Sharp Bound for Average Treatment Effects with Monotone Instrument Variables

Kyungtae Park*

January 24, 2025

Abstract

This article analytically derives the sharp bound of the Average Treatment Effect under the monotone instrumental variable and the monotone treatment response assumptions. While delivering the proof, this paper also identifies the gap in the existing results for the sharp bound of arm-specific means and provides thorough proof. The result can be extended to other inequality constraints on the outcomes such as the conditional monotone treatment selection assumption. Despite its limited practical importance, the findings advance our theoretical understanding of how much identifying power different monotone assumptions have.

Keywords: Monotone Instrument Variables, Manski Bounds, Partial Identification

JEL classification: C13

*Department of Political Science, Stanford University.

1 Introduction

The partial identification literature began with various shape restrictions to narrow down possible ranges of the average treatment effect. Examples include monotone treatment response, concave monotone treatment response, monotone treatment selection, and in the instrument variable settings monotone instrumental variable (Manski 1997; Manski and Pepper 2000). Further shape restrictions can be found in the more recent literature (Giustinelli 2011; Twinam 2017; Kim et al. 2018). Despite these notable theoretical advancements, there are two points to be noted. First, many analytic results are proven for arm-specific means rather than for average treatment effects that are more of immediate interest for researchers. The bound derived from sharp bounds for arm-specific means might not be sharp enough for average treatment effects. Second, the proof for sharpness results is sometimes incomplete. The best way to show sharpness is to explicitly construct a population that is compatible with the observed data, satisfies all assumptions and achieves the stated bound, which is seldom employed in theoretical works due to cumbersome notations and complicated logic.

This paper complements the foundational paper of Manski and Pepper (2000) by providing the sharp bound of the Average Treatment Effect (ATE) under the monotone instrumental variable and the monotone treatment response assumptions. This paper also fills in the gap of the literature by providing a thorough proof for the sharp bound of arm-specific means in the same settings. It will be shown below that the sharpness proof in Manski (2003) is incomplete and there is no easy fix as the monotone instrument variable condition imposes a system of linear equations subject to some inequality restrictions. Using an algorithm that guarantees the existence of solutions under the restrictions, this paper shows that any combination of conditional expectations of potential outcomes with respect to the instrumental variable can be simultaneously achieved as long as they do not violate the assumptions. The algorithm turns out to be the key to the results for the ATE bounds as well since they require to construct a population out of an arbitrary combination of conditional expectations. The sharp bound for the ATE decomposes the two expected outcome into conditional expectations each, and applies monotone treatment response to the pairs that share a common realized treatment.

These results easily extend to the conditional monotone treatment selection assumption. The monotone instrument variable is often used in conjunction with the monotone treatment selection in empirical analyses (Gonzalez 2005; De Haan 2011; Kreider et al. 2012). Lafférs (2013) notes the confusion in the literature around the monotone treatment selection and the conditional monotone treatment selection in instrumental variable settings, which have different meanings and one of which is not subsumed by the other. The conditional monotone treatment selection states that potential outcomes are monotone in the realized treatment conditionally on the instrument assignment. This assumption mathematically imposes additional rank restrictions on the conditional expectations. The above lemma can be modified to accommodate these additional rank restrictions, leading to sharp bounds of arm-specific means and average treatment effects under the conditional monotone treatment selection on top of the other assumptions previously considered.

Computational methods have become increasingly popular in the partial identification literature more broadly as they facilitate handling with multiple assumptions with ease (Mogstad, Santos and Torgovitsky 2018; Duarte et al. 2024). Especially, Laffére (2013, 2019) derive the bounds for the ATE under both sets of assumptions using linear programming approaches. The analytical results here might not offer a huge practical contribution if the computation is not particularly costly. However, computationally derived bounds often lack intuition in terms of why one assumption has a stronger bounding effect than another if one goal of partial identification is to understand how different assumptions have different implications in analyses. Moreover, approaches in Laffére (2013, 2019) assume binary treatments or continuous treatments but with linear effects, while the bounds here accommodate finite treatments and finite instruments. This difference is substantial in partial identification because the resulting bound may not be sharp when estimated with a subset data, unlike in point identification settings. This paper addresses a gap in our understanding of the monotone instrument variable that has remained unresolved since the early stages of the literature. At the end of the paper, I empirically illustrate the methods by reanalyzing Gonzalez (2005) that studies the wage effect of language skills with the monotone instrumental variable, but with the sharp bounds and the additional conditional monotone treatment selection assumption.

2 Theoretical Results

Each observation is drawn *i.i.d* from a population of $(\{Y(d, z) : d \in \mathcal{D}, z \in \mathcal{Z}\}, \{D(z) : z \in \mathcal{Z}\}, Z)$ where Y is the outcome, D is the treatment, and Z is the instrument. Assume $\mathcal{Z}$ and $\mathcal{D}$ are discrete and their elements are totally ordered under a relation $\succeq$. We only observe $(Y_i(D_i), D_i(Z_i), Z_i) \equiv (Y_i, D_i, Z_i)$ in an infinitely large data. The goal is to find the range of $\mathbb{E}[Y(d)]$ and $\mathbb{E}[Y(d) - Y(d')]$ for given d and d' among the populations that satisfy:

Assumption 1 (SUTVA)**.** *$Y_i(\boldsymbol{d}, \boldsymbol{z}) = Y_i(d_i, z_i)$ for all i and treatment and instrument vectors $\boldsymbol{d}$ and $\boldsymbol{z}$.*

Assumption 2 (Exclusion restriction)**.** *$Y_i(d, z) = Y_i(d)$ for all i, d and z.*

Assumption 3 (Bounded outcomes)**.** *$Y_i(d) \in [\underline{y}, \overline{y}]$ for all i and d.*

Assumption 4 (MTR: Monotone treatment response)**.** *$Y_i(d) \geq Y_i(d')$ if $d \succeq d'$.*

Assumption 5 (MIV: Monotone instrument variable)**.** *$\mathbb{E}[Y(\tilde{d}) \mid Z = z] \geq \mathbb{E}[Y(\tilde{d}) \mid Z = z']$ for all $\tilde{d}$ if $z \succeq z'$.*

Assumption 4 states that the treatment always has a nonnegative effect. Assumption 5 assumes that the potential outcome is not mean-independent of the instrument, and the expected potential outcome increases in the realized instrument. This is a relaxation of identifying assumptions in the modern instrumental variable settings. Manski and Pepper (2009) point out that there are two ways to do so. One is keeping the exclusion restriction and imposing restrictions on the assignment process such as the monotone instrumental variable. The other is to allow the instrumental variable to influence the outcome and directly impose restrictions on the individual potential outcomes: $Y_i(d, z) \geq Y_i(d, z')$ if $z \succeq z'$. I take the former approach following the original paper (Manski and Pepper 2000).

