

# Partial Identification of the Distribution of Treatment Effects

Brigham R. Frandsen\*      Lars J. Lefgren\*

October 5, 2016

## Abstract

This article develops bounds on the distribution of treatment effects under plausible and testable assumptions on the joint distribution of potential outcomes, namely that potential outcomes are stochastically increasing. We show how to test the empirical restrictions implied by those assumptions. The resulting bounds substantially sharpen the classical bounds based on Frechet-Hoeffding limits. We present an application in which we identify bounds on the distribution of effects of attending a Knowledge is Power Program (KIPP) charter school on student academic achievement.

## 1 Introduction

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

---

\*Department of Economics, Brigham Young University

then improve the efficiency of enrollment decisions and minimize the probability that the student is harmed by placement in an ill-suited program. Policy makers may be rightly concerned about a program that harms a substantial fraction of participants despite a positive average effect.

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 potential outcomes at different quantiles, or average treatment effects, but not parameters that depend on the joint distribution of potential outcomes, such as the fraction of subjects harmed by the treatment, the expected treatment effect given a subject's outcome in the control distribution, the median treatment effect, or other features of 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 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 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 individual level treatment effects 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 \(1997\)](#) derives bounds on the distribution of treatment effects assuming treatment response is monotone. [Fan and Park \(2009\)](#) show the bounds can sometimes be tightened if one assumes that a certain dependence measure between between potential outcomes is known. These restrictions allow the bounds to be substantially tightened, but may be too strong to be plausible in many empirical settings.

We propose partially identifying restrictions that are plausible in many economic contexts, are testable, and lead to much sharper bounds on the distribution of treatment effects than classical inequalities. The restriction we propose is that potential outcomes are weakly stochastically increasing: the distribution of outcomes under treatment among individuals who would have realized a higher outcome without treatment (weakly) stochastically dominates that among individuals who would have realized a lower outcome without treatment.

This assumption implies bounds on the distribution of treatment effects conditional on each point of the control distribution. These bounds can be substantially tightened with the use of covariates. Bounds on the aggregate (marginal) distribution of treatment effects are obtained by integrating the conditional bounds across the distribution of control outcomes. These aggregate bounds can be further tightened by imposing conditional rank exchangeability, a condition more difficult to motivate economically, but which nevertheless has testable implications.

We propose a test of stochastic increasingness that probes the implication that

potential outcomes are non-negatively correlated. Although the correlation between potential outcomes cannot be computed directly, when the projections of potential outcomes on a sufficiently predictive set of covariates are sufficiently highly correlated, the Cauchy-Schwarz inequality allows us to sign the correlation between potential outcomes themselves.

We calculate our bounds in the context 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. The average benefits are largest for students who would have scored in the bottom of the control distribution. Such students also have the highest probability of benefitting from treatment. The bounds we calculate on the cdf of treatment effects are substantially narrower than the Williamson-Downs bounds that do not impose the assumption of stochastic increasingness. When examining English language arts, the average treatment effect is much smaller. Consequently the bounds tend to be less informative than in mathematics. However, we can still conclude that students at the bottom of the control distribution of achievement benefitted on average from attendance. In language arts, our bounds on the cdf of treatment effects are narrower than the Williamson-Downs bounds and are further tightened if we impose the additional assumption of exchangeability.

The next section 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 3 illustrates the bounds and the finite-sample size and power of the test of the partially identifying restrictions via Monte Carlo simulations. Section 4 applies the bounds and the testing procedure to an empirical example that is yet to be determined. Section 5 concludes.

## 2 Econometric Framework

Consider a binary treatment,  $D$ , that possibly affects a continuously distributed outcome  $Y$ . Let  $Y(1)$  and  $Y(0)$  be potential outcomes with and without treatment, with cdfs  $F_1$  and  $F_0$ . The observed outcome is  $Y = Y(D)$ . In addition to outcomes and treatment, we observe vectors of pre-treatment variables  $S$  and  $W$ . In the case of an endogenous treatment, let  $Z$  be instrumental variable that is independent of potential outcomes. The variables in  $S$  and  $W$  are taken to be in addition to any covariates  $X$  that may be included in the analysis for identification. For exposition, we suppress  $X$ , but all results below continue to hold conditional on  $X$ , if necessary. Denote the conditional cdf of potential outcomes given  $S$  as  $F_{1|S}$  and  $F_{0|S}$ .

The parameters of interest in this paper are features of the distribution of treatment effects  $\Delta := Y(1) - Y(0)$ , including the cdf,  $F_\Delta$ ; the conditional cdf given  $Y(0)$ ,  $F_{\Delta|Y(0)}$ ; and the expectation conditional on  $Y(0)$ ,  $E[\Delta|Y(0)]$ . 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, untestable assumptions (Heckman et al., 1997). The bounds we construct here sharpen 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.

## 2.1 Bounding the Distribution of Treatment Effects

The separate distributions of  $Y(0)$  and  $Y(1)$  (either marginal or conditional on  $S$ ) themselves imply bounds on the joint distribution of  $(Y(1), Y(0))$  and also the distribution of  $Y(1) - Y(0)$ . The well-known Frchet-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 treatment effects:

$$F_{\Delta|S}^L(t|S) = \sup_y \max \{F_{1|S}(y) - F_{0|S}(y - t), 0\}, \quad (1)$$

$$F_{\Delta|S}^U(t|S) = 1 + \inf_y \min \{F_{1|S}(y) - F_{0|S}(y - t), 0\}. \quad (2)$$

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  $Y(0)$ . Further restrictions are required to provide more informative bounds.

The restriction we propose assumes that potential outcomes are mutually stochastically increasing conditional on  $S$ :

**Definition 1**  $Y(1)$  is *weakly stochastically increasing* in  $Y(0)$  conditional on  $S$  if  $\Pr(Y(1) \leq t | Y(0) = y, S)$  is nonincreasing in  $y$  almost everywhere.

[Lehmann \(1966\)](#) described this property, referring to it as positive regression dependence. It means that individuals with higher  $Y(0)$  have a conditional distribution of  $Y(1)$  that weakly stochastically dominates the distribution of  $Y(1)$  among those with lower  $Y(0)$ . 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.

Stochastically increasing potential outcomes should be a plausible assumption in many economic settings. This condition rules out negative dependence between potential outcomes conditional on  $S$ , a plausible restriction in many settings, and can be tested, as we discuss below in Section 2.3.

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

**Theorem 2** *Suppose  $Y(1)$  is weakly stochastically increasing in  $Y(0)$  conditional on  $S$ . Then  $F_{\Delta|Y(0),S}(t|Y(0), S) := \Pr(\Delta \leq t|Y(0), S)$  is bounded from below by*

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

and from above by

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

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

**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 conditional distributions of potential outcomes, which are

identified. The bounds themselves are proper probability distributions.

Bounds on the distribution of treatment effects conditional on  $Y(0)$  only can be obtained by taking the conditional expectation given  $Y(0)$ :

$$F_{\Delta|Y(0)}^L(t|Y(0)) = E[F_{\Delta|Y(0),S}^L(t|Y(0),S)|Y(0)] \quad (5)$$

$$F_{\Delta|Y(0)}^U(t|Y(0)) = E[F_{\Delta|Y(0),S}^U(t|Y(0),s)|Y(0)]. \quad (6)$$

Bounds on the overall distribution of treatment effects can be constructed by taking the expectation of the conditional bounds (3) and (4):

$$F_{\Delta}^L(t) = E[F_{\Delta|Y(0),S}^L(t|Y(0),S)] \quad (7)$$

$$F_{\Delta}^U(t) = E[F_{\Delta|Y(0),S}^U(t|Y(0),S)]. \quad (8)$$

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  $\Delta$  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. For example, the bounds on the treatment effect cdf given by (3) and (4) also imply bounds on the average treatment effect conditional on  $Y(0)$ , a quantity that is frequently of great interest in applications, but not point identified. Let the average treatment effect conditional on  $Y(0)$  and  $S$  be denoted  $\Delta(Y(0), S) := E[Y(1) - Y(0)|Y(0), S]$ . By definition, bounds on the conditional expectation are given by integrating the derivative of the

cdf bounds:

$$\Delta^L(Y(0)) = \int t dF_{\Delta|Y(0)}^U(t|Y(0)), \quad (9)$$

$$\Delta^U(Y(0)) = \int t dF_{\Delta|Y(0)}^L(t|Y(0)). \quad (10)$$

The bounds (3) and (4) on the conditional distribution of treatment effects are by construction sharp pointwise in  $Y(0)$  and  $S$ , but not uniformly. Thus integrating (3) and (4) over  $Y(0)$  and  $S$  yields conservative bounds on the unconditional distribution of treatment effects. The unconditional bounds can be further tightened by assuming that potential ranks  $U(1) := F_1(Y(1))$  and  $U(0) := F_0(Y(0))$  are exchangeable conditional on  $S$ :

**Definition 3**  $U(0)$  and  $U(1)$  are *exchangeable* conditional on  $S$  if  $G_{01|S}(u, v|s) = G_{01|S}(v, u|s)$  for all  $u, v$  in the support of  $U(0)$  and  $U(1)$  conditional on  $S$ , where  $G_{01|S}$  is the joint distribution function of  $U(0)$  and  $U(1)$  conditional on  $S$ .

Like stochastic increasingness, the exchangeability condition also includes constant treatment effects and rank invariance as special cases. It implies rank similarity with respect to  $S$  (Frandsen and Lefgren, 2015). Unlike stochastic increasingness, however, it imposes a kind of symmetry on the joint distribution of potential outcomes whose economic meaning is less clear. Nevertheless, it, too has testable implications, and can therefore be falsified, as discussed in Section (2.3).

