Bounding Average Treatment Effects using Linear Programming*

Lukáš Laffers†

September 13, 2014

Abstract

This paper presents a method of calculating sharp bounds on the average treatment effect using linear programming under identifying assumptions commonly used in the literature. This new method provides a sensitivity analysis of the identifying assumptions and missing data in an application regarding the effect of parent's schooling on children's schooling. Even a mild departure from identifying assumptions may substantially widen the bounds on average treatment effects. Allowing for a small fraction of the data to be missing also has a large impact on the results.

Keywords: Partial identification; Bounds; Average treatment effect, Sensitivity analysis.

1 Introduction and Literature Review

The recent literature on the average effect of parent's schooling on children's schooling appears inconclusive. Identification strategies based on twins, adoptees or instrumental variables lead to results that differ in size and statistical significance in terms of

*This paper is a chapter from my dissertation at the Norwegian School of Economics. I would like to thank to Monique de Haan for generously providing me with the data used in this paper. I also thank Christian Brinch, Andrew Chesher, Christian Dahl, Gernot Doppelhofer, Charles Manski, Peter Molnar, Adam Rosen, Erik Sorensen, Ivan Sutoris and Alexey Tetenov for valuable feedback.

†Department of Mathematics, Faculty of Natural Sciences, Matej Bel University, Tajovského 40, 97411 Banská Bystrica, Slovakia. E-mail: lukas.laffers@gmail.com, Web: http://www.lukaslaffers.com

the average treatment effect and that lead to conflicting policy recommendations on educational reform. An attempt to address this problem was made in de Haan (2011), who studied the nonparametric bounds on the average treatment effect and relied on weaker nonparametric assumptions that have clear economic interpretations. Nevertheless, these assumptions may and should be challenged. This study discusses the validity and the importance of these assumptions. Moreover, this paper presents a method that allows some assumptions to be relaxed and an examination of how fragile or robust the reported bounds are to some mild violations of these assumptions. This paper also looks at how missing data may affect the results, and it imposes no structure on the missingness mechanism. Knowing what drives the results, and which assumptions are important, may sharpen the discussion about the underlying identifying assumptions, and also that about the economic problem at hand.

The contribution of this paper is twofold. First, this paper presents a flexible way of calculating the sharp bounds on the average treatment effect using a linear program. If all the variables are discrete, it is often practical to achieve identification by conducting a search of the set of joint probability distributions of the observed and unobserved variables. Second, this paper uses the linear programming method to compute the bounds on the average treatment effect when some or all of the identifying assumptions are relaxed, also allowing for the presence of missing data, in the context of the effect of parents' schooling on children's schooling. The linear programming formulation helps to clarify why one presumably irrelevant identifying assumption becomes important once another assumption is relaxed, and therefore, the two assumptions work as substitutes for each other.

There are two opposing explanations of how a parent's schooling affects a child's schooling. One relates to causation and the other to selection. Either the parents change during their education process (and this changes the way that they approach the education of their children) or the child's education merely reflects the transmission of the high-ability genes from his or her parents. An understanding of the intergenerational transmission of education has very important policy consequences. First, policy makers care about the return on investment to schooling. If the link between parents' schooling and children's schooling is causal, the beneficial spillover effect has to be taken into account when devising an educational policy. Second, if the effect is purely related to the transmission of genes, then the inequality in opportunities may simply be a consequence of the distribution of high-ability genes, and inequality-reducing policy is unlikely to be beneficial.

There are three main identifying strategies in the literature for estimating the *causal* effect of parents' education on children's education, as presented in a comprehensive overview in de Haan (2011).

The first approach is based on twins data in Behrman and Rosenzweig (2002, 2005) and Antonovics and Goldberger (2005). Children of identical twins should not differ much in the unobservable genetic endowments that they inherit from their parents, and this helps to remove an important source of correlation between parents' and children's schooling. This approach assumes that any differences other than genetic between the schooling levels of identical twins are exogenous.

The second method is based on adopted children (Björklund et al., 2006), where there clearly is no genetic link between the parents and the adopted children. This method assumes that the way the parents raise their children is unrelated to their schooling level.

The last approach is based on an instrumental variable. The strategy is to find a variable that provides a source of variation in parents' schooling that is unrelated to children's schooling. Black et al. (2005) use a school reform in Norway that changed the number of compulsory years of education from seven to nine. Chevalier (2004) use a law that changed the minimum school leaving age in the 1970s in Britain. Oreopoulos et al. (2006) also use the timing of the compulsory-schooling law changes as an instrument for completed parents' education. College availability is used as an instrument for maternal education in Currie and Moretti (2003) for US data. Carneiro et al. (2013) instruments maternal educational attainment with schooling cost during the mother's adolescence. Maurin and McNally (2008) is based on the series of events in May 1968 that led to the lowering of thresholds in the education system and enabled students to remain longer in the higher education system. Validity of the results from these papers hinges upon the validity and relevance of the instruments in use and may be challenged. It is also known that instrumental variable models only estimate the average treatment effect for a subpopulation of individuals (LATE of Imbens and Angrist (1994)).

The results from all these analyses are mixed. They differ in the size and statistical significance of the potential effect of the intergenerational transmission of human capital. The analysis in Holmlund et al. (2011) compares the three different identification strategies using Swedish data and finds similar patterns to the previous literature. They conclude that the differences follow from the identification, not from the different data sources. These findings stress the importance of the careful inspection of the identification strategy. As a solution to the diverging results, the analysis in de Haan (2011) studies the bounds on the average treatment effect rather than a point-identified model, and the analysis is based on weaker identifying assumptions.[1] This paper will discuss the validity and the importance of these assumptions. The analysis will consider the sensitivity of the results to some mild deviations from the identifying assumptions and to the missing data, and why the sensitivity analysis is relevant.

This paper also contributes to the literature on bounds analysis advocated by Manski (1990, 1995, 1997, 2003, 2007, 2008) by providing a way to conduct a sensitivity analysis. This paper uses the linear programming identification framework presented in Laffers (2013a), which is based on Galichon and Henry (2009). Not only is it possible to determine which assumptions are important and drive the results but also the linear programming formulation helps to *quantify* how sensitive the results are. Note that there are other papers that consider partially identified models using linear programming; most notably, Balke and Pearl (1997, 1994), Honore and Tamer (2006), Manski (2007), Chiburis (2010) and Freyberger and Horowitz (2012). The growing empirical literature that bounds rather than point identifies the average treatment effects includes [TO BE ADDED].

Section 2 introduces the setup and notation, and how an identification problem can be captured within a linear programming framework. Section 3 presents data and results, and a sensitivity analysis on the effect of mothers' schooling on children's schooling follows in Section 4. Section 5 concludes.

[1] One may argue that these assumptions are not weaker, they are just different. "Weaker" means that these assumptions are not strong enough to deliver point identification.

2 Method and Identifying Assumptions

2.1 Notation

Following the notation of Manski (1990), child j from population J has a specific response function $y_j(.)$ that maps the schooling of parent $t \in T$ (a *treatment*) to the child's schooling $y_j(t) \in Y$ (an *outcome*). For every child, we observe the schooling of his or her parent z_j (a *realized treatment*), schooling $y_j \equiv y_j(z_j)$ (a *realized outcome*), and other parent's (or grandparent's) schooling $v_j \in V$ (a *monotone instrument*), but we do not observe the child's schooling $y_j(t)$ for parents' schooling $t \neq z_j$ (a *counterfactual outcome*). The data reveal the probability distribution $P(y,z,v)$ (realized outcomes, realized treatments and instruments), yet the probability distribution of the counterfactual potential outcomes $P(y(t_1),\ldots,y(t_k))$ remains unknown.[2] The goal of the analysis is to uncover some features of the unobserved probability distribution of counterfactual outcomes $P(y(t_1),\ldots,y(t_k))$. The feature of interest may be an expectation of the child's schooling if his or her parents' schooling is equal to t ($E[y(t)]$), or it may be the average treatment effect of the change of parents' schooling from s to t on the child's schooling ($\Delta(s,t) = E[y(t)] - E[y(s)]$).

Under exogenous selection, the average treatment response to treatment t ($E[y(t)]$) is point identified, but this assumption is often not plausible, as discussed later. Depending on the strength of the maintained identifying assumptions, the expectation of children's schooling with parents' education equal to t may be set rather than point identified. There may exist an interval of values for $E[y(t)]$ so that all the values in this interval are compatible with the observed probability distribution $P(y,z,v)$ and with the identifying assumptions.

2.2 Method

The method of obtaining the bounds for average treatment effects follows in this subsection. For a given set of assumptions, instead of analytically deriving the bounds, we translate all the assumptions into restrictions on the joint probability distribution of the unobserved component $(y(t_1), y(t_2), ..., y(t_m))$ and the observed compo-

[2]Formally, the population forms a probability space $(I, \mathcal{F}, \mathcal{P})$, where the population of individuals I is the sample space, $\mathcal{F}$ is a set of events and P is a probability measure. Hence, the only source of randomness is the choice of individual. The individual's behavior is deterministic.

nent (y, z, v). The joint probability distribution carries complete information about the probabilistic behavior of all the variables in the model; there is nothing more that could possibly be learned.

If the outcome is a child's college attendance, and the treatment is the college attendance of a parent, so that it takes two different values (0 - no college, 1 - college), and we are interested in identifying the probability that a child will obtain a college degree if his or her parent has a college education ($E[y(1)]$), we will search in the space of probability distributions of $(y(0), y(1), y, z, v)$ that are compatible with the observed probabilities of (y, z, v), that satisfy all the identifying assumptions, and that minimize (maximize) $E[y(1)]$, which would give the lower (upper) bound. If both the assumptions and the feature of interest are linear in the joint probability distribution $(y(0), y(1), y, z, v)$, then finding a lower or upper bound corresponds to solving one linear program.

The approach of the presented identification scheme follows that of Galichon and Henry (2009) and Ekeland et al. (2010), which was further extended in Laffers (2013a).

The linear programming method presented in this paper offers flexible identification. It is easy to add, remove or change assumptions. This paper will use this method to explore how sensitive the bounds are to some mild violations of the identifying assumptions.

The following subsections discuss how different identifying assumptions translate into restrictions on the joint probability distribution $(y(0), y(1), y, z, v)$ in the light of the following specific example from de Haan (2011).

- $y_i \in Y = \{0, 1\}$ - child's college (0 - no college, 1 - college).
- $z_i \in Z = \{0, 1\}$ - mother's (father's) college (0 - no college, 1 - college).
- $v_i \in V = \{1, 2, 3, 4\}$ - other parent's (grandparent's) schooling level (high school or less ($\leq$ 12 years), some college (13–15 years), bachelor's degree (16 years), master's degree or more ($\geq$ 17 years)).

The aim is to learn about the average treatment effect of an increase in mother's college attendance on a child's college attendance ($\Delta(0, 1) = E[y(1)] - E[y(0)]$).

2.3 Identifying assumptions

This subsection explains how the linear program whose extremes are the bounds on the ATE is created. The presentation of identifying assumptions begins with a discussion of how the unobserved component $(y(0), y(1))$ must be linked to the observed component (y, z, v), and it is called the **correct specification**. The marginal distribution of the joint probability distribution of $(y(0), y(1), y, z, v)$ must be the probability distribution of the observed component, and this is called **compatibility with observed probabilities**. Furthermore, the **monotone treatment response**, the **monotone treatment selection**, the **conditional monotone treatment selection** and the **monotone instrumental variable** assumptions are presented and explained. The figures associated with these assumptions elucidate how they translate into restrictions on the joint probability distribution $(y(0), y(1), y, z, v)$.