A few auxiliary variables are defined to present the first results: $m^-_{d,z} = \mathbb{E}[Y \mid D \leq d, Z = z]$, $m^+_{d,z} = \mathbb{E}[Y \mid D \geq d, Z = z]$, $\pi^-_{d,z} = \mathrm{P}[D \leq d \mid Z = z]$, $\pi^+_{d,z} = \mathrm{P}[D \geq d \mid Z = z]$. These are all point identifiable from the observed data. The following proposition gives the sharp bounds for the arm-specific mean, $\mathbb{E}[Y(d)]$, and the ATE, $\mathbb{E}[Y(d)-Y(d')]$, under the above assumptions.

Proposition 1. *Under Assumptions 1-5, the sharp bounds for arm-specific means and average treatment effects are*

$$\mathbb{E}[Y(d)] \in [L_d, U_d], \qquad \mathbb{E}[Y(d)-Y(d')] \in [L_{d,d'}, U_d - L_{d'}]$$

where $d \succeq d'$, and

$$\begin{aligned} s_{d,z} &= \sup_{z' \leq z}\{m^-_{d,z'}\pi^-_{d,z'} + \underline{y}(1-\pi^-_{d,z'})\}, \\ i_{d,z} &= \inf_{z' \geq z}\{m^+_{d,z'}\pi^+_{d,z'} + \overline{y}(1-\pi^+_{d,z'})\}, \\ L_d &= \sum_z \mathrm{P}[Z=z]\cdot s_{d,z}, \\ U_d &= \sum_z \mathrm{P}[Z=z]\cdot i_{d,z}, \\ L_{d,d'} &= \sum_z \mathrm{P}[Z=z]\cdot \max(s_{d,z} - i_{d',z}, 0). \end{aligned}$$

Proof. We only prove the validity of the bound for arm-specific means $\mathbb{E}[Y(d)]$ here. All omitted proofs can be found in Appendix. To avoid confusion, $\tilde{d}$ is used for potential treatments and d, d' for the specific treatments for which we find the bound. Define $n_{\tilde{d},z} = \mathbb{E}[Y(\tilde{d})\,|\,Z=z]$. Then,

$$\begin{aligned} n_{\tilde{d},z} &= \sum_e \mathbb{E}[Y(\tilde{d})\,|\,D=e, Z=z]\cdot \mathrm{P}[D=e\,|\,Z] \\ &= \sum_{e \succ \tilde{d}} \mathbb{E}[Y(\tilde{d})\,|\,D=e, Z=z]\cdot \mathrm{P}[D=e\,|\,Z] + \sum_{e \preceq \tilde{d}} \mathbb{E}[Y(\tilde{d})\,|\,D=e, Z=z]\cdot \mathrm{P}[D=e\,|\,Z] \\ &\geq \underline{y}\cdot \mathrm{P}[D \succ \tilde{d}\,|\,Z] + \sum_{e \preceq \tilde{d}} \mathbb{E}[Y(e)\,|\,D=e, Z=z]\cdot \mathrm{P}[D=e\,|\,Z] \\ &= m^-_{\tilde{d},z}\pi^-_{\tilde{d},z} + \underline{y}(1-\pi_{\tilde{d},z}). \end{aligned}$$

Similarly, $n_{\tilde{d},z} \leq m^+_{\tilde{d},z}\pi^+_{\tilde{d},z} + \overline{y}(1-\pi^+_{\tilde{d},z})$. Since $n_{\tilde{d},z}$ increases in z by MIV, we have $\sup_{z' \leq z}\{m^-_{d,z'}\pi^-_{d,z'} + \underline{y}(1-\pi^-_{d,z'})\} \leq n_{\tilde{d},z} \leq \inf_{z' \geq z}\{m^+_{d,z'}\pi^+_{d,z'} + \overline{y}(1-\pi^+_{d,z'})\}$, and $\mathbb{E}[Y(d)] = \sum_z \mathrm{P}[Z=z]\cdot n_{\tilde{d},z} \in [L_d, U_d]$. An observable implication of Assumptions 1-5 is

$$\sup_{z' \leq z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + \underline{y}(1-\pi^-_{\tilde{d},z'})\} \leq \inf_{z' \geq z}\{m^+_{\tilde{d},z'}\pi^+_{\tilde{d},z'} + \overline{y}(1-\pi^+_{\tilde{d},z'})\}. \tag{1}$$

■

The bound for the arm-specific mean was first reported in Manski and Pepper (2000), although the proof was given only later in Manski (2003, p. 146). The analytic bound for the ATE is original. Showing the validity of the bounds is starightforward. The first part of the proof, $n_{\tilde{d},z} \geq m^-_{\tilde{d},z}\pi^-_{\tilde{d},z} + \underline{y}(1-\pi_{\tilde{d},z})$, is the result from MTR applied conditionally on z. MIV imposes monotonicity on $n_{\tilde{d},z}$, hence sup and inf. The bound somewhat mechanically synthesizes the two assumptions.

The difficulty lies in showing its sharpness. The proof in Manski (2003) says that the lower bound is achieved if the conditional expectations are simultaneously set at their lower bound: $n_{\tilde{d},z} = \sup_{z' \le z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + \underline{y}(1-\pi^-_{\tilde{d},z'})\}$ for all $\tilde{d}$ and z. This is unfortunately not obvious. Under MTR only, the sharpness can be shown by setting unobserved outcomes at either the lower bound or the upper bound. However, the unobserved outcomes cannot be always at either extreme due to the MIV conditions so they should be somewhere in the middle for the most of the time, and it is not obvious how that can be reconciled with MTR that must hold at the individual observation level. One might think that since only one specific group of potential outcomes matters in the arm-specific mean case, the unobserved outcomes can be set so that only these conditional expectations are set at their lower bounds: $n_{d,z} = \sup_{z' \le z}\{m^-_{d,z'}\pi^-_{d,z'} + \underline{y}(1-\pi^-_{d,z'})\}$ for all z. However, this approach also requires some control over potential outcomes other than $Y(d)$ since MIV must hold for all potential treatments $\tilde{d}$.

The following lemma addresses this problem by guaranteeing a population that simultaneously achieves the conditional expectation conditions for all treatment assignments under MTR, provided that the given values for conditional expectations are compatible with MTR.

