

NBER WORKING PAPER SERIES

PARTIAL IDENTIFICATION OF THE DISTRIBUTION OF TREATMENT EFFECTS WITH
AN APPLICATION TO THE KNOWLEDGE IS POWER PROGRAM (KIPP)

Brigham R. Frandsen
Lars J. Lefgren

Working Paper 24616
<http://www.nber.org/papers/w24616>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2018

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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Brigham R. Frandsen and Lars J. Lefgren. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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

Brigham R. Frandsen and Lars J. Lefgren

NBER Working Paper No. 24616

May 2018

JEL No. C01,I21

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 academic 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.

Brigham R. Frandsen
Department of Economics
Brigham Young University
Provo, UT 84602
frandsen@byu.edu

Lars J. Lefgren
Department of Economics
Brigham Young University
130 Faculty Office Building
Provo, UT 84602
and NBER
lars_lefgren@byu.edu

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

Brigham R. Frandsen* Lars J. Lefgren*

April 19, 2018

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 academic 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 and 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, a point made by [Heckman et al. \(1997\)](#). 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 distribution.

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 Frechet-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 distribution, student A would likely have outperformed student B in the treated distribution, and vice versa. The assumption should be plausible in many economic settings, and indeed is even implied by some models of program self-selection, as shown by [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 [Djebbari and Smith, 2008](#) and [Bitler et al., 2014](#) 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, as our approach can. Similarly, indi-

viduals 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 observables* 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 treatment 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 increasingness.

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 distribution 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, derives the implied bounds on the distribution of treatment effects, shows how they are identified in the data, and shows how they may be tested. 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 \(1990\)](#) and [Heckman et al. \(1997\)](#) 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 Frechet-Hoeffding bounds. Unfortunately, these bounds, which place no additional restrictions on the joint distribution of outcomes, tend to be uninformative. 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. These restrictions allow the bounds to be substantially tightened, but may be too strong to be plausible in many empirical settings. [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 between 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 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.

3 Econometric Framework

Consider a binary treatment indicator, D , that possibly affects a continuously distributed outcome 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 $\Delta := Y(1) - Y(0)$, 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 Frechet-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 Frechet-Hoeffding 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 Frechet-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. They also provide no information on the distribution of treatment effects conditional on $Y(d)$. 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 [Chernozhukov and Hansen \(2005\)](#). The condition 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. The condition also 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, pretreatment 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.

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. In such cases as these, 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 mutually stochastically increasing. Then $F_{\Delta|Y(0)}(t|Y(0)) := \Pr(\Delta \leq t|Y(0))$ is bounded from below by*

$$F_{\Delta|Y(0)}^L(t|Y(0)) := \begin{cases} 0 & , Y(0) + t < \tilde{Y}(1) \\ \frac{F_1(Y(0)+t) - F_0(Y(0))}{1 - F_0(Y(0))} & , Y(0) + t \geq \tilde{Y}(1) \end{cases} \quad (3)$$

and from above by

$$F_{\Delta|Y(0)}^U(t|Y(0)) := \begin{cases} \frac{F_1(Y(0)+t)}{F_0(Y(0))} & , Y(0) + t \leq \tilde{Y}(1) \\ 1 & , Y(0) + t \geq \tilde{Y}(1) \end{cases}, \quad (4)$$

where $\tilde{Y}(1) := F_1^{-1}(F_0(Y(0)))$.

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 stochas-

tic increasingness also implies analogous bounds on the conditional distribution of treatment effects given $Y(1)$, denoted $F_{\Delta|Y(1)}^L$ and $F_{\Delta|Y(1)}^U$, expressions for which are provided in the proof.

Figure 1 provides graphical intuition for the bounds. The expression for the upper bound of the cdf in equation (4) corresponds to the worst-case conditional distribution for $Y(1)$ given $Y(0)$ that is still consistent with stochastic increasingness. As the left panel of Figure 1 shows, 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, but instead draws from the truncated distribution of $Y(1)$ ranks below his or her rank in the control state. The 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).

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:

$$\Delta^L(Y(d)) = \int t dF_{\Delta|Y(d)}^U(t|Y(d)), \quad (5)$$

$$\Delta^U(Y(d)) = \int t dF_{\Delta|Y(d)}^L(t|Y(d)). \quad (6)$$

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:

$$\tilde{F}_{\Delta}^L(t) = \max_{d \in \{0,1\}} E [F_{\Delta|Y(d)}^L(t|Y(d))] \quad (7)$$

$$\tilde{F}_{\Delta}^U(t) = \min_{d \in \{0,1\}} E [F_{\Delta|Y(d)}^U(t|Y(d))]. \quad (8)$$

Integrating over the conditional bounds in this way will yield conservative bounds on the overall distribution, however, since the conditional bounds $F_{\Delta|Y(d)}^L$ and $F_{\Delta|Y(d)}^U$ 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 principal 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 overall distribution of treatment effects as

$$F_{\Delta}^L(t) = \inf_{H(\cdot, \cdot) \in \mathcal{C}^{SI}} \int \int 1(F_1^{-1}(v) - F_0^{-1}(u) \leq t) H(u, v) dudv, \quad (9)$$

$$F_{\Delta}^U(t) = \sup_{H(\cdot, \cdot) \in \mathcal{C}^{SI}} \int \int 1(F_1^{-1}(v) - F_0^{-1}(u) \leq t) H(u, v) dudv. \quad (10)$$

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 11 and 12 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

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

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

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 methods, 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. 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 bound estimators (11) and (12). 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 bound estimates 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:

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

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 $\hat{H} :=$

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

Theorem 3 *Let $\{Y_i, D_i\}_{i=1}^n$ be an iid sample. Then $\sqrt{n}(\hat{H}(v) - H(v))$ converges uniformly 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*

$$\begin{aligned}J(v) & : = \begin{bmatrix} -\frac{1-F_1(y+t)}{(1-F_0(y))^2} & (1-F_0(y))^{-1} \\ -\frac{F_1(y+t)}{F_0(y)^2} & F_0(y)^{-1} \end{bmatrix}, \\ \gamma(v) & : = \begin{bmatrix} \frac{1}{1-p} & -\frac{1}{1-p} & 0 & \frac{F_0(y)}{(1-p)} \\ 0 & 0 & \frac{1}{p} & -\frac{F_1(y+t)}{p} \end{bmatrix},\end{aligned}$$

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

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

Proof. The result follows from an application of standard Donsker and functional delta method results in, for example, Theorem 3.9.5 in [van der Vaart and Wellner \(1996\)](#). ■

Inference on $F_{\Delta|0}(t|y)$ can then be performed following [Imbens and Manski \(2004\)](#),

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}_α satisfies

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

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 that $Y(1)$ and $Y(0)$ are positively correlated. 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 5 in the appendix) that a necessary condition for $Cov(Y(1), Y(0)) \geq 0$ is that

$$Corr(\hat{Y}(0), \hat{Y}(1)) \geq -\sqrt{\frac{(1 - R_0^2)(1 - R_1^2)}{R_0^2 R_1^2}}, \quad (13)$$

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 (13) is only nontrivial when the covariates S strongly predict potential outcomes: the respective R^2 s between S and each potential outcome must geometrically average at least .5 in order for the right-hand side of (14) 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 (13). When treatment is endogenous, but responds monotonically to an exogenous instrument Z_i , the projections of potential outcomes on covariates and the calculation of the correlation should be performed using Abadie’s (2003) κ weights. In this case the procedure tests stochastic increasingness among compliers.

The test described above may have little power 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.

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 observable factors move both treated and untreated outcomes in the same direction lending plausibility to the belief that unobservable factors act in the same manner. This is similar in spirit to how selection on unobservables can be assessed by examining selection on observables (Altonji et al., 2013). If the correlation between predicted outcomes is not only positive but also satisfies the more stringent condition

$$Corr(\hat{Y}(0), \hat{Y}(1)) \geq \sqrt{\frac{(1 - R_0^2)(1 - R_1^2)}{R_0^2 R_1^2}}, \quad (14)$$

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

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.

[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 7th grade. Hence our estimate captures the cumulative effect of KIPP attendance for up to three years of attendance. For this reason, our estimated effect sizes will tend to be somewhat larger than those estimated by prior researchers.

In Table 1 we present summary statistics for our sample. 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. Examining the 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. Seventh 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 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 observable 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.

In order to construct our bounds, we make the assumption of stochastic increasingness. We employ the test we develop to see if observable 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. Table 3 shows the results. The correlation between predicted outcomes in treated and control states 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 lend 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 methods showing the lower and upper bound on the fraction of students experiencing negative treatment effects. Examining the second column, we see that the 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.

These 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 covariates. 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 between 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.

In Figure 4 we show the 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 minority 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.