Exchangeability potentially tightens the bounds on the distribution of treatment effects dramatically, as the following result shows:

**Theorem 4** Suppose  $U(0)$  and  $U(1)$  are exchangeable conditional on  $S$ . Then (1)  $F_{1|S}(y|s) \leq F_{0|S}(y - t|s)$  for all  $y$  implies  $\Pr(\Delta \leq t|S = s) \leq 1/2$ ; and (2)  $F_{1|S}(y|s) \geq F_{0|S}(y - t|s)$  for all  $y$  implies  $\Pr(\Delta \leq t|S = s) \geq 1/2$ .

**Proof.** See the appendix. ■

Case (1) of the result means that if the (observed) conditional distribution of  $Y(1)$  is shifted in a stochastically dominant sense relative to  $Y(0)$  by at least some distance  $t$ , then the upper bound of the distribution of treatment effects given by (4) evaluated up through  $t$  can be tightened to  $1/2$ . Case (2) means the reverse: if  $Y(0)$  stochastically dominates  $Y(1)$  by at least some distance  $t$  then the lower bound given by (3) evaluated at  $t$  and above can be tightened to  $1/2$ .

As the simulations and application below show, these bounds can be dramatically tighter than those based on classical inequalities in [Williamson and Downs \(1990\)](#) and the bounds based on weak stochastic increasingness given by (3) and (4). These bounds based on exchangeability are most useful when the treatment has a modest, stochastically dominant shifting effect on outcomes.

## 2.2 Estimating the Bounds

The conditional cdf bounds (3) and (4) can be consistently estimated by plugging in consistent estimators for the conditional cdfs  $F_{1|S}$  and  $F_{0|S}$ . 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  $S_i$  in the untreated subsample and construct predicted value  $\hat{F}_{Y(0)|S}(Y_j|S_j)$
2. Nonparametrically regress an indicator  $1(Y_i \leq Y_j(0) + t)$  on  $S_i$  in the treated subsample and construct predicted value  $\hat{F}_{Y(1)|S}(Y_j(0) + t|S_j)$

### 3. Form estimates of the bounds

$$\begin{aligned} \hat{F}_{\Delta|Y(0),S}^L(t|Y_j(0), S_j) & : = \max \left\{ 0, \frac{\hat{F}_{Y(1)|S}(Y_j(0) + t|S_j) - \hat{F}_{Y(0)|S}(Y_j(0)|S_j)}{1 - \hat{F}_{Y(0)|S}(Y_j(0)|S_j)} \right\} \\ \hat{F}_{\Delta|Y(0),S}^U(t|Y_j(0), S_j) & : = \min \left\{ 1, \frac{\hat{F}_{Y(1)|S}(Y_j(0) + t|S_j)}{\hat{F}_{Y(0)|S}(Y_j(0)|S_j)} \right\}. \end{aligned} \quad (12)$$

The bounds (5) and (6) on the conditional distribution of treatment effects given  $Y(0)$  can be constructed by nonparametrically regressing the estimates (11) and (12). Bounds on the overall cdf of treatment effects can be constructed by taking the sample averages of (11) and (12). Finally, bounds (9) and (10) on the conditional expectation of treatment effects given  $Y(0)$  can be computed by numerically integrating the estimates for (5) and (6) on a discrete grid.

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 (5) and (6). 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  $\kappa$ -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.

### 2.2.1 Asymptotic Distribution

The procedure described above leads to consistent and asymptotically normal estimates of the bounds on the treatment effect distribution. 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 bounds.

The bound estimates are themselves functions of estimators for potential outcome conditional cdfs,  $\hat{F}_{Y(0)|S}$  and  $\hat{F}_{Y(1)|S}$ . 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  $S$  has continuous components (but not too many dimensions), Yu and Jones's (1998) double-kernel-weighted local polynomial method may be most appropriate. When  $S$  is discrete, standard logit regressions where treatment is fully interacted with  $S$  may be used. For our purposes, we assume only that the chosen estimator is uniformly consistent and asymptotically Gaussian. Let  $\hat{W}(v) := \left( \hat{F}_{Y(0)|S}(y|s), \hat{F}_{Y(1)|S}(y+t|s) \right)'$  be the vector of conditional cdf estimators, and let  $W(v)$  be the corresponding vector of the conditional cdfs themselves, both indexed by the vector  $v := (y, s, t)'$ . We consider the estimators as stochastic processes indexed by  $v \in \mathcal{V} := \mathcal{Y}_0 \times \mathcal{S} \times \mathcal{D}$ , where  $\mathcal{Y}_0$  is the support of  $Y(0)$ ,  $\mathcal{Y}_1$  is the support of  $Y(1)$ ,  $\mathcal{S}$  is the support of  $S$ , and  $\mathcal{D} = (\inf \mathcal{Y}_1 - \sup \mathcal{Y}_0, \sup \mathcal{Y}_1 - \inf \mathcal{Y}_0)$ . Then we require that

**Condition 5**  $r(n) \left( \hat{W}(v) - W(v) \right)$  weakly converges to a Gaussian process with zero mean function and covariance function  $C(v, v')$ , where  $\lim_{n \rightarrow \infty} r(n) = \infty$ .

The primitive conditions implying this condition will depend on the specific empirical setting, but should be met in typical cases, including the examples described above. The condition accommodates nonparametric estimators that may converge at a slower rate (e.g.  $r(n) = n^{2/5}$ ) or (semi)parametric estimators that converge

faster. Given the conditional cdf estimators' limiting distribution, we now establish the limiting distribution of the bound estimators. 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

$$\begin{aligned}\hat{H}^L & : = \frac{\hat{F}_{Y(1)|S}(y+t|s) - \hat{F}_{Y(0)|S}(y|s)}{1 - \hat{F}_{Y(0)|S}(y|s)} \\ \hat{H}^U & : = \frac{\hat{F}_{Y(1)|S}(y+t|s)}{\hat{F}_{Y(0)|S}(y|s)},\end{aligned}$$

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

**Theorem 6** *Suppose Condition 5 holds. Then  $r(n) (\hat{H}(v) - H(v))$  converges uniformly to a Gaussian process with zero mean function and covariance function  $J(v) C(v, v') J(v)'$  where the Jacobian  $J(v)$  is given by*

$$J(v) := \begin{bmatrix} -\frac{1 - F_{Y(1)|S}(y+t|s)}{(1 - F_{Y(0)|S}(y|s))^2} & (1 - F_{Y(0)|S}(y|s))^{-1} \\ -\frac{F_{Y(1)|S}(y+t|s)}{F_{Y(0)|S}(y|s)^2} & F_{Y(0)|S}(y|s)^{-1} \end{bmatrix}.$$

**Proof.** See the appendix. ■

Inference on the conditional bounds (3) and (4) based on the limiting distribution given in the theorem is straightforward. The hypothesis  $H_0 : F_{\Delta|Y(0),S}^L(t|y, s) = 0$ , for example is equivalent to the hypothesis  $H^L(y, s, t) \leq 0$ , of which an asymptotically valid and uniformly consistent test can be performed via a standard one-sided test of the simple hypothesis  $H^L(y, s, t) = 0$  against the alternative  $H^L(y, s, t) > 0$ .

## 2.3 Testing the Restrictions

The restrictions proposed in the previous section, stochastic increasingness and exchangeability, have testable implications. This section derives those implications and shows how they can be tested.

Conditional stochastic increasingness implies that  $Y(1)$  and  $Y(0)$  are positively correlated conditional on  $S$ . 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  $W$  (conditional on  $S$ ). Specifically, the Cauchy-Schwarz inequality implies (see Theorem 8 in the appendix) that a necessary condition for  $Cov(Y(1), Y(0) | S) \geq 0$  is that for all  $s$  in the support of  $S$

$$Corr(\hat{Y}(0), \hat{Y}(1) | S = s) \geq -\sqrt{\frac{(1 - R_{0,s}^2)(1 - R_{1,s}^2)}{R_{0,s}^2 R_{1,s}^2}}, \quad (13)$$

where  $\hat{Y}(0)$  and  $\hat{Y}(1)$  are linear projections of potential outcomes on  $W$  conditional on  $S = s$  with corresponding coefficients of determination  $R_{0,s}^2$  and  $R_{1,s}^2$ . Condition (13) is only nontrivial when the covariates  $W$  strongly predict potential outcomes: the respective  $R^2$ s between  $W$  and each potential outcome must geometrically average at least .5 (conditional on  $S = s$ ) in order for the right-hand side of (13) to be greater than than negative one. A practical procedure for verifying this condition when treatment is exogenous is within subgroups defined by  $S$  to estimate the conditional expectation of regress  $Y_i$  on  $W_i$  in the treated and untreated subsamples, calculate the correlation coefficient between the predicted values, and compare it to the right hand side of (13). Then 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)  $\kappa$  weights. In this case the procedure tests stochastic increasingness among compliers.

In order to have power, a test based on condition (13) requires covariates  $W$  that sufficiently strongly predict outcomes even conditional on  $S$ , a luxury not always available to researchers. When such covariates are not available, the plausibility of the stochastic increasingness assumption can still be assessed by examining the sign and strength of the correlation between  $\hat{Y}(0)$  and  $\hat{Y}(1)$ . This correlation reflects the extent to which potential outcomes move together based on observables, and if it is positive it lends support to potential outcomes moving together in unobservable ways as well, similar in spirit to how selection on unobservables can be assessed by examining selection on observables (Altonji et al., 2013).

