcomes are continuously distributed. Outcomes with mass points can be accommodated by replacing the cdf inverses with generalized left inverses such that

$$F_0^{-1}(\tau) = \inf \{x : F_0(x) \geq \tau\}.$$

Figure 1 provides graphical intuition for the bounds. The top panels show the density of potential outcomes in the control state. Consider an individual with $Y(0) = y$ in the control distribution. The conditional cdf given in the lower left panel of the figure corresponds to the worst-case conditional distribution for $Y(1)$ given $Y(0) = y$ that is still consistent with stochastic increasingness. The worst case occurs when an individual with a given $Y(0)$ has zero probability of drawing a $Y(1)$ that exceeds his or her rank in the control state—that is, rank invariance holds above the individual's rank—and instead draws from the truncated distribution of $Y(1)$ below his or her rank in the control state—that is, rank independence holds below the individual's rank. This satisfies stochastic increasingness because the individual's distribution of counterfactual outcomes is strictly worse than everyone with an outcome in the control state better than hers and weakly better (exactly the same in this case) than everyone with a control outcome below hers. Even under this worst case, bounds on the distribution of treatment effects are made tighter by the fact that expected outcomes in the treated state are increasing in the control state outcome.

The corresponding best case, shown in the right panel of Figure 1, is just the opposite: an individual with a given $Y(0)$ has zero probability of drawing a $Y(1)$ *below* his or her rank in the control state, but instead draws from the truncated distribution of $Y(1)$ ranks *above* his or her rank in the control state. The best case corresponds to the lower-bound cdf given in equation (3).

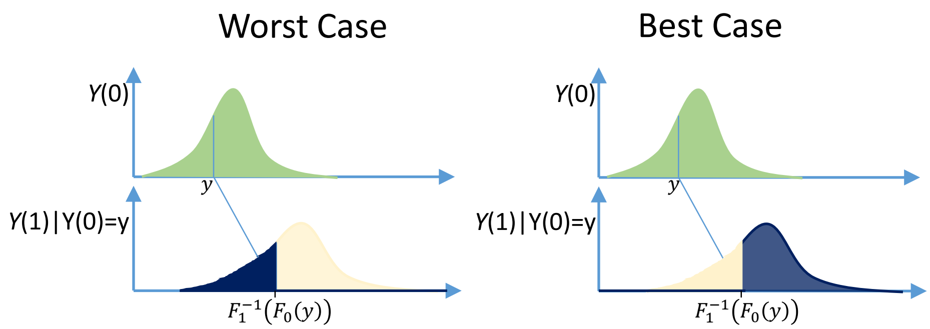

Figure 1: The figure illustrates intuitively the bounds implied by stochastic increasingness of potential outcomes on the conditional distribution of treated potential outcomes given the untreated potential outcome. The graphs in the top row plot the density of untreated potential outcomes, $Y(0)$. The graphs in the bottom row plot the worst- and best-case conditional density of the treated potential outcomes $Y(1)$ given $Y(0) = y$ as the dark shaded region, rescaled and superimposed on the marginal density of $Y(1)$ (light shaded). The worst-case plot on the left corresponds to the upper-bound cdf of treatment effects in equation (4). The best-cast plot on the right corresponds to the lower-bound cdf of treatment effects in equation (3).

These bounds on the treatment effect cdf also imply bounds on the average treatment effect conditional on $Y(0)$ (or $Y(1)$), a quantity that is frequently of great interest in applications, but not point identified. Let the average treatment effect conditional on $Y(d)$ be denoted $\Delta(Y(d)) := E[Y(1) - Y(0) | Y(d)]$. By definition, bounds on the conditional expectation are given by integrating the derivative of the cdf bounds:

$$\Delta^L(Y(d)) = \int t dF^U_{\Delta|Y(d)}(t|Y(d)), \tag{5}$$

$$\Delta^U(Y(d)) = \int t dF^L_{\Delta|Y(d)}(t|Y(d)). \tag{6}$$

3.1.2 Bounds on the overall treatment effect distribution

Bounds on the overall distribution of treatment effects can be constructed by taking the expectation of the conditional bounds:

$$\begin{aligned}\tilde{F}^{L}_{\Delta}(t) &= \max_{d\in\{0,1\}} E\left[F^{L}_{\Delta|Y(d)}(t|Y(d))\right] && (7)\\ \tilde{F}^{U}_{\Delta}(t) &= \min_{d\in\{0,1\}} E\left[F^{U}_{\Delta|Y(d)}(t|Y(d))\right]. && (8)\end{aligned}$$

Integrating over the conditional bounds in this way will yield conservative bounds on the overall distribution, however, since the conditional bounds $F^{L}_{\Delta|Y(d)}$ and $F^{U}_{\Delta|Y(d)}$ are by construction sharp pointwise in $Y(d)$, but not uniformly.

Sharp bounds on the overall distribution of treatment effects based on conditional stochastic increasingness can be obtained in principle by searching over the set of joint distributions of $(Y(0), Y(1))$ that satisfy mutual stochastic increasingness, and of course yield the observed marginal distribution distributions of $Y(0)$ and $Y(1)$. Defining $\mathcal{C}^{SI}$ to be the set of bivariate copula functions H that satisfy mutual stochastic increasingness, we can define sharp bounds on the overall distribution of treatment effects as

$$\begin{aligned}F^{L}_{\Delta}(t) &= \inf_{H(\cdot,\cdot)\in\mathcal{C}^{SI}} \int\int 1\left(F_1^{-1}(v) - F_0^{-1}(u) \leq t\right) H(u,v)\,dudv, && (9)\\ F^{U}_{\Delta}(t) &= \sup_{H(\cdot,\cdot)\in\mathcal{C}^{SI}} \int\int 1\left(F_1^{-1}(v) - F_0^{-1}(u) \leq t\right) H(u,v)\,dudv. && (10)\end{aligned}$$

As functions of the observed F_1 and F_0 only, these bounds are in principal identified, although the infinite dimensional optimization problem that defines them may complicate estimation, as described below.

These results can be applied directly to bound quantities such as the fraction of individuals who are harmed by treatment (i.e., the cdf of Δ evaluated at zero), but can also be used to construct sharp bounds on any feature of the distribution of treatment effects that is monotonic in the cdf in a stochastically dominant sense, such as the expectation or any quantile of the treatment effect.

Our bounds are substantially tighter than the bounds based on classical limits. Appendix Figures A5 and A6 report numerical simulations of the bounds on the fraction hurt by treatment, and in all cases our bounds are much narrower than the Williamson-Downs bounds.

When are the bounds tightest? The numerical simulations reported in the appendix illustrate factors that determine the tightness of the bounds. First, the bounds on the fraction hurt by treatment are more informative the larger in magnitude the average treatment effect (or other central measure of the treatment effect size). The intuition for this is that with such minimal restrictions on heterogeneity, even a treatment with a small average effect is nevertheless consistent with a large fraction of individuals being either slightly helped or harmed by treatment. Second, the bounds can be substantially tightened by introducing covariates that predict outcomes, which we show how to incorporate in the appendix. A practical implication is that researchers will benefit from collecting a rich set of covariates, especially lagged outcomes, as part of the study design.

3.2 Estimating the Bounds

The conditional cdf bounds (3) and (4) can be consistently estimated by plugging in consistent estimators for the conditional cdfs F_1 and F_0. Here we give details for the simplest case where D_i is independent of potential outcomes. See the appendix for estimation details when covariates are available or instrumental variables methods are necessary.

Bounds on the distribution of treatment effects given some untreated potential outcome value y can be constructed via the following steps:

1. Construct $\hat{F}_0(y)$ as the sample mean of the indicator $1(Y_i \leq y)$ in the untreated subsample
2. Construct $\hat{F}_1(y+t)$ as the sample mean of the indicator $1(Y_i \leq y+t)$ in the treated subsample
3. Plug in to form estimates of the bounds

$$\hat{F}^L_{\Delta|0}(t|Y(0)=y) \quad : \quad = \max\left\{0, \frac{\hat{F}_1(y+t) - \hat{F}_0(y)}{1 - \hat{F}_0(y)}\right\} \tag{11}$$

$$\hat{F}^U_{\Delta|0}(t|Y(0)=y) \quad : \quad = \min\left\{1, \frac{\hat{F}_1(y+t)}{\hat{F}_0(y)}\right\}. \tag{12}$$

Bounds (5) and (6) on the conditional expectation of treatment effects given $Y(0)$ can be computed by integrating over the numerical derivative of the cdf estimates (11) and (12) on a discrete grid. Analogous steps can be followed for the distribution of treatment effects given $Y(1)$.