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 .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.

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 would have performed extremely well in the control distribution 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.

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 benefitted 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 distribution.

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 few suggestions. First, 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 and when possible collect additional covariates, even if such variables are not required to consistently estimate an average treatment effect.

References

- 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 variables 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.

- 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. Working Paper 20142, National Bureau of Economic Research, May 2014.
- 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.
- Victor Chernozhukov, Iván Fernández-Val, and Blaise Melly. Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268, 2013.
- 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.
- Habiba Djebbari and Jeffrey Smith. Heterogeneous impacts in PROGRESA. 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*, 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, 6 2010.
- Yanqin Fan, Robert Sherman, and Matthew Shum. Identifying treatment effects under data combination. *Econometrica*, 82(2):811–822, 2014.
- Sergio Pinheiro Firpo and Geert Ridder. Bounds on functionals of the distribution treatment effects. Textos para discussão 201, Escola de Economia de So Paulo, Getulio Vargas Foundation (Brazil), June 2010.
- 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(56):823 – 848, 2007.
- 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 Charles F. Manski. Confidence intervals for partially identified parameters. *Econometrica*, 72(6):1845–1857, 2004.

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

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

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

A.W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer-Verlag, New York, 1996.

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

Appendix A: Proofs

The following result is crucial to Theorem 2:

Lemma 4 *Let X and Y be random variables with marginal distributions F_X and F_Y , where \mathcal{Y} is the support of Y . Suppose continuously distributed random variable X is weakly stochastically increasing in Y . Then*

$$\underline{F}_{X|Y}(x|y) \leq \Pr(X \leq x|Y = y) \leq \bar{F}_{X|Y}(x|y),$$

where

$$\underline{F}_{X|Y}(x|y) = \begin{cases} 0 & , x < F_X^{-1}(F_Y(y)) \\ \frac{F_X(x) - F_Y(y)}{1 - F_Y(y)} & , x \geq F_X^{-1}(F_Y(y)) \end{cases}$$

and

$$\bar{F}_{X|Y}(x|y) = \begin{cases} \frac{F_X(x)}{F_Y(y)} & , \quad x \leq F_X^{-1}(F_Y(y)) \\ 1 & , \quad x \geq F_X^{-1}(F_Y(y)) \end{cases}.$$

Proof. Take the lower bound first. Assume $x \geq F_X^{-1}(F_Y(y))$ since the bound is trivially satisfied otherwise. The lower bound for $\Pr(X \leq x|Y = y)$, $\underline{F}_{X|Y}(x|y)$, solves the following minimization problem for a given x, y :

$$\min_{F(x|\cdot)} F(x|y)$$

subject to

$$\begin{aligned} (1) & : F(x|y) \leq F(x|y'), \quad y' < y, \\ (2) & : F(x|y) \geq F(x|y''), \quad y'' \geq y, \end{aligned}$$

(since X is stochastically increasing in Y) and

$$(3) : \int_y F(x|s) dF_Y(s) = F_X(x)$$

(since the conditional must integrate to the marginal). The second constraint will clearly bind at the lower bound, which implies $\underline{F}_{X|Y}(x|y'') = K(x)$ for $y'' \geq y$, where $K(x)$ is some function that does not depend on y'' . The first constraint is maximally relaxed by setting $\underline{F}_{X|Y}(x|y') = 1$ for $y' < y$. The third constraint then implies the

result:

$$\begin{aligned}
F_X(x) &= \int_{\mathcal{Y}} \underline{F}_{X|Y}(x|s) dF_Y(s) \\
&= F_Y(y) + \int_{[y,\infty) \cap \mathcal{Y}} K(x) dF_Y(s) \\
&= F_Y(y) + K(x) \int_{[y,\infty) \cap \mathcal{Y}} dF_Y(s) \\
&= F_Y(y) + K(x) (1 - F_Y(y)) \\
\Leftrightarrow K(x) &= \frac{F_X(x) - F_Y(y)}{1 - F_Y(y)}.
\end{aligned}$$

Now take the upper bound. Assume $x \leq F_X^{-1}(F_Y(y))$ since the bound is trivially satisfied otherwise. The upper bound for $\Pr(X \leq x|Y = y)$, $\bar{F}_{X|Y}(x|y)$, solves the following maximization problem:

$$\max_{F(x|\cdot)} F(x|y)$$

subject to

$$\begin{aligned}
(1) &: F(x|y) \leq F(x|y'), \quad y' \leq y, \\
(2) &: F(x|y) \geq F(x|y''), \quad y'' > y,
\end{aligned}$$

(since X is stochastically increasing in Y) and

$$(3) : \int_{\mathcal{Y}} F(x|s) dF_Y(s) = F_X(x)$$

(since the conditional must integrate to the marginal). The first constraint will clearly bind at the upper bound, which implies $\bar{F}_{X|Y}(x|y') = G(x)$ for $y' \leq y$, where $G(x)$ is some function that does not depend on y' . The second constraint is maximally

relaxed by setting $\bar{F}_{X|Y}(x|y') = 0$ for $y'' > y$. The third constraint then implies the result:

$$\begin{aligned}
F_X(x) &= \int_{\mathcal{Y}} \bar{F}_{X|Y}(x|s) dF_Y(s) \\
&= \int_{(-\infty, y] \cap \mathcal{Y}} G(x) dF_Y(s) \\
&= G(x) \int_{(-\infty, y] \cap \mathcal{Y}} dF_Y(s) \\
&= G(x) F_Y(y) \\
&\Leftrightarrow G(x) = \frac{F_X(x)}{F_Y(y)}.
\end{aligned}$$

■

Proof of Theorem 2. Note that by definition

$$\Pr(\Delta \leq t | Y(0), X) = \Pr(Y(1) \leq Y(0) + t | Y(0), X).$$

Since $Y(1)$ is conditional stochastically increasing in $Y(0)$, Lemma 4 applies to this case conditionally on X , taking $x = Y(0) + t$; $y = Y(0)$; $F_X = F_{Y(1)|X}$; $F_Y = F_{Y(0)|X}$. Making these substitutions in the lemma's result gives the result in the theorem. The argument for the lower bound is similar. Bounds on the conditional distribution of Δ given $Y(1)$ and X can be obtained by imposing that $Y(0)$ is stochastically increasing in $Y(1)$ in an analogous manner. Note that by definition

$$\begin{aligned}
\Pr(\Delta \leq t | Y(1), X) &= \Pr(Y(0) \geq Y(1) - t | Y(1), X) \\
&= 1 - \Pr(Y(0) \leq Y(1) - t | Y(1), X)
\end{aligned}$$

since $Y(0)$ is assumed to be continuously distributed. We can thus simply interchange $Y(1)$ and $Y(0)$ in (3) and take the complement to obtain an upper bound:

$$F_{\Delta|Y(1),X}^U(t|Y(1),X) := \begin{cases} 1 & , Y(1) + t < \tilde{Y}(0|X) \\ \frac{1-F_{0|X}(Y(1)+t|X)}{1-F_{1|X}(Y(1)|X)} & , Y(1) + t \geq \tilde{Y}(0|X) \end{cases}, \quad (15)$$

and similarly with (4) to obtain a lower bound:

$$F_{\Delta|Y(1),X}^L(t|Y(1),X) := \begin{cases} \frac{F_{1|X}(Y(1)|X)-F_{0|X}(Y(1)+t|X)}{F_{1|X}(Y(1)|X)} & , Y(1) + t \leq \tilde{Y}(0|X) \\ 0 & , Y(1) + t \geq \tilde{Y}(0|X) \end{cases}, \quad (16)$$

where $\tilde{Y}(0|X) := F_{0|X}^{-1}(F_{1|X}(Y(1)|X)|X)$. ■

Theorem 5 *Let $\hat{Y}(0)$ and $\hat{Y}(1)$ be linear projections of potential outcomes on X with corresponding coefficients of determination R_0^2 and R_1^2 . Then $Cov(Y(1), Y(0)) \geq 0$ implies*

$$Corr(\hat{Y}(1), \hat{Y}(0)) \geq -\sqrt{(1-R_0^2)(1-R_1^2)/(R_0^2 R_1^2)}.$$

Proof. Define $\varepsilon(1) = Y(1) - \hat{Y}(1)$ and $\varepsilon(0) = Y(0) - \hat{Y}(0)$. Note that by construction $Cov(\hat{Y}(1), \varepsilon(0)) = Cov(\hat{Y}(0), \varepsilon(1)) = 0$. Also, note that $Var(\varepsilon(1)) = (1-R_1^2)Var(Y(1))$ and $Var(\varepsilon(0)) = (1-R_0^2)Var(Y(0))$. Since $\varepsilon(0)$ is orthogonal to $\hat{Y}(1)$ and $\varepsilon(1)$ is orthogonal to $\hat{Y}(0)$, the covariance between potential outcomes can be written:

$$Cov(Y(0), Y(1)) = Cov(\hat{Y}(0), \hat{Y}(1)) + Cov(\varepsilon(0), \varepsilon(1)). \quad (17)$$