Correct Specification

The observed component (y, z, v) has to be compatible with the unobserved component $(y(0), y(1))$; that is, they are linked by $\forall j : z_j = t \implies y_j(t) = y_j$. If this assumption fails, it means that either the child's schooling level or the mother's schooling level is not correctly measured or that child j's schooling is not determined by mother's education.[3]

Figure 1 depicts the support of the joint probability distribution of $(y(0), y(1), y, z, v) \in Y^3 \times T \times V$. Every point in the figure represents a subpopulation of individuals. The circle denotes children with a college degree ($y = 1$), with a college-educated mother ($z = 1$), and with a grandparent with a high school education ($v = 1$). The unobserved counterfactual outcomes for these children are $y(0) = 1$ and $y(1) = 1$. The observed component $(y = 1, z = 1, v = 1)$ implies that $y(1) = 1$, which is compatible with the counterfactual outcomes. For a child that belongs to the subpopulation denoted by a triangle, the observed component $(y = 0, z = 1, v = 1)$ implies that $y(1) = 0$. At the same time, the unobserved counterfactual outcomes are $y(0) = 0$ and $y(1) = 1$, and therefore, not compatible with the observed component. There must be no such

[3] The assumption that outcome is a deterministic function of a treatment is intrinsic in the potential outcome framework of Rubin (1974).

children, and the probability of the point $(0,1,0,1,1)$ must be equal to zero. Figure 2 shows all the points that can be assigned nonzero probabilities.

Compatibility with Observed Probabilities

The joint probability distribution of $(y(0), y(1), y, z, v)$ must be compatible with the observed probabilities of (y, z, v). In the data, 39.6% of the children do not have college education ($y = 0$), their mother has a college degree ($z = 1$), and their father has a high school education ($v = 1$) so that the probabilities in the column of $(0,1,1)$ sum to 0.396 as depicted in Figure 3.

Monotone Treatment Response

There seems to be a consensus that a child's schooling does not decrease with mother's schooling. The monotone treatment response (MTR) assumption (Manski, 1997) interprets this statement such that for every child, the schooling level is an increasing function of mother's schooling, specifically $\forall j, t_2 \geq t_1 : y_j(t_2) \geq y_j(t_1)$. The MTR assumption is a strong assumption and guarantees that the average treatment effect is nonnegative.

The MTR assumption rules out all the rows of unobservables for which $y(1) \geq y(0)$ does not hold (that is if $(y(0), y(1)) = (1,0)$) as shown in Figure 4. Given the MTR assumption, there must exist no children who would obtain a college degree if their mother had not, and who would not finish college if their mother had finished college.

Monotone Treatment Selection

The assumption of monotone treatment selection (MTS) (Manski and Pepper, 2000) provides another interpretation of how a child's schooling increases with mother's schooling. Instead of assuming the selection bias away by imposing exogenous treatment selection ($\forall t_1, t_2 : E[y(t)|z = t_1] = E[y(t)|z = t_2]$) that delivers point identification, the MTS assumption restricts the direction of the selection bias.[4] The MTS assumption states that for a fixed potential mother's college attendance, children with observed college-educated parents have a weakly higher probability of graduating

[4] Ordinary least squares regression analysis assumes ETS, and it point identifies the average potential outcome: $E[y(t)] = E[y(t)|z = t]P(z = t) + E[y(t)|z \neq t]P(z \neq t) = E[y(t)|z = t]$.

from college. That is, the probability that a child with a college-educated mother obtains a college degree ($E[y(1)|z = 1]$) is higher than the *potential* probability of a child with a mother without a college degree if (counterfactually) this mother had a college education ($E[y(1)|z = 0]$). Moreover, the probability that a child with a less-educated mother finishes college ($E[y(0)|z = 0]$) is not as high as it would be for a child with a more-educated mother if (counterfactually) this mother does not have a college education ($E[y(0)|z = 1]$). The differences in these probabilities may stem from the fact that higher-educated parents tend to have higher abilities, and these can be transmitted to their children, and that these parents with higher abilities create a more stimulating environment for their children. Formally, the MTS assumption is $\forall t_2 \geq t_1 : E[y(t)|z = t_2] \geq E[y(t)|z = t_1]$.

The MTS assumption restricts the space of the joint probability distribution functions of $(y(0), y(1), y, z, v)$ to those that are compatible with the corresponding set of linear constraints. Figure 5 shows that the probability of graduating from a college if the mother's school attainment is equal to t conditional on her having a college degree ($E[y(t)|z = 1]$, which is calculated using the probabilities in solid rectangles) is greater than or equal to the probability conditional on her not having a college degree ($E[y(t)|z = 0]$, which is calculated using the probabilities in dashed rectangles). The MTS assumption states that given a mother's schooling, any difference in unobserved characteristics between college-educated and non-college-educated mothers does not make a child's probability of graduating from a college lower than that of children with higher-educated mothers.

Conditional Monotone Treatment Selection

The conditional monotone treatment selection (cMTS) assumption, formally $\forall i, t_2 \geq t_1 : E[y(t)|z = t_2, v = i] \geq E[y(t)|z = t_1, v = i]$, also states that a child's potential probability of getting into college increases with the mother's education but conditional on (and hence regardless of) the father's (or grandparent's) schooling level. The father's (or grandparent's) education is, therefore, restricted to have no impact on the direction of the selection bias due to mother's education.

Restricting the space of the joint probability distribution functions is similar to the MTS assumption, with the conditioning on events $[z = t, v = i]$ instead of $[z = t]$. Figure 6 illustrates the effect of the cMTS assumption for a subpopulation with $v = 1$.

The difference between the MTS and the cMTS assumption is explained in Laffers (2013c). The distinction is similar to whether or not to include father's education into a regression as an explanatory variable as discussed in Holmlund et al. (2011). The inclusion (similar to the cMTS assumption) would imply that the effect of mother's schooling is net of assortative mating effects. On the other hand, not including father's schooling as an explanatory variable (similar to the MTS assumption) means capturing both direct effects of mother's education and indirect effects of assortative mating. As was pointed out in Laffers (2013c), when considering higher-educated mothers that "married down" to less-educated men, we have to consider any observed or unobserved factors that made these mothers self-select into such marriages. These mothers might have compensated for unobserved low ability, or the cost of finding a partner might have been high, which is true especially for older women (Lichter, 1990), and children of older women have lower cognitive skills on average (Zybert et al., 1978).

Monotone Instrumental Variable

The monotone instrumental variable (MIV) assumption (Manski and Pepper, 2000) is a weakened version of the instrumental variable assumption ($\forall i_1, i_2 : E[y(t)|v = i_1] = E[y(t)|v = i_2]$). It ensures that a child's mean potential schooling is weakly increasing in its grandparent's schooling. The MTS assumption is, in fact, a special case of the MIV assumption.

The restrictions on $(y(0), y(1), y, z, v)$ implied by the MIV assumption work in a similar way as for MTS as depicted in Figure 7. Given the mother's college attainment, a child's probability of graduating from college is greater for children with higher-educated grandparents.

The average treatment effect is a linear function of the joint probability distribution of $(y(0), y(1), y, z, v)$. To find the upper and lower bounds on ATE, we conduct a search in the joint probability distributions that maximizes and minimizes the average treatment effect under linear identifying constraints, which is a linear program.

The resulting bounds on ATE are sharp by construction, and the identifying assumptions translate one-to-one to restrictions on the joint probability distribution; therefore, there is no information gain or loss. If there is no joint probability distribution that satisfies the constraints imposed by the identifying assumptions and is compatible with the data, then the linear program has no feasible solution, and the model can be refuted. The linear program that leads to the upper bound on ATE of an increase in mother's college education on the probability that the child finishes college is depicted in Figure 8, and the joint probability distribution that maximizes the ATE under the MTR+cMTS+MIV assumption is shown in Figure 9. Lemma 1 shows that if the identification problem takes the form of a linear program, then the identified set is an interval between the lower and upper bound.[5] The average treatment effect is a linear function of the joint probability distribution of $(y(0), y(1), y, z, v)$.

Lemma 1. *The identified set for the ATE is an interval.*

Proof. Let p denote the probability vector of the observed variables. Let $\Pi(p)$ denote the set of all joint probability distributions of the observed and unobserved components that are compatible with p and with the identifying assumptions, and let $ATE(\pi)$ be the average treatment effect when the joint probability distribution is π. Furthermore, let $ub(p) \equiv \max_{\pi\in\Pi(p)} ATE(\pi)$ be the upper bound, and let $lb(p) \equiv \max_{\pi\in\Pi(p)} ATE(\pi)$ be the lower bound on ATE under the set of identifying assumptions.

Consider a nontrivial case where $lb(p)$ and $ub(p)$ exist and $lb(p) \neq ub(p)$. It is sufficient to show that $\forall a \in (lb(p), ub(p)) : \exists \pi \in \Pi(p) : ATE(\pi) = a$. For every a, there must exist γ so that $a = \gamma lb(p) + (1-\gamma)ub(p)$. Let $\pi_{ub} = \arg\max_{\pi\in\Pi(p)} ATE(\pi)$ and $\pi_{lb} = \arg\min_{\pi\in\Pi(p)} ATE(\pi)$ denote the joint probability distributions that maximize and minimize the ATE, respectively. For $\pi_a = \gamma\pi_{ub} + (1-\gamma)\pi_{lb}$, it must hold that $\pi_a \in \Pi(p)$, because $\Pi(p)$ is defined as a set of vectors that satisfy a finite number of linear equalities and inequalities. Finally, $ATE(\pi_a) = ATE(\gamma\pi_{ub} + (1-\gamma)\pi_{lb}) = \gamma ATE(\pi_{ub}) + (1-\gamma)ATE(\pi_{lb}) = a$ because $ATE(\pi)$ is a linear function, and this completes the proof. □

The identified interval is finite, because the feasible set is bounded.

[5] The proof is very similar to that in Freyberger and Horowitz (2012).

Manski (1990, 1995, 2003) study the bounds on $E[y(t)]$ and $\Delta(s,t)$ under various combinations of these assumptions (apart from cMTS), and de Haan (2011) explains these assumptions in great detail in the context of the presented schooling application.

3 Data and Results

3.1 Data

The Wisconsin Longitudinal Study (WLS) involves a random sample of 10317 high school graduates in Wisconsin in 1957.[6] WLS also collects information from parents, spouses, and siblings of the original graduates. Similarly to de Haan (2011), this paper uses the data from the most recent surveys (2004: original respondents or their parents, 2005: siblings, 2006: spouses) and restricts the sample to the parents that have children from their first marriage, because spouses are not linked to children. Children that might still be at school (1.5%) are eliminated from the sample. Overall, the data consist of information on 21545 children.

3.2 Results

This paper employs the 90% confidence sets based on the bias-corrected bootstrap method of Imbens and Manski (2004), which considers the situation where the aim is to cover the unknown parameter with a fixed probability asymptotically.[7] The confidence sets are based on 500 bootstrap replications. Different statistical inference schemes, when the identified set follows from a linear programming formulation, are compared in Laffers (2013b). There is no clear winner, but the method that is used here performed well in most scenarios.

Our discussion of the results begin with the bounds on the effect of an increase in a mother's (father's) education on the probability that the child has a college degree. Table 1 presents the bounds for two different monotone instruments: other parent's and grandparent's schooling level, under different sets of identifying assumptions. The no-assumption bounds are not very informative, and the length of the identified

[6] Available at http://www.ssc.wisc.edu/wlsresearch/.

[7] The confidence sets that cover the whole identified set asymptotically are generally larger and may be preferable for a policy maker concerned with robust decisions as is argued in Henry and Onatski (2012).

interval is equal to one. The MTR assumption only affects the lower bound and sets it equal to zero with the exception that the effect of a father's attendance at college increases when the grandparent's schooling level is used as the monotone instrument, but the lower bound is not significantly different from zero under the 90% confidence level. The MTS assumption reduces the upper bound from 64.1% to 36.5% for the mother's college education and from 68.1% to 39.3% for the father's college education. The cMTS is much stronger for an increase in the mother's college attendance than for an increase in the father's, and it reduces the upper bound on the probability that the child obtains a college degree to 21.4% compared with 37.2% for the father's. The MIV only slightly affects the lower bound for the father's college attendance and has no effect on the upper bound. The monotone instrument affects the upper bound indirectly, via conditioning when the cMTS assumption is assumed. None of the sets of assumptions yields a lower bound significantly different from zero.