The exchangeability condition—which is not required for the main results—implies rank similarity with respect to  $S$ , a restriction imposed in many econometric models. Thus tests of rank similarity such as those developed in Dong and Shen (2015) and Frandsen and Lefgren (2015) can also be considered tests of exchangeability.

### 3 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  $Y_i(0) = \beta S_i + \varepsilon_i$ . The treated potential outcome is  $Y_i(1) = Y_i(0) + \delta$ . The treatment indicator  $D_i$  is assigned independently of  $S_i$  and  $\varepsilon_i$  by random lottery whereby half the sample receives  $D_i = 1$  and half receive  $D_i = 0$ . The unobservables are generated according to

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

In the simulated model, the  $R^2$  between  $Y_i(0)$  and  $S_i$  is  $R^2 = \beta^2 \sigma_S^2 / (\beta^2 \sigma_S^2 + \sigma_\varepsilon^2)$ . The simulations set  $\sigma_S^2 = 1$ . The simulations vary  $\sigma_\varepsilon^2$  from .01 to 1, corresponding to an  $R^2$  between  $Y_i(0)$  and  $S$  from .99 to zero, and  $\beta$  is set accordingly to  $\sqrt{R^2 / \sigma_S^2}$  to ensure the variance of  $Y_i(0)$  remains equal to one. The simulations also vary the treatment effect size  $\delta$  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  $S_i$  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 1 plots the bounds (9) and (10) 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  $S$ . These simulations set the treatment effect  $\delta = 1$  and vary the  $R^2$  between  $Y_i(0)$  and  $S_i$  from zero to .99. Figure 2 plots the bounds (9) and (10) 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  $S$  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  $\delta$ . As above, these simulations set the  $R^2$  between  $Y_i(0)$  and  $S_i$  to 0.7. Figure 3 plots the bounds (5) and (6) evaluated

at zero as a function of  $\delta$  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  $S$ . These simulations set the treatment effect  $\delta$  equal to one, and vary the  $R^2$  between  $Y_i(0)$  and  $S_i$  from zero to .99. Figure 4 plots the bounds (5) and (6) 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  $\delta$ . Again,  $R^2$  is set to 0.7 for these simulations. Figure 5 plots the bounds (7) and (8) evaluated at zero as a function of  $\delta$ . The figure also plots the tighter bounds that result from imposing exchangeability (darker gray), and the wider bounds (1) and (2) that impose no restrictions (lighter gray). The bounds are reasonably tight when the treatment effect is large in magnitude (to the left and right ends of the plot) and are substantially tighter than the bounds that impose no restrictions. The bounds without exchangeability are quite wide, however, when the treatment effect is small in magnitude. Imposing exchangeability substantially tightens the bounds for modest treatment effect sizes, since exchangeability implies that either the upper bound must be no greater than .5 or the lower bound no less than .5.

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  $\delta$  is set to one, and  $R^2$  varies from zero to .99. Figure 6 plots the bounds (7) and (8) evaluated at zero as a function of  $R^2$ . The figure also plots the wider bounds with no restrictions (lighter gray). 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  $S$  increases, the bounds tighten substantially.

## 4 Empirical Example: Distributional Effects of KIPP

### Lynn

A substantial literature has found 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 interesting. 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 increase in math achievement of 0.35 standard deviations. The corresponding effect size in English language arts (ELA) is 0.12 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 variables are math and ELA 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 and 0.47 standard deviations below the mean in ELA. Seventh grade performance suggests the program was efficacious given that applicants performed just above the state mean in mathematics and only 0.21 standard deviations below the mean in language arts. This is confirmed when observing the substantial difference in performance between admitted and non-admitted students, particularly in mathematics. 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. We see that 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 academic performance in math and language arts. To do so, we simply employ two-stage least squares in which the dependent variable is either math or ELA 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. 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 300. The estimated effect of enrollment on math achievement for applicants who choose to enroll is 0.69 standard deviations”an effect that is both large and highly statistically significant. The corresponding effect for ELA is only 0.07 standard deviations, which is statistically insignificant. 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 separately regress outcomes in the treated and control state on a spline in prior math or ELA performance. [Table 3](#) shows the results. The correlation between predicted outcomes in treated and control states is virtually 1 for mathematics. This exceeds the threshold correlation of 0.94 required for ensuring that predicted outcomes in both states are sufficiently highly correlated that realized outcomes are also positively correlated, a necessary condition for stochastic increasingness. Examining ELA, we see that the correlation of predicted values in the treated and control state is also very close to one. However, the threshold correlation is 1.66, suggesting that even this high correlation is insufficient to ensure that realized outcomes in the two states are positively correlated. This high threshold is a consequence of the fact that prior ELA achievement has only moderate predictive power ( $r\text{-squared}=0.24$ ) in the treated state. While the threshold correlation in ELA is higher than we would prefer, the positive correlation between predicted ELA outcomes in the two states 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.

To tighten our bounds, in some cases it is helpful to further assume exchangeability between ranks in the treated and control states. We test this assumption using the procedure of [Frandsen and Lefgren \(2015\)](#). The p-values for the test are 0.79 in mathematics and 0.32 in ELA, suggesting that the assumption may be warranted.

We now employ the methodology we developed to bound the average effect of treatment by outcome in the control state. Figure 7 shows these bounds for mathematics with and without the use of covariates. In the specification with covariates, we include a spline with a single knot in the prior math score. We see that in both cases the bounds are informative. Without using additional covariates, even the lower bound of treatment effects is positive for all students who would have performed below the state level mean in the control state. Given the disadvantaged population, this suggests that average effects were positive for most students. Once we condition on prior performance, we see that the bounds are even tighter and we can rule out negative average effects for all but the best-performing students in the control state.

Figure 8 shows the corresponding bounds in ELA. Because the overall treatment effect is smaller, our bounds cannot rule out negative average effects for a substantial fraction of students. The upper bound of effects, however, are always above zero while the lower bound is above zero for students who would have performed poorly in the control state. The inclusion of covariates tightens the bounds substantially at the lower end of the control achievement distribution but has relatively smaller effects near the top of the distribution. Notice that the bounds with covariates do not lie uniformly within those without covariates. This highlights the fact that the bounds are constructed on the basis of non-nested assumptions. In other words, conditional stochastic increasingness could hold even if unconditional stochastic increasingness does not hold and vice-versa.

We next examine the probability that a subject was hurt as a function of their outcome in the control distribution. Figure 9 shows this with and without covariates for math while Figure 10 shows the corresponding information for ELA. Focusing on math, 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. Both figures

make 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. The bounds are generally tighter with the inclusion of covariates but at the very bottom of the control distribution, the bounds without covariates are actually more informative.

Switching to ELA, we generally cannot rule out a substantial risk of a negative treatment effect. The bounds are also substantially different depending on whether we include covariates or not. The bounds are tighter at the bottom of the distribution when we condition on prior performance but the opposite holds when we do not condition on any covariates. Note that the lower bound exceeds zero near to the top of the control outcome distribution. This is due to the fact that there happen to be more very high achieving students in the control sample than the treatment sample.

In our final set of figures, we show bounds on the cdf of treatment effects. Figure 11 shows bounds for the cdf in mathematics with and without covariates. We see that the bounds are narrower with covariates. These bounds suggest that the 40th percentile treatment effect was positive—suggesting a clear majority of students’ math scores benefited from attending KIPP. This confirms that the treatment was efficacious for the substantial majority of students. Figure 12 shows our bounds in comparison to the Williamson-Downs bounds. For both we condition on prior math achievement. We see that our bounds are substantially narrower than the Williamson-Downs bounds. Assuming exchangeability does not further narrow the bounds in mathematics.

Moving to ELA, Figure 13 shows the bounds for the cdf in ELA with and without covariates. As expected, the bounds are narrower with covariates. Figure 14 shows the Williamson-Downs bounds which impose no restrictions, the bounds implied by stochastic increasingness, and the bounds implied by stochastic increasingness and exchangeability. The Williamson-Downs bounds are quite uninformative: even above the 85th percentile negative effects cannot be ruled out. The stochastic increasingness

bounds are also generally less informative than we would like. We cannot rule out negative effects until we approach the 80th percentile of treatment effects. However, imposing exchangability allows to substantially narrow the bounds, allowing us to determine that the median treatment effect was non-negative. This implies that a majority of students did in fact benefit from admission to KIPP.

Reviewing what we learn from the KIPP charter school experiment, we confirm that attendance increased math achievement substantially but had only a much smaller effect on ELA performance. Our bounding methodology shows that in mathematics, the average treatment effect was very large and positive for students with poor outcomes in the control distribution and still unambiguously positive for all except the students with the very best control outcomes. Furthermore, the worst case bounds suggest extremely small probabilities that low achieving students would have been harmed by treatment. Even in the worst case, we know that treatment increased achievement for the majority of students. 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.

Examining the ELA results, while we can say that students with very poor control outcomes would likely have benefited substantially from treatment, we generally cannot rule out negative (or positive) effects for most students. Bounds on the fraction of students hurt by treatment are also not particularly informative, unless one is willing to make the assumption of exchangability, which allows us to say that no more than half of students were harmed by treatment, and potentially far fewer. The ELA example highlights the limitations of our (and other) bounding methods when treatment effects are comparatively small. It also highlights, however, the value of the exchangability assumption for calculations regarding the fraction of subjects hurt

by treatment.

## 5 Conclusion

This paper proposed 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 marginal distributions. Our bounds are further tightened with the use of covariates or by making the additional testable assumption of rank exchangeability.