The Cauchy-Schwarz inequality implies

$$Cov(\varepsilon(0), \varepsilon(1)) \leq \sqrt{(1-R_0^2)Var(Y(0))(1-R_1^2)Var(Y(1))}.$$

Inserting this into (17) yields an upper bound on the covariance between potential outcomes:

$$\begin{aligned} \text{Cov}(Y(0), Y(1)) &\leq \text{Cov}(\hat{Y}(0), \hat{Y}(1)) \\ &\quad + \sqrt{(1 - R_0^2) \text{Var}(Y(0)) (1 - R_1^2) \text{Var}(Y(1))}. \end{aligned}$$

This upper bound is nonnegative when

$$\text{Cov}(\hat{Y}(0), \hat{Y}(1)) \geq -\sqrt{(1 - R_0^2) \text{Var}(Y(0)) (1 - R_1^2) \text{Var}(Y(1))},$$

or, equivalently,

$$\text{Corr}(\hat{Y}(0), \hat{Y}(1)) \geq -\sqrt{\frac{(1 - R_0^2)(1 - R_1^2)}{R_0^2 R_1^2}}.$$

■

Appendix B: Incorporating Covariates and Instrumental Variables

For expositional simplicity the bounds above were developed without additional covariates and assuming exogenous treatment assignment. In practice, the bounds may be substantially tightened by incorporating additional covariates X , and, in the case of endogenous treatments, identification may require the use of instrumental variables. In this section we show how the framework can be extended to these cases.

When additional covariates are available that predict the outcome Y , the bounds may be tightened by adopting the following conditional version of the stochastic

increasingness assumption:

Definition 6 $Y(0)$ and $Y(1)$ are *mutually stochastically increasing conditional on X* if $\Pr(Y(1) \leq t | Y(0) = y, X)$ and $\Pr(Y(0) \leq t | Y(1) = y, X)$ are each nonincreasing in y almost everywhere.

The following bounds on the treatment effect cdf conditional on $Y(0)$ and X are conditional versions of (3) and (4) and follow from the conditional mutual stochastic increasingness condition above:

$$F_{\Delta|Y(0),X}^L(t|Y(0), X) : = \begin{cases} 0 & , Y(0) + t < \tilde{Y}(1|X) \\ \frac{F_{1|X}(Y(0)+t|X) - F_{0|X}(Y(0)|X)}{1 - F_{0|X}(Y(0)|X)} & , Y(0) + t \geq \tilde{Y}(1|X) \end{cases}$$

$$F_{\Delta|Y(0),X}^U(t|Y(0), X) : = \begin{cases} \frac{F_{1|X}(Y(0)+t|X)}{F_{0|X}(Y(0)|X)} & , Y(0) + t \leq \tilde{Y}(1|X) \\ 1 & , Y(0) + t \geq \tilde{Y}(1|X) \end{cases} ,$$

where $\tilde{Y}(1|X) := F_{1|X}^{-1}(F_{0|X}(Y(0)|X))$; expressions for the treatment effect cdf conditional on $Y(1)$ and X are similar.

Bounds on the distribution of treatment effects conditional on $Y(0)$ (or $Y(1)$) alone—which are frequently of greater interest than the distribution conditional on $Y(0)$ and X —can be obtained by integrating the conditional bounds over the conditional distribution of X given $Y(0)$:

$$F_{\Delta|Y(d)}^L(t|Y(d)) = E[F_{\Delta|Y(d),X}^L(t|Y(d), X) | Y(d)] \quad (18)$$

$$F_{\Delta|Y(d)}^U(t|Y(d)) = E[F_{\Delta|Y(d),X}^U(t|Y(d), X) | Y(d)] , \quad (19)$$

for $d = 0$ or 1 . As before, bounds on the average treatment effect conditional on $Y(d)$

can be formed by integrating over the cdf bounds:

$$\begin{aligned}\Delta^L(Y(d)) &= \int t dF_{\Delta|Y(d)}^U(t|Y(d)), \\ \Delta^U(Y(d)) &= \int t dF_{\Delta|Y(d)}^L(t|Y(d)).\end{aligned}$$

A simple method for computing bounds on the marginal cdf of the treatment effect is to average over the bounds on the conditional cdf as before, though these bounds may not be sharp:

$$\tilde{F}_{\Delta}^L(t) = \max_{d \in \{0,1\}} E[F_{\Delta|Y(d),X}^L(t|Y(d),X)] \quad (20)$$

$$\tilde{F}_{\Delta}^U(t) = \min_{d \in \{0,1\}} E[F_{\Delta|Y(d),X}^U(t|Y(d),X)]. \quad (21)$$

Sharp bounds may be computed by directly searching over the space of bivariate copulae that satisfy mutual stochastic increasingness condition to find bounds conditional on X

$$\begin{aligned}F_{\Delta|X}^L(t|X) &= \inf_{H(\cdot,\cdot) \in \mathcal{C}^{SI}} \int \int 1(F_{1|X}^{-1}(v|X) - F_{0|X}^{-1}(u|X) \leq t) H(u,v) dudv, \\ F_{\Delta|X}^U(t|X) &= \sup_{H(\cdot,\cdot) \in \mathcal{C}^{SI}} \int \int 1(F_{1|X}^{-1}(v|X) - F_{0|X}^{-1}(u|X) \leq t) H(u,v) dudv,\end{aligned}$$

which can then be averaged to produce bounds on the marginal cdf:

$$F_{\Delta}^L(t) = E[F_{\Delta|X}^L(t|X)], \quad (22)$$

$$F_{\Delta}^U(t) = E[F_{\Delta|X}^U(t|X)]. \quad (23)$$

The bounds (22) and (23) are sharp, but in practice are prohibitively costly to calculate as they require an infinite-dimensional optimization at each covariate value X . In the application we estimate the integrated conditional bounds (20) and (21), which are slightly less tight but computationally feasible.

The conditional cdf bounds (18) and (19) can be consistently estimated by plugging in consistent estimators for the conditional cdfs $F_{1|X}$ and $F_{0|X}$. For the case where D_i is exogenous, the bounds can be constructed via the following steps for each untreated observation j :

1. Nonparametrically regress an indicator $1(Y_i \leq Y_j)$ on X_i in the untreated subsample and construct predicted value $\hat{F}_{0|X}(Y_j|X_j)$
2. Nonparametrically regress an indicator $1(Y_i \leq Y_j(0) + t)$ on X_i in the treated subsample and construct predicted value $\hat{F}_{1|X}(Y_j(0) + t|X_j)$
3. Form estimates of the bounds

$$\hat{F}_{\Delta|0,X}^L(t|Y_j(0), X_j) \quad : \quad = \max \left\{ 0, \frac{\hat{F}_{1|X}(Y_j(0) + t|X_j) - \hat{F}_{0|X}(Y_j(0)|X_j)}{1 - \hat{F}_{0|X}(Y_j(0)|X_j)} \right\} \quad (24)$$

$$\hat{F}_{\Delta|0,X}^U(t|Y_j(0), X_j) \quad : \quad = \min \left\{ 1, \frac{\hat{F}_{1|X}(Y_j(0) + t|X_j)}{\hat{F}_{0|X}(Y_j(0)|X_j)} \right\}. \quad (25)$$

The bounds (18) and (19) on the conditional distribution of treatment effects given $Y(0)$ can be constructed by nonparametrically regressing the estimates (24) and (25). Bounds (5) and (6) on the conditional expectation of treatment effects given $Y(0)$ can be computed by numerically integrating the estimates for (18) and (19) on a discrete grid. Analogous steps can be followed for the distribution of treatment effects given $Y(1)$.

When treatment status is exogenous, standard nonparametric regression methods

such as local polynomial regression or spline regression suffice in steps 1 and 2 and in constructing the bounds (18) and (19). When treatment is endogenous, instrumental variables methods will be required. The particular instrumental variables method to be used depends on which assumptions are appropriate in the empirical setting, and the interpretation of the bounds may depend on those assumptions. For example, in settings where individuals' treatment status can be assumed to respond monotonically to the instrument Z_i , the nonparametric regressions above can be estimated via Abadie's (2003) semiparametric κ -weighted estimator. The resulting estimates (11) and (12) would then identify bounds on the distribution of treatment effects among compliers, those individuals whose treatment status is affected by the instrument.