Lemma 1. *Assume $m^-_{\tilde{d},z}\pi^-_{\tilde{d},z} + \underline{y}(1-\pi_{\tilde{d},z}) \le m^+_{\tilde{d},z}\pi^+_{\tilde{d},z} + \overline{y}(1-\pi^+_{\tilde{d},z})$ for all $\tilde{d}$ and z. Let $n_{\tilde{d},z}$ be constants that satisfy*

$$m^-_{\tilde{d},z}\pi^-_{\tilde{d},z} + \underline{y}(1-\pi_{\tilde{d},z}) \le n_{\tilde{d},z} \le m^+_{\tilde{d},z}\pi^+_{\tilde{d},z} + \overline{y}(1-\pi^+_{\tilde{d},z}),$$

and $n_{\tilde{d},z} \le n_{\tilde{d}',z}$ whenever $\tilde{d} \preceq \tilde{d}'$. Then, there exists a population that satisfies Assumptions 1-4 such that $\mathbb{E}[Y(\tilde{d})\,|\,Z=z] = n_{\tilde{d},z}$ for all $\tilde{d}$ and z.

Lemma 1 has two conditions. The first is that the values for the conditional expectations lie within the data-driven bound, and the second is that the values for the conditional expectation increases in the potential treatment $\tilde{d}$. Both represent the observable implications of MTR. The lemma establishes the sufficient and necessary condition for the possible values of the conditional expectations under MTR. Then, we can construct a population that attains the lower and upper bound in the original proposition by setting the conditional expectations to values that satisfy MIV. Lemma 1 justifies the aforementioned mechanical synthesis of the two conditions in the bound for arm-specific means by separating MTR and MIV in constructing a population. The proof proceeds by algorithmically finding conditional expectations $\mathbb{E}[Y(\tilde{d})\,|\,D=d, Z=z]$ such that $\mathbb{E}[Y(\tilde{d})\,|\,Z=z] = n_{\tilde{d},z}$ using induction on $\tilde{d}$.

Circling back to Proposition 1, the upper bound for ATEs is the same as the upper bound constructed from the individual bounds. Although showing its sharpness is a nontrivial task, more interesting is the lower bound. Its validity can be checked by observing that $\mathbb{E}[Y(d)]$ and $\mathbb{E}[Y(d')]$ can be decomposed into conditional expectations $\mathbb{E}[Y(d)\,|\,Z]$, and MTR guarantees $\mathbb{E}[Y(d)\,|\,Z] \ge \mathbb{E}[Y(d')\,|\,Z]$ for all Z. It seems that the lower bound has not been refined in the literature to my knowledge despite the simplicity of its derivation. When is the new bound different from the old bound $\max(L_d - U'_d, 0)$? The two bounds differ whenever there exists z such that $s_{d,z} > i_{d',z}$, or the lower bound of the conditional expectation at $\tilde{d} = d$, $\mathbb{E}[Y(d) \mid Z = z]$, is greater than the upper bound of the conditional expectation at $\tilde{d} = d'$, $\mathbb{E}[Y(d') \mid Z = z]$. One necessary condition is that the inner arguments cannot be increasing over all z. If they did, then $s_{d,z} = m^{-}_{d,z}\pi^{-}_{d,z} + \underline{y}(1 - \pi^{-}_{d,z}) < m^{+}_{d,z}\pi^{+}_{d,z} + \overline{y}(1 - \pi^{+}_{d,z}) = i_{d,z}$. This is the same argument that MTR alone cannot rule out the zero effect. Therefore, the lower bound is a large positive number if the individual conditional expectations $\mathbb{E}[Y(d) \mid D = d, Z = z]$ are highly nonmonotonic in z, although their weighted averages are constrained to be monotonic by MIV.

This ATE bound is also sharp. The sharpness requires some conditional expectations $\mathbb{E}[Y(d) \mid Z]$ to be at their lower bound, others at their upper bound, and in the case of the lower bound, still others at a random point in between. This implies that the arbitrariness of the values for conditional expectations in Lemma 1 play a crucial role in the sharpness argument.

MIV was combined only with MTR in Manski and Pepper (2000), but the subsequent empirical literature has often employed MIV in conjunction with another assumption that was also introduced in the same paper (Gonzalez 2005; De Haan 2011; Kreider et al. 2012). Monotone Treatment Selection (MTS) is a relaxation of the exogenous treatment assumption and can be thought of a special case of MIV when the treatment and the instrument concur. Formally, $\mathbb{E}[Y(\tilde{d}) \mid D = d] \geq \mathbb{E}[Y(\tilde{d}) \mid D = D']$ for all $\tilde{d}$ if $d \succeq d'$. This assumption states that units with a higher absolute level of potential outcomes select into higher treatments, thus largely implying a positive selection effect.

Nevertheless, Lafférs (2013) notes the confusion around the joint use of MIV and MTR in instrumental variable settings. The two assumptions restrict the conditional expectations $\mathbb{E}[Y(d) \mid D = d, Z = z]$ along different dimensions, one on their weighted averages with respect to D and the other on their weighted averages with respect to Z. Lafférs (2013) points out that researchers sometimes implicitly assume that the selection into treatment is positive after conditioning on the instrumental variable, which leads to the following definition of conditional monotone treatment selection.

Assumption 6 (cMTS: conditional monotone treatment selection)**.** $\mathbb{E}[Y(\tilde{d}) \mid D = d, Z = z] \geq \mathbb{E}[Y(\tilde{d}) \mid D = d', Z = z]$ *for all* $\tilde{d}$ *and* z *if* $d \succeq d'$*.*

Assumption 6 conceptually extends one step beyond the typical MTS but does not mathematically subsume it. MTS is a linear combination of the conditional expectation $\mathbb{E}[Y(\tilde{d}) \mid D = d, Z = z]$ and the conditional probability $\mathrm{P}[Z = z \mid D = d]$. Since the conditional probability component differs across the realized treatment D, the dominance of individual conditional expectations does not imply the dominance of their weighted average by Simpson's paradox. Lafférs (2013) gives an example where this assumption can be substantively justifiable.

Since cMTS is designed to emulate MTS after conditioning on the instrumental variable, MTR and cMTS jointly have an observable implication that mirrors that of MTR and MTS in non-instrumental variable settings. Let $m_{d,z} = \mathbb{E}[Y \mid D = d, Z = z]$, which is point identifiable from the observed data. Then, $m_{d,z}$ increases in d since if $d \succeq d'$, then $m_{d,z} = \mathbb{E}[Y \mid D = d, Z = z] = \mathbb{E}[Y(d) \mid D = d, Z = z] \geq \mathbb{E}[Y(d') \mid D = d, Z = z] \geq \mathbb{E}[Y(d') \mid D = d', Z = z] = m_{d',z}$. The inequalities hold by MTR and cMTS respectively. The monotonicity of $m_{d,z}$ is crucial in the proof of the following proposition that provides the sharp bounds under MIV, MTR and cMTS.