Bounds on the overall cdf of treatment effects can be constructed in either of two ways, following the discussion in Section 3.1.2. The computationally simpler method takes the sample averages of (11) and (12) evaluated at the observed outcomes in the untreated sample:

$$\hat{F}^L_{\Delta}(t) \quad : \quad = \frac{1}{n_0} \sum_{i:D_i=0} \hat{F}^L_{\Delta|0}(t|Y_i) \tag{13}$$

$$\hat{F}^U_{\Delta}(t) \quad : \quad = \frac{1}{n_1} \sum_{i:D_i=0} \hat{F}^U_{\Delta|0}(t|Y_i), \tag{14}$$

where n_0 and n_1 are the number of untreated and treated observations, respectively. Appendix E provides additional estimation details and describes a Stata software implementation, which also allows for additional covariates and instrumental variables.

While computationally simple, these bounds may be conservative. The second method computes a numerical approximation to the sharp uniform bounds defined in (9) and (10) by adapting Chetty et al.'s (2016) computational algorithm for optimizing over the space of discrete copulae defined on a $k \times k$ grid, subject to the mutual stochastic increasingness constraints. This approximation is the solution to a linear programming problem, and can be computed relatively quickly for grid sizes on the order of one hundred. The grid approximation error can be made arbitrarily small for large k. The appendix provides details on the algorithm.

3.3 Inference

The procedures described above for bounds conditional on $Y(0)$ or $Y(1)$ and for the overall bounds when obtained by integrating the conditional bounds lead to consistent and asymptotically normal estimates. This subsection gives the limiting distribution of the conditional bound estimators (11) and (12) and unconditional bound estimators (13) and (14). The limiting distributions provide the basis for asymptotically valid inference on the parameters of interest, by applying Imbens and Manski's (2004) method for inference on partially identified parameters.

The conditional bound estimates (11) and (12) are themselves functions of estimators for potential outcome cdfs, $\hat{F}_0$ and $\hat{F}_1$:

$$\begin{pmatrix} \hat{F}_0(y) \\ \hat{F}_1(y+t) \end{pmatrix} = \begin{pmatrix} \frac{n^{-1}\sum_{i=1}^{n} 1(Y_i \leq y) - n^{-1}\sum_{i=1}^{n} 1(Y_i \leq y) D_i}{1 - n^{-1}\sum_{i=1}^{n} D_i} \\ \frac{n^{-1}\sum_{i=1}^{n} 1(Y_i \leq y+t) D_i}{n^{-1}\sum_{i=1}^{n} D_i} \end{pmatrix},$$

which in turn are (differentiable) functions of the following vector of sample means:

$$\hat{W}(v) = \left(\begin{array}{c} n^{-1}\sum_{i=1}^{n} 1(Y_i \leq y) \\ n^{-1}\sum_{i=1}^{n} 1(Y_i \leq y) D_i \\ n^{-1}\sum_{i=1}^{n} 1(Y_i \leq y+t) D_i \\ n^{-1}\sum_{i=1}^{n} D_i \end{array} \right),$$

with corresponding vector of population expectations $W(v)$, where $v = (y,t)'$. Collect the arguments of the max and min in expressions (11) and (12) in the vector $\hat{H} := \left(\hat{H}^L, \hat{H}^U\right)'$, where

$$\begin{aligned} \hat{H}^L &:= \frac{\hat{F}_1(y+t) - \hat{F}_0(y)}{1 - \hat{F}_0(y)} \\ \hat{H}^U &:= \frac{\hat{F}_1(y+t)}{\hat{F}_0(y)}, \end{aligned}$$

with corresponding probability limits $H := \left(H^L, H^U\right)'$. The following theorem establishes the limiting distribution of $\hat{H}$:

Theorem 3 *Let $\{Y_i, D_i\}_{i=1}^{n}$ be an iid sample, let $p := \Pr(D_i = 1)$, and let $\bar{\mathcal{Y}} = \left[F_0^{-1}(\delta), F_0^{-1}(1-\delta)\right] \subset \mathcal{Y}$ for some $\delta > 0$. Assume there exists $\kappa > 0$ such that $\kappa \leq p \leq 1-\kappa$. Then $\sqrt{n}\left(\hat{H}(v) - H(v)\right)$ converges uniformly in $v \in \bar{\mathcal{Y}} \times \mathcal{D}$ to a Gaussian process with zero mean function and covariance function $\Omega(v,\tilde{v}) := J(v)\gamma(v)\Sigma(v,\tilde{v})\gamma(\tilde{v})' J(\tilde{v})'$ where the Jacobians $J(v)$ and $\gamma(v)$ are given by*

$$\begin{aligned} J(v) &:= \left[\begin{array}{cc} -\frac{1-F_1(y+t)}{(1-F_0(y))^2} & (1-F_0(y))^{-1} \\ -\frac{F_1(y+t)}{F_0(y)^2} & F_0(y)^{-1} \end{array} \right], \\ \gamma(v) &:= \left[\begin{array}{cccc} \frac{1}{1-p} & -\frac{1}{1-p} & 0 & \frac{F_0(y)}{(1-p)} \\ 0 & 0 & \frac{1}{p} & -\frac{F_1(y+t)}{p} \end{array} \right], \end{aligned}$$

and the covariance function $\Sigma(v,\tilde{v})$ *is given by*

$$\Sigma(v,\tilde{v}) := E\left[W_i(v)\,W_i(\tilde{v})' - W(v)\,W(\tilde{v})'\right].$$

Proof. See appendix. ■

The theorem shows that the arguments of the max and min functions in the conditional bound estimators (11) and (12) have limiting normal distribution. The conditional bound estimators themselves will not necessarily have a normal limit because of the max and min. Nevertheless, valid confidence intervals based on $\hat{H}$ can be the basis for valid confidence intervals for the treatment effect distribution itself by taking the intersection of intervals based on $\hat{H}$ with the unit interval. Specifically, inference on $F_{\Delta|0}(t|y)$ can be performed following Imbens and Manski (2004) and Stoye (2009), whose method applied to our setting yields the following confidence interval:

$$CI(1-\alpha) := \left[\hat{H}^L - \bar{C}_\alpha\sqrt{n^{-1}\hat{\Omega}(v,v)_{[1,1]}}, \hat{H}^U + \bar{C}_\alpha\sqrt{n^{-1}\hat{\Omega}(v,v)_{[2,2]}}\right] \cap [0,1],$$

where the critical value $\bar{C}_\alpha$ satisfies

$$\Phi\left(\bar{C}_\alpha + \sqrt{n}\frac{\hat{H}^U - \hat{H}^L}{\sqrt{\max\left(\hat{\Omega}(v,v)_{[1,1]}, \hat{\Omega}(v,v)_{[2,2]}\right)}}\right) - \Phi\left(-\bar{C}_\alpha\right) = \alpha.$$

Note that $\hat{H}^L \leq \hat{H}^U$ with probability one, which satisfies the superconsistency condition given in Stoye (2009) for the bounds to apply. Constructing the confidence intervals requires estimates of the covariance matrix, $\hat{\Omega}(v,v)$. This can be estimated by plugging in empirical counterparts to the analytical formula in the theorem, or, fixing a value for y and t, the variance can be estimated by the bootstrap.[1]

The unconditional bound estimators (13) and (14) likewise converge in distribution to a normal limit, as they are (Hadamard) differentiable functions of $\hat{H}$, which itself has a Gaussian limit as shown above. These results are formalized in the following theorem:

Theorem 4 *Let $\{Y_i, D_i\}_{i=1}^n$ be an iid sample with $Y_i = Y_i(0)(1-D_i) + Y_i(1) D_i$ and $Y_i(0)$ and $Y_i(1)$ continuously distributed and let $p := \Pr(D_i = 1)$ with $\kappa \leq p \leq 1-\kappa$ for some $\kappa > 0$. Define $F(t) := \left(F_\Delta^L(t), F_\Delta^U(t)\right)'$ and $\hat{F}(t) := \left(\hat{F}_\Delta^L(t), \hat{F}_\Delta^U(t)\right)'$ and assume $\Pr(F_1(Y_i + t) = F_0(Y_i)) = 0$. Then $\sqrt{n}\left(\hat{F}(t) - F(t)\right)$ converges uniformly in $t \in \mathcal{D}$ to $\phi'_{\theta_0}(\mathbb{G}_0)$ where ϕ'_{θ_0} is a linear and continuous map defined in the proof and $\mathbb{G}_0$ is a tight Gaussian element in $\ell^\infty\left(\bar{\mathcal{Y}} \times \mathcal{D}\right)^2 \times \ell^\infty(\mathcal{M}) \times \mathbb{R}$.*

Proof. See appendix. ■

The theorem means that given the stated conditions the unconditional bound estimators are asymptotically normally distributed. Further, the results in Fang and Santos (2018) imply that their asymptotic variance can be consistently estimated via the bootstrap, and thus the Stoye (2009) approach for inference on the unconditional treatment effect cdf can be applied as above for the conditional cdf. The condition that $\Pr(F_1(Y_i + t) = F_0(Y_i)) = 0$ is reasonable in settings with a continuously distributed outcome and non-vanishing treatment effect heterogeneity.