Table 2 shows the bounds on the effect of a parent's college degree on a child's years of finished schooling. No assumption bounds are not informative. The results show a similar pattern for both mother's and father's college as treatments and other parent's or grandparent's schooling level as monotone instruments. The MTR assumption increases the lower bound to zero. The MTS assumption reduces the upper bound from 10.8 years to 1.8 years when mother's college is a treatment and from 11.6 years to 1.9 years for father's college. The MIV assumption affects the upper bound only in connection with the cMTS assumption and if grandparent's schooling is used as the monotone instrument. Other parent's schooling level has greater identifying power than grandparent's schooling, and the resulting bounds are narrower. Finally, under the MTR+cMTS+MIV assumption, the effect of mother's education on child's years of completed schooling is between zero and 1.08 years or 1.52 years, respectively, when father's and grandparent's schooling level is used as the monotone instrument. The effect of father's college degree increases child's schooling by 0 to 1.43 years if mother's education is used as the MIV and 0.008 years (three days) to 1.7 years with grandparent's schooling level as the MIV. The lower bound is not statistically significant.

4 Sensitivity Analysis

This section studies the sensitivity of the results to relaxed identifying assumptions. The flexibility of the linear programming identification framework allows this in a straightforward manner. The identifying assumptions are relaxed in the following ways.

- Mismeasurement of Outcomes or Treatments (MOT):
  $P[z_j = t \Rightarrow y_j = y_j(t)] \geq 1 - \alpha_{MOT}$.
- Relaxed monotone treatment response (rMTR):
  $P[t_2 \geq t_1 \Rightarrow y_j(t_2) \geq y_j(t_1)] \geq 1 - \alpha_{MTR}$.
- Relaxed monotone treatment selection (rcMTS):
  $\forall z_2 \geq z_1 : E[y(t)|z = z_1] - E[y(t)|z = z_2] \leq \alpha_{cMTS}$.
- Relaxed monotone instrumental variable (rMIV):
  $\forall v_2 \geq v_1 : E[y(t)|v = v_1] - E[y(t)|v = v_2] \leq \alpha_{MIV}$.
- Missing data (MISS): at most α_{MISS}-fraction of the sample is not observed, and nothing is assumed about the nature of the missingness.

Mismeasurement of outcomes or treatments (MOT) says that for α_{MOT} fraction of the population, observed outcome y_i may not be equal to the outcome of the actual treatment z_i, either because y_i or z_i is mismeasured or because individual i's outcome is not a deterministic function of the treatment. As the data were collected mostly via phone interviews, it is reasonable to expect that some entries were not recorded correctly, although the probability of mismeasurement is likely to be low. The joint distribution that maximizes the upper bound on the ATE of mother's college degree on child's college completion under the MTR+cMTS+MIV assumption with the MOT relaxed by $\alpha_{MOT} = 0.001$ is shown in Figure 10.

The assumption of **relaxed monotone treatment response** (rMTR) states that α_{MTR} proportion of the population is allowed to have the outcome function that is not monotone in the treatment. The assumption that children's education is weakly increasing in mother's education is consistent with a wide range of studies. Behrman and Rosenzweig (2002) suggest that one possible channel that works in the other direction is that a more educated woman spends less time with her children.[8] The results in the literature deal with the *average* response to mother's schooling; however, it is not unreasonable to think of a small proportion of children whose schooling would not be increasing in mother's schooling. Figure 11 shows how relaxing the MTR assumption by $\alpha_{MTR} = 0.01$ allows up to 1% of children to respond negatively to the treatment: mother's college degree.

Relaxed conditional monotone treatment selection (rcMTS) says that the difference in mean potential outcomes between subpopulations with lower and higher observed treatments cannot be larger than δ_{cMTS} when conditioning on a value of the monotone instrument. An argument that goes against this assumption is that the outcome (child's college degree) only reflects the benefits and does not consider the cost of finishing college for the mother. A mother's college degree is an investment. If the cost of studying is very high, it may be optimal for the future mother to give up college education, and she may eventually earn more and be able to support the child's education better.

Similarly, **relaxed monotone instrumental variable** (rMIV) states that the difference in mean potential outcomes between subpopulations with lower and higher instrument values cannot be larger than δ_{MIV}.

So far, all the relaxed assumptions are straightforward modifications of the original assumptions and still linear in the joint probability distribution. This is not the case when considering the **missing data**. Even though the survey's responsiveness' rates are very good, around 90%, the fact that the data are not missing-at-random may lead to potential problems. Hauser (2005) argues that there is a systematic non-responsiveness in the studied dataset and that the missingness mechanism therefore cannot be ignored. We remain agnostic about the actual process that drives the missingness. Let $\mathcal{P}$ denote the space of all probability distributions of observed variables. If no assumptions are made about the missing data, the probability distribution of the missing component p_{MISS} can be any element in $\mathcal{P}$. The data reveals $\hat{p}_n \in \mathcal{P}$, where n is the sample size. Let α_{MISS} be the fraction of the missing part, and let $\mathcal{P}_{MISS}$ be the

[8]This analysis was challenged by Antonovics and Goldberger (2005), who claim that their results are driven by a specific data coding. In a reply, Behrman and Rosenzweig (2005) argue that their story is supported by an additional data source.

space of all probability vectors that are convex combinations of the data component $\hat{p}_n$ and arbitrary probability vector of the missing component p_{MISS}.

$$\mathcal{P}_{MISS} = \{(1 - \alpha_{MISS})\hat{p}_n + \alpha_{MISS} p_{MISS} | p_{MISS} \in \mathcal{P}\} \quad (1)$$

To find bounds on ATE under the MISS assumption, it is necessary to calculate the minimum and the maximum ATE across all probability vectors in $\mathcal{P}_{MISS}$. The linear program takes the vector of observables p from $\mathcal{P}_{MISS}$ as fixed. The outer loop is an optimization in $\mathcal{P}_{MISS}$, which is a convex set. Note that there are different ways to model the missing data. Here the interpretation is that α_{MISS} proportion of the data is missing. No assumptions are made about the missing subpopulation separately; the identifying assumptions must hold for the whole population.[9] The following lemma states that the identified set is an interval under the MISS assumption.

Lemma 2. *Suppose that the matrices and the vector that define the equalities and the inequalities in $\Pi(p)$ are continuous in p element-wise. Then the identified set for the ATE under the missing data assumption is an interval.*

Proof. This proof uses the notation from the proof of Lemma 1. Further define $p_{\max} = \arg\max_{p \in \mathcal{P}_{MISS}} ub(p)$ and $p_{\min} = \arg\min_{p \in \mathcal{P}_{MISS}} lb(p)$. It is sufficient to show that $\forall a \in (lb(p_{\min}), ub(p_{\max}))\ \exists p \in \mathcal{P}_{\mathcal{MISS}} : \exists \pi \in \Pi(p) : ATE(p) = a$.

Firstly, note that $\mathcal{P}_{MISS}$ defined in equation 1 is a convex set. Consider any $p_1, p_2 \in \mathcal{P}_{MISS}$. From 1, there exist p_1^M and p_2^M such that $p_1 = (1 - \alpha_{MISS})\hat{p}_n + \alpha_{MISS} p_1^M$ and $p_2 = (1 - \alpha_{MISS})\hat{p}_n + \alpha_{MISS} p_2^M$. For any $0 < \lambda < 1$, it must hold that $\lambda p_1 + (1 - \lambda)p_2 = (1 - \alpha_{MISS})\hat{p}_n + \alpha_{MISS}(\lambda p_1^M + (1 - \lambda)p_2^M) \in \mathcal{P}_{MISS}$ as $\lambda p_1^M + (1 - \lambda)p_2^M \in \mathcal{P}$.

Secondly, Theorem 1.1 in Martin (1975) shows that $ub(p)$ (and $lb(p)$) is a continuous function of p on $\mathcal{P}_{MISS}$.

Finally, by virtue of the Intermediate Value Theorem (Munkres, 2000), the image set $ub(\mathcal{P}_{MISS})$ must contain the interval $(ub(\hat{p}_n), ub(p_{\max}))$, and the image set $lb(\mathcal{P}_{MISS})$ must contain $(lb(p_{\min}), lb(\hat{p}_n))$, and this, together with Lemma 1, completes the proof.

□

[9] Note that the nature of some identifying assumptions (e.g., the MTR assumption) are such that they must also hold for every subpopulation.

All sets of assumptions used in this paper satisfy the assumption of the continuity of the equalities and of the inequalities that define the set of feasible joint probability distributions of $(y(0), y(1), y, z, v)$.

We will now look more closely at the effect of the increase in mother's college education on the probability that the child has a college degree with father's schooling level as a monotone instrument for the sake of simplicity. The results with child's years of schooling as the treatment are qualitatively similar, and the average treatment effect has an appealing interpretation of a probability increase that a child has a college degree. Figure 12 illustrates the sensitivity of the bounds to different deviations from the MOT, MTR, cMTS and MIV assumptions.

Relaxing the (**MOT**) leads to the lower bound under the MTR, and the MTR+cMTS +MIV assumption remains at zero. The lower bound under the MIV assumption is linear in the relaxation parameter α_{MOT}. The upper bound under the benchmark MTR+cMTS+MIV assumption jumps from 21% to 35% when 1% of the outcomes are allowed to be mismeasured. The shape of the upper bound curve is convex. The already large upper bounds under the MTR assumption and under the MIV assumption do not respond to α_{MOT} as steeply. It seems that the stronger assumptions make the results more fragile to mild deviation from MOT.

The **MTR** assumption does not affect the upper bounds on ATE at all. The lower bound shows the same linear pattern for all studied models. This is not surprising because allowing 1% of children to respond negatively to mother's college increase cannot lead to an ATE smaller than 1%. Deviation from the **cMTS** assumption only affects the upper bound on ATE and in an exactly linear way.

The **MIV** assumption itself has weak identifying power and only affects the upper bound. If the potential probability that a child gets a college degree is not greater than 2% for children with less-educated fathers ($\alpha_{MIV} = 0.02$), then this assumption is irrelevant, and the upper bound increases to the no-assumption bound. The upper MTR+cMTS+MIV bound is not affected at all.

Figure 13 shows how the results are sensitive to **missing data**. The lower bound stays at zero if the MTR assumption is made. Under the MIV assumption, the lower bound is linear in the proportion of missing observations. The upper bound under the MTR+MIV assumption and under the MIV assumption is similar and is linearly increasing in α_{MISS}. The upper bound under the benchmark MTR+cMTS+MIV assumption gets less sensitive with increases in the amount of missing data, and the shape of this sensitivity curve is convex as it was for the MOT assumption.

So far, this analysis has considered the different relaxations one by one. Two different scenarios illuminate how the identifying assumptions interact. In the first, "optimistic", scenario, the assumptions are relaxed slightly. It is assumed that 1% of children may respond to mother's college negatively, that for up to 0.1% of children, the data on mother's or child's college attendance may be mismeasured and also that the potential probability of a child's getting a college degree cannot be greater by more than 1% for a child with a lower-educated mother (cMTS) and father (MIV). Such relaxations lead to bounds on the effect of mother's college on child's college from -1% to 24.36% as shown in Table 3. Adding the assumption that 1% of the data are missing shifts the upper bound to 28.62%. It is apparent that the missing data assumption is the most important determinant of the change in the upper bound. Assuming that 1% of the data are missing, the additional relaxations only change the upper bound from 27.31% to 28.62%.

Considering the more realistic ("pessimistic") scenario with 5% of children potentially responding negatively to mother's college increase, 1% of mismeasured data, 5% relaxation of the cMTS assumption and the MIV assumption, the effect is between -5% and 44.1%, so that the upper bound more than doubles from 21.44%, which is the upper bound for the benchmark specification. Adding that up to 10% of the data may be missing, which is the actual rate of survey responsiveness, the upper bound jumps to 53.27%.