The bound estimates are themselves functions of estimators for potential outcome conditional cdfs, $\hat{F}_{0|X}$ and $\hat{F}_{1|X}$. Several methods exist for estimating conditional cdfs; which is most suitable will depend on the specific empirical setting. For example, when treatment is exogenous and X has continuous components, the semiparametric distribution regression approach of Chernozhukov et al. (2013) may be most appropriate. When X is discrete, standard least squares regressions where treatment is fully interacted with X may be used.

Appendix C: Algorithm for approximating overall bounds

Let $\{C[i, j]\}_{i=1..k, j=1..k}$ be the elements of a $k \times k$ matrix which discretely approximates a bivariate copula function. By definition, each marginal distribution is uniform,

which implies the following constraints:

$$\left\{ \sum_{s=1}^k C[s, j] = 1 \right\}_{j=1}^k, \\ \left\{ \sum_{s=1}^k C[i, s] = 1 \right\}_{i=1}^{k-1}.$$

Stochastic increasingness means each conditional cdf is decreasing in the conditioning dimension, which implies the following set of constraints:

$$\left\{ \left\{ \sum_{s=1}^i C[s, j] \geq \sum_{s=1}^i C[s, j+1] \right\}_{j=1}^{k-1} \right\}_{i=1}^{k-1},$$

and

$$\left\{ \left\{ \sum_{s=1}^j C[i, s] \geq \sum_{s=1}^j C[i+1, s] \right\}_{i=1}^{k-1} \right\}_{j=1}^{k-1}.$$

Let the set of discrete copulae satisfying the above $k^2 + (k-1)^2$ constraints be denoted \mathcal{C}_k . Given estimates of the separate conditional distributions of $Y(0)$ and $Y(1)$ given X (obtained possibly via the methods described in Section 3.2) the lower bound on $F_{\Delta|X}$ can be approximated by solving the following linear program:

$$\min_{\{C[i,j]\}} \sum_{j=1}^k \sum_{i=1}^k 1 \left(F_{1|X}^{-1}(r(i)) - F_{0|X}^{-1}(r(j)) \leq t \right) C[i, j] \\ \text{subject to } C[\cdot, \cdot] \in \mathcal{C}_k,$$

where

$$r(i) = \frac{i}{n} - \frac{1}{2n}.$$

The upper bound can be approximated by replacing the min with a max. Since the objective function and all constraints are linear, the program can be solved using efficient dual-simplex linear programming routines. Unconditional bounds can then be obtained by integrating the conditional bounds over X . In practice the algorithm works well for $k \approx 100$ and discrete X with a moderate number of cells.

Appendix D: Simulations

This section illustrates the bounds on the distribution of treatment effects derived above using numerical simulations. The simulations adopt the following data generating process. Untreated potential outcomes are generated as $Y_i(0) = \beta X_i + \varepsilon_i$. The treated potential outcome is $Y_i(1) = Y_i(0) + \delta$. The treatment indicator D_i is assigned independently of X_i and ε_i by random lottery whereby half the sample receives $D_i = 1$ and half receive $D_i = 0$. The unobservables are generated according to

$$\begin{pmatrix} X_i \\ \varepsilon_i \end{pmatrix} \sim N \left(0, \begin{bmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix} \right).$$

In the simulated model, the R^2 between $Y_i(0)$ and X_i is $R^2 = \beta^2 \sigma_X^2 / (\beta^2 \sigma_X^2 + \sigma_\varepsilon^2)$. The simulations set $\sigma_X^2 = 1$. The simulations vary σ_ε^2 from .01 to 1, corresponding to an R^2 between $Y_i(0)$ and X from .99 to zero, and β is set accordingly to $\sqrt{R^2 / \sigma_X^2}$ to ensure the variance of $Y_i(0)$ remains equal to one. The simulations also vary the treatment effect size δ from -1 to 1 .

The first set of simulations illustrates how the bounds on the average treatment effect conditional on $Y_i(0)$ vary by across the values of $Y_i(0)$. These simulations set the R^2 between $Y_i(0)$ and X_i to 0.7, corresponding to $\sigma_\varepsilon^2 = 0.3$ and $\beta = \sqrt{.7} \approx 0.84$ and set the treatment effect size to $\delta = 1$. Figure 7 plots the bounds (5) and (6) as

a function of $Y(0)$. The bounds always include the true treatment effect $\delta = 1$, and are tightest in the middle of the $Y(0)$ distribution, and widen in the tails. Notice that although in the simulated model the treatment effect is positive across the entire distribution of $Y(0)$, the bounds reach into negative territory for very high values of $Y(0)$, since the stochastic increasingness assumption allows for mean reversion; individuals with high values of $Y(0)$ have a larger probability of drawing a value of $Y(1)$ lower than $Y(0)$.

The second set of simulations shows how these bounds on the average treatment effect conditional on $Y_i(0)$ depend on the informativeness of the covariate X . These simulations set the treatment effect $\delta = 1$ and vary the R^2 between $Y_i(0)$ and X_i from zero to .99. Figure 8 plots the bounds (5) and (6) at $Y(0) = 0$ (i.e., at the median) as a function of the R^2 . They show that the bounds tighten dramatically as the covariate X more strongly predicts outcomes.

The next set of simulations illustrates how bounds on the fraction of individuals harmed by treatment (i.e., the treatment effect cdf evaluated at zero) conditional on $Y(0)$ depends on the size of the treatment effect δ . As above, these simulations set the R^2 between $Y_i(0)$ and X_i to 0.7. Figure 9 plots the bounds (18) and (19) evaluated at zero as a function of δ for $Y(0) = 0$. Since the simulated model has constant treatment effects, the true fraction is one on the left side of the graph (where the treatment effect is negative) and zero on the right side. When the treatment effect is sufficiently large in magnitude, the bounds are quite tight. When the treatment effect is zero or slightly positive, the bounds are completely uninformative, spanning zero and one.

The next set of simulations shows how the bounds on the fraction of individuals hurt conditional on $Y_i(0)$ depend on the informativeness of the covariate X . These simulations set the treatment effect δ equal to one, and vary the R^2 between $Y_i(0)$

and X_i from zero to .99. Figure 10 plots the bounds (18) and (19) evaluated at zero as a function of R^2 for $Y(0) = 0$. Since the (constant) treatment effect in this simulation is positive, the true fraction is zero. On the far left, where the covariate has no predictive power, the bounds are quite wide, the upper bound reaching .3, but the bounds tighten dramatically as R^2 increases.

The next set of simulations shows how the bounds on the overall fraction of individuals hurt by treatment vary with the treatment effect size δ . Again, R^2 is set to 0.7 for these simulations. Figure 11 plots the Williamson-Downs bounds (which make no restrictions), our stochastic increasingness bounds calculated by integrating the conditional bounds, and the stochastic increasingness bounds calculated by searching over the space of copula functions, each evaluated at zero, for a range of treatment effects sizes δ . All bounds are tightest when the treatment effect is large in magnitude. The stochastic increasingness bounds are much tighter than the classical Williamson-Downs bounds.

The final set of simulations shows how the bounds on the overall fraction of individuals hurt by treatment vary with the predictive power of the covariate S . Again, the treatment effect δ is set to one, and R^2 varies from zero to .99. Figure 12 plots the Williamson-Downs bounds (which make no restrictions), our stochastic increasingness bounds calculated by integrating the conditional bounds, and the stochastic increasingness bounds calculated by searching over the space of copula functions, each evaluated at zero, for a range of values of R^2 . Since the treatment effect is positive, the true fraction is zero. On the left side of the plot, where the covariate has little explanatory power, the bounds we propose are quite wide, spanning zero to .35. However, even these are much tighter than the bounds that impose no restrictions, which span zero to over .6. As the R^2 between $Y(0)$ and X increases, the bounds tighten substantially.

Table 1: KIPP Lynn Summary Statistics

	Full Sample	Non-Admitted	Admitted	P-Value Equal Means
Offered	0.65 (0.48)	0.00 (0.00)	1.00 (0.00)	
Enrolled	0.55 (0.50)	0.00 (0.00)	0.85 (0.36)	
7th Grade Math	0.03 (1.02)	-0.36 (1.09)	0.23 (0.93)	
4th Grade Math	-0.39 (1.06)	-0.43 (1.16)	-0.37 (1.01)	0.70
Female	0.46 (0.50)	0.46 (0.50)	0.45 (0.50)	0.73
Black	0.20 (0.40)	0.17 (0.38)	0.21 (0.41)	0.46
Hispanic	0.57 (0.50)	0.61 (0.49)	0.56 (0.50)	0.60
Asian	0.02 (0.16)	0.01 (0.12)	0.03 (0.17)	0.71
Other	0.01 (0.10)	0.00 (0.00)	0.02 (0.12)	0.30
Special Education	0.19 (0.40)	0.18 (0.39)	0.20 (0.40)	0.63
Limited English Proficiency	0.19 (0.39)	0.24 (0.43)	0.16 (0.37)	0.16
Free or Reduced Price Lunch	0.83 (0.38)	0.85 (0.36)	0.82 (0.38)	0.83
Observations	202	71	131	

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 Enrollment	0.67 (0.17)
First Stage Partial F-Statistic	572.2
Observations	176

Notes: The first row shows the estimated impact of enrollment in KIPP on mathematics in the 7th 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 gender, ethnicity, 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.

