

Weak Stochastic Increasingness, Rank Exchangeability, and Partial Identification of The Distribution of Treatment Effects

Brigham R. Frandsen* Lars J. Lefgren*

December 16, 2015

Abstract

This article develops bounds on the distribution of treatment effects under testable assumptions on the joint distribution of potential outcomes, namely weak stochastic increasingness, and rank exchangeability, and shows how to test the empirical restrictions implied by those assumptions. The resulting bounds sharpen the classical bounds based on Fréchet-Hoeffding limits. An empirical application on the impacts of charter schools shows the bounds are informative.

1 Introduction

In experimental data, researchers can identify the effect of treatment on the distribution of outcomes. Consequently, researchers can identify the average treatment effect or effects on quantiles of the outcome distribution. However, experimental data only identifies the marginal distributions, not the joint distribution, of outcomes in the treated and untreated state. As a consequence, the distribution of treatment effects is not identified. This includes parameters such as the fraction of subjects harmed by the treatment and the expected treatment effect conditional upon a subject's outcome in the control distribution.

*Department of Economics, Brigham Young University

These characteristics of the distribution of treatment effects are of significant policy and research interest. In particular, agents may have private information regarding their likely outcome in the absence of treatment. In this case, providing separate estimates of treatment effects for individuals likely to do well and do poorly in the absence of treatment can improve the efficiency of treatment assignment and minimize the probability that individuals are harmed by inappropriate treatment. Also, it is only by exploring how treatment effects vary across the distribution of control outcomes that we can estimate not only the average effect of treatment but also the fraction of individuals that was benefited by the treatment. This is of great importance in that the average benefit of treatment may not be robust to the scaling of the outcome variable. Furthermore, subjects might be rightly wary of treatment if it exposes them to a significant risk of harm, even if the average outcome is improved. From an economic perspective, identifying the distribution of treatment effects sheds insight into whether agents self-select into treatment on the basis of the benefit they are likely to receive.

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. [Fan and Park \(2010\)](#) show how to perform inference on these bounds. 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 size. 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. 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.

When researchers are willing to make this assumption, it is simple to derive bounds on the distribution of treatment effects for each quantile of the control distribution. The lower bound corresponds to the case in which rank invariance holds for all quantiles higher than the reference quantile and rank independence holds for all lower quantiles. The upper bound corresponds to the opposite case. These bounds can be substantially tightened with the use of covariates. Bounds on the aggregate fraction of individuals harmed is obtained by integrating the pointwise bounds across the distribution of control outcomes. These aggregate bounds can be further tightened by imposing conditional rank exchangeability, which can be tested using the method discussed by [Frandsen and Lefgren \(2015\)](#).

We propose a test of unconditional stochastic increasingness that relies on the intuition that under this assumption we would expect observed covariates to move outcomes in the same direction for both treated and control observations. We demonstrate that when covariates have sufficient predictive power for outcomes, our test indicates whether the covariance between treatment and control outcomes is positive—a necessary condition for stochastic increasingness. When covariates lack sufficient

power for this interpretation, our test remains informative in the context of models with a single index of heterogeneity.

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 example that examines the effect of charter school enrollment on student outcomes. 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 a vector of pre-treatment variables S and, in the case of an endogenous treatment, an instrumental variable Z that is independent of potential outcomes. The variables in S 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_Δ ; 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 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 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 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. 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

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

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 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.

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.

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 6 in the appendix) that a sufficient condition for $Cov(Y(1), Y(0)) \geq 0$ is

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

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) can only be satisfied when the covariates S strongly predict potential outcomes: the respective R^2 s between S and each potential outcome must average at least .5 in order for the right-hand side of (13) to be less than one. A practical procedure for verifying this condition is to regress Y_i on S_i in the treated and untreated samples, calculate the correlation coefficient between the predicted values, and compare it to the right hand side of (13).

Verifying condition (13) requires covariates that sufficiently strongly predict outcomes, 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 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 unobservables (Altonji et al., 2013). Furthermore, if potential outcomes are jointly determined by a scalar index of heterogeneity, a positive correlation between $\hat{Y}(0)$ and $\hat{Y}(1)$ implies the potential outcomes themselves are positively correlated.