3.4 Testing for Stochastic Increasingness

Stochastic increasingness has testable implications. This section illustrates these implications and shows how they can be tested. Stochastic increasingness implies

[1]The conditions given in Theorem 3, for a fixed y and t, satisfy the conditions for the bootstrap delta method in van der Vaart (1998), Theorem 23.9.

that $Y(1)$ and $Y(0)$ are positively correlated (Lehmann, 1966). This implication cannot be tested directly, since we do not observe the joint distribution of potential outcomes, but we can test it indirectly by examining how $Y(1)$ and $Y(0)$ move with observed variables S. Specifically, the Cauchy-Schwarz inequality implies (see Theorem 2 in the appendix) that an implication of $Cov\left(Y(1), Y(0)\right) \geq 0$ and thus an implication of stochastic increasingness is that

$$Corr\left(\hat{Y}(0), \hat{Y}(1)\right) \geq -\sqrt{\frac{\left(1-R_0^2\right)\left(1-R_1^2\right)}{R_0^2 R_1^2}}, \tag{15}$$

where $\hat{Y}(0)$ and $\hat{Y}(1)$ are linear projections of potential outcomes on S with corresponding coefficients of determination R_0^2 and R_1^2. Condition (15) is only nontrivial when the covariates S strongly predict potential outcomes: the respective R^2s between S and each potential outcome must geometrically average at least 0.5 in order for the right-hand side of (16) to be greater than negative one. A practical procedure for verifying this condition when treatment is exogenous is to estimate the conditional expectation by regressing Y_i on S_i in the treated and untreated subsamples, calculate the correlation coefficient between the predicted values, and compare it to the right hand side of (15).

The test described heuristically above may provide little information for either of two reasons. The first is that unless covariates S sufficiently strongly predict outcomes, the inequality will trivially be satisfied, since the right-hand side will be less than negative one. The second is that predicted potential outcomes must be very negatively correlated in order for the null to be rejected, meaning that moderate violations of stochastic increasingness will not be detected, no matter how large the sample size.

Even in cases in which the formal test is underpowered, the test statistic can still provide useful evidence regarding the plausibility of the stochastic increasingness assumption. In particular if the correlation between $\hat{Y}(0)$ and $\hat{Y}(1)$ is positive, it indicates that observed factors move both treated and untreated outcomes in the same direction lending plausibility to the belief that unobserved factors act in the same manner. This is similar in spirit to how selection on observed variables can be used to inform a prior about selection on unobserved variables in Altonji et al. (2013). If the correlation between predicted outcomes is not only positive but also satisfies the more stringent condition

$$Corr\left(\hat{Y}(0), \hat{Y}(1)\right) \geq \sqrt{\frac{(1-R_0^2)(1-R_1^2)}{R_0^2 R_1^2}}, \tag{16}$$

then it must be the case that $Cov(Y(1), Y(0)) \geq 0$, a necessary condition for stochastic increasingness.

3.5 Bounding the distribution of complier effects of an endogenous treatment

To this point we have assumed for simplicity that treatment, D, is exogenous. In many applications, including our empirical example below, treatment is endogenous. In this subsection we show how the framework extends to endogenous treatments when an instrumental variable that satisfies the conditions of the local average treatment effects framework of Imbens and Angrist (1994).

Suppose treatment status is correlated with potential outcomes so that treatment-control comparisons do not reflect causal effects. Identification relies on a binary instrumental variable, Z; for example, Z may be an indicator for being selected in a lottery to be eligible to enroll in a charter school. Let $D(z)$ be the (potential) treatment status that would be realized if Z were set to $z \in \{0,1\}$. Likewise, let $Y(d,z)$ be the potential outcome that would be realized if D were $d \in \{0,1\}$ and Z were z. We adopt the familiar exclusion and monotonicity conditions from the local average treatment effects framework of Imbens and Angrist (1994). The exclusion restriction means the instrument has no direct causal effect on outcomes, that is, $Y(d,z) = Y(d,z')$, so that we can continue to index potential outcomes by treatment status alone: $Y(0), Y(1)$. The monotonicity condition means the instrument never induces individuals not to receive treatment: $D(1) \geq D(0)$ almost surely. Finally, we assume that the instrument, Z, is independent of $(Y(1), Y(0), D(1), D(0))$. These conditions allow one to identify the distributions of potential outcomes conditional on $D(1) > D(0)$ (Abadie, 2002). We denote the so-called "complier" distributions of potential outcomes $F_{0|C}(y) := \Pr(Y(0) \leq y | D(1) > D(0))$ and $F_{1|C}(y) := \Pr(Y(1) \leq y | D(1) > D(0))$.

The key condition for bounding the treatment effect distribution when treatment is endogenous is *local* mutual stochastic increasingness:

Definition 5 *Potential outcomes $Y(0)$ and $Y(1)$ are* ***locally mutually stochastically increasing*** *if*

$$\Pr(Y(1) \leq t | Y(0) = y, D(1) > D(0))$$

and

$$\Pr(Y(0) \leq t | Y(1) = y, D(1) > D(0))$$

are each nonincreasing in y almost everywhere.

Local mutual stochastic increasingness means that among the subpopulation of compliers potential outcomes are positively related. This condition implies that the distribution of treatment effects conditional on compliers can be bounded by the distributions of potential outcomes conditional on compliers, as the following theorem formalizes.

Theorem 6 *Suppose $Y(1)$ and $Y(0)$ are locally mutually stochastically increasing. Then $F_{\Delta|Y(0),C}(t|Y(0)) := \Pr(\Delta \leq t|Y(0), D(1) > D(0))$ is bounded from below pointwise in t and $Y(0)$ by*

$$F^L_{\Delta|Y(0),C}(t|Y(0)) := \begin{cases} 0 & , \quad Y(0) > F^{-1}_{0|C}\left(F_{1|C}(Y(0)+t)\right) \\ \frac{F_{1|C}(Y(0)+t)-F_{0|C}(Y(0))}{1-F_{0|C}(Y(0))} & , \quad Y(0) \leq F^{-1}_{0|C}\left(F_{1|C}(Y(0)+t)\right) \end{cases} \tag{17}$$

and from above by

$$F^U_{\Delta|Y(0),C}(t|Y(0)) := \begin{cases} \frac{F_{1|C}(Y(0)+t)}{F_{0|C}(Y(0))} & , \quad Y(0) \geq F^{-1}_{0|C}\left(F_{1|C}(Y(0)+t)\right) \\ 1 & , \quad Y(0) < F^{-1}_{0|C}\left(F_{1|C}(Y(0)+t)\right) \end{cases}. \tag{18}$$

Proof. See the appendix. ■

The theorem shows that when treatment is endogenous, the conditional distribution of treatment effects among compliers can be bounded by functions of the distributions of potential outcome among compliers, which Abadie (2002) shows are identified given the exclusion and monotonicity conditions described above.