This paragraph looks more closely at the last interesting result that the MIV does not affect the bounds if the cMTS assumption is made. The linear program formulation allows us to inspect which assumptions are most important by examining the values of the Lagrange multipliers corresponding to the identifying assumptions. Figure 14 shows the Lagrange multipliers that correspond to the linear restrictions that the cMTS assumption and the MIV assumption induce on the joint probability distribution. The cMTS multipliers sum to one, so these numbers also show the relative importance.[10] The cMTS with $v = 1$—that is, for the subpopulation of children with

[10]Figure 12 and Table 4 show that relaxation of α_{cMTS} translates to the upper bound one by one.

high-school educated fathers—drives the result most, and it accounts for 58.08% of the change in the upper bound. In the situation where the cMTS assumption holds only for children with fathers that have at least some college education ($v \geq 2$), the MIV assumption actually matters. Table 4 shows that the MIV assumption has a big impact on the upper bound by shrinking it from 46.71% to 27.54%. The Lagrange multipliers provide some insight into the source of the identifying power. Figure 15 indicates that the MIV restriction, which says that the potential probability of getting a college degree for a father with some college ($v = 2$) is greater than that of a child with a high-school educated father ($v = 1$) if their mother had a college degree ($E[y(1)|v = 2] \geq E[y(1)|v = 1]$), now takes the role of the omitted cMTS for children with less-educated fathers with the value of Lagrange multiplier of 0.582. Therefore, not only is it possible to see that the MIV is now important but also this highlights which part of the MIV assumption is relevant. The reason for this is that once the cMTS is not assumed for children with lower-educated fathers ($v = 1$), nothing is assumed about this large proportion of data, 58.21%, which is exactly the value of the Lagrange multiplier for the part of the MIV assumption that binds. Therefore, in this situation, the cMTS and the MIV assumptions are substitutes for each other.

5 Conclusion

de Haan (2011) provides a novel attempt to address an identification problem in the context of intergenerational transmission of education. The minimal identifying assumptions that she imposes do not deliver point identification, yet the bounds on the treatment effects are still informative. This paper has presented a method for finding sharp bounds on the average treatment effect via linear programming and has then used this method to show how sensitive the bounds are to mild violations of the identifying assumptions. The sensitivity analysis provides insights into the determinants of the identification. The bounds on ATE are very sensitive to missing data and possible mismeasurement of treatments or outcomes. Realistic relaxations of identifying assumptions double the upper bound on the effect of mother's college increase on the probability that a child finishes college.

The findings in this paper stress the importance of discussing the identification assumptions in great detail. Special care should be exercised with the assumptions with the greatest identifying power, and this paper has presented a method of identifying and analyzing them.

References

ANTONOVICS, K. L. AND A. S. GOLDBERGER (2005): "Does Increasing Women's Schooling Raise the Schooling of the Next Generation? Comment," *American Economic Review*, 95, 1738–1744.

BALKE, A. AND J. PEARL (1994): "Counterfactual Probabilities: Computational Methods, Bounds, and Applications," in *Uncertainty in Artificial Intelligence 10*, ed. by L. R. de Mantaras and D. Poole, Morgan Kaufmann, 46–54.

——— (1997): "Bounds on treatment effects from studies with imperfect compliance," *Journal of the American Statistical Association*, 439, 1172–1176.

BEHRMAN, J. R. AND M. R. ROSENZWEIG (2002): "Does Increasing Women's Schooling Raise the Schooling of the Next Generation?" *American Economic Review*, 92, 323–334.

——— (2005): "Does Increasing Women's Schooling Raise the Schooling of the Next Generation? Reply," *American Economic Review*, 95, 1745–1751.

BJÖRKLUND, A., M. LINDAHL, AND E. PLUG (2006): "The origins of intergenerational associations: Lessons from Swedish adoption data," *The Quarterly Journal of Economics*, 121, 999–1028.

BLACK, S., P. DEVEREUX, AND K. SALVANES (2005): "Why the Apple DoesnÂ't Fall Far: Understanding Intergenerational Transmission of Human Capital," *American Economic Review*, 95, 437–42.

CARNEIRO, P., C. MEGHIR, AND M. PAREY (2013): "Maternal education, home environments, and the development of children and adolescents," *Journal of the European Economic Association*, 11, 123–160.

CHEVALIER, A. (2004): "Parental education and child's education: A natural experiment," IZA discussion paper.

CHIBURIS, R. C. (2010): "Bounds on Treatment Effects Using Many Types of Monotonicity," Working paper, Department of Economics, University of Texas at Austin.

Currie, J. and E. Moretti (2003): "Mother's education and the intergenerational transmission of human capital: Evidence from college openings," *The Quarterly Journal of Economics*, 118, 1495–1532.

de Haan, M. (2011): "The Effect of Parents' Schooling on Child's Schooling: A Nonparametric Bounds Analysis," *Journal of Labor Economics*, 29, 859–892.

Ekeland, I., A. Galichon, and M. Henry (2010): "Optimal transportation and the falsifiability of incompletely specified economic models," *Economic Theory*, 42, 355–374.

Freyberger, J. and J. Horowitz (2012): "Identification and shape restrictions in nonparametric instrumental variables estimation," CeMMAP working papers CWP15/12, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Galichon, A. and M. Henry (2009): "A test of non-identifying restrictions and confidence regions for partially identified parameters," *Journal of Econometrics*, 152, 186–196.

Hauser, R. M. (2005): "Survey response in the long run: The Wisconsin Longitudinal Study," *Field Methods*, 17, 3–29.

Henry, M. and A. Onatski (2012): "Set Coverage and Robust Policy," *Economics Letters*, 115, 256–257.

Holmlund, H., M. Lindahl, and E. Plug (2011): "The Causal Effect of Parents' Schooling on Children's Schooling: A Comparison of Estimation Methods," *Journal of Economic Literature*, 49, 615–51.

Honore, B. E. and E. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica*, 74, 611–629.

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

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

Laffers, L. (2013a): “Identification in Models with Discrete Variables,” NHH Dept. of Economics Discussion Paper No. 01/2013, available at http://ssrn.com/abstract=2205827.

——— (2013b): “Inference in Partially Identified Models with Discrete Variables,” Working paper.

——— (2013c): “A Note on Bounding Average Treatment Effects,” *Economics Letters*, 120, 424–428.

Lichter, D. (1990): “Delayed Marriage, Marital Homogamy, and the Mate Selection Process Among White Women,” *Social Science Quarterly*, 71, 802–811.

Manski, C. F. (1990): “Nonparametric Bounds on Treatment Effects,” *American Economic Review*, 80, 319–23.

——— (1995): *Identification Problems in the Social Sciences*, Cambridge, Harvard University Press.

——— (1997): “Monotone Treatment Response,” *Econometrica*, 65, 1311–1334.

——— (2003): *Partial Identification of Probability Distributions*, New York: Springer-Verlag.

——— (2007): “Partial Indentification of Counterfactual Choice Probabilities,” *International Economic Review*, 48, 1393–1410.

——— (2008): “Partial Identification in Econometrics,” in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf and L. E. Blume, Basingstoke: Palgrave Macmillan.

Manski, C. F. and J. V. Pepper (2000): “Monotone Instrumental Variables, with an Application to the Returns to Schooling,” *Econometrica*, 68, 997–1012.

Martin, D. (1975): “On the continuity of the maximum in parametric linear programming,” *Journal of Optimization Theory and Applications*, 17, 205–210.

Maurin, E. and S. McNally (2008): “Vive la Révolution! Long-Term Educational Returns of 1968 to the Angry Students,” *Journal of Labor Economics*, 26, 1–33.

Munkres, J. R. (2000): *Topology*, Prentice Hall, second ed.

Oreopoulos, P., M. E. Page, and A. H. Stevens (2006): "The intergenerational effects of compulsory schooling," *Journal of Labor Economics*, 24, 729–760.

Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

Zybert, P., Z. Stein, and B. Lillian (1978): "Demography of the marriage market in the United States," *Perceptual and Motor Skills*, 47, 815–818.

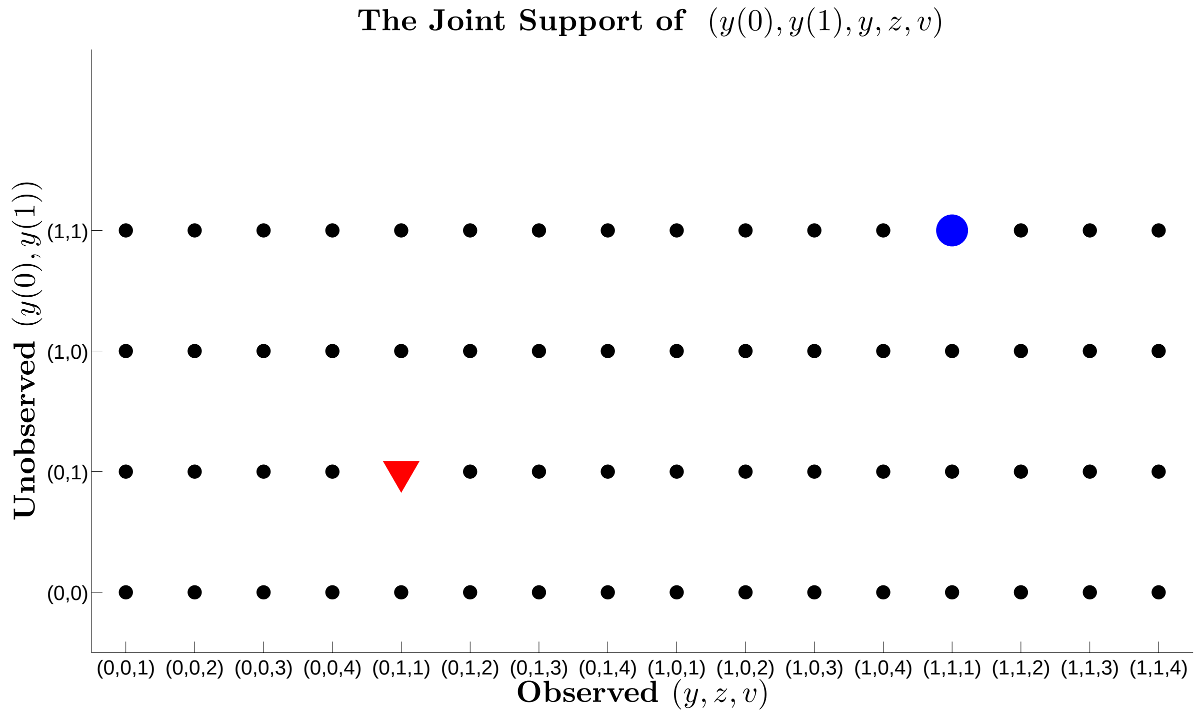

Figure 1: The joint support of the observed (y, z, v) and the unobserved component $(y(0), y(1))$. The large circle corresponds to the population with the observed $y = 1$, $z = 1$, $v = 1$, and the unobserved outcomes $y(0) = 1$ and $y(1) = 1$. The triangle stands for the individuals with $y = 0$, $z = 1$, $v = 1$, but $y(0) = 0$ and $y(1) = 1$ so the unobserved component is not compatible with the observed component.