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 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

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

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 σ_ε^2 from .01 to 1, corresponding to an R^2 between $Y_i(0)$ and S from .99 to zero, and β 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 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 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 δ . 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 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 S . These simulations set the treatment effect δ 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 δ . 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 δ . 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 δ 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: Charter Schools and Student Performance

5 Conclusion

This paper showed how to construct bounds on the distribution of individual-level treatment effects when an individual's potential ranks are each weakly stochastically increasing in the other, and showed how to test the empirical restrictions implied by the assumption. The bounds can be constructed from standard estimates of the conditional distributions of potential outcomes.

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.
- Victor Chernozhukov and Christian Hansen. An IV model of quantile treatment effects. *Econometrica*, 73(1):245–261, January 2005.
- 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. 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.
- Brigham R. Frandsen and Lars J. Lefgren. Testing rank similarity. Unpublished working paper, 2015.
- 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.
- 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

The following result is crucial to Theorem 2:

Lemma 5 *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

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

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

- (1) : $\underline{F}_{X|Y}(x|y) \leq \underline{F}_{X|Y}(x|y')$, $y' < y$,
- (2) : $\underline{F}_{X|Y}(x|y) \geq \underline{F}_{X|Y}(x|y'')$, $y'' \geq y$,

(since X is stochastically increasing in Y) and

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

(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:

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

(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:

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

■

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 5 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

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

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. ■

Theorem 6 *Let $\hat{Y}(0)$ and $\hat{Y}(1)$ be linear projections of potential outcomes on S with corresponding coefficients of determination R_0^2 and R_1^2 . Then $\text{Corr}(\hat{Y}(1), \hat{Y}(0)) \geq \sqrt{(1 - R_0^2)(1 - R_1^2) / (R_0^2 R_1^2)}$ implies $\text{Cov}(Y(1), Y(0)) \geq 0$.*

Proof. Define $\varepsilon(1) = Y(1) - \hat{Y}(1)$ and $\varepsilon(0) = Y(0) - \hat{Y}(0)$. Note that by construction $\text{Cov}(\hat{Y}(1), \varepsilon(0)) = \text{Cov}(\hat{Y}(0), \varepsilon(1)) = 0$. Also, note that $\text{Var}(\varepsilon(1)) = (1 - R_1^2) \text{Var}(Y(1))$ and $\text{Var}(\varepsilon(0)) = (1 - R_0^2) \text{Var}(Y(0))$. 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:

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

The Cauchy-Schwarz inequality implies

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

Inserting this into (15) yields a lower bound on the covariance between potential outcomes:

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

This lower 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,

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

■

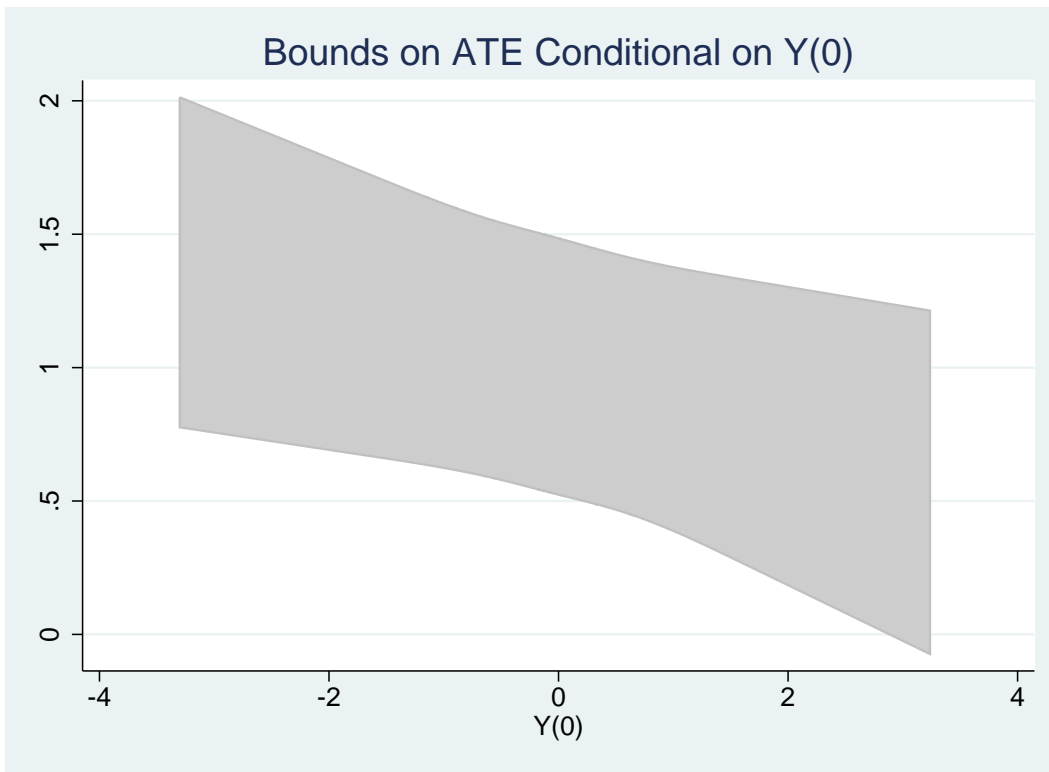


Figure 1: Simulated bounds on the average treatment effect conditional on untreated potential outcome. The true treatment effect is one for all values of $Y(0)$.

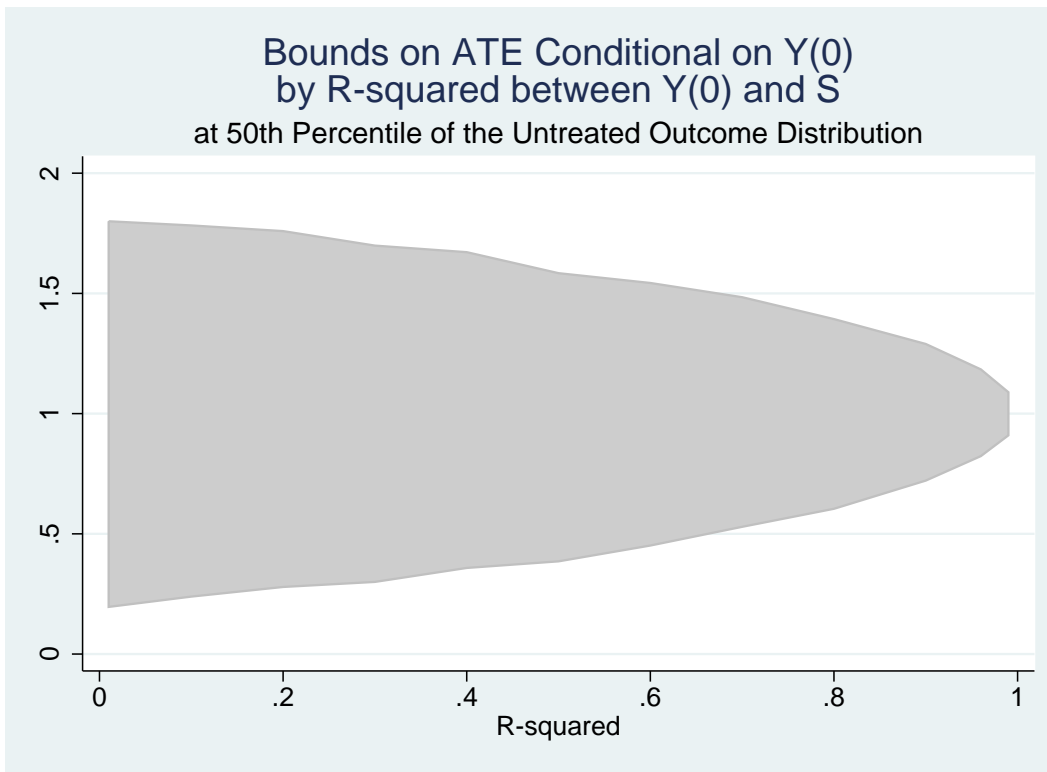


Figure 2: Simulated bound on the average treatment effect conditional on $Y(0) = 0$ as a function of the R^2 between $Y(0)$ and S . The true treatment effect is one.