Proposition 2. *Under Assumptions 1, 2, 4-6, the sharp bounds for arm-specific means and average treatment effects are*

$$\mathbb{E}[Y(d)] \in [L_d, U_d], \qquad \mathbb{E}[Y(d) - Y(d')] \in [L_{d,d'}, U_d - L_{d'}]$$

where $d \succeq d'$, and

$$\begin{aligned}
s_{d,z} &= \sup_{z' \leq z} \{ m^-_{d,z'} \pi^-_{e,z'} + m_{d,z'}(1 - \pi^-_{d,z'}) \}, \\
i_{d,z} &= \inf_{z' \geq z} \{ m^+_{d,z'} \pi^+_{e,z'} + m_{d,z'}(1 - \pi^+_{d,z'}) \}, \\
L_d &= \sum_z \mathrm{P}[Z = z] \cdot s_{d,z}, \\
U_d &= \sum_z \mathrm{P}[Z = z] \cdot i_{d,z}, \\
L_{d,d'} &= \sum_z \mathrm{P}[Z = z] \cdot \max(s_{d,z} - i_{d',z}, 0).
\end{aligned}$$

Proposition 2 mirrors the previous results under MTR and MIV. There are two differences. First, Assumption 3 is dropped because potential outcomes are now bounded by the minimum and the maximum of $m_{d,z}$ for each z. Second, the extreme outcomes $\overline{y}$ and $\underline{y}$ that were in the individual upper and lower bounds are now replaced by the observed conditional mean $m_{d,z}$. This is due to cMTS. In fact, the arguments in sup and inf functions are the sharp bound under MTR and MTS if we split the observed data by z. The bounds thus mechanically synthesize MIV on one hand and MTR and cMTS on the other hand. This is possible due to the following lemma that plays a similar role to Lemma 1.

Lemma 2. *Assume $m^-_{\tilde{d},z} \pi^-_{\tilde{d},z} + m_{\tilde{d},z}(1 - \pi_{\tilde{d},z}) \leq m^+_{\tilde{d},z} \pi^+_{\tilde{d},z} + m_{\tilde{d},z}(1 - \pi^+_{\tilde{d},z})$ for all $\tilde{d}$ and z, and $m_{d,z} \geq m_{d',z}$ if $d \succeq d'$. Let $n_{\tilde{d},z}$ be constants that satisfy*

$$m^-_{\tilde{d},z} \pi^-_{\tilde{d},z} + m_{\tilde{d},z}(1 - \pi_{\tilde{d},z}) \leq n_{\tilde{d},z} \leq m^+_{\tilde{d},z} \pi^+_{\tilde{d},z} + m_{\tilde{d},z}(1 - \pi^+_{\tilde{d},z}),$$

and $n_{\tilde{d},z} \leq n_{\tilde{d}',z}$ whenever $\tilde{d} \preceq \tilde{d}'$. Then, there exists a population that satisfies Assumptions 1, 2, 4 and 6 such that $\mathbb{E}[Y(\tilde{d}) \mid Z = z] = n_{\tilde{d},z}$ for all $\tilde{d}$ and z.

We have another rank restriction on $m_{d,z}$ here, but this is to ensure the minimal compatibility between the data and the MTR and cMTS assumptions rather than to restrict the possible values of conditional expectations. Although the proof employs a different algorithm than before as the current rank restrictions on the conditional expectations are two-way, the proof for the proposition is structured around the lemma in the same manner.

3 Empirical Example

This section replicates the findings of Gonzalez (2005) that studies the wage effect of language skills for Hispanic workers in the US. The literature struggled to disentangle the effect of language skills from the effect of unobserved worker capability that is potentially correlated with language skills. The study suggested that the Manski-style bounds can produce conservative estimates of the desired effect and proposed using the age at which the worker arrived in the US for a monotone instrumental variable. The language skills and the age at arrival are each broken down into five and four categories.

The above identifying assumptions map to this empirical setting as follows. Assumption 1 states that the wage dose not depend on the treatment and instrument distribution of other individuals. Assumption 2 further restricts that the instrument cannot affect the wage independently of the language skills. Assumption 1 may be violated in general equilibrium analyses. Assumption 2 may be violated if the age at arrival affects unobserved worker capability that is not perfectly correlated with the language skills and, in turn, the wage. Since the potential correlation between the two motivated the use of the partial identification approach in this study, the alternative model of partial instrumental variables, monotone direct effect of the instrumental variable, seems more suitable here.

Assumption 3 says that the wage cannot be infinitely large, and Assumption 4 rules out the possibility that the wage decreases as the same individual has higher language skills. These assumptions seem reasonable although where to set the upper bound of the wage might not be straightforward. Assumption 5 implies that the individuals who arrived the US early have higher potential outcomes than those who arrived the US late at every potential language skill level. This potentially depends on the factors that affect the migration decision of the families. Finally, Assumption 6 posits that for the individuals who arrived at the US at the same age, those who have acquired higher language skills have higher potential outcomes than those who have not at every potential language skill level. The interpretation of cMTS is arguably easier than MTS as it conditions out the selection process in migration. Overall, the biggest threat to identification is the violation of the exclusion restriction. I proceed under this assumption for the comparison of the loose bounds and the sharp bounds.

The analysis uses the 1% Public Use Micro Sample (PUMS) of the 1990 US Census. The data considers all individuals with a Hispanic origin between the age of 16 and 64. Unemployed individuals are dropped. The outcome variable is the natural log of hourly wage, and the upper bound and the lower bound are set at 5 and 1. The treatment is the five answer categories to the question on ability to speak English: not at all, not well, well, very well, only English at home. The instrument is constructed from the question on the year of birth and the year of entry in the United States, which takes four possible values: US born, arrived as a child (0 to 11), arrived as a teenager (12 to 17), arrived as an adult (18 or older). Since the data does not record the exact year of the entry, I take the midpoint of each two-to-four year interval to calculate the age at arrival. Details of data construction can be found in Appendix.

Since the sample does not cover the entire population, bounds have to be estimated to make inference at the population level. However, naive plug-in analogues that replace expectations in the bounds with sample means are known to be biased (Manski and Pepper 2009), and there are more principled finite-sample bias correction methods as well as bootstrap techniques (Andrews and Shi 2013; Chernozhukov, Lee and Rosen 2013). I abstract from this estimation problem and use the biased plug-in estimators for the purpose of the term paper due to a lack of time. I note that this is perfectly fine for the comparison because Gonzalez (2005) also did not correct the finite-sample bias for MIV results.