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 bounds were less informative in the context of English Language Arts but employing the additional assumption of rank exchangeability we can conclude that a majority of students benefited from attendance in this subject as well.

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

- Yingying Dong and Shu Shen. Testing for rank invariance or similarity in program evaluation: The effect of training on earnings revisited. Unpublished working paper, 2015.
- 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 São Paulo, Getulio Vargas Foundation (Brazil), June 2010.
- Brigham R. Frandsen and Lars J. Lefgren. Testing rank similarity. Unpublished working paper, 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(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.
- 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.

Keming Yu and M. C. Jones. Local linear quantile regression. *Journal of the American Statistical Association*, 93(441):228–237, 1998.

## Appendix

The following result is crucial to Theorem 2:

**Lemma 7** *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

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

**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  $\Pr(X \leq x|Y = y)$  minimizes  $\underline{F}_{X|Y}(x|y)$  subject to the following constraints:

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

(since  $X$  is stochastically increasing in  $Y$ ) and

$$(3) : \int_y \underline{F}_{X|Y}(x|s) dF_Y(s) = F_X(x)$$

(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:

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

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

subject to the following constraints:

$$(1) : \bar{F}_{X|Y}(x|y) \leq \bar{F}_{X|Y}(x|y'), \quad y' \leq y,$$

$$(2) : \bar{F}_{X|Y}(x|y) \geq \bar{F}_{X|Y}(x|y''), \quad y'' > y,$$

(since  $X$  is stochastically increasing in  $Y$ ) and

$$(3) : \int_y \bar{F}_{X|Y}(x|s) dF_Y(s) = F_X(x)$$

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

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

■

**Proof of Theorem 2.** Note that by definition

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

Since  $Y(1)$  is conditional stochastically increasing in  $Y(0)$ , Lemma 7 applies to this case conditionally on  $S$ , taking  $x = Y(0) + t$ ;  $y = Y(0)$ ;  $F_X = F_{Y(1)|S}$ ;  $F_Y = F_{Y(0)|S}$ . Making these substitutions in the lemma's result gives the result in the theorem. The argument for the lower bound is similar. ■

**Proof of Theorem 4.** The theorem's premise  $F_{1|S}(y|S) \leq F_{0|S}(y - t|S)$  is by definition equivalent to  $F_{1|S}^{-1}(\tau|S) \geq F_{0|S}^{-1}(\tau|S) + t$  which is in turn equivalent to

$$F_{1|S}\left(F_{0|S}^{-1}(\tau|S) + t|S\right) \leq \tau. \quad (14)$$

Let  $U(d, S) := F_{d|S}(Y(d)|S)$  be the conditional rank of  $Y(d)$  conditional on  $S$ . Then

$$\begin{aligned} & \Pr(Y(1) - Y(0) < t|S) \\ &= \Pr\left(U(1, S) < F_{1|S}\left(F_{0|S}^{-1}(U(0, S)|S) + t|S\right) | S\right) \\ &\leq \Pr(U(1, S) < U(0, S) | S) \\ &\leq 1/2, \end{aligned}$$

where the first equality is by definition, the following inequality follows from (14) and the final inequality follows from the definition of conditional exchangeability. ■

**Proof of Theorem 6.** Regarded as a function of  $W$ , the function  $H$  is continuously differentiable with Jacobian

$$J(v) := \begin{bmatrix} -\frac{1 - F_{Y(1)|S}(y+t|s)}{(1 - F_{Y(0)|S}(y|s))^2} & (1 - F_{Y(0)|S}(y|s))^{-1} \\ -\frac{F_{Y(1)|S}(y+t|s)}{F_{Y(0)|S}(y|s)^2} & F_{Y(0)|S}(y|s)^{-1} \end{bmatrix}.$$

Condition 5 and the functional delta method (van der Vaart and Wellner, 1996, Theorem 3.9.4) imply the result. ■

**Theorem 8** Let  $\hat{Y}(0)$  and  $\hat{Y}(1)$  be linear projections of potential outcomes on  $W$  conditional on  $S = s$  with corresponding coefficients of determination  $R_{0,s}^2$  and  $R_{1,s}^2$ . Then  $Cov(Y(1), Y(0) | S = s) \geq 0$  implies

$$Corr(\hat{Y}(1), \hat{Y}(0) | S) \geq -\sqrt{(1 - R_{0,s}^2)(1 - R_{1,s}^2) / (R_{0,s}^2 R_{1,s}^2)}.$$

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

$$Cov(Y(0), Y(1) | S = s) = Cov(\hat{Y}(0), \hat{Y}(1) | S = s) + Cov(\varepsilon(0), \varepsilon(1) | S = s). \quad (15)$$

The Cauchy-Schwarz inequality implies

$$Cov(\varepsilon(0), \varepsilon(1) | S = s) \leq \sqrt{(1 - R_{0,s}^2) Var(Y(0) | S = s) (1 - R_{1,s}^2) Var(Y(1) | S = s)}.$$

Inserting this into (15) yields an upper bound on the covariance between potential outcomes:

$$Cov(Y(0), Y(1) | S = s) \leq Cov(\hat{Y}(0), \hat{Y}(1) | S = s) + \sqrt{(1 - R_{0,s}^2) Var(Y(0) | S = s) (1 - R_{1,s}^2) Var(Y(1) | S = s)}.$$

This upper bound is nonnegative when

$$Cov(\hat{Y}(0), \hat{Y}(1) | S = s) \geq -\sqrt{(1 - R_{0,s}^2) Var(Y(0) | S = s) (1 - R_{1,s}^2) Var(Y(1) | S = s)},$$

or, equivalently,

$$\text{Corr}(\hat{Y}(0), \hat{Y}(1) | S = s) \geq -\sqrt{\frac{(1 - R_{0,s}^2)(1 - R_{1,s}^2)}{R_{0,s}^2 R_{1,s}^2}}.$$

■

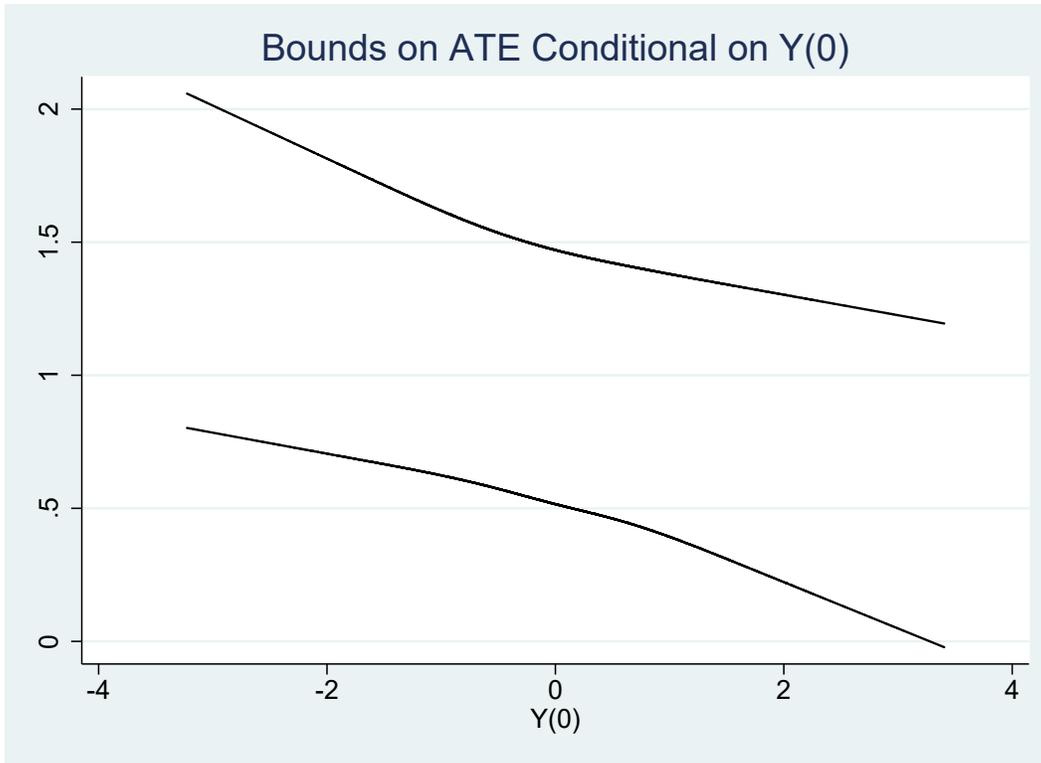


Figure 1: Simulated bounds on the average treatment effect conditional on untreated potential outcome. The true treatment effect is one for all values of  $Y(0)$ .

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)	
7th Grade English Language Arts	-0.21 (0.97)	-0.26 (1.08)	-0.18 (0.92)	
4th Grade Math	-0.39 (1.06)	-0.43 (1.16)	-0.37 (1.01)	0.70
4th Grade English Language Arts	-0.47 (1.07)	-0.45 (1.12)	-0.48 (1.04)	0.80
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

	Mathematics	English Language Arts
Effect of Enrollment	0.69 (0.15)	0.07 (0.15)
First Stage F-Statistic	372.88	372.88
Observations	177	177

Notes: The first row shows the estimated impact of enrollment in KIPP on mathematics and ELA performance in the 7<sup>th</sup> grade. This estimate comes from a two-stage least squares regression in which the instrument is admission into KIPP through the lottery. Robust standard errors are in parentheses. Controls include 4<sup>th</sup> grade achievement as well as 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 F-statistic of the instrument.