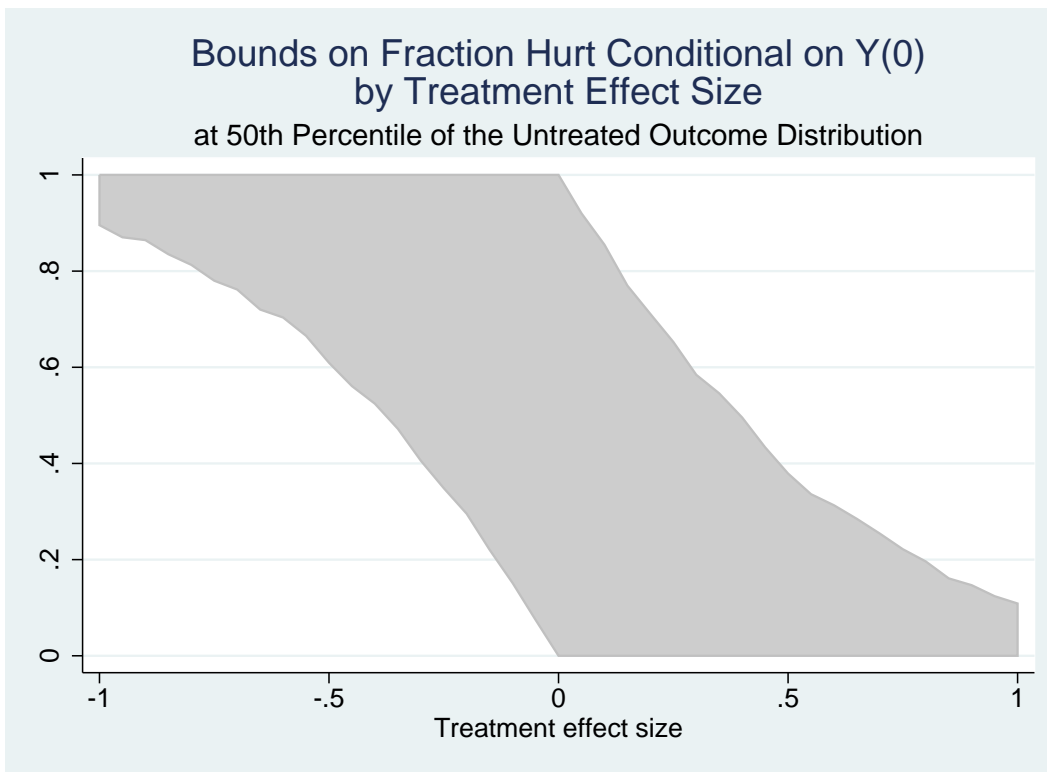


Figure 3: Simulated bound on the fraction hurt by treatment conditional on $Y(0) = 0$ as a function of the treatment effect. The true fraction is one when the treatment effect is negative (left side of the plot) and zero when the treatment effect is positive.