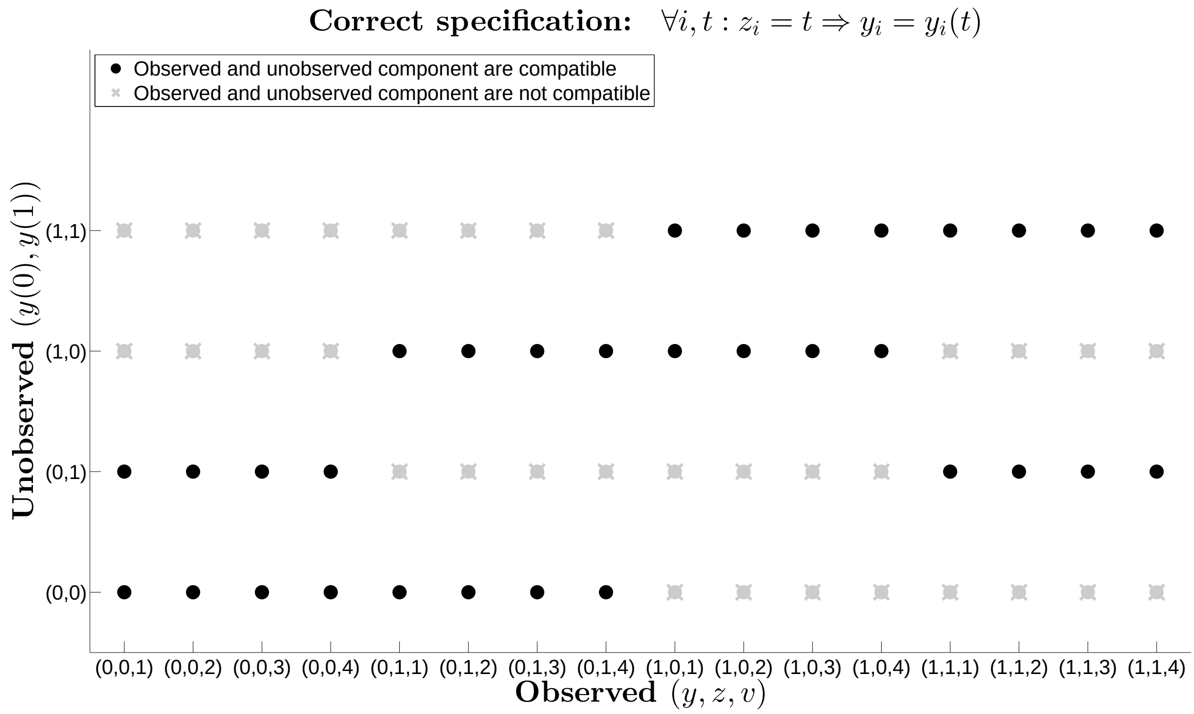

Figure 2: The joint support of the observed (y, z, v) and the unobserved component $(y(0), y(1))$. The grey points correspond to populations for which the unobserved component is incompatible with the observed component.

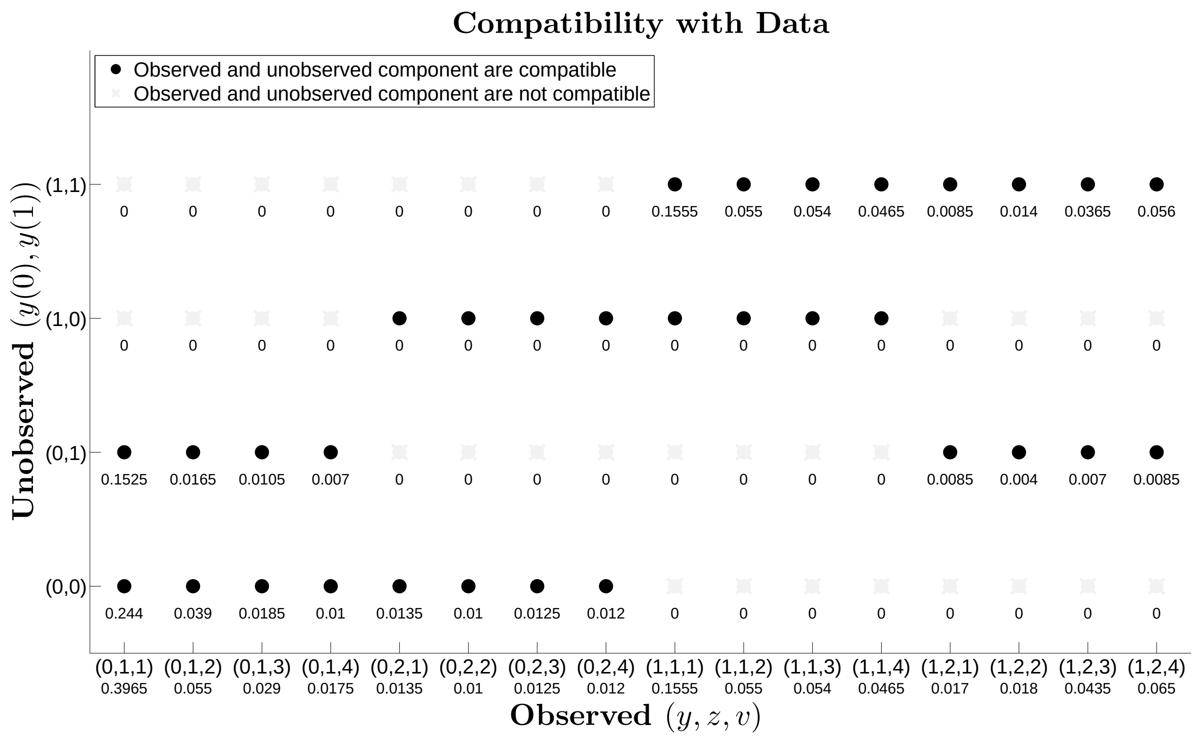

Figure 3: An example of the joint probability distribution of $(y(0), y(1), y, z, t)$. We observe one of its marginal distributions from the data (numbers on the horizontal axis).

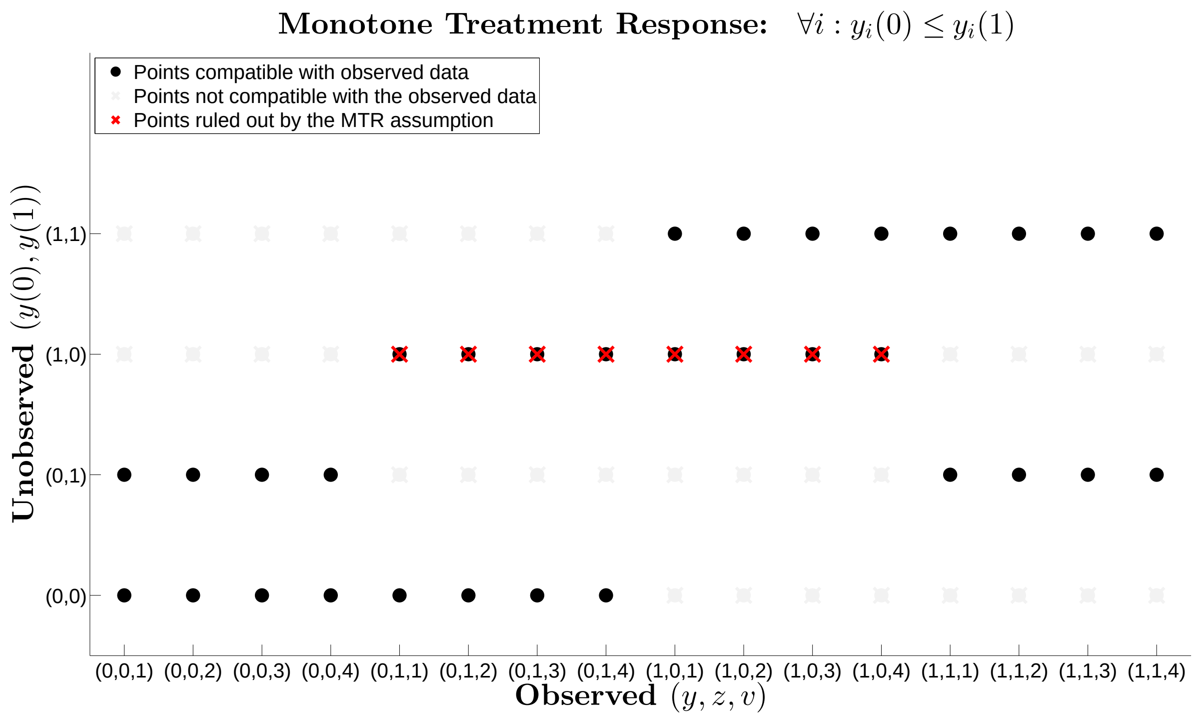

Figure 4: The joint support of the observed (y, z, v) and the unobserved component $(y(0), y(1))$. The MTR assumption rules out points for which $y(0) \leq y(1)$ is violated.

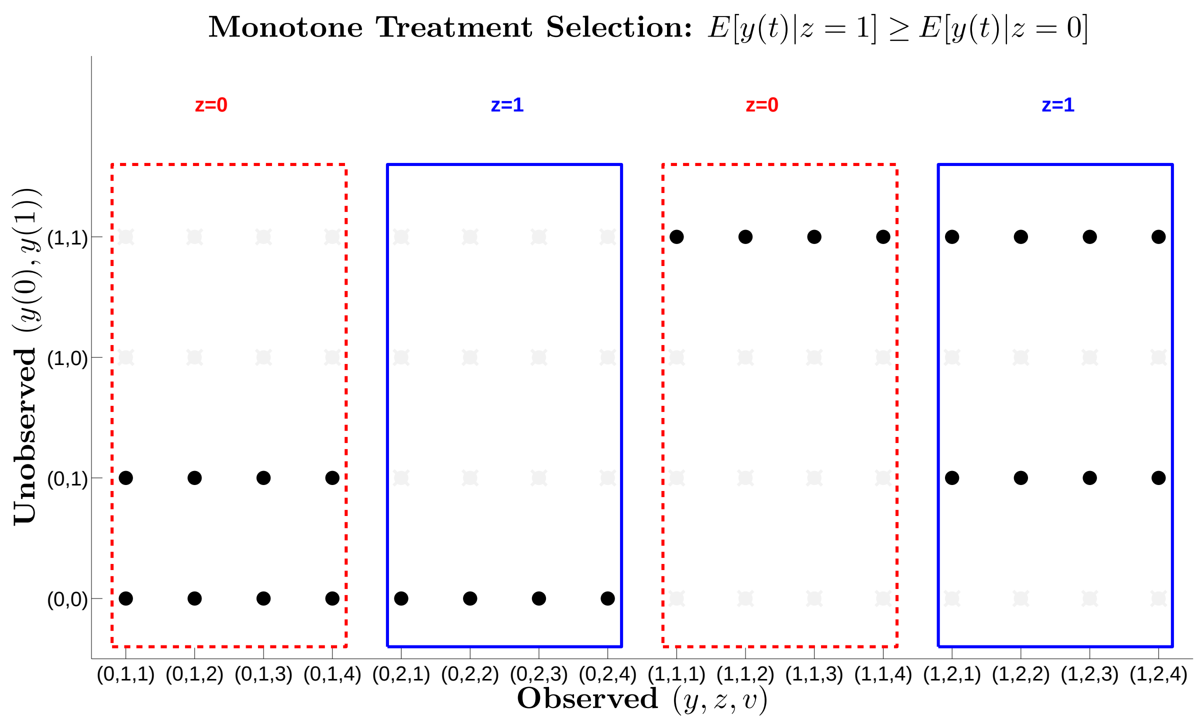

Figure 5: The joint support of the observed (y, z, v) and the unobserved component $(y(0), y(1))$. The MTS assumption states that the expectation of $y(t)$ based on the conditional distribution of the solid region is greater than that based on the dashed region.