Table 3: Test of Positive Correlation of Potential Outcomes

	Mathematics	English Language Arts
R <sup>2</sup> Treated Outcomes	0.47	0.24
R <sup>2</sup> Control Outcomes	0.56	0.54
Correlation between Predicted Treated and Control Outcomes	1.00	0.97
Threshold Correlation	0.94	1.66

Notes: The R<sup>2</sup> for the treated outcomes comes from a regression of math or reading outcomes on a one knot spine in 4<sup>th</sup> grade math or reading achievement.

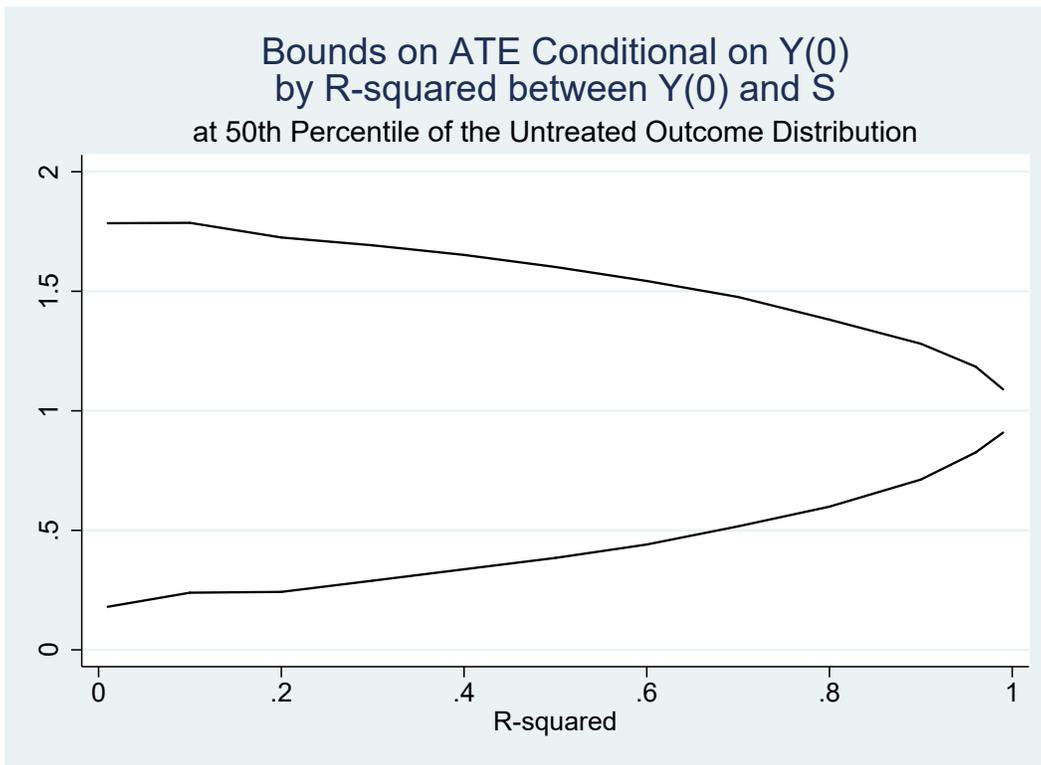


Figure 2: Simulated bound on the average treatment effect conditional on  $Y(0) = 0$  as a function of the  $R^2$  between  $Y(0)$  and  $S$ . The true treatment effect is one.

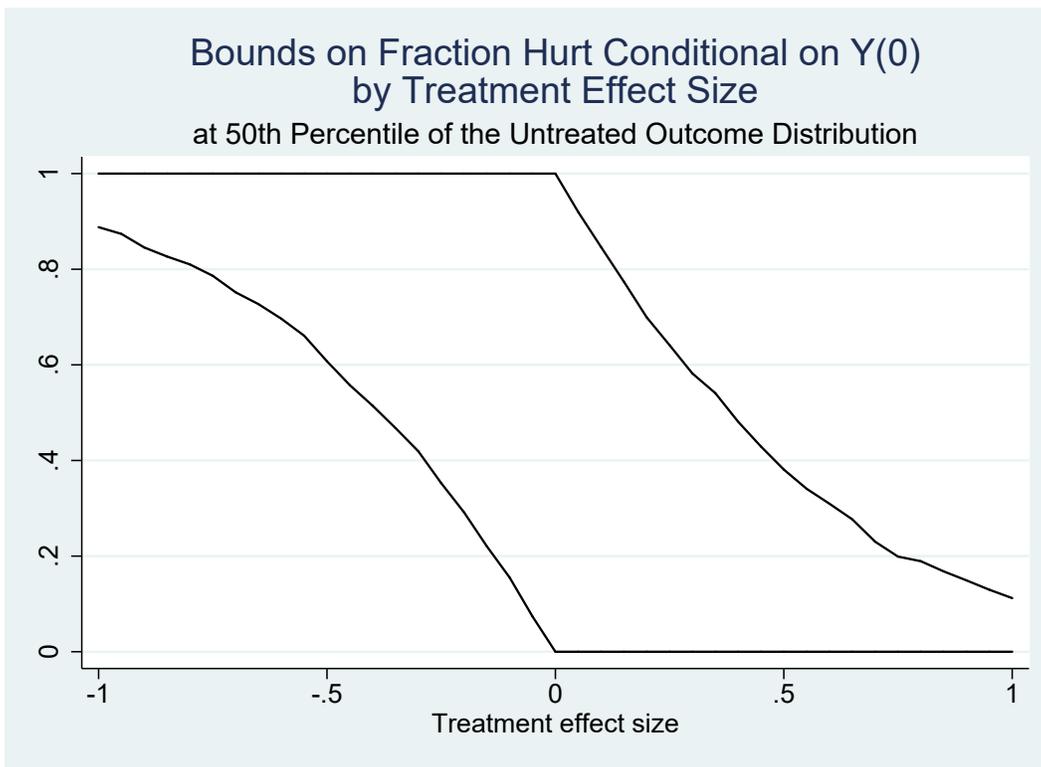


Figure 3: Simulated bound on the fraction hurt by treatment conditional on  $Y(0) = 0$  as a function of the treatment effect. The true fraction is one when the treatment effect is negative (left side of the plot) and zero when the treatment effect is positive.

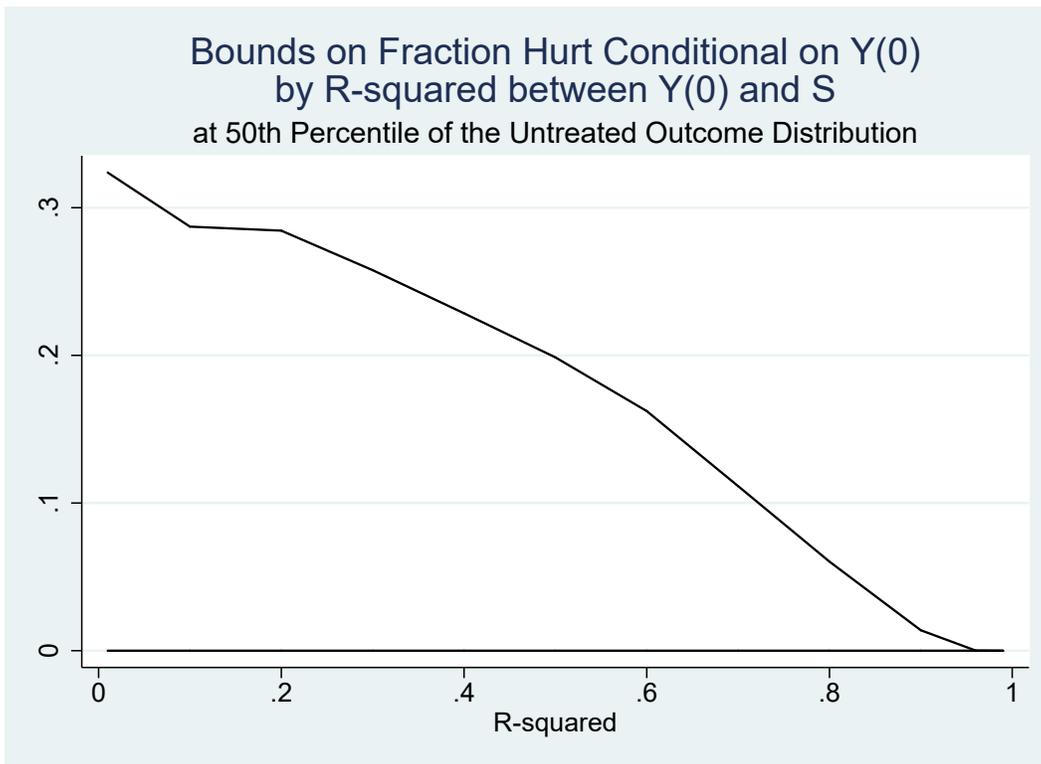


Figure 4: Simulated bound on the fraction hurt by treatment conditional on  $Y(0) = 0$  as a function of the  $R^2$  between  $Y(0)$  and  $S$ . The true fraction in the simulation is zero.