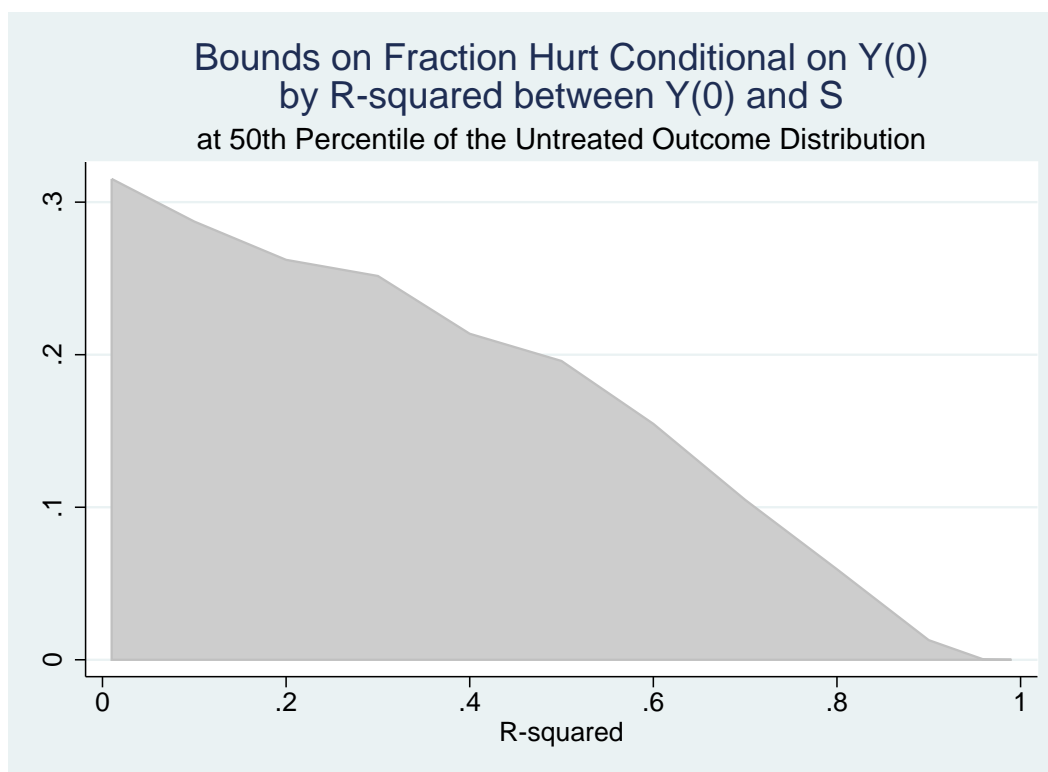


Figure 4: Simulated bound on the fraction hurt by treatment conditional on $Y(0) = 0$ as a function of the R^2 between $Y(0)$ and S . The true fraction in the simulation is zero.