As in the exogenous case, bounds on the distribution of compliers' treatment effects can be constructed by integrating the conditional bounds defined in Theorem 6 over the compliers' distribution of potential outcomes:

$$\tilde{F}^L_{\Delta|C}(t) = \max_{d\in\{0,1\}} E\left[F^L_{\Delta|Y(d),C}(t|Y(d))\,|D(1) > D(0)\right] \tag{19}$$

$$\tilde{F}^U_{\Delta|C}(t) = \min_{d\in\{0,1\}} E\left[F^U_{\Delta|Y(d),C}(t|Y(d))\,|D(1) > D(0)\right]. \tag{20}$$

These integrated bounds are expecations of a function of potential outcomes conditional on compliers, of the sort that can be computed using Abadie's (2003) κ weights.

Simarly to the exogenous case, bounds on the marginal distribution of compliers' treatment effects can be further tightened, albeit at significant computational cost, by directly searching over the set of bivariate copulae that satisfy stochastic increasingness:

$$F^{L}_{\Delta|C}(t) = \inf_{H(\cdot,\cdot)\in\mathcal{C}^{SI}} \int\int 1\left(F^{-1}_{1|C}(v) - F^{-1}_{0|C}(u) \leq t\right) H(u,v)\,dudv, \quad (21)$$

$$F^{U}_{\Delta|C}(t) = \sup_{H(\cdot,\cdot)\in\mathcal{C}^{SI}} \int\int 1\left(F^{-1}_{1|C}(v) - F^{-1}_{0|C}(u) \leq t\right) H(u,v)\,dudv. \quad (22)$$

Estimation of the bounds takes the form of (11) and (12) as in the exogenous case, except empirical cdfs $\hat{F}_0$ and $\hat{F}_1$ are replaced by Abadie's (2002) estimates of complier potential outcome cdfs, $\hat{F}_{0|C}$ and $\hat{F}_{1|C}$. Inference likewise proceeds as in the exogenous case. The asymptotic normality result in Theorem 3 holds when the underlying potential outcome cdf estimators are instrumental variables estimates of the complier cdfs proposed in Abadie (2002), but with $\gamma(v)\,\Sigma(v,\tilde{v})\,\gamma(\tilde{v})'$ in Theorem 3 replaced by $v_{0,1}(y, y+t)$ defined in Frandsen (2015), Theorem 4.

Finally, testing stochastic increasingness when treatment is endogenous closely follows the procedure described in Section 3.4, with one modification: the linear projections $\hat{Y}(0)$ and $\hat{Y}(1)$ of outcomes on covariates S, the associated R^2s, and the correlation between the projections that appear in condition (15) should be computed using Abadie's (2003) κ weights.

4 Empirical Example: Distributional Effects of KIPP Lynn

A substantial literature has found that charter schools have widely varying effects on student achievement (see Hanushek et al., 2007; Bettinger, 2005; Dobbie and Fryer, 2013). In many cases, students who attend charter school appear to perform no better than students attending traditional public schools. However, Dobbie and Fryer (2013) show that charter schools that focus on increased instructional time, tutoring, high expectations, effective use of data, and frequent teacher feedback are effective at increasing student achievement. Specific examples such as Harlem Children's Zone and the Knowledge is Power Program (KIPP) have been shown to close or dramatically narrow the achievement gaps between white and minority students (see Dobbie and Fryer, 2011; Angrist et al., 2010, 2012). While these studies suggest that effective charter schools may boost disadvantaged students' academic achievement on average, understanding the distribution of effects is also important. In particular, parents may be more comfortable enrolling their students in charter schools if a large majority of students benefit from attendance than if only a minority of students do. Additionally, by understanding how the effects of achievement vary across the distribution of control outcomes, parents and policy makers may have a better sense of the types of children who would most benefit from charter school attendance.

For these reasons, we estimate our bounds in the context of KIPP, which is an organization that manages a set of "No Excuses" charter schools. Relative to many other traditional and charter schools, KIPP schools employ a longer school day and school year. They seek to maintain high behavioral standards and focus instruction on math and reading skills. This setting is also well-suited for evaluation since the KIPP Lynn charter school is over-subscribed and admission is rationed through a lottery.

As a consequence, admission can be used as a valid instrument for KIPP attendance. We focus on the effects of attendance rather than the effects of admission, since by assumption the latter only affects outcomes through attendance.

In the KIPP context, mutual stochastic increasingness implies that if student A performed better than student B in a traditional school, then student A's counterfactual outcome in the KIPP Lynn charter school would be drawn from a distribution that weakly stochastically dominates student B's counterfactual distribution. Unlike rank invariance, this does not rule out student B leapfrogging student A in KIPP Lynn, but it does mean student B has a lower probability of outperforming student A in KIPP Lynn than would a student who performed better than student A in a traditional school. Similarly, performing better in the charter school implies a stochastically dominant counterfactual outcome distribution in traditional schools.

This assumption seems plausible in the KIPP context. In particular, unobserved drivers of performance including latent ability, effort, and parental support are likely to be helpful for academic achievement both in charter as well as traditional schools. Consistent with our assumption, we show that observed characteristics predict performance in similar ways in both types of schools. Indeed, it is difficult to think of factors that would be helpful for academic achievement in one of these settings but not the other.

Angrist et al. (2010, 2012) provide an evaluation of this program utilizing data from students who applied to the KIPP Academy in Lynn, Massachusetts from 2005 to 2008. Student outcomes are observed prior to application in 4th grade and then in subsequent grades up to 8th grade. Taking advantage of the fact that admission to this location was rationed through a lottery, the authors find that each year of attendance leads to an average increase in math achievement of 0.35 standard deviations.

We utilize the data from these earlier studies. In contrast to prior work, our treatment is a binary variable for whether the student attended KIPP academy. Our outcome variable is math performance in the sixth grade. Hence our estimate captures the cumulative effect of KIPP attendance for up to two years of attendance. For this reason, our estimated effect sizes will tend to be somewhat larger than those estimated by prior researchers. We view this effect of treatment on the treated as the policy relevant parameter for parents considering enrolling their children in KIPP and for policy makers ascertaining the efficacy of the program.

In Table 1 we present summary statistics for our sample, which consists of 176 students who entered the lottery for 5th grade KIPP admission in 2006-2007.[2] Similar to prior studies, we find that approximately 65 percent of students are admitted into KIPP and 55 percent of all applicants eventually enrolled. This implies that 85 percent of admitted students attended for at least one year. Admitted students attended 1.63 years out of a maximum of 2 years. One student who was not admitted and did not initially enroll eventually attended KIPP. Given the near zero probability of attending among those in our sample not admitted via lottery, the local average treatment effect of attending coincides with the average effect of treatment on the treated. Examining student performance prior to application, we see that in fourth grade the students performed 0.39 standard deviations below the state-level mean in mathematics. Sixth grade performance suggests the program was efficacious given that applicants performed just above the state mean in mathematics. This is confirmed when observing the substantial difference in performance between admitted and non-admitted students. Examining demographics, we see that the sample is disproportionately male and Hispanic. Roughly 20 percent of students are categorized as special education and the same fraction are limited English proficient. Over 80

[2]We chose this sample because Figure 1 in Angrist et al. (2012) suggests average effects are largest, and thus bounds are most likely to be informative in the 2006-2007 cohorts. The trade off of course is the smaller sample size, which decreases precision.

percent of applicants qualify for free or reduced price lunch. In the same table, we show characteristics of students who won the lottery for admission and those who did not. All of the observed characteristics prior to application appear balanced across admitted and non-admitted students suggesting randomization was successful.

Prior to estimating our bounds, it is helpful to estimate the average effect of attendance on math achievement. To do so, we simply employ two-stage least squares in which the dependent variable is math performance and our binary attendance measure is instrumented by an indicator variable that takes of a value of 1 if the student was admitted to KIPP. We control for covariates including indicator variables for gender, ethnicity, special education status, limited English proficiency, and free or reduced price lunch receipt. Table 2 shows the results. We see that the first stage has very high power with an F-statistic of the instrument in excess of 500. The estimated effect of enrollment on math achievement for applicants who choose to enroll is 0.67 standard deviations, an effect that is both large and highly statistically significant. While our specification differs from those in Angrist et al. (2010, 2012), the results are broadly consistent with those that they found.[3]

We employ the test of stochastic increasingness we develop to see if observed characteristics are associated with outcomes in the treated and control state in a manner consistent with this assumption. To perform this test, we perform Abadie-κ-weighted regressions of outcomes on a spline in prior math score and the demographic characteristics described above. In Appendix E we provide additional estimation details and describes the Stata software implementation of our procedure. Table 3 shows the results. The correlation between predicted outcomes in the treated and control states

[3]The quantile treatment effects (not reported) range between .1 standard deviations and 1.15 standard deviations, implying that charter school attendance first-order stochastically shifted the distribution of math scores. The rank invariance assumption in this setting would therefore imply that no students suffered a reduction in math achievement due to attending a charter school. Rank invariance is a much stronger assumption than the stochastic increasingess assumption we consider.