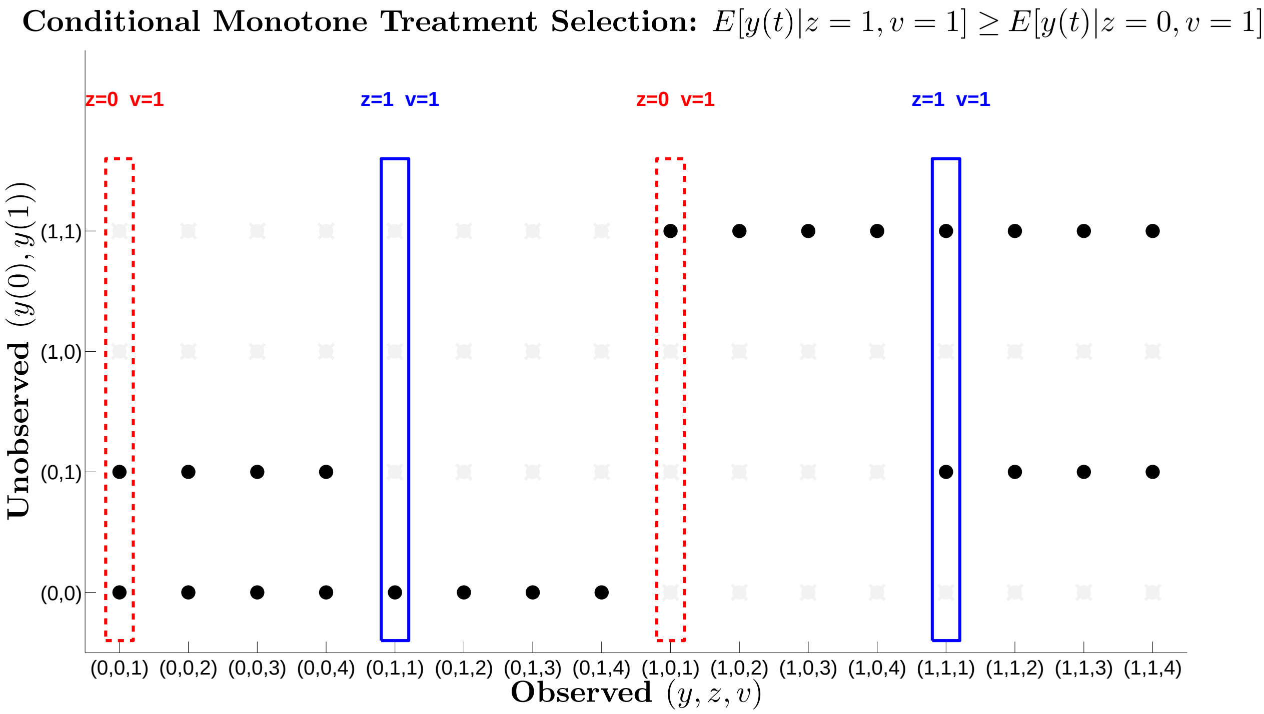

Figure 6: The joint support of the observed (y, z, v) and the unobserved component $(y(0), y(1))$. The cMTS assumption states that the expectation of $y(t)$ based on the conditional distribution of the solid region is greater than that based on the dashed region if we condition on $v = 1$.

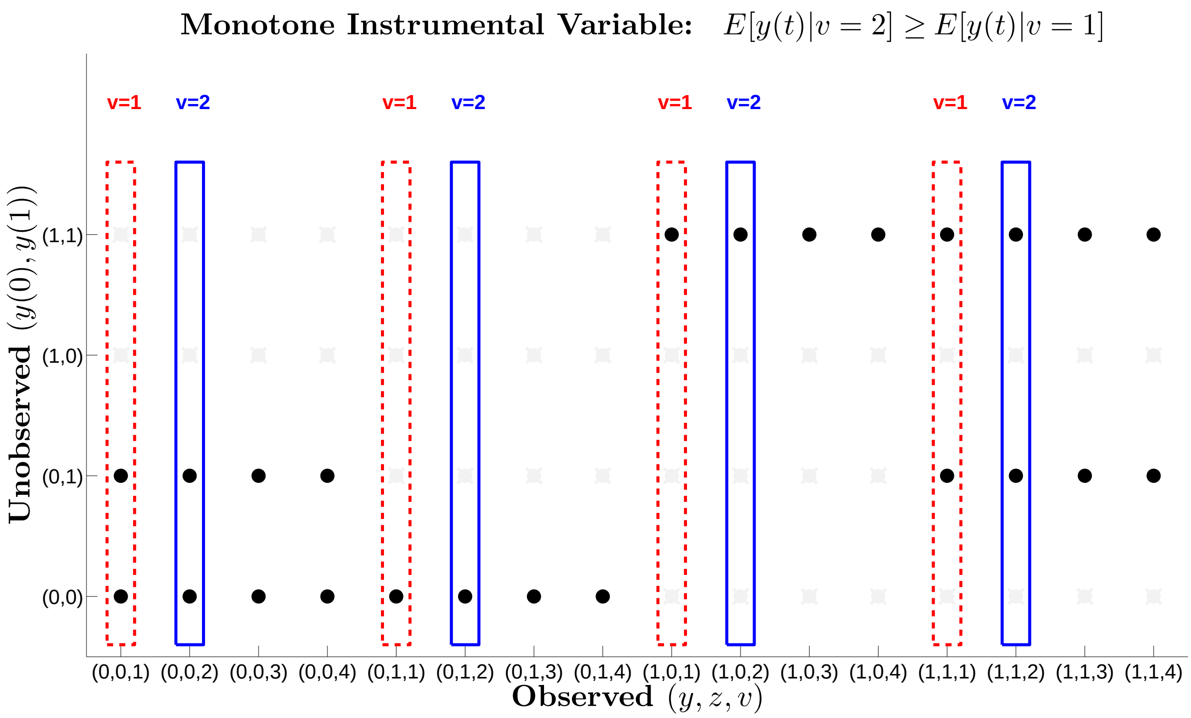

Figure 7: The joint support of the observed (y, z, v) and the unobserved component $(y(0), y(1))$. The MIV assumption states that the expectation of $y(t)$ based on the conditional distribution of the solid region is greater than that based on the dashed region.

Bounds on the Effect of an Increase in the Mother's (Father's) College Education on the Probability the Child has a College Degree				
Outcome	Child's college			
Treatment	Mother's college		Father's college	
Instrument	Father's	Grandparent's	Mother's	Grandparent's
No Assumption	[-35.9%, 64.1%] (-36.5%, 64.7%)	[-35.9%, 64.1%] (-36.5%, 64.7%)	[-31.9%, 68.1%] (-32.6%, 68.7%)	[-31.9%, 68.1%] (-32.6%, 68.7%)
MTR	[0%, 64.1%] (0%, 64.7%)	[0%, 64.1%] (0%, 64.7%)	[0%, 68.1%] (0%, 68.8%)	[0%, 68.1%] (0%, 68.7%)
MTS	[-35.9%, 36.5%] (-36.5%, 37.9%)	[-35.9%, 36.5%] (-36.6%, 37.9%)	[-31.9%, 39.3%] (-32.6%, 40.6%)	[-31.9%, 39.3%] (-32.5%, 40.6%)
cMTS	[-35.9%, 21.4%] (-36.5%, 23.7%)	[-35.9%, 33.7%] (-36.5%, 35.4%)	[-31.9%, 30%] (-32.6%, 31.6%)	[-31.9%, 37.2%] (-32.5%, 38.7%)
MTR+MTS	[0%, 36.5%] (-0%, 37.9%)	[0%, 36.5%] (-0%, 37.9%)	[0%, 39.3%] (0%, 40.6%)	[0%, 39.3%] (-0%, 40.5%)
MTR+cMTS	[0%, 21.4%] (0%, 23.7%)	[0%, 33.7%] (0%, 35.3%)	[0%, 30%] (0%, 31.7%)	[0%, 37.2%] (0%, 38.6%)
MTR+MTS+MIV	[0%, 36.5%] (0%, 37.9%)	[0%, 36.5%] (-0.1%, 37.9%)	[0%, 39.3%] (-0.1%, 40.7%)	[0.1%, 39.3%] (-0.8%, 40.6%)
MTR+cMTS+MIV	[0%, 21.4%] (0%, 23.6%)	[0%, 30.6%] (-0.1%, 33.3%)	[0%, 30%] (-0.1%, 31.7%)	[0.1%, 34.7%] (-0.7%, 37.1%)
Sample size	16912		14614	
90% confidence intervals in parentheses using the method of Imbens and Manski (2004)				

Table 1: Bounds on the effect of an increase in the parent's college education on the probability that the child has a college degree under different identifying assumptions.

Bounds on the Effect of an Increase in the Mother's (Father's) College Education on the Years of Child' schooling				
Outcome	Child's years of schooling			
Treatment	Mother's college		Father's college	
Instrument	Father's	Grandparent's	Mother's	Grandparent's
No Assumption	[-12.164, 10.836] (-12.203, 10.874)	[-12.164, 10.836] (-12.204, 10.872)	[-11.387, 11.613] (-11.43, 11.656)	[-11.387, 11.613] (-11.43, 11.652)
MTR	[0, 10.836] (0, 10.872)	[0, 10.836] (0, 10.874)	[0, 11.613] (0, 11.653)	[0, 11.613] (0, 11.655)
MTS	[-12.164, 1.809] (-12.204, 1.881)	[-12.164, 1.809] (-12.203, 1.873)	[-11.387, 1.943] (-11.43, 2.004)	[-11.387, 1.943] (-11.432, 2.002)
cMTS	[-12.164, 1.088] (-12.204, 1.184)	[-12.164, 1.651] (-12.202, 1.723)	[-11.387, 1.437] (-11.429, 1.508)	[-11.387, 1.83] (-11.428, 1.892)
MTR+MTS	[-0, 1.809] (-0, 1.875)	[0, 1.809] (0, 1.87)	[0, 1.943] (-0, 2.002)	[0, 1.943] (-0.151, 2.003)
MTR+cMTS	[0, 1.088] (-0, 1.185)	[-0, 1.651] (-0, 1.72)	[-0, 1.437] (-0, 1.513)	[0, 1.83] (-0, 1.898)
MTR+MTS+MIV	[-0, 1.809] (-0, 1.872)	[-0, 1.809] (-0.139, 1.872)	[-0, 1.943] (-0, 2.005)	[0.008, 1.943] (-0.03, 2.007)
MTR+cMTS+MIV	[0, 1.088] (-0.114, 1.185)	[0, 1.523] (-0.111, 1.658)	[0, 1.437] (-0.147, 1.509)	[0.008, 1.702] (-0.202, 1.815)
Sample size	16912		14614	
90% confidence intervals in parentheses using the method of Imbens and Manski (2004)				

Table 2: Bounds on the effect of an increase in the parent's college education on the years of the child's schooling under different identifying assumptions.

$$\max_\pi \overbrace{\begin{bmatrix} 0\,1\,0\,1\,0\,1\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,1\,0\,1\,0\,1\,0 \end{bmatrix}}^{\text{Average Treatment Effect}} \times \pi$$

subject to

$$\text{DATA}\left\{\begin{array}{c}\begin{bmatrix}
1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\\
0\,0\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\\
0\,0\,0\,0\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\\
0\,0\,0\,0\,0\,0\,1\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\\
0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\\
0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\,0\\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,0\,0\,0\,0\,0\,0\,0\,0\\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,0\,0\,0\,0\,0\,0\\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,0\,0\,0\,0\\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1\,0\,0\\
0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,1\,1
\end{bmatrix}\end{array}\right. \times \pi = \left.\begin{bmatrix}0.397\\0.055\\0.029\\0.017\\0.013\\0.01\\0.013\\0.012\\0.155\\0.055\\0.054\\0.047\\0.017\\0.018\\0.043\\0.065\end{bmatrix}\right\}\begin{array}{c}\text{Observed}\\\text{probabilities}\end{array}$$

$$\begin{array}{l}
\text{cMTS}\left\{\begin{bmatrix}
0&0&0&0&0&0&0&0&0&0&0&0&.03&0&0&0&0&-.55&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&.03&0&0&0&0&-.11&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&.06&0&0&0&0&0&0&-.08&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&.08&0&0&0&0&0&0&0&-.06\\
0&.03&0&0&0&0&0&0&0&0&0&0&.03&0&0&0&-.55&-.55&0&0&0&0&0&0\\
0&0&0&.03&0&0&0&0&0&0&0&0&0&.03&0&0&0&0&-.11&-.11&0&0&0&0\\
0&0&0&0&0&.06&0&0&0&0&0&0&0&0&.06&0&0&0&0&0&-.08&-.08&0&0\\
0&0&0&0&0&0&0&.08&0&0&0&0&0&0&0&.08&0&0&0&0&0&0&-.06&-.06
\end{bmatrix}\right. \\
\text{MIV}\left\{\begin{bmatrix}
0&0&0&0&0&0&0&0&0&0&0&0&.13&-.58&0&0&0&.13&0&-.58&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&.13&-.13&0&0&0&0&.13&0&-.13&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&.14&-.13&0&0&0&0&0&.14&0&-.13\\
0&.13&0&-.58&0&0&0&0&0&0&0&0&.13&-.58&0&0&.13&.13&-.58&-.58&0&0&0&0\\
0&0&0&.13&0&-.13&0&0&0&0&0&0&0&.13&-.13&0&0&0&.13&.13&-.13&-.13&0&0\\
0&0&0&0&0&.14&0&-.13&0&0&0&0&0&0&.14&-.13&0&0&0&0&.14&.14&-.13&-.13
\end{bmatrix}\right.
\end{array} \times \pi \leq \begin{array}{l}\begin{bmatrix}0\\0\\0\\0\\0\\0\\0\\0\end{bmatrix}\\ \begin{bmatrix}0\\0\\0\\0\\0\\0\end{bmatrix}\end{array}$$

$$\pi \geq \begin{bmatrix}0\\ \vdots \\ 0\end{bmatrix},$$

$$\pi^* = \begin{array}{l} [0.244\ \ 0.152\ 0.039\ 0.016\ 0.019\ 0.010\ 0.010\ 0.007\ 0.013\ 0.010\ 0.013\ \ 0.012\ \ \ldots \\ \ldots 0.155\ 0.055\ 0.054\ 0.047\ 0.008\ 0.009\ 0.004\ 0.014\ 0.007\ 0.036\ 0.009\ 0.056]'. \end{array}$$

Figure 8: This linear program searches in the space of the joint probability distributions assigned to all combinations of the observed component (y, z, v) and the unobserved component $(y(0), y(1))$ that are compatible $(\forall i, t : z_i = t \rightarrow y_i = y_i(t))$ and satisfy the MTR assumption (as depicted in Figure 4). The space of the joint distributions is further restricted to satisfy the cMTS assumption and the MIV assumption, and to be compatible with the observed probabilities. The optimal solution π^* maximizes the average treatment effect.