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 R² for the treated outcomes comes from a kappa-weighted regression 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.

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

Bounding method:	No covariates		Covariates	
	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.

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.

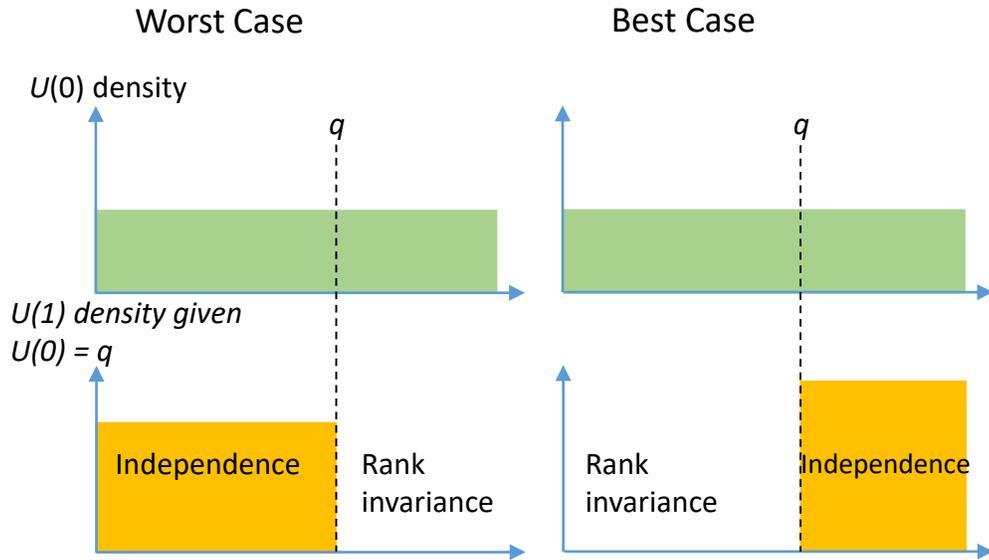


Figure 1: The figure illustrates intuitively the bounds implied by stochastic increasingness of potential outcomes on the conditional distribution of treated potential outcomes given the untreated potential outcome. The graphs in the top row plot the density of untreated ranks, $U(0) := F_0(Y(0))$, which is by definition uniform. The graphs in the bottom row plot the worst- and best-case conditional density of the treated rank $U(1) := F_1(Y(1))$ given the untreated rank is equal to q . The worst-case plot on the left corresponds to the upper-bound cdf of treatment effects in equation (4). The best-case plot on the right corresponds to the lower-bound cdf of treatment effects in equation (3).

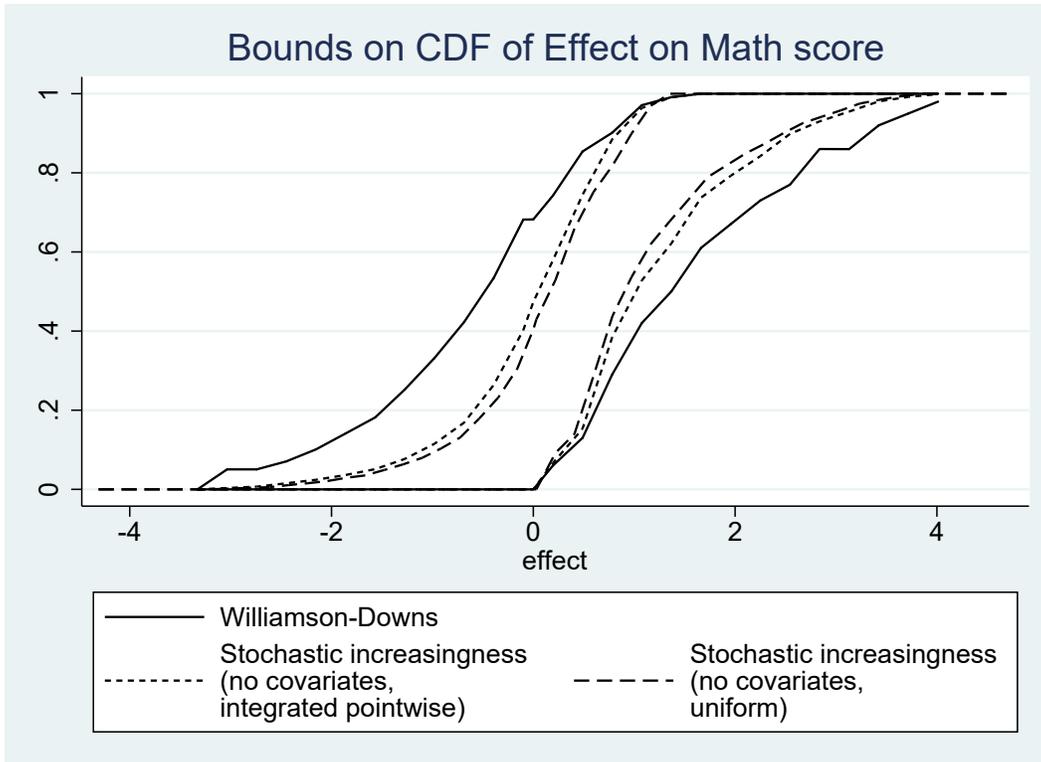


Figure 2: Estimated bounds on the cdf of effects on 7th grade math score. The solid curves show the Williamson-Downs bounds. The short-dash curves impose stochastic increasingness by integrating over the conditional (pointwise) bounds. The long-dash curves impose stochastic increasingness uniformly by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

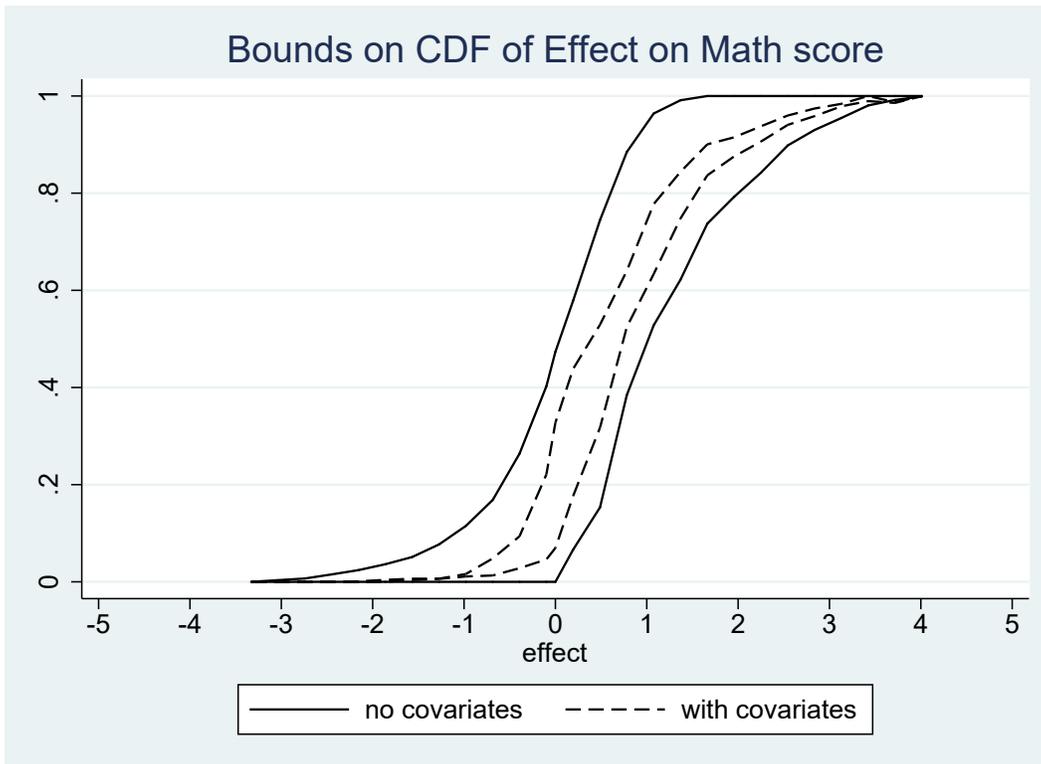


Figure 3: Estimated bounds on the cdf of effects on 7th grade math score. The solid bounds include no covariates. The dashed bounds use 4th grade math score and demographic characteristics described in the text.