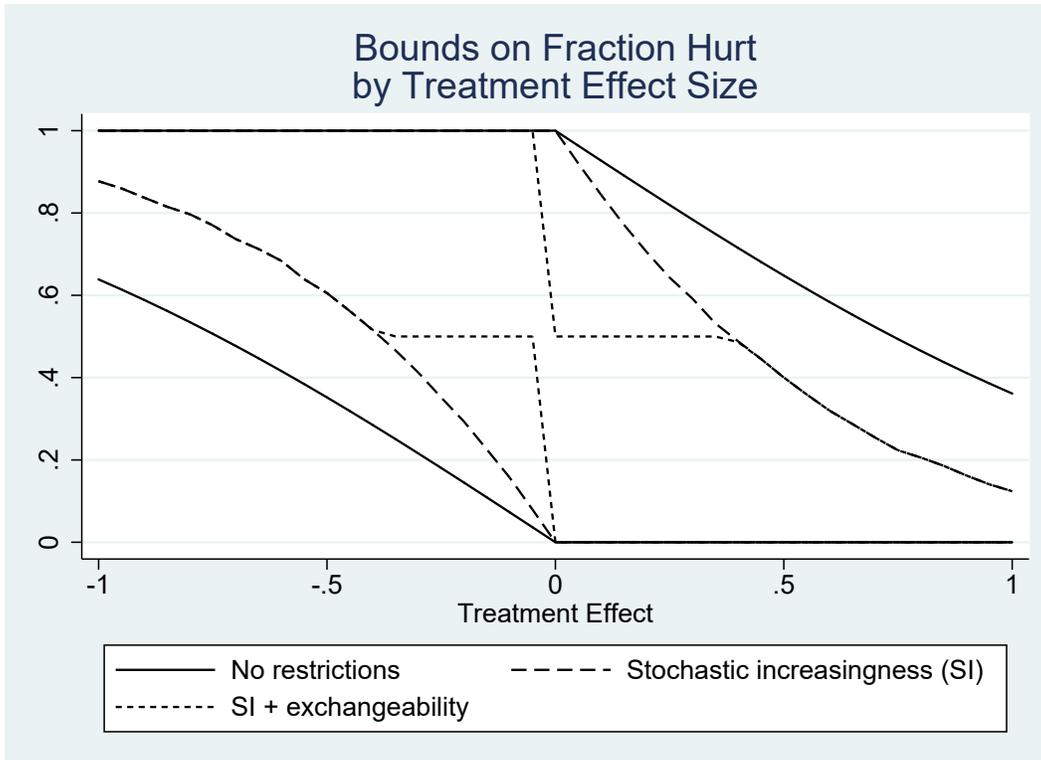


Figure 5: Simulated bound on the fraction hurt by treatment as a function of the treatment effect. The true fraction is one when the treatment effect is negative (left side of the plot) and zero when the treatment effect is positive. The lightest gray bounds impose no restrictions. The medium gray bounds impose stochastic increasingness. The darker gray bounds impose stochastic increasingness and conditional rank exchangeability.

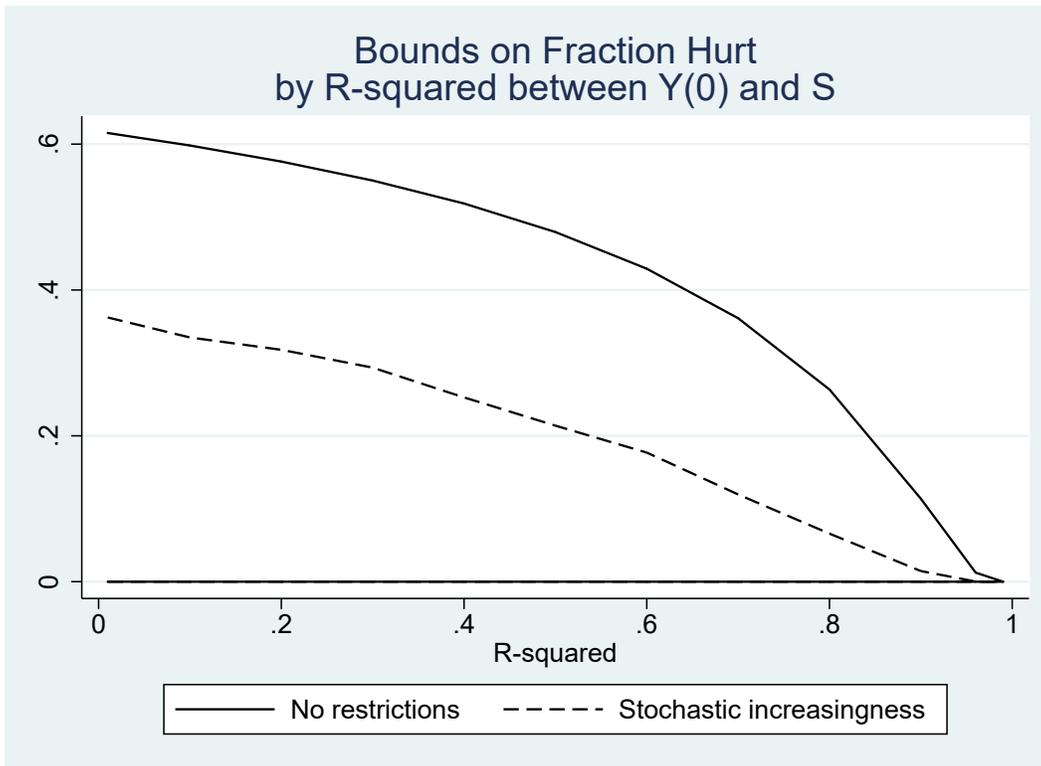


Figure 6: Simulated bound on the fraction hurt by treatment as a function of the  $R^2$  between  $Y(0)$  and  $S$ . The true fraction is zero. The lightest gray bounds impose no restrictions. The medium gray bounds imposing stochastic increasingness. The darker gray bounds impose stochastic increasingness and conditional rank exchangeability.

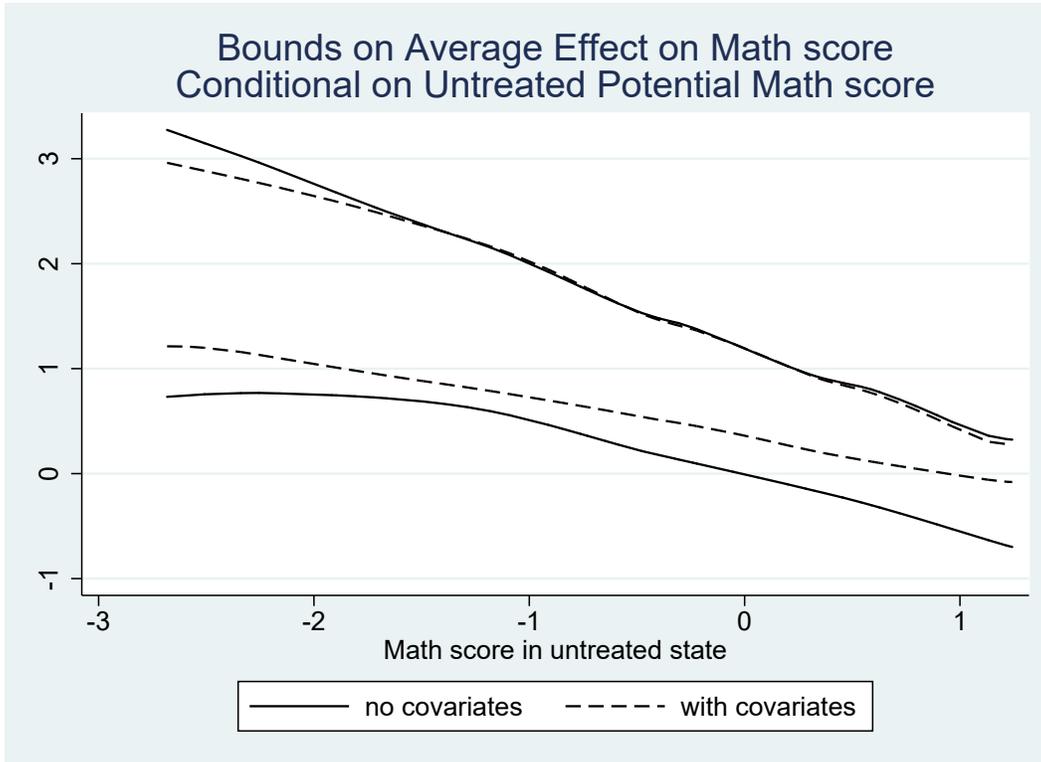


Figure 7: Estimated bounds on the average effect of KIPP attendance on 7th grade math scores conditional on 7th grade math score in the untreated state. The darker region includes no covariates in computing the bounds. The lighter region uses 4th grade math score and demographic characteristics described in the text in computing the bounds.