Table 1: KIPP Lynn Summary Statistics

	Full Sample	Non-Admitted	Admitted	P-Value Equal Means
Offered	0.66	0.00	1.00	
	(0.47)	(0.00)	(0.00)	
Enrolled	0.56	0.00	0.85	
	(0.50)	(0.00)	(0.36)	
Years Attended	1.10	0.03	1.63	
	(0.98)	(0.26)	(0.75)	
6th Grade Math	.04	-0.33	0.23	
	(1.01)	(1.07)	(0.93)	
4th Grade Math	-0.35	-0.35	-0.36	0.960
	(1.02)	(1.10)	(.99)	
Female	0.45	0.46	0.45	0.954
	(0.50)	(0.50)	(0.50)	
Black	0.19	0.19	0.20	0.873
	(0.40)	(0.39)	(0.40)	
Hispanic	0.59	0.63	0.56	0.426
	(0.49)	(0.49)	(0.50)	
Asian	0.03	0.02	0.03	0.519
	(0.17)	(0.13)	(0.18)	
Other	0.01	0.00	0.02	0.315
	(0.11)	(0.00)	(0.13)	
Special Education	0.20	0.19	0.21	0.771
	(0.40)	(0.39)	(0.41)	
Limited English Proficiency	0.19	0.24	0.16	0.232
	(0.39)	(0.43)	(0.37)	
Free or Reduced Price Lunch	0.83	0.83	0.83	0.981
	(0.38)	(0.38)	(0.38)	
Observations	176	59	117	

Notes: The table shows summary statistics for the entire sample as well as for admitted and non-admitted students. Standard deviations are in parentheses. The right column contains p-values of an F-test of equal means between the admitted and non-admitted students.

Table 2: Estimated Effects of KIPP Lynn Attendance on Math Score

Effect of Attendance	0.67**
	(0.17)
First Stage Partial F-Statistic	572.2
Observations	176

Notes: The first row shows the estimated impact of attendance at KIPP on mathematics in the 6th grade. This estimate comes from a two-stage least squares regression in which the excluded instrument is an offer of admission into KIPP through the lottery. Robust standard errors are in parentheses. Controls include indicator variables for female, ethnicity categories, special education status, limited English proficiency, and free or reduced price lunch receipt. The second row shows the first stage partial F-statistic of the instrument. ** Indicates statistical significance at the 5 percent level.

is 0.947. This satisfies the necessary condition for positive correlation of potential outcomes (and thus for stochastic increasingness) and also satisfies the sufficient condition for positive correlation, shown in the bottom row of Table 3 labeled "threshold correlation." The positive correlation between predicted potential outcomes lends plausibility to the assumption of stochastic increasingness. Naturally, when using covariates to tighten the bounds, we must make the assumption of conditional stochastic increasingness, which is untestable without other predictive variables in addition to those in the conditioning set.

We now examine bounds on the cdf of treatment effects. Figure 2 shows the Williamson-Downs bounds, our integrated pointwise bounds, and the uniform bounds, all calculated without incorporating covariates. Note that the integrated pointwise bounds are much tighter than the Williamson-Downs bounds. The uniform bounds tend to be even somewhat narrower. This figure suggests that the assumption of stochastic increasingness is very valuable for narrowing the bounds on the distribution of treatment effects. Table 4 provides a numerical comparison between these three

Table 3: Test of Positive Correlation of Potential Outcomes

R^2 Treated Outcomes	0.539
R^2 Control Outcomes	0.628
Correlation between Predicted Treated and Control Outcomes	0.947
Threshold Correlation	0.71

Notes: The R^2 values for the treated and control outcomes come from kappa-weighted regressions of math outcomes on a cubic spline in 4th grade math achievement with three knots and variables for gender, ethnicity, special education status, limited English proficiency, and free or reduced price lunch receipt.

methods showing the lower and upper bound on the fraction of students experiencing negative treatment effects. Examining the second column, we see that the upper bound on the fraction hurt using the Williamson-Downs bounds is 0.68, while our integrated pointwise and uniform bounds are 0.47 and 0.40 respectively.

The Williamson-Downs bounds allow for top students in the control state systematically to be the bottom students in the treated state. Our bounds are tighter than the Williamson-Downs bounds because we rule out this systematic switching of positions. Our integrated pointwise bounds assume the worst (or best) counterfactual distribution consistent with stochastic increasingness for each student in the control group. However, the worst case bounds for all students in general cannot all obtain simultaneously, and so these bounds are conservative. Our uniform bounds find a worst (or best) case that is internally consistent across all individuals in both the control and treated distribution, leading to further tightening of the bounds. This comes at the cost of a much higher computational burden, however.

Our bounds, while substantially tighter than the Williamson-Downs bounds, can be tightened further by incorporating covariates. Figure 3 shows the integrated pointwise bounds on the overall distribution of treatment effects with and without covari-

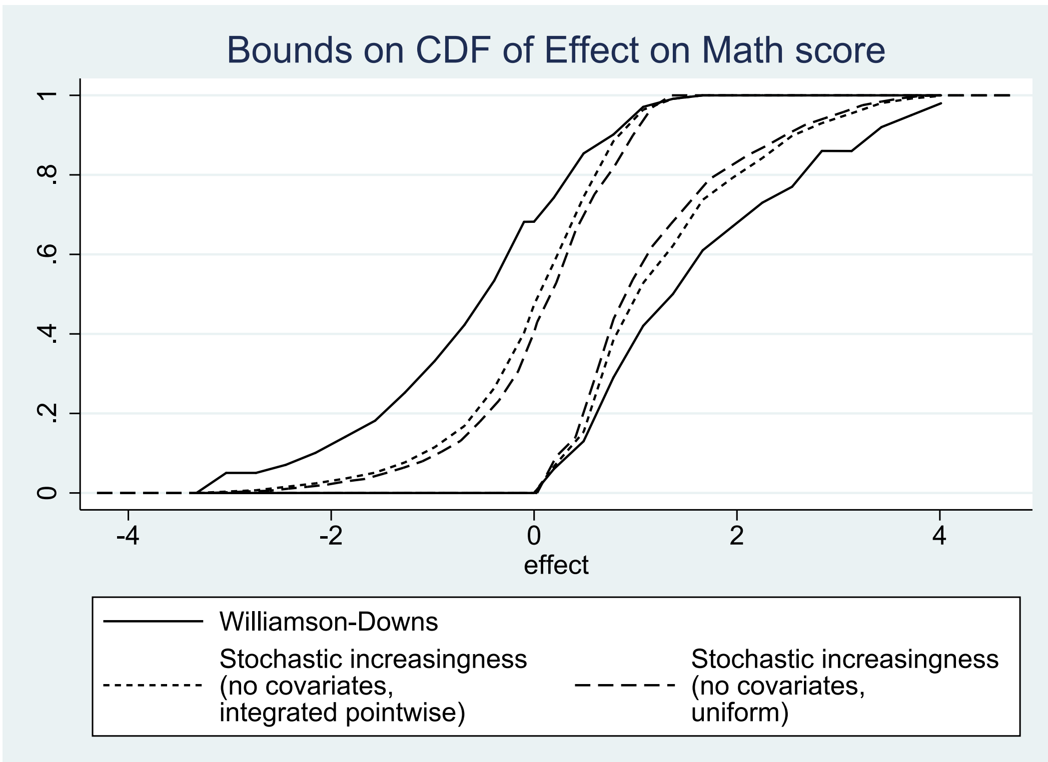

Figure 2: Estimated bounds on the cdf of effects on 7th grade math score. The solid curves show the Williamson-Downs bounds. The short-dash curves impose stochastic increasingness by integrating over the conditional (pointwise) bounds. The long-dash curves impose stochastic increasingness uniformly by searching over the space of copulae that satisfy stochastic increasingness. No set of bounds uses covariates.