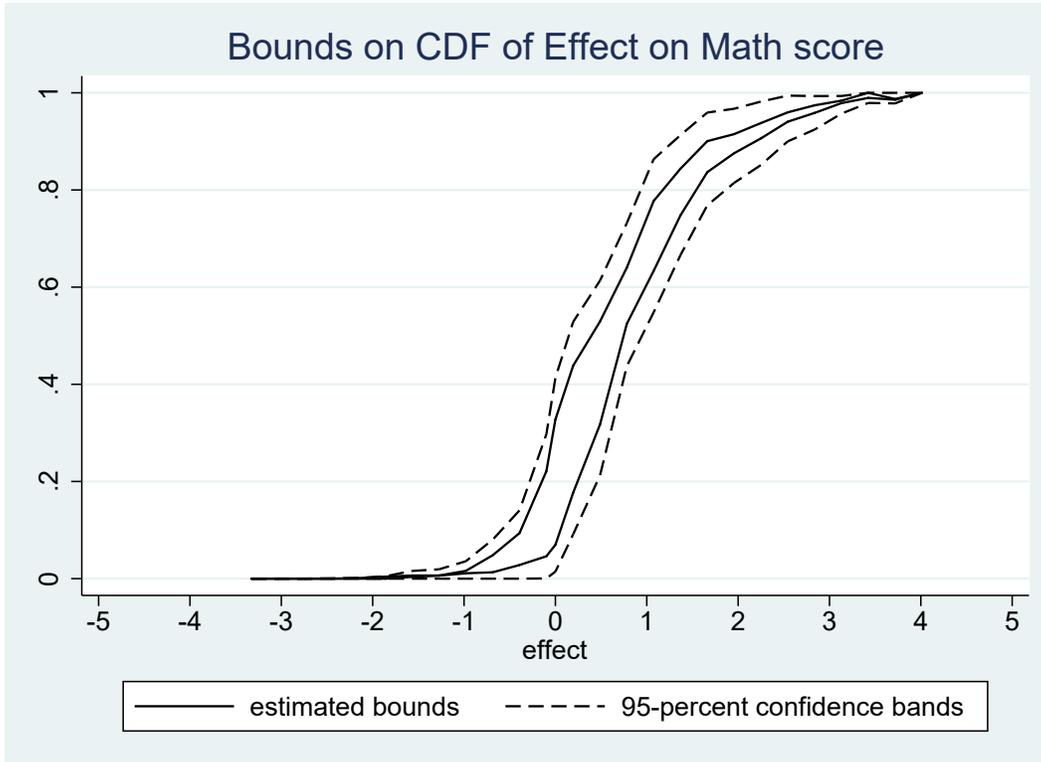


Figure 4: Estimated bounds on the cdf of effects on 7th grade math score. The solid curves show the bounds imposing stochastic increasingness by integrating over the conditional (pointwise) bounds incorporating covariates. The dashed curves show 95-percent confidence bands for the treatment effect cdf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

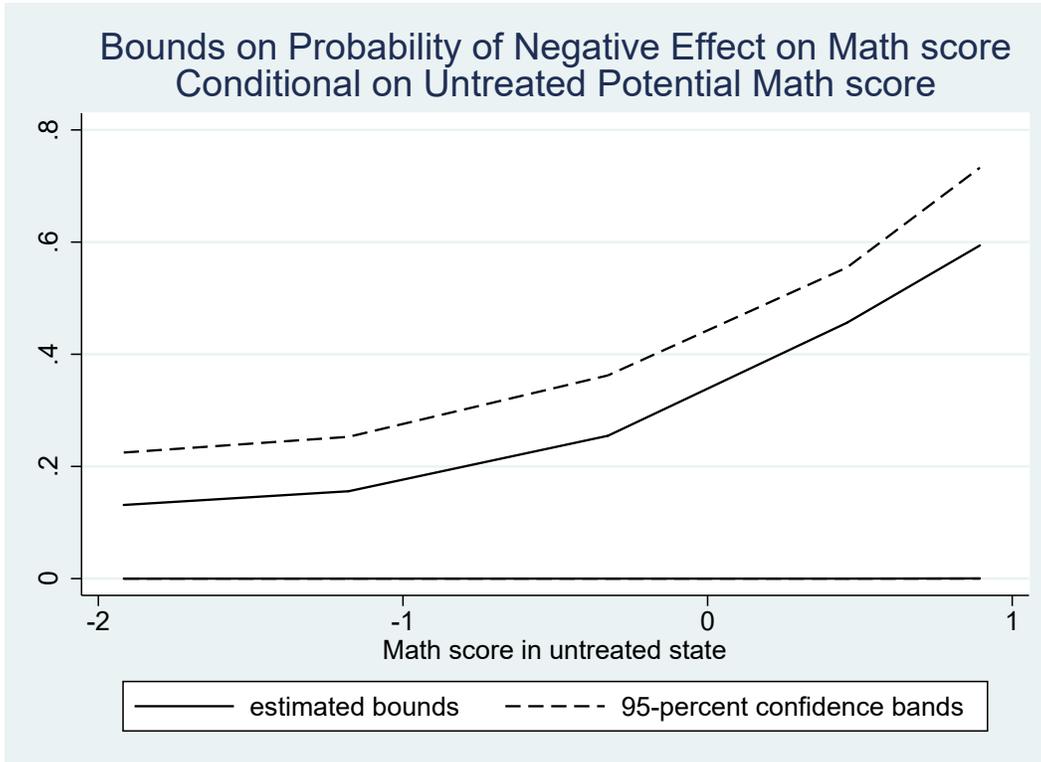


Figure 5: Estimated bounds on the probability of a negative effect of KIPP attendance on 7th grade math scores conditional on 7th grade math score in the untreated state. The solid curves show the estimated bounds using 4th grade math score and demographic characteristics described in the text. The dashed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

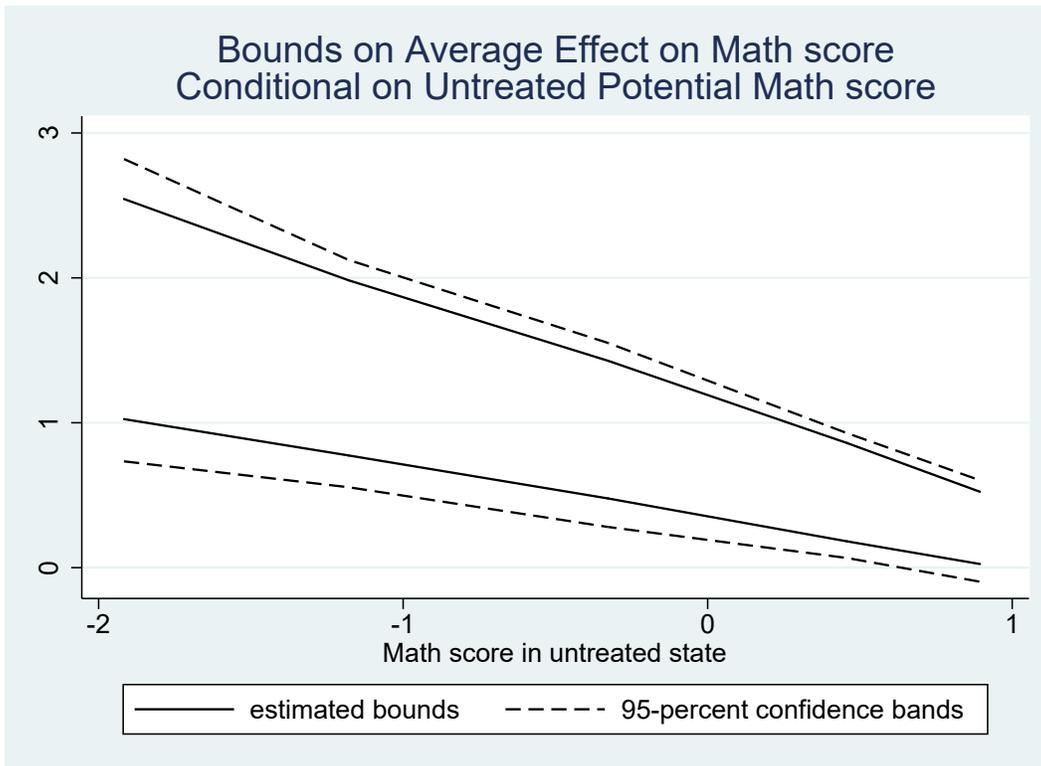


Figure 6: Estimated bounds on the average effect of KIPP attendance on 7th grade math scores conditional on 7th grade math score in the untreated state. The solid curves show the estimated bounds using 4th grade math score and demographic characteristics described in the text. The dashed curves show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