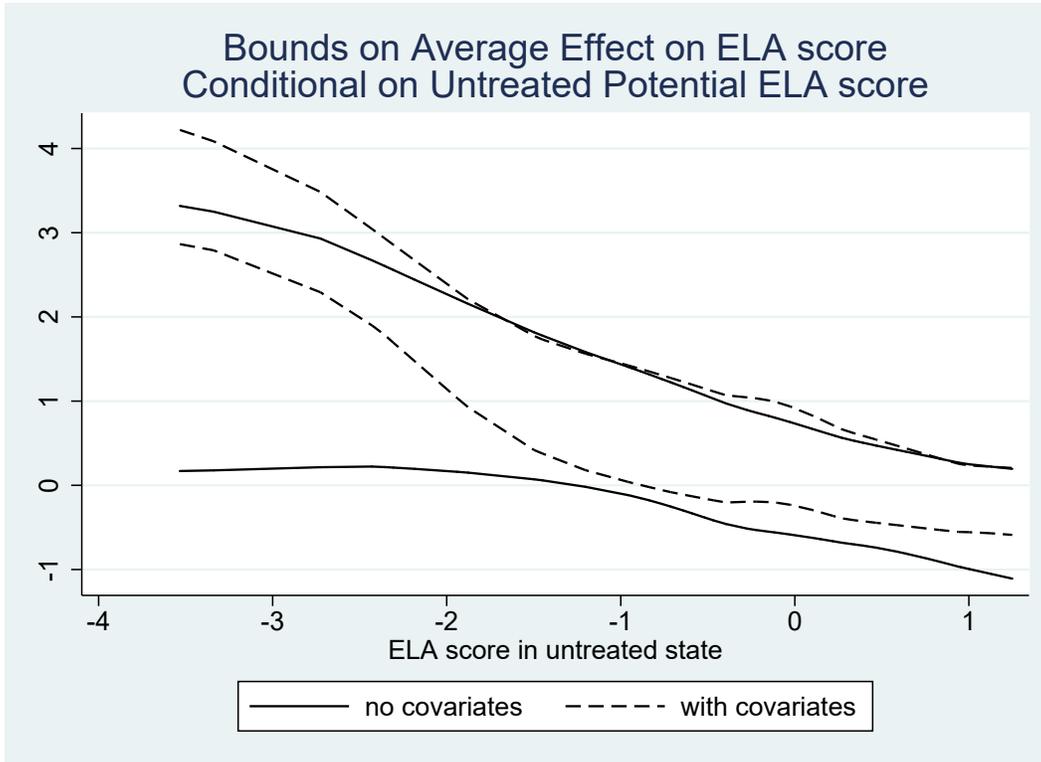


Figure 8: Estimated bounds on the average effect of KIPP attendance on 7th grade ELA scores conditional on 7th grade ELA score in the untreated state. The darker region includes no covariates in computing the bounds. The lighter region uses 4th grade ELA score and demographic characteristics described in the text in computing the bounds.

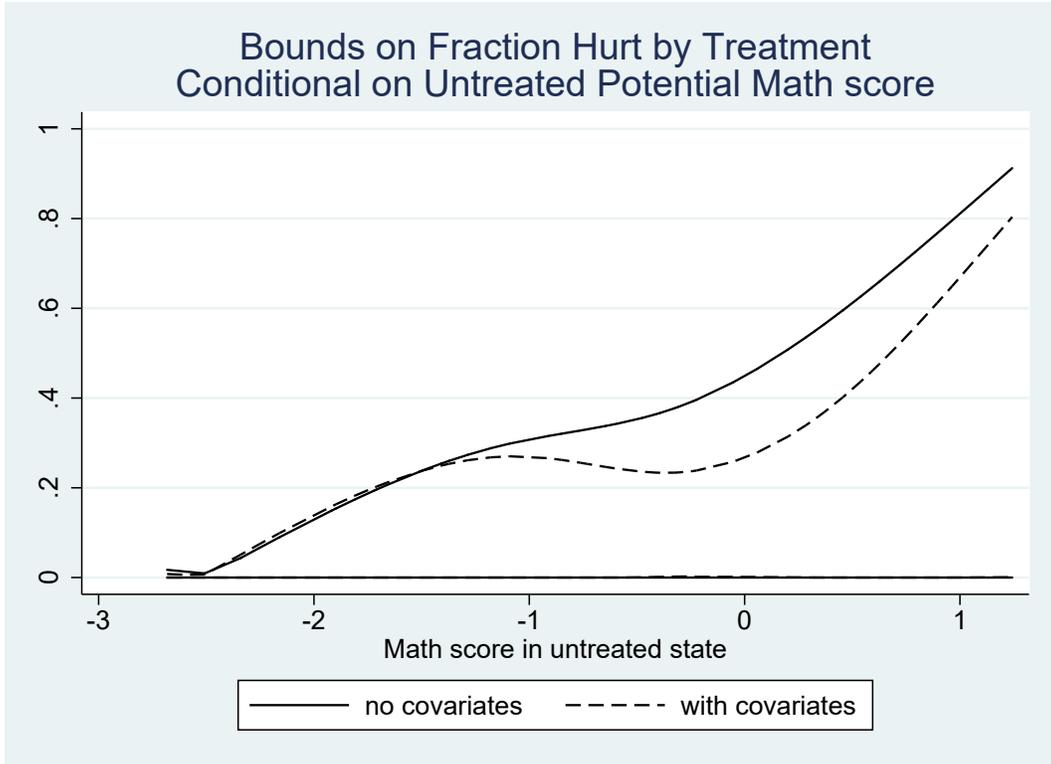


Figure 9: Estimated bounds on the fraction of students whose 7th grade math score was hurt by KIPP attendance conditional on 7th grade math score in the untreated state. The darker region includes no covariates in computing the bounds. The lighter region uses 4th grade math score and demographic characteristics described in the text in computing the bounds.

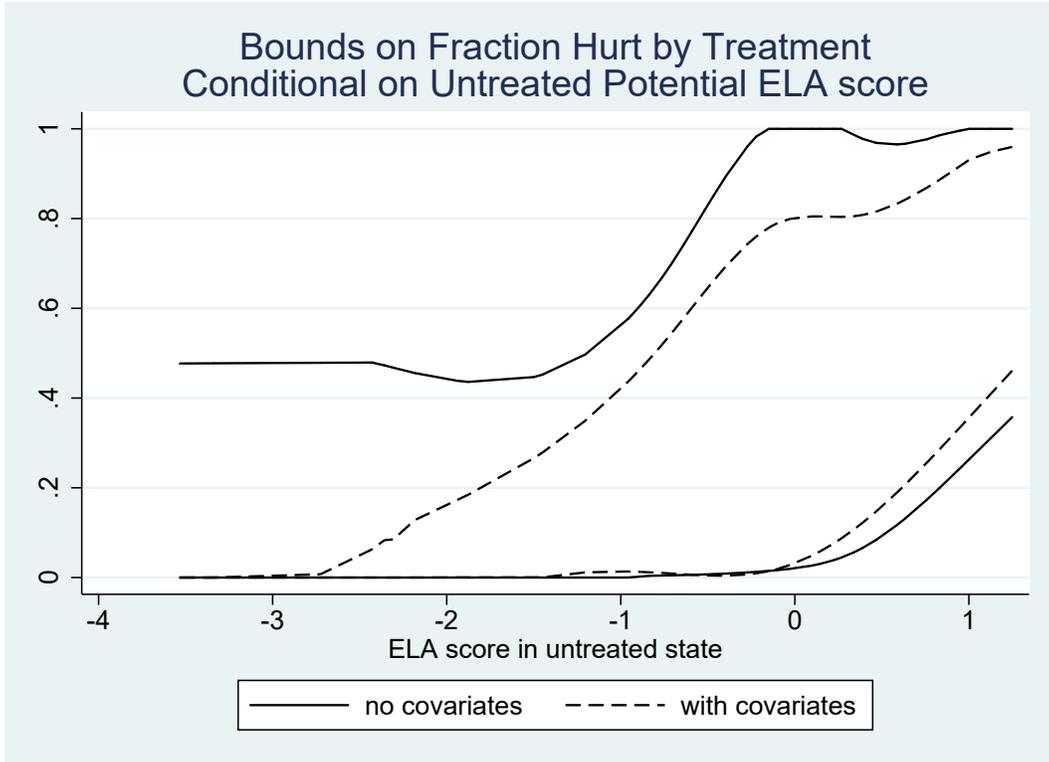


Figure 10: Estimated bounds on the fraction of students whose 7th grade ELA score was hurt by KIPP attendance conditional on 7th grade ELA score in the untreated state. The darker region includes no covariates in computing the bounds. The lighter region uses 4th grade ELA score and demographic characteristics described in the text in computing the bounds.

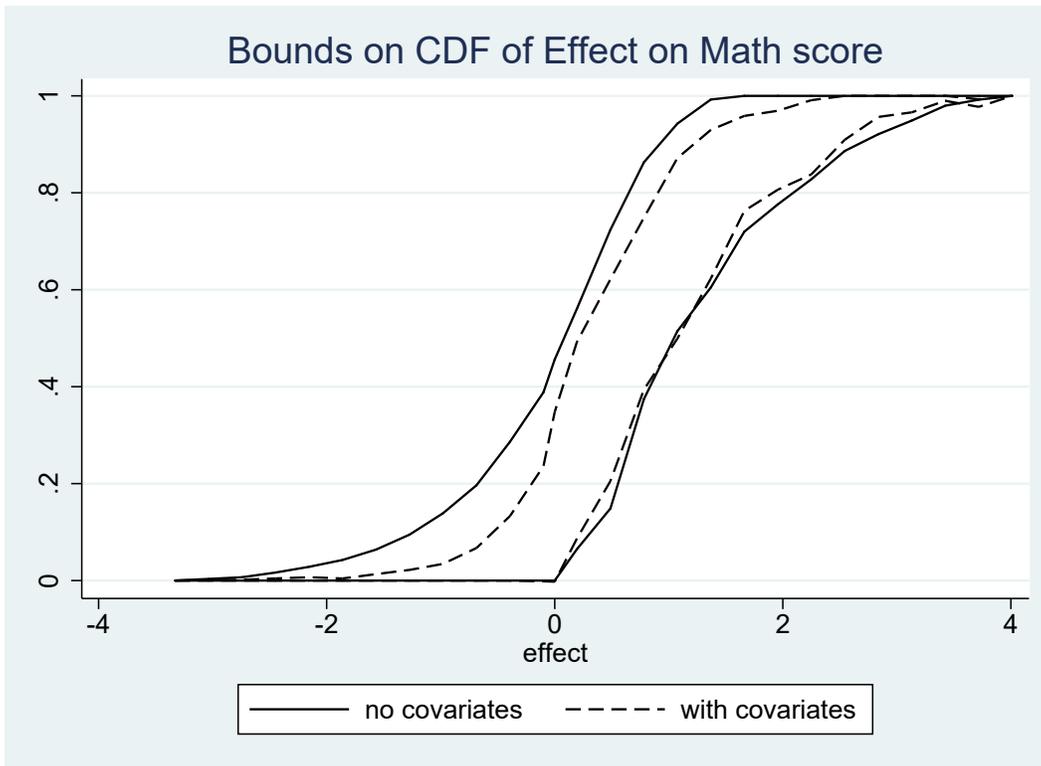


Figure 11: Estimated bounds on the cdf of effects on 7th grade math score. The darker region includes no covariates in computing the bounds. The lighter region uses 4th grade math score and demographic characteristics described in the text in computing the bounds.