Table 4: Bounds on the Fraction with a Negative Effect of Charter School Enrollment on Math Score

	No covariates		Covariates	
Bounding method:	lower	upper	lower	upper
Williamson-Downs	0.00	0.68	0.05	0.49
	(0.00	, 0.81)	(0.01	, 0.57)
Stochastic Increasingness (integrated pointwise)	0.00	0.47	0.07	0.33
	(0.00	, 0.63)	(0.02	, 0.41)
Stochastic Increasingness (uniform)	0.00	0.40	-	-
	-	-	-	-

Notes: Estimated lower and upper bounds for the fraction of students whose 7th grade math score was hurt by enrollment in KIPP Lynn, among students whose enrollment was determined by the lottery outcome. 95-percent confidence intervals for the fraction hurt are reported in parentheses below the estimated bounds, calculated via Imbens and Manski's (2004) method. The covariates used in computing the bounds in the two right-hand columns include fourth-grade math score and indicators for female, black, hispanic, other race or ethnicity, special education, limited English proficiency, free or reduced-price lunch status, and the interaction of female and minority status. Confidence intervals and covariate-adjusted bounds corresponding to Stochastic Increasingness (uniform) were not estimated due to computational infeasibility. N = 176.

ates. Our covariates include fourth grade math score and indicators for female, black, Hispanic, other race or ethnicity, special education, limited English proficiency, free or reduced-price lunch status, and the interaction bertween female and minority status. We show the integrated pointwise bounds since they are nearly as tight as the uniform bounds and calculating uniform bounds when using continuous covariates takes orders of magnitude more computational time. The figure shows that covariates dramatically tighten the bounds. Referring back to Table 4 we see that using covariates tighten our bounds on the fraction hurt from 0.47 to 0.33.

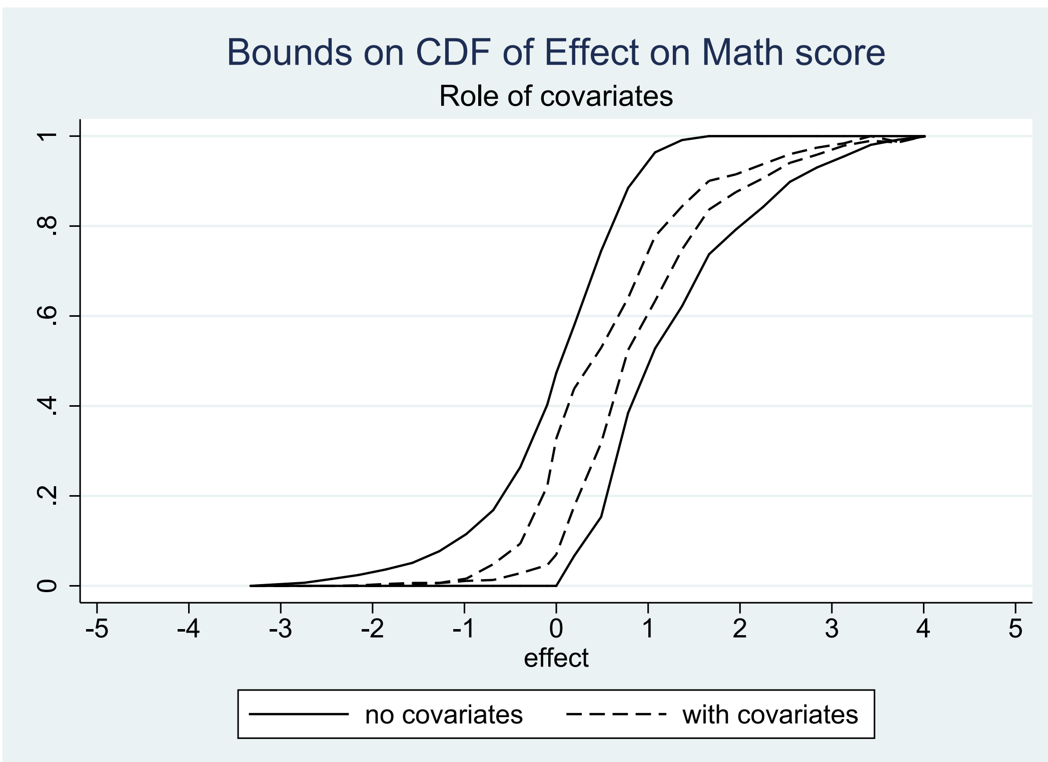

Figure 3: Estimated bounds (13) and (14) on the cdf of effects on 7th grade math score. The solid bounds include no covariates. The dashed bounds use 4th grade math score and demographic characteristics described in the text.

In Figure 4 we show the (pointwise) 95 percent confidence intervals corresponding to our integrated pointwise bounds incorporating covariates. Even looking at the top of the confidence interval on the upper bound of fraction of students hurt by treatment, we can still infer that even in the worst case only a minorty of students could have been harmed by treatment. Indeed, Table 4 shows that the top of the confidence interval on the upper bound of students who could have been harmed by treatment is 0.41.

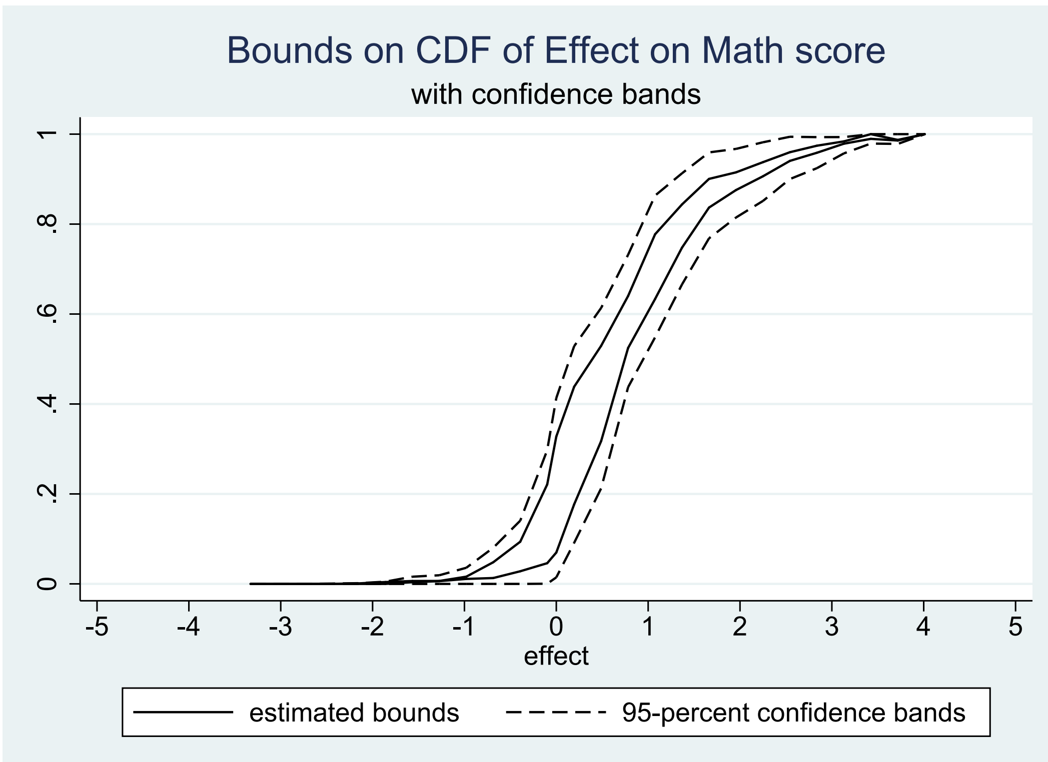

Figure 4: Estimated bounds on the cdf of effects on 7th grade math score. The solid curves show the bounds imposing stochastic increasingness by integrating over the conditional (pointwise) bounds incorporating covariates. The dashed curves show 95-percent confidence bands for the treatment effect cdf calculated via Imbens and Manski's (2004) method.

We now employ the methodology we developed to bound the fraction hurt by outcome in the control state. Figure 5 shows this relationship incorporating covariates. We see that the upper bound on the fraction of students hurt is increasing in the outcome in the control distribution. The lower bound is uniformly zero. The figure makes clear that the probability of being hurt by KIPP attendance was lower than 0.5 for students performing below the state average level in the control distribution, even taking into account the 95-percent confidence interval. Table 5 shows these results in table form. Once covariates are used, our bounds suggest that even students expected to perform in the 75th percentile of the control distribution likely benefitted from attending KIPP. Collectively, these results suggest that students who would have performed poorly in regular public schools overwhelmingly benefitted from treatment.