Table 1: Sharp bounds for arm-specific means and ATEs under various assumptions

	Original paper		MIV+MTR		MIV+MTR+cMTS	
Quantity	UB	LB	UB	LB	UB	LB
$\mathbb{E}[y(1)]$	1.046	1.932	1.054	1.812	1.679	2.008
$\mathbb{E}[y(2)]$	1.166	2.209	1.189	2.113	1.838	2.014
$\mathbb{E}[y(3)]$	1.339	2.681	1.36	2.257	1.948	2.052
$\mathbb{E}[y(4)]$	1.767	3.221	1.777	2.646	2.014	2.1
$\mathbb{E}[y(5)]$	2.016	4.433	2.034	4.41	2.034	2.178
$\mathbb{E}[y(5) - y(4)]$	0	1.162	0	2.633	0	0.164
$\mathbb{E}[y(4) - y(3)]$	0	1.515	0	1.286	0	0.153
$\mathbb{E}[y(3) - y(2)]$	0	1.882	0	1.068	0	0.215
$\mathbb{E}[y(2) - y(1)]$	0	2.666	0	1.059	0	0.335
$\mathbb{E}[y(4) - y(1)]$	0	2.175	0	1.593	0.007	0.421
$\mathbb{E}[y(5) - y(1)]$	0.084	3.387	0.222	3.357	0.026	0.499

Note: Finite-sample biases are not corrected. Columns 3-6 report the sharp bounds derived in this paper. The last two columns drops all observations with $Z = 2$, or those arrives as a teenager as they do not satisfy the observable implications.

Table 1 reports the bounds under MIV and MTR in the original paper and their replicated values. The arm-specific means are replicated closely enough. The error seems to be due to the choices made in preprocessing as the uploaded replication file does not contain the final variables used in their analysis. The difference in the upper bounds is harder to understand because the sharp bound can be directly derived from the bounds for arm-specific means. The author does not describe what bound they used in the paper for average treatment effects. Our interest is whether there is any improvement in the lower bound. The original paper ruled out zero effect in only one ATE, and the effect was marginal. The replicated bounds also ruled out zero effect in only one ATE. However, this is not due to the theoretical improvement in the sharp lower bound because 0.222 is the difference between the lower bound of $y(5)$ and the upper bound of $y(1)$. The improvement is not present because the individual bound for the conditional expectation $\mathbb{E}[Y(\tilde{d}) \mid D, Z]$ was fairly wide, ranging one to four for each expectation.

Table 2 reports the observed conditional means $m_{d,z} = \mathbb{E}[Y(d) \mid D = d, Z = z]$. As pointed out earlier, these quantities must increase in d if MIV, MTR nad cMTS jointly hold. Note that the expectations are monotonic except when $Z = 2$. This suggests that cMTS might not hold in conjunction with the other two assumptions in this data because the observable implications of MIV and MTR is that the derived upper bound is larger than the derived lower bound, which the above results corroborate. The last two columns in Table 1 reports the bounds after dropping these individuals, and the derived upper bound is

Table 2: Observed means conditional on the treatment and the instrument

	$Z = 1$	$Z = 2$	$Z = 3$	$Z = 4$
$D = 1$	1.679	1.647	1.646	1.556
$D = 2$	1.849	1.755	1.778	1.788
$D = 3$	1.983	1.931	1.984	1.994
$D = 4$	2.083	1.796	2.093	2.105
$D = 5$	2.166	1.539	2.212	2.236

greater than the derived lower bound for every specification. However, cMTS in this data set was not strong enough to further reduce the lower bound from the difference between the arm-specific bounds. The upper bound for $\mathbb{E}[y(5) - y(1)]$ is greater under MIV and MTR than MIV, MTR and cMTS. This is because we lost individuals with high identification power in trimming the sample. Overall, this data shows that the sharp lower bounds derived above do not have much practical significance if the individual conditional expectations are not tightly bounded enough.

References

Andrews, Donald WK and Xiaoxia Shi. 2013. "Inference based on conditional moment inequalities." *Econometrica* 81(2):609–666.

Chernozhukov, Victor, Sokbae Lee and Adam M Rosen. 2013. "Intersection bounds: Estimation and inference." *Econometrica* 81(2):667–737.

De Haan, Monique. 2011. "The effect of parents' schooling on child's schooling: a nonparametric bounds analysis." *Journal of Labor Economics* 29(4):859–892.

Duarte, Guilherme, Noam Finkelstein, Dean Knox, Jonathan Mummolo and Ilya Shpitser. 2024. "An automated approach to causal inference in discrete settings." *Journal of the American Statistical Association* 119(547):1778–1793.

Giustinelli, Pamela. 2011. "Non-parametric bounds on quantiles under monotonicity assumptions: with an application to the Italian education returns." *Journal of Applied Econometrics* 26(5):783–824.

Gonzalez, Libertad. 2005. "Nonparametric bounds on the returns to language skills." *Journal of Applied Econometrics* 20(6):771–795.

Kim, Wooyoung, Koohyun Kwon, Soonwoo Kwon and Sokbae Lee. 2018. "The identification power of smoothness assumptions in models with counterfactual outcomes." *Quantitative Economics* 9(2):617–642.

Kreider, Brent, John V Pepper, Craig Gundersen and Dean Jolliffe. 2012. "Identifying the effects of SNAP (food stamps) on child health outcomes when participation is endogenous and misreported." *Journal of the American Statistical Association* 107(499):958–975.

Laffers, Lukáš. 2013. "A note on bounding average treatment effects." *Economics Letters* 120(3):424–428.

Lafférs, Lukáš. 2019. "Bounding average treatment effects using linear programming." *Empirical economics* 57:727–767.

Manski, Charles F. 1997. "Monotone treatment response." *Econometrica: Journal of the Econometric Society* pp. 1311–1334.

Manski, Charles F. 2003. *Partial identification of probability distributions.* Springer Science & Business Media.

Manski, Charles F and John V Pepper. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica* 68(4):997–1010.

Manski, Charles F and John V Pepper. 2009. "More on monotone instrumental variables." *The Econometrics Journal* 12(suppl_1):S200–S216.

Mogstad, Magne, Andres Santos and Alexander Torgovitsky. 2018. "Using instrumental variables for inference about policy relevant treatment parameters." *Econometrica* 86(5):1589–1619.

Twinam, Tate. 2017. "Complementarity and identification." *Econometric Theory* 33(5):1154–1185.

Proof of Lemma 1

Define the population as

$$Y(\tilde{d})\,|\,D=e,\,Z=z=\begin{cases}Y_{e,z}+w^{+}_{\tilde{d},e,z}(\overline{y}-Y_{e,z}) & \tilde{d}\succ e\\ Y_{e,z} & \tilde{d}=e\\ Y_{e,z}+w^{-}_{\tilde{d},e,z}(\underline{y}-Y_{e,z}) & \tilde{d}\prec e.\end{cases} \tag{2}$$

where $w^{+}_{\tilde{d},e,z}$, $w^{-}_{\tilde{d},e,z}\in[0,1]$, and $w^{+}_{\tilde{d},e,z}$ is increasing in $\tilde{d}$ and $w^{-}_{\tilde{d},e,z}$ is decreasing in $\tilde{d}$. This population satisfies Assumptions 1-4. It suffices to show that it is possible to find weights $w^{+}_{\tilde{d},e,z}$, $w^{-}_{\tilde{d},e,z}$ that satisfy the conditional expectation condition.