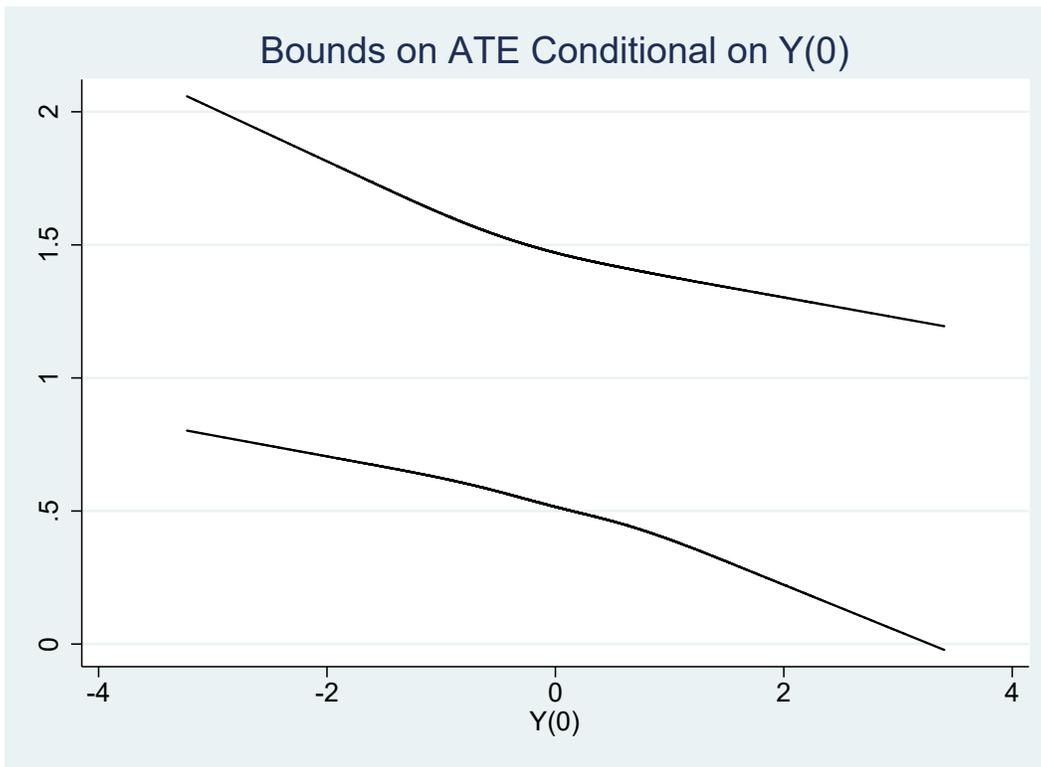


Figure 7: Simulated bounds on the average treatment effect conditional on untreated potential outcome. The true treatment effect is one for all values of $Y(0)$.

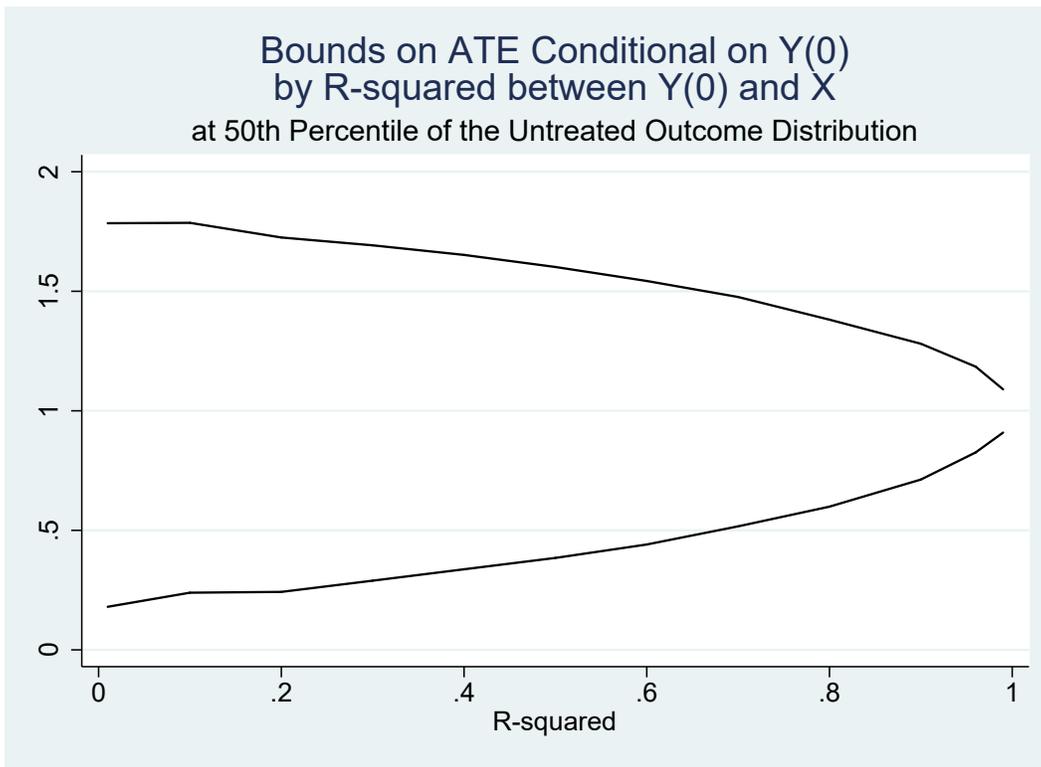


Figure 8: Simulated bound on the average treatment effect conditional on $Y(0) = 0$ as a function of the R^2 between $Y(0)$ and X . The true treatment effect is one.

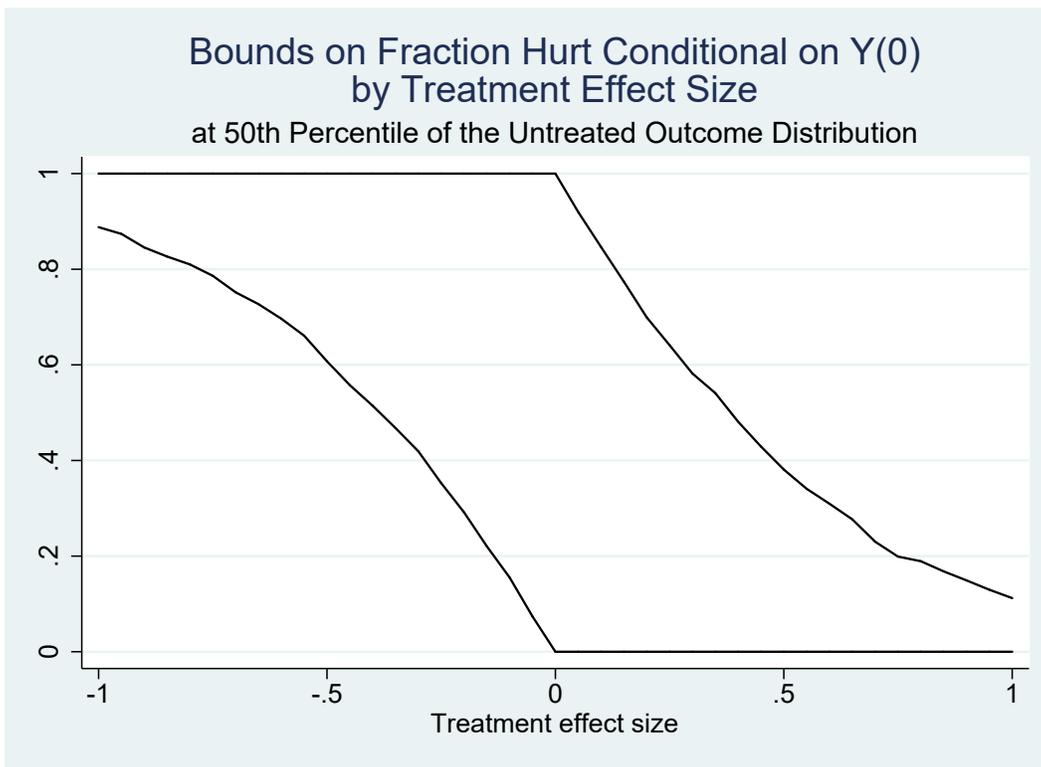


Figure 9: Simulated bound on the fraction hurt by treatment conditional on $Y(0) = 0$ as a function of the treatment effect. The true fraction is one when the treatment effect is negative (left side of the plot) and zero when the treatment effect is positive.