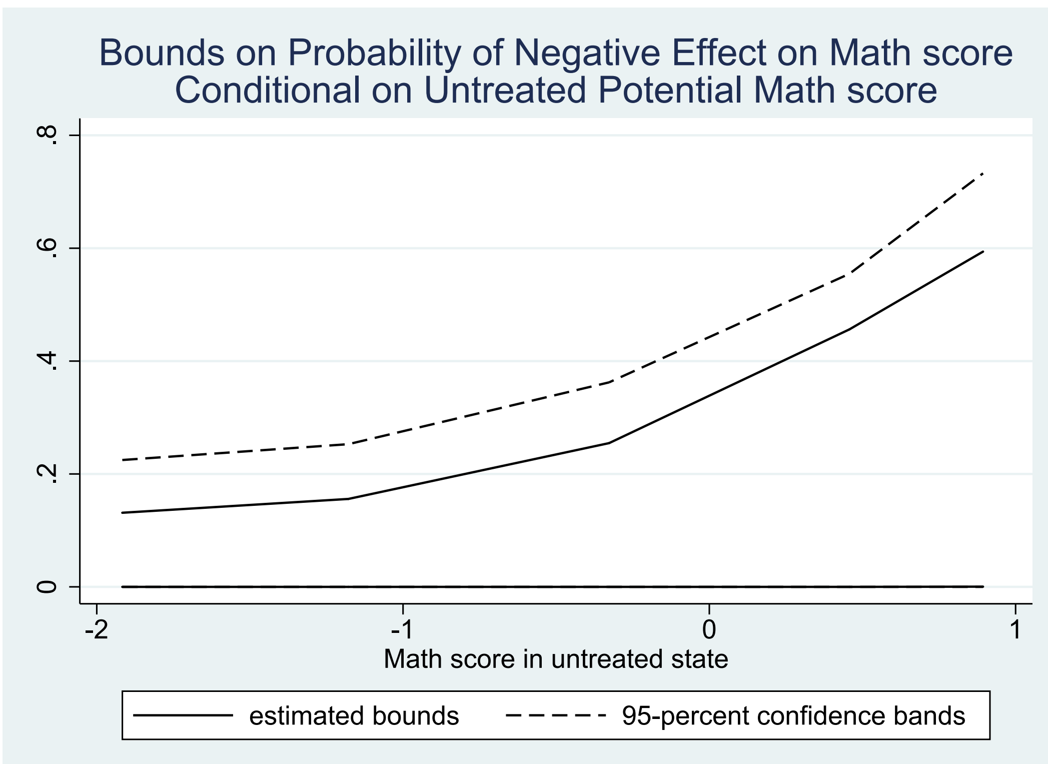

Figure 5: Estimated bounds on the probability of a negative effect of KIPP attendance on 7th grade math scores conditional on 7th grade math score in the untreated state. The solid curves show the estimated bounds using 4th grade math score and demographic characteristics described in the text. The dashed curves show 95-percent confidence bands.

In Figure 6 we trace out bounds on the average treatment effect as a function of the outcome in the control state. The lower bound suggests that all except students who

Table 5: Effects of Charter School Enrollment Conditional on Math Score in Untreated State

Untreated math score		No covariates		Covariates	
Percentile in control group	St.dev. relative to MA average	lower	upper	lower	upper
A. Bounds on fraction with negative treatment effect					
10	-1.92	0.00	0.16	0.00	0.13
		(0.00 ,	0.39)	(0.00 ,	0.22)
25	-1.18	0.00	0.25	0.00	0.16
		(0.00 ,	0.43)	(0.00 ,	0.25)
50	-0.33	0.00	0.40	0.00	0.25
		(0.00 ,	0.58)	(0.00 ,	0.36)
75	0.46	0.00	0.62	0.00	0.46
		(0.00 ,	0.74)	(0.00 ,	0.55)
90	0.89	0.00	0.76	0.00	0.59
		(0.00 ,	0.86)	(0.00 ,	0.73)
B. Bounds on conditional average treatment effect					
10	-1.92	0.65	2.65	1.03	2.54
		(0.15 ,	2.82)	(0.73 ,	2.82)
25	-1.18	0.46	2.01	0.77	1.98
		(0.05 ,	2.14)	(0.55 ,	2.12)
50	-0.33	0.16	1.43	0.48	1.43
		-(0.18 ,	1.60)	(0.28 ,	1.55)
75	0.46	-0.24	0.89	0.18	0.86
		-(0.46 ,	0.97)	(0.07 ,	0.93)
90	0.89	-0.49	0.58	0.03	0.52
		-(0.66 ,	0.69)	-(0.10 ,	0.60)

Notes: Estimated lower and upper bounds for the conditional fraction of students whose 7th grade math score was hurt by enrollment in KIPP Lynn (Panel A) and the conditional average effect of KIPP Lynn attendance (Panel B), among students whose enrollment was determined by the lottery outcome, by level of the untreated potential math score. 95-percent confidence intervals for the fraction hurt are reported in parentheses below the estimated bounds, calculated via Imbens and Manski's (2004) method. The covariates used in computing the bounds in the two right-hand columns include fourth-grade math score and indicators for female, black, hispanic, other race or ethnicity, special education, limited English proficiency, free or reduced-price lunch status, and the interaction of female and minority status. N = 176.

would have performed extremely well in the control state enjoyed a positive expected treatment effect from KIPP attendance. This provides additional evidence that students who would have performed poorly in their regular schools likely benefitted from KIPP attendance. This is true even looking at the bottom of the confidence interval of the lower bound of average treatment effects. We show these results numerically in Table 5.

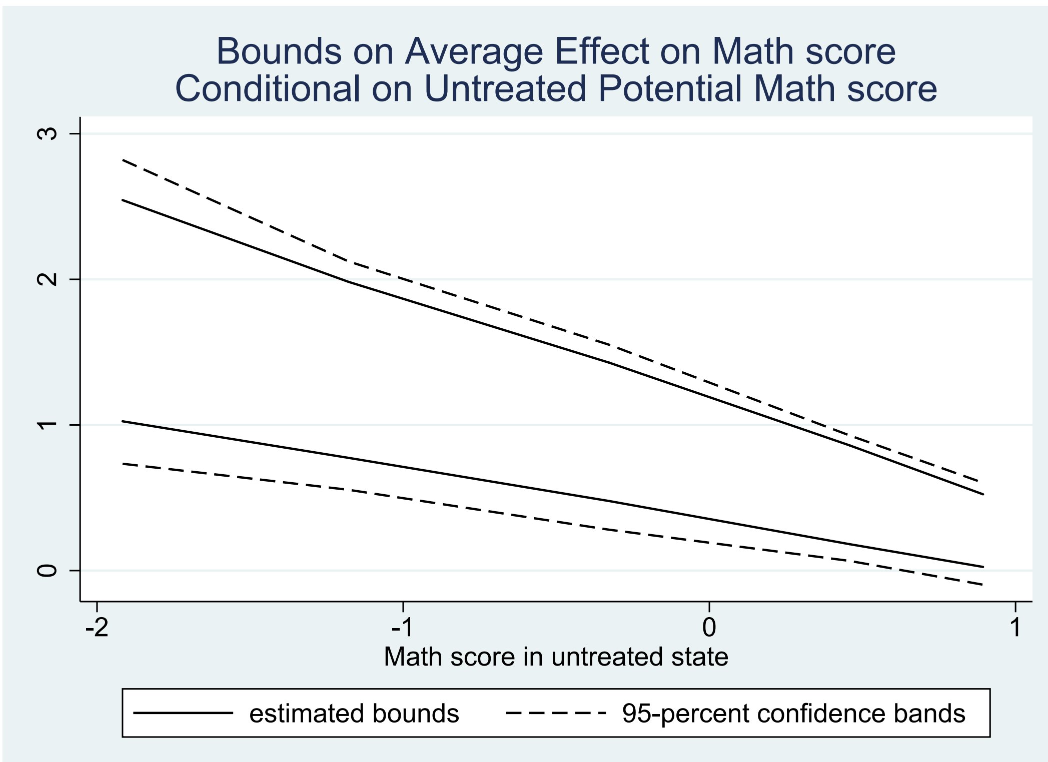

Figure 6: Estimated bounds on the average effect of KIPP attendance on 7th grade math scores conditional on 7th grade math score in the untreated state. The solid curves show the estimated bounds using 4th grade math score and demographic characteristics described in the text. The dashed curves show 95-percent confidence bands.