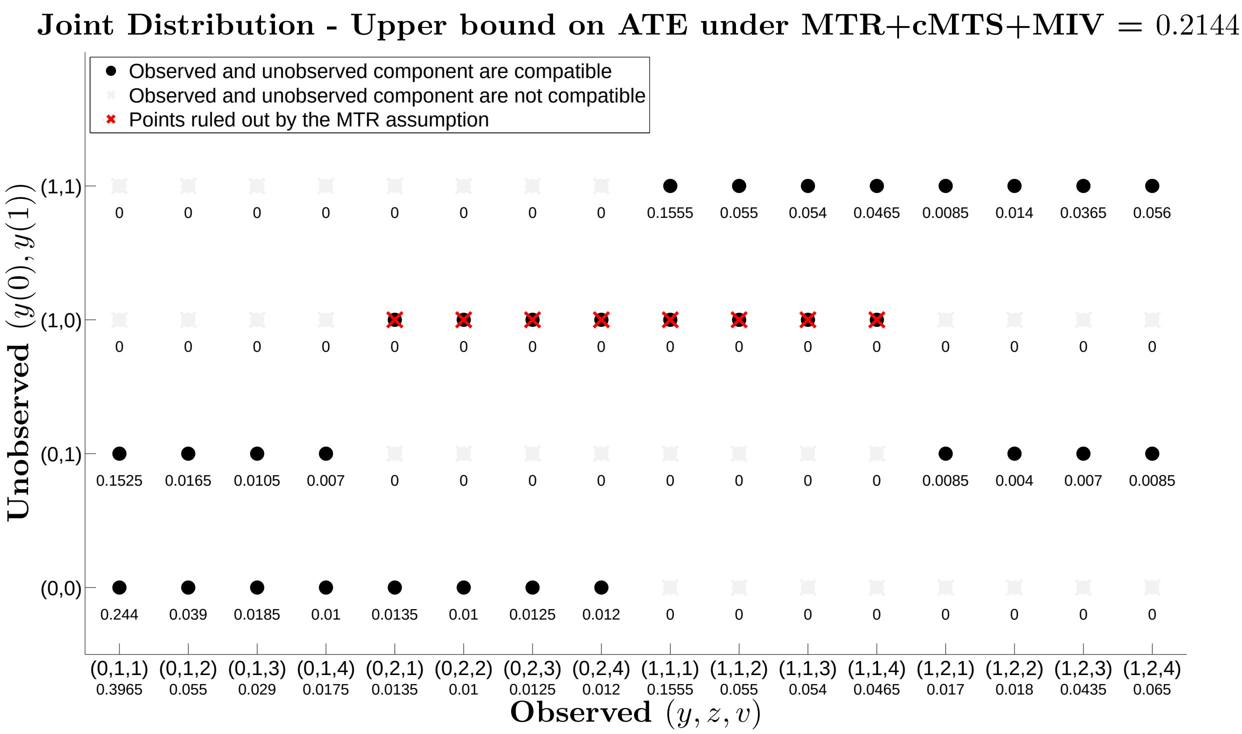

Figure 9: The joint probability distribution that maximizes the ATE of mother's college increase on child's probability of getting a college degree using other parent's schooling as a monotone instrumental variable under the MTR+cMTS+MIV assumption. Numbers on the horizontal axis are probabilities of the observed variables.

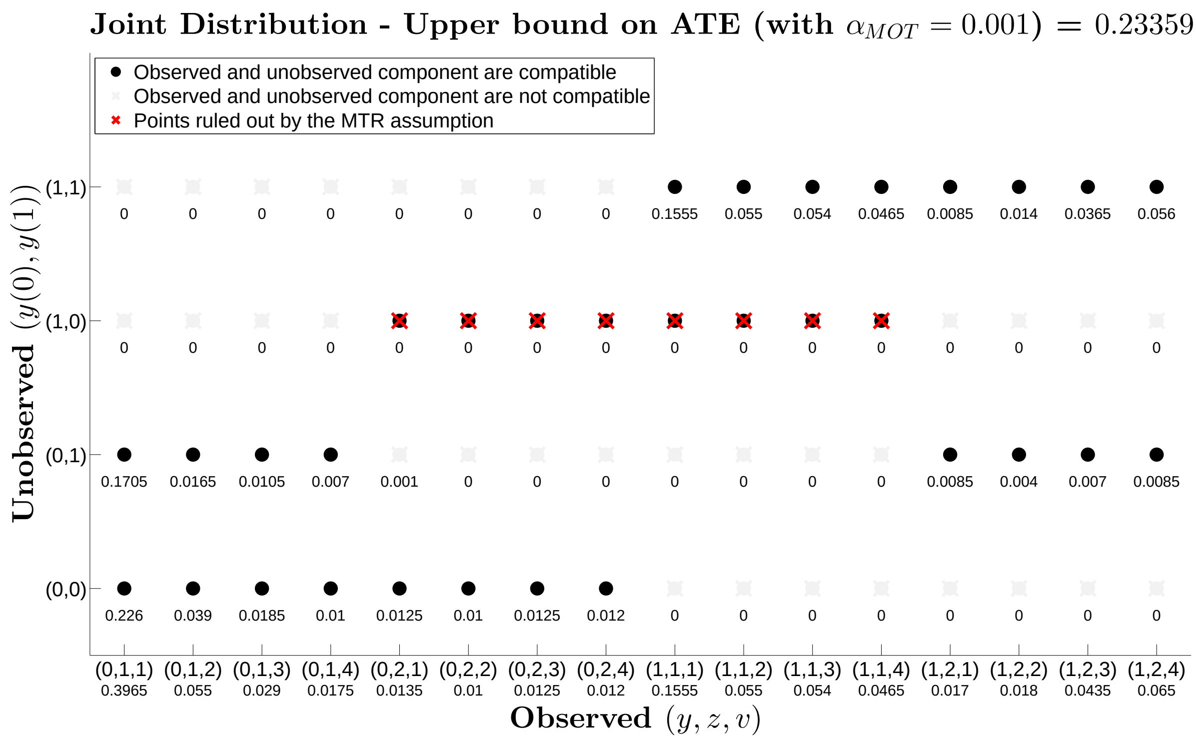

Figure 10: The joint probability distribution that maximizes the ATE of mother's college increase on child's probability of getting a college degree using other parent's schooling as a monotone instrumental variable. The MOT assumption is relaxed by $\alpha_{MOT} = 0.001$. We can see that this probability was assigned to the point $(0, 1, 0, 2, 1)$.

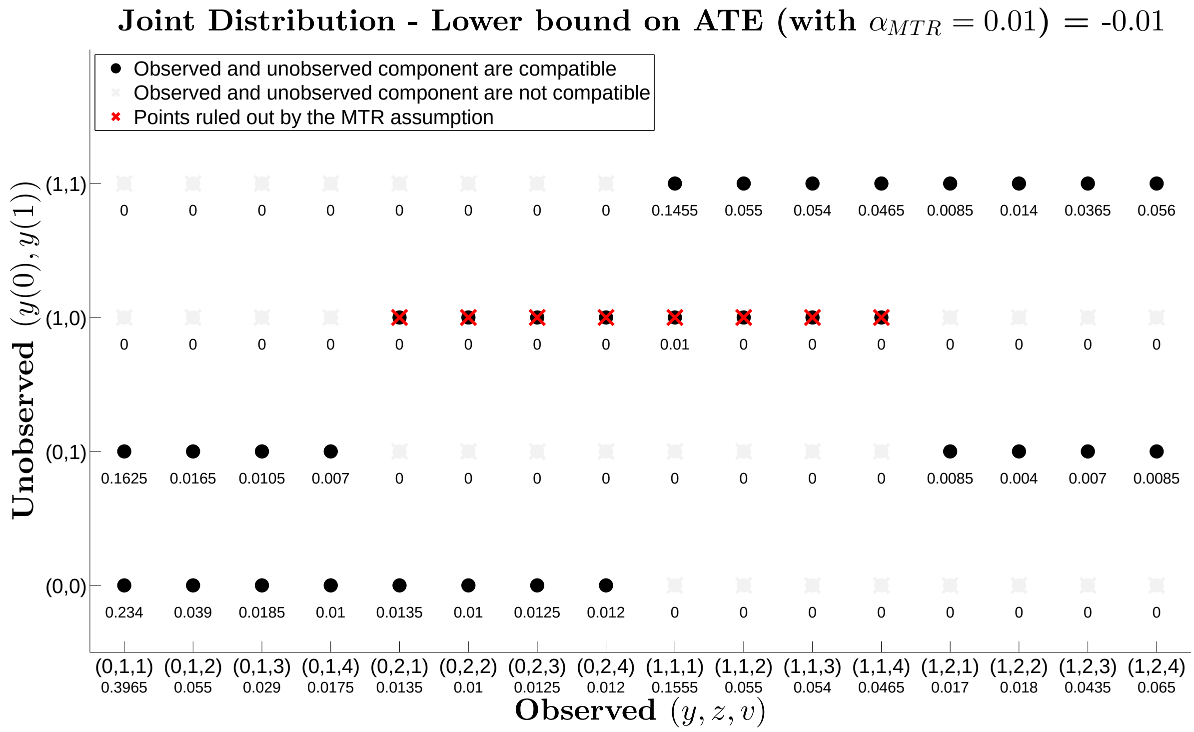

Figure 11: The joint probability distribution that minimizes the ATE of mother's college increase on child's probability of getting a college degree using other parent's schooling as a monotone instrumental variable. The MTR assumption does not need to hold for 1% of the children $\alpha_{MTR} = 0.01$. This 1% of children was assigned to the point $(1, 0, 1, 1, 1)$ and decreased the lower bound of ATE accordingly by 0.01.

Bounds on Effect of Mother's College Increase on the Probability that the Child has a College Degree					
MTR+cMTS+MIV					
[Lower bound, Upper bound] = **[0, 21.44%]**					
Confidence Set = (0, 23.74%)					
	Lower bound	Upper bound			
	α_{MTR}	α_{MOT}	α_{cMTS}	α_{MIV}	α_{MISS}
Optimistic	0.01	0.001	0.01	0.01	0.01
	-1%	23.36%	22.44%	21.44%	27.31%
	(-1.46%)	(25.63%)	(24.71%)	(23.71%)	(29.64%)
Pessimistic	0.05	0.01	0.05	0.05	0.10
	-5%	35.66%	26.44%	21.44%	38.15%
	(-5%)	(37.74%)	(28.71%)	(23.71%)	(40.67%)
Optimistic	0.01	0.001	0.01	0.01	0
			$[-1\%, 24.36\%]$		
			$(-1\%, 26.63\%)$		
	0.01	0.001	0.01	0.01	0.01
			$[-1\%, 28.62\%]$		
			$(-1\%, 29.66\%)$		
Pessimistic	0.05	0.01	0.05	0.05	0
			$[-5\%, 41.54\%]$		
			$(-5\%, 43.67\%)$		
	0.05	0.01	0.05	0.05	0.10
			$[-5\%, 53.25\%]$		
			$(-5\%, 55.08\%)$		

Note: Estimates are not bias corrected, $n = 16912$
90% confidence intervals in parentheses using the method of Imbens and Manski (2004)

Table 3: Sensitivity analysis of the bounds on the effect of mother's college degree on the probability that the child gets a college degree. Father's education level was used as a monotone instrumental variable.