Express $w^{+}_{\tilde{d},e,z}=\sum_{e\prec k\preceq\tilde{d}}v^{+}_{k,e,z}$ and $w^{-}_{\tilde{d},e,z}=\sum_{\tilde{d}\preceq k\prec e}v^{-}_{k,e,z}$. Then, each $v^{+}_{k,e,z}$ and $v^{-}_{k,e,z}$ is non-negative and

$$\sum_{e\prec k}v^{+}_{k,e,z}\leq 1,\qquad \sum_{k\prec e}v^{-}_{k,e,z}\leq 1$$

for all e. Denote the complements of the sum of the coefficients as $v^{-}_{0,D,z}\equiv 1-\sum_{k\prec D}v^{-}_{k,D,z}$ and $v^{+}_{\overline{d},D,z}\equiv 1-\sum_{k\succ D}v^{+}_{k,D,z}$. Defining $\mathbb{E}[Y\,|\,D=e,\,Z=z]=y_{e,z}$, we obtain

$$\mathbb{E}[Y(\tilde{d})\,|\,D=e,\,Z=z]=\begin{cases}y_{e,z}+\sum_{e\prec k\preceq\tilde{d}}v^{+}_{k,e,z}(\overline{y}-y_{e,z}) & \tilde{d}\succ e\\ y_{e,z} & \tilde{d}=e\\ y_{e,z}+\sum_{\tilde{d}\preceq k\prec e}v^{-}_{k,e,z}(\underline{y}-y_{e,z}) & \tilde{d}\prec e\end{cases}$$

and

$$n_{\tilde{d},z}=\mathbb{E}\left[y_{D,z}+\mathbb{I}[\tilde{d}\succ D]\sum_{D\prec k\preceq\tilde{d}}v^{+}_{k,D,z}(\overline{y}-y_{D,z})+\mathbb{I}[\tilde{d}\prec D]\sum_{\tilde{d}\preceq k\prec D}v^{-}_{k,D,z}(\underline{y}-y_{D,z})\,\middle|\,Z=z\right].$$

With abuse of notation, let $\tilde{d}=1,,\,2,\,\cdots,\,\overline{d}$ and rewrite the above equation as

$$n_{\tilde{d}+1,z}-n_{\tilde{d},z}=\mathbb{E}\left[\mathbb{I}[\tilde{d}+1\succ D]\cdot v^{+}_{\tilde{d}+1,D,z}(\overline{y}-y_{D,z})+\mathbb{I}[\tilde{d}\prec D]\cdot v^{-}_{\tilde{d},D,z}(y_{D,z}-\underline{y})\,\middle|\,Z=z\right].$$

This relation holds even for $\tilde{d}=0$ and $\tilde{d}=\overline{d}$ if we define pseudo conditional outcomes as $n_{0,z}=\underline{y}$ and $n_{\overline{d}+1,z}=\overline{y}$. The difference can be rewritten into a matrix form:

$$\begin{pmatrix}v^{-}_{0,1,z} & v^{-}_{0,2,z} & \cdots & v^{-}_{0,\overline{d},z} & 0 & 0 & \cdots & 0\\ 0 & v^{-}_{1,2,z} & \cdots & v^{-}_{1,\overline{d},z} & v^{+}_{2,1,z} & 0 & \cdots & 0\\ & & & \vdots & & & & \\ 0 & 0 & \cdots & 0 & v^{+}_{\overline{d},1,z} & v^{+}_{\overline{d},2,z} & \cdots & v^{+}_{\overline{d},\overline{d},z}\end{pmatrix}\begin{pmatrix}p_{1,z}(y_{1,z}-\underline{y})\\ \vdots\\ p_{\overline{d},z}(y_{\overline{d},z}-\underline{y})\\ p_{1,z}(\overline{y}-y_{1,z})\\ \vdots\\ p_{\overline{d},z}(\overline{y}-y_{\overline{d},z})\end{pmatrix}=\begin{pmatrix}n_{1,z}-n_{0,z}\\ \vdots\\ n_{\overline{d}+1,z}-n_{\overline{d},z}\end{pmatrix}$$

where $p_{d,z}=\mathrm{P}[D=d\,|\,Z=z]$ and each column of the coefficient matrix sums to one. We use the following algorithm to find a solution. The elements will be filled in sequentially starting the first row.

- Step 1: Assign $v^{-}_{0,1,z}=1$, and to the rest arbitrary numbers in $[0,\,1]$ so that the first equation holds.

- Step 2: Suppose the first $\tilde{d}$ rows are filled ($\tilde{d} < \overline{d}$). Assign $v^-_{\tilde{d},\tilde{d}+1,z} = 1 - \sum_{k=0}^{\tilde{d}-1} v_{k,\tilde{d}+1,z}$ and to 'the rest' arbitrary numbers in [0, 1] so that the ($\tilde{d}$ + 1)-th equation holds and the sum of each column does not exceed one.

- Step 3: The above step would be impossible if the right hand side is smaller than the left hand side when 'the rest' are all zero, or the right hand side is larger than the left hand side when 'the rest' are as large as possible.

- Step 3-1: Suppose $v^-_{\tilde{d},\tilde{d}+1,z}$ is too large. Name the column in which $v^-_{\tilde{d},\tilde{d}+1,z}$ is as the *pivot*. For the columns to the right of the pivot, reduce all elements so that the first $\tilde{d}$ elements in the pivot can inflate to satisfy the equations. This will in turn reduce $v^-_{\tilde{d},\tilde{d}+1,z}$. The sum of each column must not exceed one and all coefficients must remain non-negative in the process. Let $\min v^-_{\tilde{d},\tilde{d}+1,z} = v$. If $v = 0$, then the left hand side can be smaller than the right hand side, solved. If $v \neq 0$, this means that all the right-side elements are zero. Summing the first $\tilde{d} + 1$ rows, the minimum of the left hand side is

$$\sum_{k=1}^{\tilde{d}+1} p_{k,z}(y_{k,z} - \underline{y}),$$

  which cannot be larger than the right hand side $n_{\tilde{d}+2,z} - n_{0,z}$ by condition. Find $v^-_{\tilde{d},\tilde{d}+1,z}$ that attains the equality with 'the rest' all being zero and assign.

- Step 3-2: Suppose $v^-_{\tilde{d},\tilde{d}+1,z}$ is too small. Do the contrary of Step 3-1, which is to inflate all elements on the right of the pivot, deflate the first $\tilde{d}$ elements in the pivot and inflate $v^-_{\tilde{d},\tilde{d}+1,z}$. Let $\max v^-_{\tilde{d},\tilde{d}+1,z} = v$. If $v = 1$, then if 'the rest' are maximally large, then summing the first $\tilde{d} + 1$ rows, the maximum of the left hand side is

$$\sum_{k=1}^{\overline{d}} p_{k,z}(y_{k,z} - \underline{y}) + \sum_{k=1}^{\tilde{d}} p_{k,z}(\overline{y} - y_{k,z}).$$

  This cannot be smaller than the right hand side $n_{\tilde{d}+2,z} - n_{0,z}$ by condition. If $v \neq 1$, then this implies that the sum of the columns on the right are all ones already, so we get the same maximum of the left hand side summing the first $\tilde{d} + 1$ rows. Find $v^-_{\tilde{d},\tilde{d}+1,z}$ that attains the equality with 'the rest' all being as large as possible and assign.