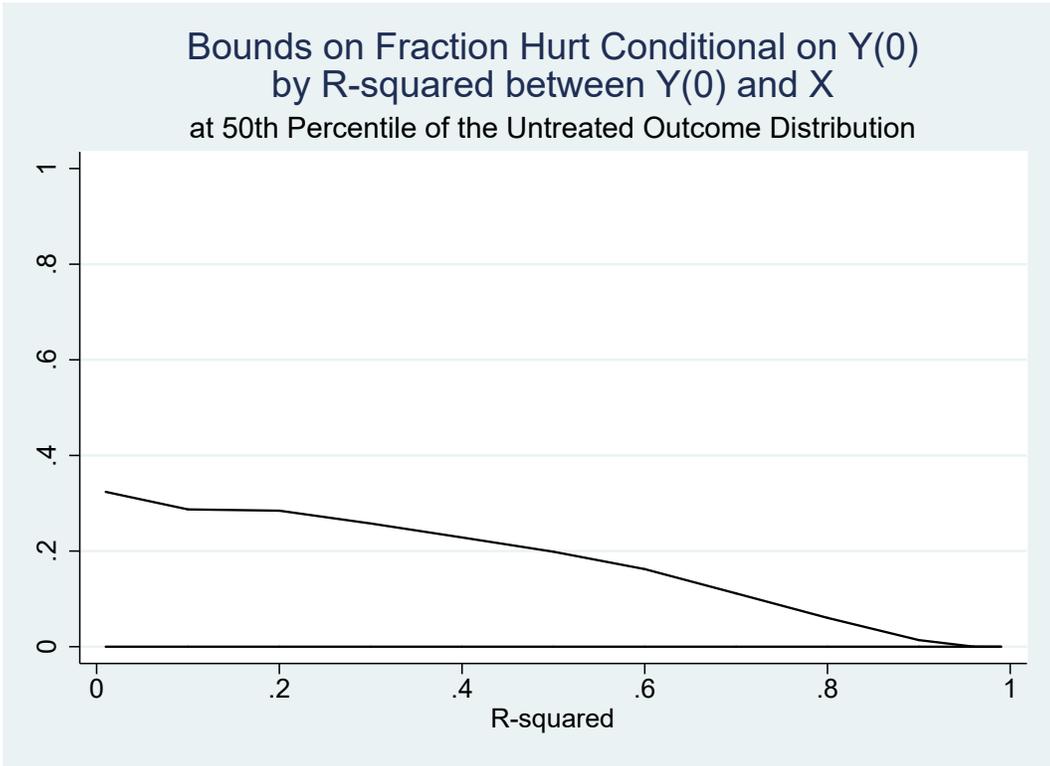


Figure 10: Simulated bound on the fraction hurt by treatment conditional on $Y(0) = 0$ as a function of the R^2 between $Y(0)$ and X . The true fraction in the simulation is zero.

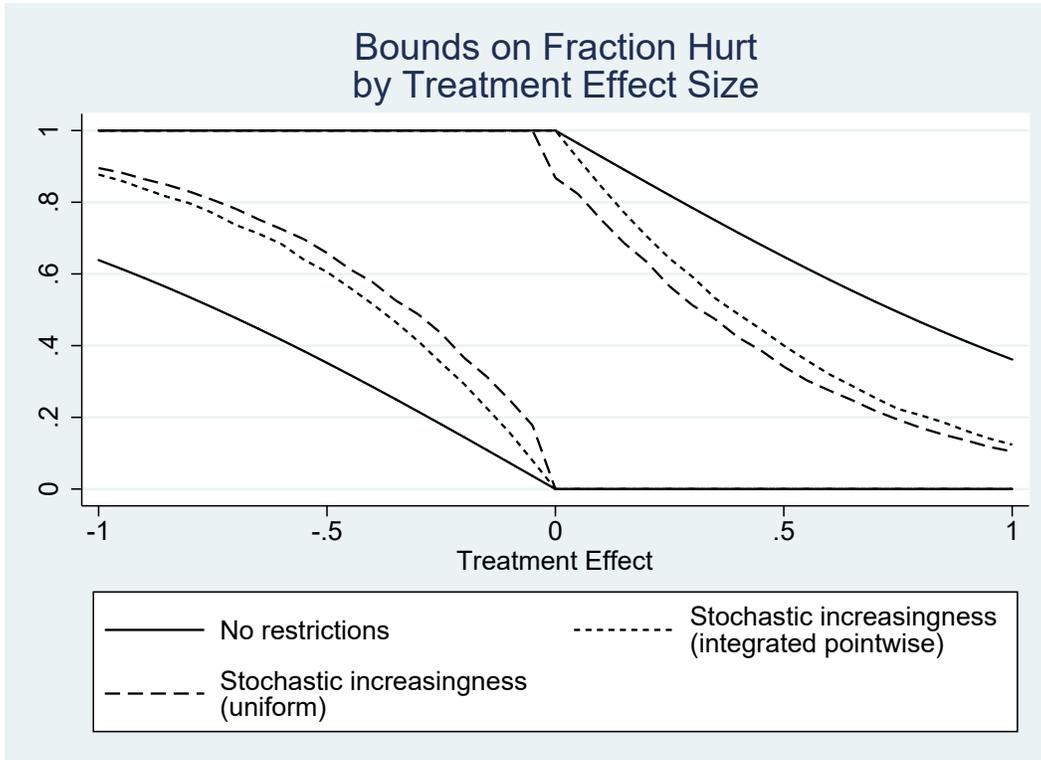


Figure 11: Simulated bounds on the fraction hurt by treatment as a function of the treatment effect. The true fraction is one when the treatment effect is negative (left side of the plot) and zero when the treatment effect is positive. The solid bounds impose no restrictions. The short-dashed bounds impose stochastic increasingness by integrating over the conditional (pointwise) bounds. The long-dashed bounds impose stochastic increasingness uniformly by searching over the space of copulae that satisfy stochastic increasingness.

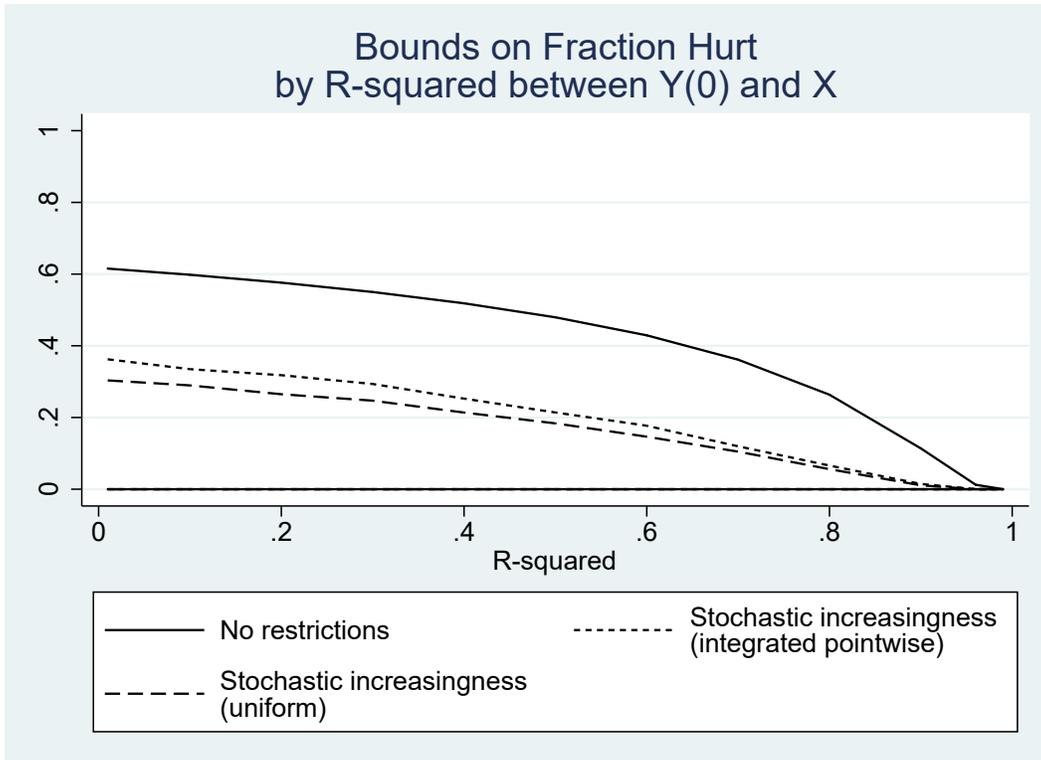


Figure 12: Simulated bound on the fraction hurt by treatment as a function of the R^2 between $Y(0)$ and X . The true fraction is zero. The solid bounds impose no restrictions. The short-dashed bounds impose stochastic increasingness by integrating over the conditional (pointwise) bounds. The long-dashed bounds impose stochastic increasingness uniformly by searching over the space of copulae that satisfy stochastic increasingness.