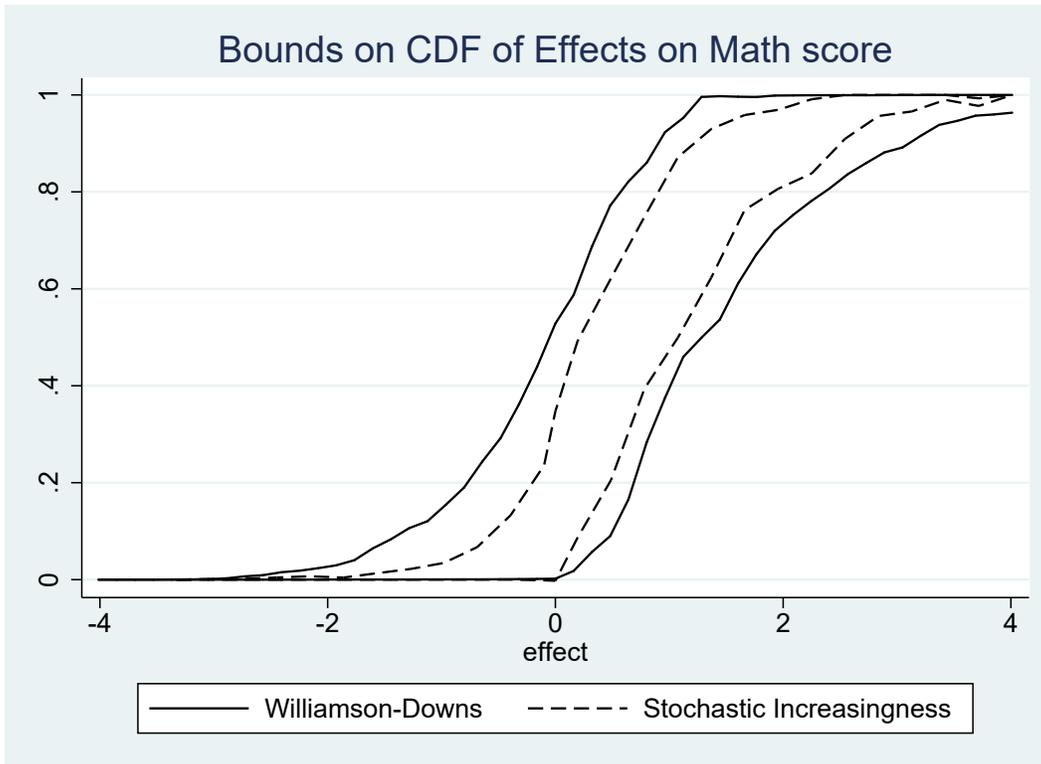


Figure 12: Estimated bounds on the cdf of effects on 7th grade math score. The darker region shows the Williamson-Downs bounds. The lighter region shows bounds that impose stochastic increasingness. Both bounds condition on 4th grade math score.

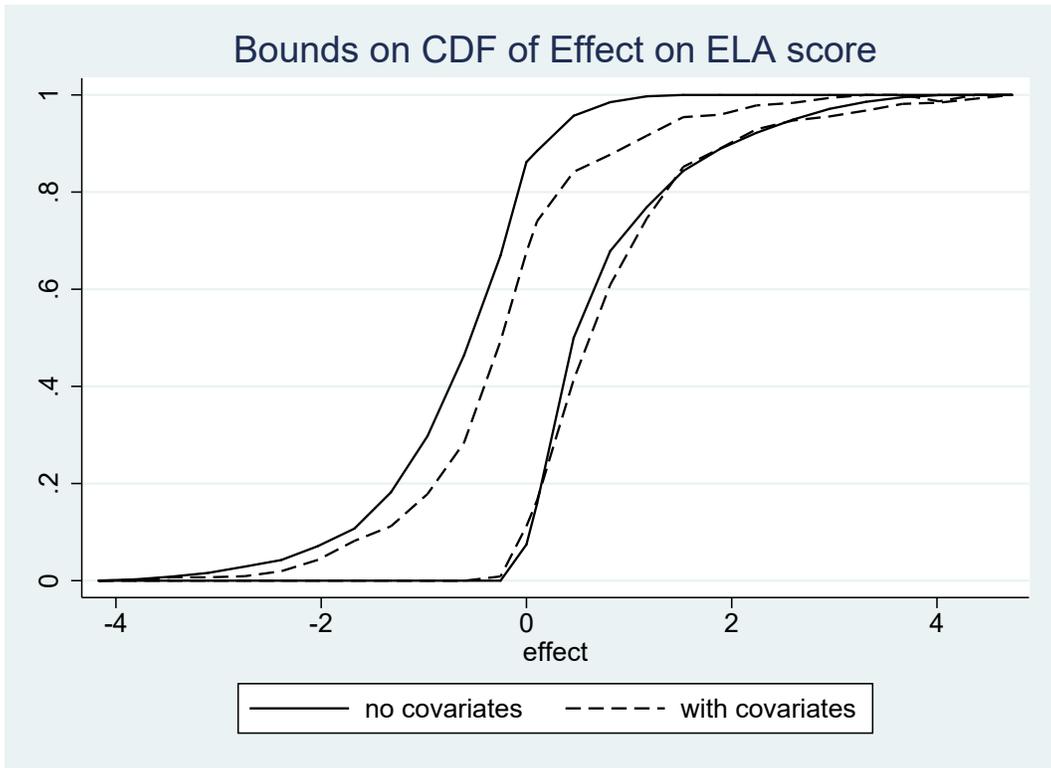


Figure 13: Estimated bounds on the cdf of effects on 7th grade ELA score. The darker region includes no covariates in computing the bounds. The lighter region uses 4th grade ELA score and demographic characteristics described in the text in computing the bounds.

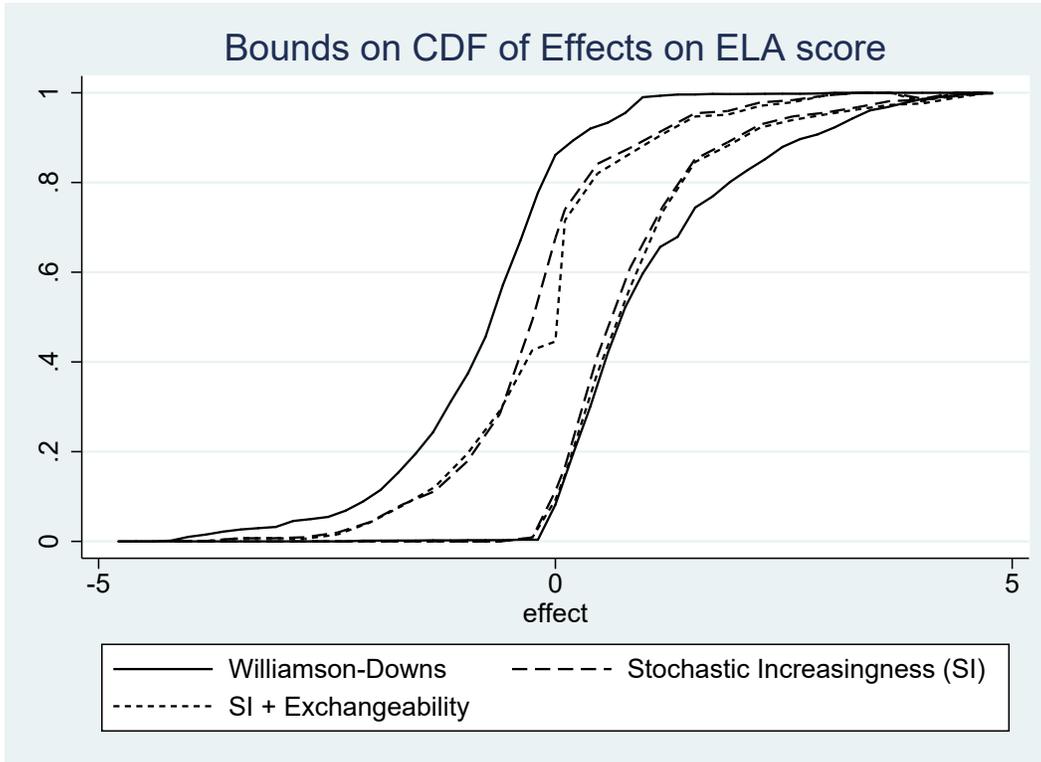


Figure 14: Estimated bounds on the cdf of effects on 7th grade ELA score. The darkest region shows the Williamson-Downs bounds. The medium region shows bounds that impose stochastic increasingness. The lightest region shows bounds that impose stochastic increasingness and exchangeability. All bounds condition on 4th grade ELA score.