- Step 4: Repeat Step 2 and 3 until the second last row is filled out. Then, the column condition automatically fills out the last row, and the equation automatically holds because the sum of all $\tilde{d} + 2$ rows should be equal on both sides.

Proof of Proposition 1

For sharpness of the lower bound, define $n_{\tilde{d},z} = \sup_{z' \leq z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + \underline{y}(1-\pi^-_{\tilde{d},z'})\}$. Since $n_{\tilde{d},z}$ replaces less observations with the lower bound as $\tilde{d}$ increases, $n_{\tilde{d},z}$ increases in $\tilde{d}$ by construction, and

$$\begin{aligned}
& m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + \underline{y}(1-\pi^-_{\tilde{d},z'}) \\
\leq & \sup_{z' \leq z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + \underline{y}(1-\pi^-_{\tilde{d},z'})\} = n_{\tilde{d},z} \\
\leq & \inf_{z' \geq z}\{m^+_{\tilde{d},z'}\pi^+_{\tilde{d},z'} + \underline{y}(1-\pi^+_{\tilde{d},z'})\} \\
\leq & m^+_{\tilde{d},z'}\pi^+_{\tilde{d},z'} + \underline{y}(1-\pi^+_{\tilde{d},z'}).
\end{aligned}$$

The second inequality is from expression (1). As the conditions of Lemma 1 are met, we can find a population that satisfies $\mathbb{E}[Y(\tilde{d})\,|\,Z=z] = n_{\tilde{d},z}$ for all $\tilde{d}$ and z. Meanwhile, the population satisfies MIV since $n_{\tilde{d},z}$ increases in z. The arm-specific mean recovers the lower bound. This concludes the proof. The sharpness for the upper bound can be analogously shown.

Next, we prove the validity and sharpness of the bound for average treatment effects $\mathbb{E}[Y(d) - Y(d')]$. Validity can be checked by observing that

$$\mathbb{E}[Y(d)] = \sum_z \mathrm{P}[Z=z] \cdot n_{\tilde{d},z}, \quad \mathbb{E}[Y(d')] = \sum_z \mathrm{P}[Z=z] \cdot n_{\tilde{d}',z},$$

and MTR guarantees $Y(d) \geq Y(d')$ in each observation, thus $n_{\tilde{d},z} \geq n_{\tilde{d}'z}$. For sharpness, first consider the upper bound. Define

$$n_{\tilde{d},z} = \begin{cases} \sup_{z' \leq z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + \underline{y}(1-\pi^-_{\tilde{d},z'})\} & \tilde{d} \preceq d' \\ \inf_{z' \geq z}\{m^+_{\tilde{d},z'}\pi^+_{\tilde{d},z'} + \overline{y}(1-\pi^+_{\tilde{d},z'})\} & \tilde{d} \succ d'. \end{cases}$$

We can find a population that satisfies $\mathbb{E}[Y(\tilde{d})\,|\,Z=z] = n_{\tilde{d},z}$ for all $\tilde{d}$ and z since $n_{\tilde{d},z}$ is increasing in $\tilde{d}$ and $n_{\tilde{d},z}$ satisfies the bound in Lemma 1, of which the proof is identical to the arm-specific mean case. The monotonicity around $\tilde{d} = d$ is guaranteed by

$$\begin{aligned}
\sup_{z' \leq z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + \underline{y}(1-\pi^-_{\tilde{d},z'})\} \leq & \sup_{z' \leq z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + \underline{y}(1-\pi^-_{\tilde{d},z'})\} \\
\leq & \inf_{z' \geq z}\{m^+_{\tilde{d},z'}\pi^+_{\tilde{d},z'} + \overline{y}(1-\pi^+_{\tilde{d},z'})\}
\end{aligned}$$

for any $\tilde{d}' \prec \tilde{d}$. B* and MTR thus follow. MIV holds by the monotocinity of $n_{\tilde{d},z}$ in z. Finally, the population attains the upper bound of the ATE.

For the lower bound, with auxiliary variables defined for fixed d, d',

$$\alpha_{d',z} = \inf_{z' \geq z}\{m^+_{d',z'}\pi^+_{d',z'} + \overline{y}(1-\pi^+_{d',z'})\}, \quad \alpha_{d,z} = \sup_{z' \leq z}\{m^-_{d,z'}\pi^-_{d,z'} + \underline{y}(1-\pi^-_{d,z'})\},$$

define

$$\begin{aligned}
& n_{d',z} = \alpha_{d',z},\ n_{d,z} = \alpha_{d,z} & & (\alpha_{d',z} \leq \alpha_{d,z}) \\
& n_{d',z} = n_{d,z} = \alpha_{d',z} & & (\alpha_{d',z} > \alpha_{d,z})
\end{aligned}$$

and the rest of $n_{\tilde{d},z}$ as

$$n_{\tilde{d},z} = \begin{cases} \sup_{z' \le z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + \underline{y}(1-\pi^-_{\tilde{d},z'})\} & \tilde{d} \prec d' \\ \min[\max[\sup_{z' \le z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + \underline{y}(1-\pi^-_{\tilde{d},z'})\}, n_{d',z}], n_{d,z}] & d' \prec \tilde{d} \prec d \\ \inf_{z' \ge z}\{m^+_{\tilde{d},z'}\pi^+_{\tilde{d},z'} + \overline{y}(1-\pi^+_{\tilde{d},z'})\} & \tilde{d} \succ d. \end{cases}$$

First, we show that $n_{d',z} \le n_{d,z}$ and $n_{d',z}$ and $n_{d,z}$ are increasing in z. The first inequality holds by definition. For the second inequality, note that $\alpha_{d',z}$ and $\alpha_{d,z}$ increase in z. Therefore, $n_{d',z} < n_{d',z'}$ and $n_{d,z} < n_{d,z'}$ if the order of the alphas is the same under z and z' ($z < z'$). Suppose $\alpha_{d',z} \le \alpha_{d,z}$ and $\alpha_{d',z} > \alpha_{d,z}$. Then, $n_{d',z} = \alpha_{d',z} \le \alpha_{d',z'} = n_{d',z'}$ and $n_{d,z} = \alpha_{d,z} < \alpha_{d,z'} < \alpha_{d,z'} = n_{d',z}$. The case is analogous when $\alpha_{d',z} > \alpha_{d,z}$ and $\alpha_{d',z} \le \alpha_{d,z}$.