Bounds on ATE	
$MTR + cMTS$	$[0, 21.44\%]$
$MTR + cMTS + MIV$	$[0, 21.44\%]$

If cMTS holds for $v \in \{2, 3, 4\}$ only:

Bounds on ATE	
$MTR + cMTS$	$[0, 46.71\%]$
$MTR + cMTS + MIV$	$[0, 27.54\%]$

Table 4: Bounds on the effect of mother's college increase on the probability that the child has a college degree using father's schooling level as a monotone instrument.

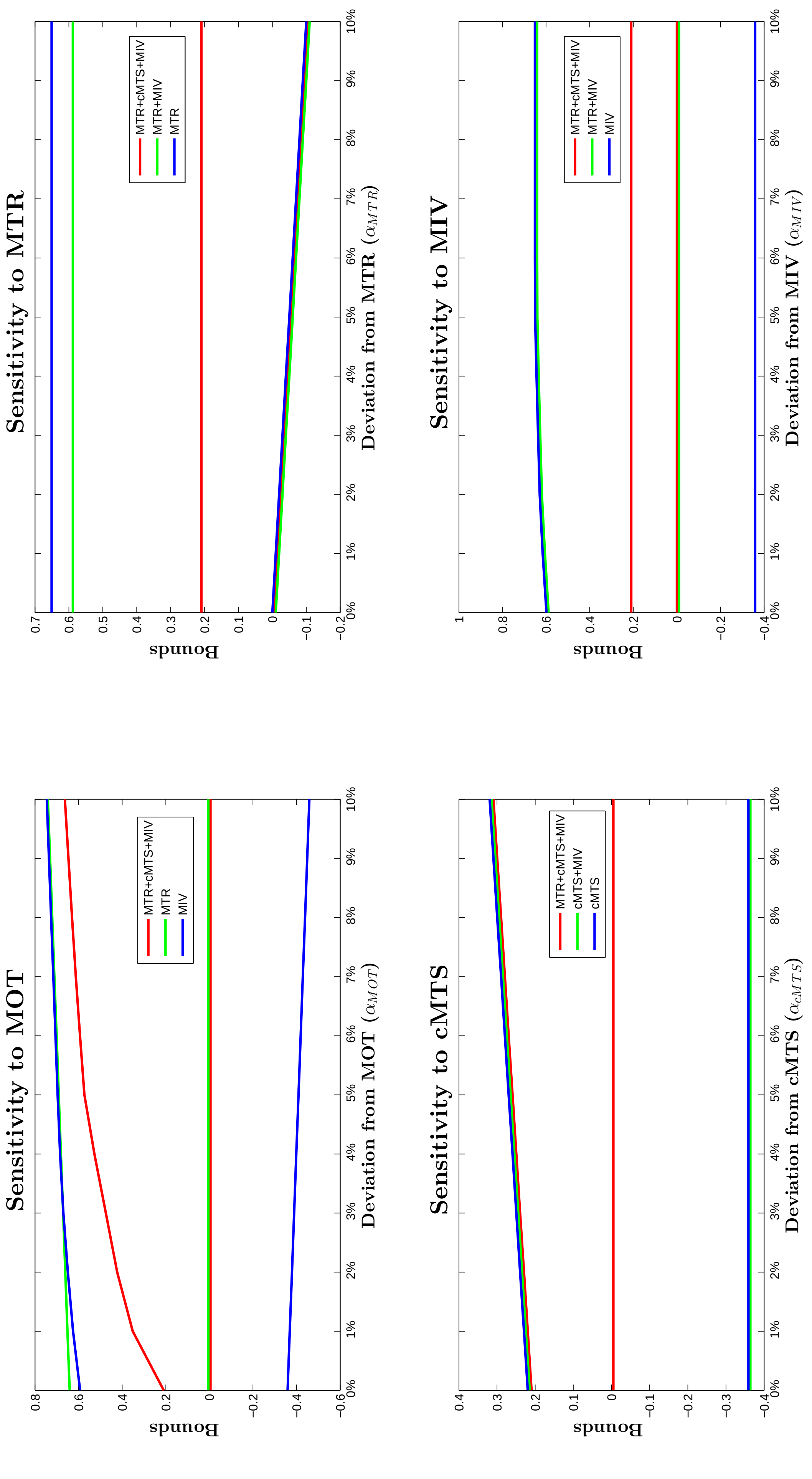

Figure 12: Sensitivity of the bounds on the effect of mother's college increase on probability change that child would graduate using father's schooling level as a monotone instrument.

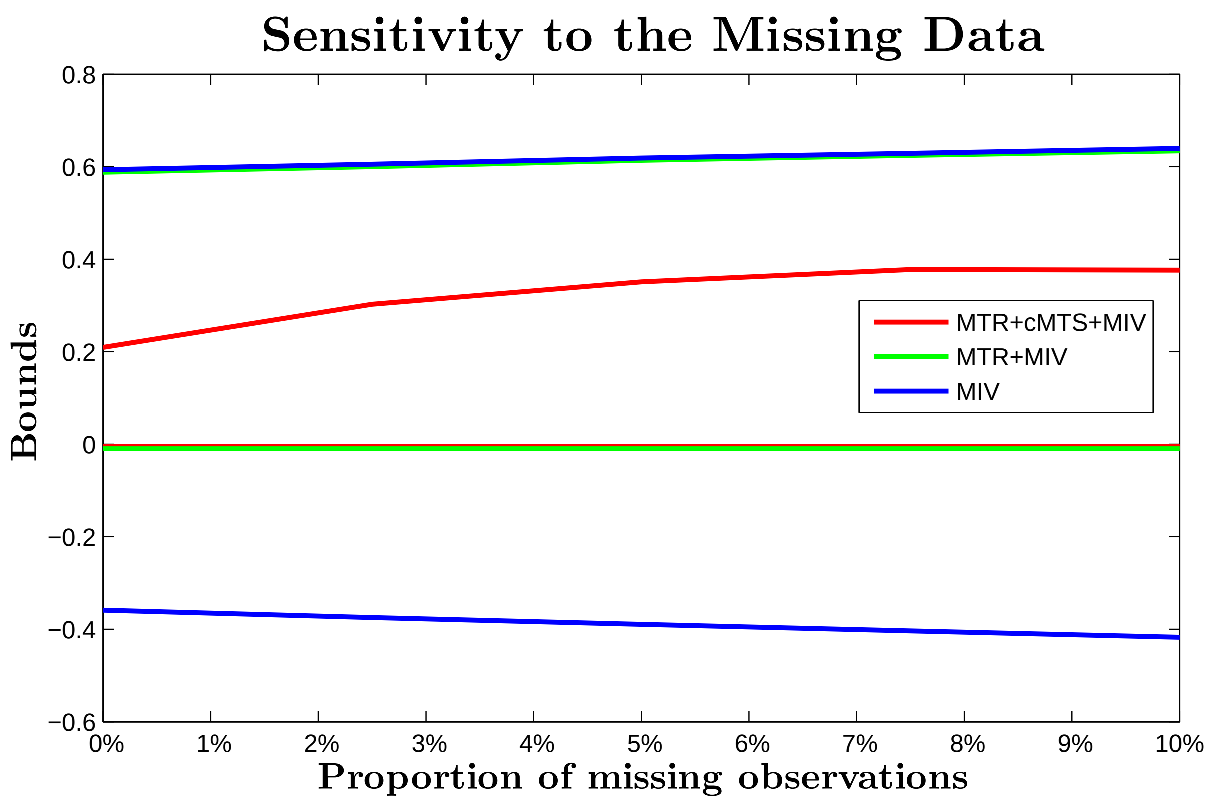

Figure 13: Sensitivity of the bounds on the effect of mother's college increase on probability change that child would graduate to missing data under different assumptions.

Binding constraints under MTR+cMTS+MIV and Lagrange multipliers:

$$cMTS \begin{cases} E[y(0)|z=1,v=1] & \geq & E[y(0)|z=0,v=1] & 0.0303 \\ E[y(1)|z=1,v=1] & \geq & E[y(1)|z=0,v=1] & 0.5505 \\ E[y(0)|z=1,v=2] & \geq & E[y(0)|z=0,v=2] & 0.0282 \\ E[y(1)|z=1,v=2] & \geq & E[y(1)|z=0,v=2] & 0.1106 \\ E[y(0)|z=1,v=3] & \geq & E[y(0)|z=0,v=3] & 0.0554 \\ E[y(1)|z=1,v=3] & \geq & E[y(1)|z=0,v=3] & 0.0823 \\ E[y(0)|z=1,v=4] & \geq & E[y(0)|z=0,v=4] & 0.0766 \\ E[y(1)|z=1,v=4] & \geq & E[y(1)|z=0,v=4] & 0.0637 \end{cases}$$

Nonbinding constraints:

$$MIV \begin{cases} E[y(0)|v=2] & \geq & E[y(0)|v=1] & 0 \\ E[y(1)|v=2] & \geq & E[y(1)|v=1] & 0 \\ E[y(0)|v=3] & \geq & E[y(0)|v=2] & 0 \\ E[y(1)|v=3] & \geq & E[y(1)|v=2] & 0 \\ E[y(0)|v=4] & \geq & E[y(0)|v=3] & 0 \\ E[y(1)|v=4] & \geq & E[y(1)|v=3] & 0 \end{cases}$$

Figure 14: Binding and nonbinding identifying constraints under the MTR+cMTS+MIV assumption with corresponding Lagrange multipliers.

Binding constraints under MTR+cMTS+MIV:
(cMTS for $v \in \{2, 3, 4\}$) and Lagrange multipliers

$$
\begin{array}{ll}
cMTS & \left\{ \begin{array}{lcll}
E[y(0)|z=1,v=2] & \geq & E[y(0)|z=0,v=2] & 0.0282 \\
E[y(1)|z=1,v=2] & \geq & E[y(1)|z=0,v=2] & 0.5768 \\
E[y(0)|z=1,v=3] & \geq & E[y(0)|z=0,v=3] & 0.0554 \\
E[y(1)|z=1,v=3] & \geq & E[y(1)|z=0,v=3] & 0.0823 \\
E[y(0)|z=1,v=4] & \geq & E[y(0)|z=0,v=4] & 0.0766 \\
E[y(1)|z=1,v=4] & \geq & E[y(1)|z=0,v=4] & 0.0637
\end{array} \right. \\
MIV & \quad E[y(1)|v=2] \geq E[y(1)|v=1] \qquad 0.5821
\end{array}
$$

Nonbinding constraints:

$$
MIV \left\{ \begin{array}{lcll}
E[y(0)|v=2] & \geq & E[y(0)|v=1] & 0 \\
\sout{E[y(1)|v=2]} & \sout{\geq} & \sout{E[y(1)|v=1]} & \\
E[y(0)|v=3] & \geq & E[y(0)|v=2] & 0 \\
E[y(1)|v=3] & \geq & E[y(1)|v=2] & 0 \\
E[y(0)|v=4] & \geq & E[y(0)|v=3] & 0 \\
E[y(1)|v=4] & \geq & E[y(1)|v=3] & 0
\end{array} \right.
$$

Figure 15: Binding and nonbinding identifying constraints under the MTR+cMTS+MIV assumption (cMTS for $v \in \{2, 3, 4\}$) with corresponding Lagrange multipliers.