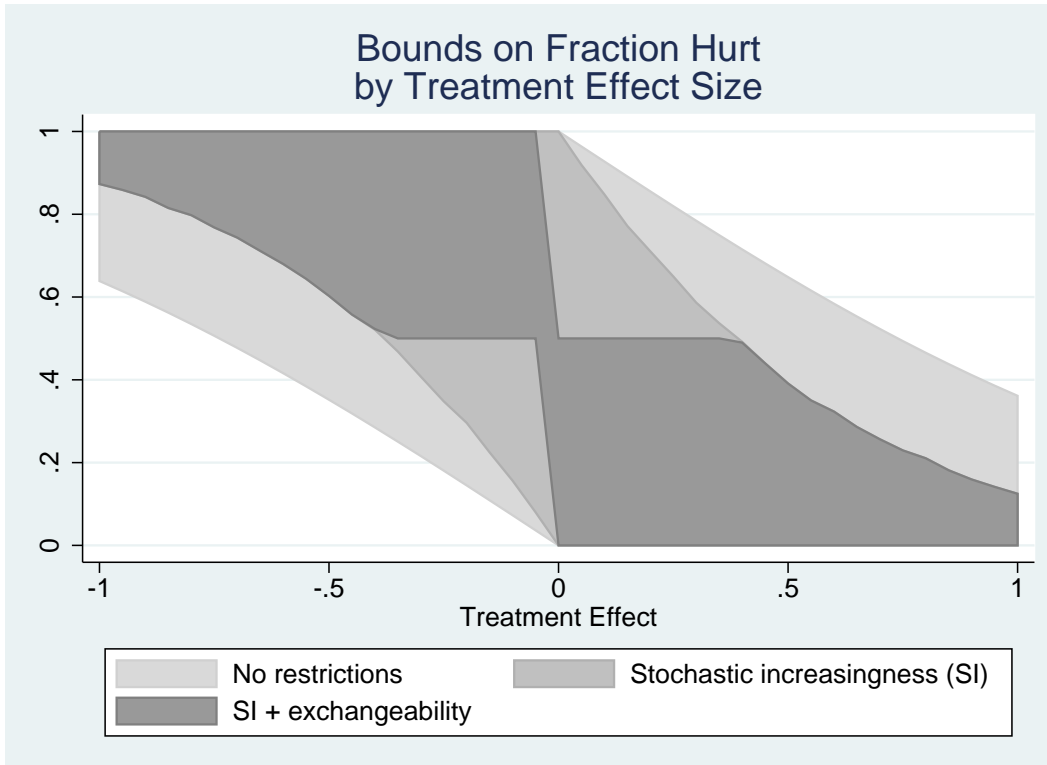


Figure 5: Simulated bound on the fraction hurt by treatment as a function of the treatment effect. The true fraction is one when the treatment effect is negative (left side of the plot) and zero when the treatment effect is positive. The lightest gray bounds impose no restrictions. The medium gray bounds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

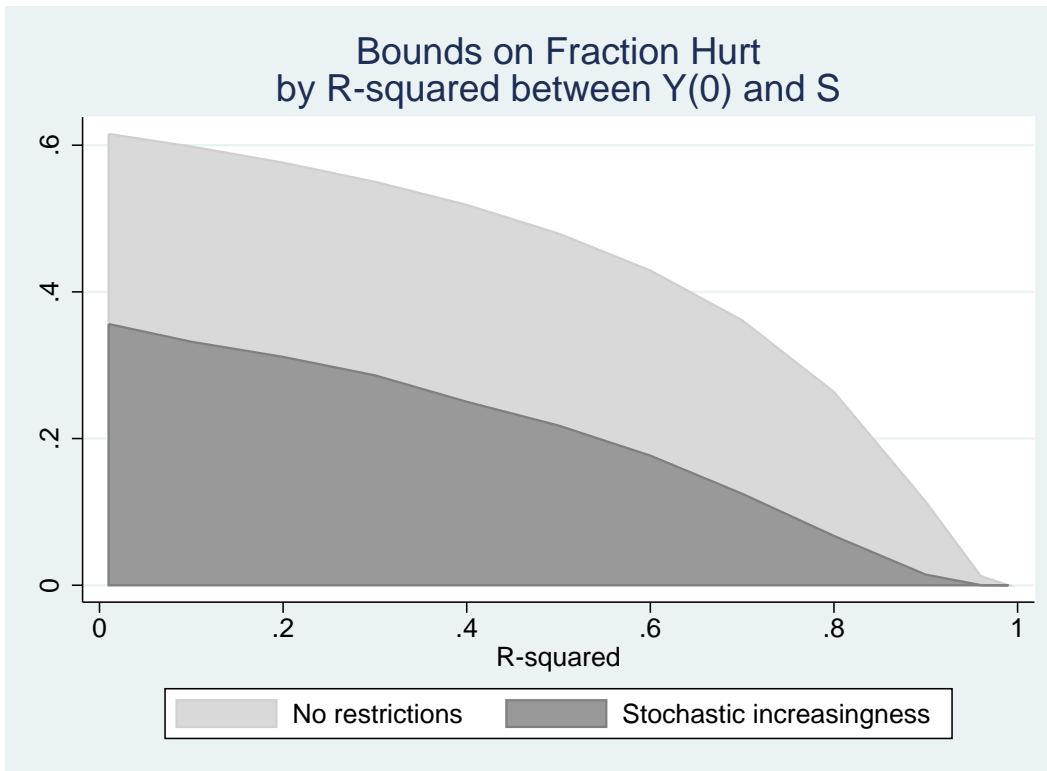


Figure 6: Simulated bound on the fraction hurt by treatment as a function of the R^2 between $Y(0)$ and S . The true fraction is zero. The lightest gray bounds impose no restrictions. The medium gray bounds imposing stochastic increasingness. The darker gray bounds impose stochastic increasingness and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