Summarizing the findings from the KIPP charter school experiment, we confirm that attendance increased math achievement substantially. Even in the worst case, we find that treatment increased math achievement for the substantial majority of students. Furthermore, worst case bounds suggest that nearly all students who would have performed poorly in the control state benefited from treatment. We also show that the average treatment effect was very large and positive for such students and still unambiguously positive for all except the students with the very best control outcomes. These results suggest that our bounds can be informative regarding the distribution of treatment effects—particularly when we have covariates that are strongly predictive of student outcomes and large treatment effects.

5 Conclusion

In this paper we propose partially identifying conditions that imply bounds on the distribution of treatment effects, an object of considerable policy and economic interest, but which is not identified under standard assumptions. The proposed condition—that an individual's potential outcomes are each weakly stochastically increasing in the other—should be plausible in many empirical settings, and has testable implications. The bounds can be constructed from standard estimates of the conditional distributions of potential outcomes.

Specifically, our results give bounds on quantities such as the fraction of individuals harmed by treatment, the median treatment effect, and the average treatment effect conditional on the untreated potential outcome. The bounds implied by our stochastic increasingness condition are substantially tighter than the Williamson-Downs bounds based only on the restrictions implied by the marginal distributions of potential outcomes. Our bounds are further tightened with the use of covariates.

We calculate our bounds in the context of a KIPP charter school. We show that not only was the impact of attendance on mathematics positive overall, but also that we can rule out that more than a small fraction of attending students were harmed. The beneficial effects were particularly strong for students with poor outcomes in the control state.

The bounding methodology we develop in this paper represents an important tool for applied researchers. To maximize the usefulness of these bounds, we offer a suggestion. Our bounds are much tighter in the presence of covariates that strongly predict outcomes in the treatment and control state. Such covariates are also useful for assessing the plausibility of the stochastic increasingness assumption. Hence, we encourage researchers to take full advantage of existing pre-treatment covariates (or other variables not affected by treatment) and when possible collect additional covariates, even if such variables are not required to consistently estimate an average treatment effect.

References

Alberto Abadie. Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association*, 97(457):284–292, mar 2002.

Alberto Abadie. Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics*, 113:231–263, 2003.

Joseph G. Altonji, Timothy Conley, Todd E. Elder, and Christopher R. Taber. Methods for using selection on observed variales to address selection on unobserved variables. Unpublished working paper, April 2013.

Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. Inputs and impacts in charter schools: KIPP Lynn. *American Economic Review*, 100(2):239–243, 2010.

Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane, Parag A. Pathak, and Christopher R. Walters. Who benefits from KIPP? *Journal of Policy Analysis and Management*, 31(4):837–860, 2012.

Eric P Bettinger. The effect of charter schools on charter students and public schools. *Economics of Education Review*, 24(2):133–147, 2005.

Jay Bhattacharya, Azeem M. Shaikh, and Edward Vytlacil. Treatment effect bounds: An application to swan-ganz catheterization. *Journal of Econometrics*, 168(2):223 – 243, 2012.

Marianne P. Bitler, Jonah B. Gelbach, and Hilary W. Hoynes. Can Variation in Subgroups' Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment. *The Review of Economics and Statistics*, 99 (4):683–697, July 2017.

German Blanco, Carlos A. Flores, and Alfonso Flores-Lagunes. Bounds on average and quantile treatment effects of Job Corps training on wages. *Journal of Human Resources*, 48(3):659–701, 2013.

Richard Blundell, Amanda Gosling, Hidehiko Ichimura, and Costas Meghir. Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica*, 75(2):323–363, 2007.

Victor Chernozhukov and Christian Hansen. An IV model of quantile treatment effects. *Econometrica*, 73(1):245–261, January 2005.

Raj Chetty, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. The fading American dream: trends in absolute income mobility since 1940. NBER Working Paper 22910, 2016.

Richard K. Crump, V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik. Nonparametric tests for treatment effect heterogeneity. *Review of Economics and Statistics*, 90:389–405, 2008.

Habiba Djebbari and Jeffrey Smith. Heterogeneous impacts in PROGRESA. *Journal of Econometrics*, 145(1-2), 2008.

Will Dobbie and Roland G. Fryer, Jr. Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone. *American Economic Journal: Applied Economics*, 3(3):158–187, 2011.

Will Dobbie and Roland G. Fryer, Jr. Getting beneath the veil of effective schools: Evidence from New York City. *American Economic Journal: Applied Economics*, 5(4):28–60, 2013.

Yanqin Fan and Sang Soo Park. Partial identification of the distribution of treatment effects and its confidence sets. In *Nonparametric Econometric Methods*, pages 3–70. 2009.

Yanqin Fan and Sang Soo Park. Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, 26:931–951, 2010.

Yanqin Fan and Jisong Wu. Partial identification of the distribution of treatment effects in switching regime models and its confidence sets. *The Review of Economic Studies*, 77(3):1002–1041, 2010.

Yanqin Fan, Robert Sherman, and Matthew Shum. Identifying treatment effects under data combination. *Econometrica*, 82(2):811–822, 2014.

Zheng Fang and Andres Santos. Inference on Directionally Differentiable Functions. *The Review of Economic Studies*, 86(1):377–412, 09 2018.

Sergio Pinheiro Firpo and Geert Ridder. Bounds on functionals of the distribution treatment effects. Textos para discussão 201, Escola de Economia de São Paulo, Getulio Vargas Foundation (Brazil), June 2010.

Brigham R. Frandsen. Treatment effects with censoring and endogeneity. *Journal of the American Statistical Association*, 110(512):1745–1752, 2015.

Eric A. Hanushek, John F. Kain, Steven G. Rivkin, and Gregory F. Branch. Charter school quality and parental decision making with school choice. *Journal of Public Economics*, 91(5–6):823 – 848, 2007.

James J Heckman and Jeffrey A Smith. Evaluating the welfare state. Working Paper 6542, National Bureau of Economic Research, May 1998.

James J. Heckman, Jeffrey Smith, and Nancy Clements. Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts. *The Review of Economic Studies*, 64(4):487–535, October 1997.

Kosuke Imai. Sharp bounds on the causal effects in randomized experiments with "truncation-by-death". *Statistics & Probability Letters*, 78(2):144 – 149, 2008.

Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994.

Guido W. Imbens and Charles F. Manski. Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857, 2004.

Ju Hyun Kim. Identifying the distribution of treatment effects under support restrictions. *Econometrics: Single Equation Models eJournal*, 2014.

Patrick Kline and Melissa Tartari. Bounding the labor supply responses to a randomized welfare experiment: A revealed preference approach. *American Economic Review*, 106(4):972–1014, April 2016.

Michael Lechner and Blaise Melly. Partial identification of wage effects of training programs. Working Paper, Brown University, Department of Economics, No. 2010-8, 2010.

Soohyung Lee and Azeem M. Shaikh. Multiple testing and heterogeneous treatment effects: Re-evaluating the effects of progresa on school enrollment. *Journal of Applied Econometrics*, 29:612–626, 2013.

E. L. Lehmann. Some concepts of dependence. *The Annals of Mathematical Statistics*, 37(5):pp. 1137–1153, 1966.

John A. List, Azeem M. Shaikh, and Yang Xu. Multiple hypothesis testing in experimental economics. *Experimental Economics*, Jan 2019.

Charles F. Manski. Monotone treatment response. *Econometrica*, 65(6):1311–1334, 1997.

Thomas M. Russell. Sharp bounds on functionals of the joint distribution in the analysis of treatment effects. *Journal of Business & Economic Statistics*, pages 1–15, 2019.

Jörg Stoye. More on confidence intervals for partially identified parameters. *Econometrica*, 77(4):1299–1315, 2009.

A.W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, New York, 1998.

R. C. Williamson and T. Downs. Probabilistic arithmetic. i. numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning*, 4(2):89–158, March 1990.