Next, we show that the entire sequence $\{n_{\tilde{d},z}\}$ increases in $\tilde{d}$. It increases in each interval because sup and inf increase in $\tilde{d}$. On $\tilde{d} = d'-$,

$$\begin{aligned} n_{d',z} &= \inf_{z' \ge z}\{m^+_{d',z'}\pi^+_{e,z'} + \overline{y}(1-\pi^+_{d',z'})\} \\ &\ge \inf_{z' \ge z}\{m^+_{\tilde{d},z'}\pi^+_{e,z'} + \overline{y}(1-\pi^+_{\tilde{d},z'})\} \\ &\ge \sup_{z' \le z}\{m^-_{\tilde{d},z'}\pi^-_{e,z'} + \overline{y}(1-\pi^-_{\tilde{d},z'})\}, \end{aligned}$$

and on $\tilde{d} = d+$,

$$n_{d,z} \le \inf_{z' \ge z}\{m^+_{d',z'}\pi^+_{d',z'} + \overline{y}(1-\pi^+_{d',z'})\} \le \inf_{z' \ge z}\{m^+_{d,z'}\pi^+_{d,z'} + \overline{y}(1-\pi^+_{edz'})\}.$$

On $\tilde{d} = d'+$ and $\tilde{d} = d-$, the monotonicity is due to the definition.

$n_{\tilde{d},z}$ also satisfy the bound in Lemma 1. $\tilde{d} \prec d'$ and $\tilde{d} \succ d$ are identical to the arm-specific mean case. For $d' \prec \tilde{d} \prec d$, since $n_{d,z} \ge \alpha_{d,z} \ge \sup_{z' \le z}\{m^-_{d,z}\pi^-_{d,z} + \underline{y}(1-\pi^-_{d,z})\}$,

$$\begin{aligned} n_{\tilde{d},z} &\ge \min\left[\sup_{z' \le z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + \underline{y}(1-\pi^-_{\tilde{d},z'})\}, \sup_{z' \le z}\{m^-_{d,z}\pi^-_{d,z} + \underline{y}(1-\pi^-_{d,z})\}\right] \\ &= \sup_{z' \le z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + \underline{y}(1-\pi^-_{\tilde{d},z'})\} \\ &\ge m^-_{\tilde{d},z}\pi^-_{\tilde{d},z} + \underline{y}(1-\pi^-_{\tilde{d},z}). \end{aligned}$$

The other side can be analogously shown. By Lemma 1, we can find a population that satisfies $\mathbb{E}[Y(\tilde{d}) \mid Z = z] = n_{\tilde{d},z}$ for all $\tilde{d}$ and z. This population also satisfies MIV because all arguments in the definition of $n_{\tilde{d},z}$ increase in z. Finally, it attains the lower bound.

Proof of Lemma 2

For any numbers $y_{\tilde{d},d,z} \in [\underline{y}, \overline{y}]$ ($\tilde{d} \neq d$), we can always find a population that satisfies $y_{\tilde{d},d,z} = \mathbb{E}[Y(\tilde{d}) \mid D = d, Z = z]$ simultaneously for all $(\tilde{d}, d, z)$ by shrinking observed outcomes $Y(d) \mid D = d, Z = z$ towards either $\underline{y}$ or $\overline{y}$ similarly as in equation (2). It suffices to show that there exist conditional expectations $y_{\tilde{d},d,z}$ that are (i) compatible with data: $y_{d,d,z} = \mathbb{E}[Y \mid D = d, Z = z] = m_{d,z}$, (ii) achieve the overall conditional expectation: $\mathbb{E}[y_{\tilde{d},D,z} \mid Z = z] = n_{\tilde{d},z}$, and (iii) increase in $\tilde{d}$ and d conditionally on z.

With abuse of notation, express the overall conditional expectation conditions in a matrix form:

$$\begin{pmatrix} y_{1,1,z} & y_{1,2,z} & \cdots & y_{1,\overline{d},z} \\ y_{2,1,z} & y_{2,2,z} & \cdots & y_{2,\overline{d},z} \\ & & \cdots & \\ y_{\overline{d},1,z} & y_{\overline{d},2,z} & \cdots & y_{\overline{d},\overline{d},z} \end{pmatrix} \begin{pmatrix} p_{1,z} \\ p_{2,z} \\ \vdots \\ p_{\overline{d},z} \end{pmatrix} = \begin{pmatrix} n_{1,z} \\ n_{2,z} \\ \vdots \\ n_{\overline{d},z} \end{pmatrix}$$

where $p_{d,z} = \mathrm{P}[D = d \mid Z = z]$. Compatibility fixes the diagonal terms of the coefficient matrix, and MTR and cMTS constraints the horizontal and vertical order relationship of the coefficients. We will find a similar algorithmic solution. The off-diagonal elements will be filled in sequentially starting the first row.

- Step 1: Assign to the first row arbitrary numbers so that the first equation holds.
- Step 2: Suppose the first $\tilde{d}$ rows are filled ($\tilde{d} < \overline{d}$). The next row can be filled if the $(\tilde{d}+1)$th element is at its upper bound, *i.e.*, $y_{\tilde{d},\tilde{d}+1,z} = y_{\tilde{d}+1,\tilde{d}+1,z}$, or the first $\tilde{d}$ elements of the $\tilde{d}$th row are at their lower bound and the rest of them are smaller than $y_{\tilde{d}+1,\tilde{d}+1,z}$, *i.e.*, $y_{\tilde{d},k,z} = y_{k,k,z}$ for all $k \leq \tilde{d}$ and $y_{\tilde{d},k,z} < y_{\tilde{d}+1,\tilde{d}+1,z}$ for all $k > \tilde{d}$. Let us call this condition (*). The crucial key is $\max y_{\tilde{d},\tilde{d}+1,z}$. We will see that if the $\tilde{d}$th row increases to the right of $y_{\tilde{d},\tilde{d}+1,z}$, or $y_{\tilde{d},k,z} > y_{\tilde{d},\tilde{d}+1,z}$ for some $k > \tilde{d}+1$, then $y_{\tilde{d},\tilde{d}+1,z}$ is either not maximized or is already at $y_{\tilde{d}+1,\tilde{d}+1,z}$.
- Step 2-1: Assume that the $\tilde{d}$th row increases to the right of $y_{\tilde{d},\tilde{d}+1,z}$. Find the smallest k such that $y_{\tilde{d},k,z} > y_{\tilde{d},k-1,z}$. Since $y_{\tilde{d},k-1,z} \geq y_{\tilde{d}-1,k-1,z}$, at least one of $y_{\tilde{d},k,z} > y_{\tilde{d}-1,k,z}$ and $y_{\tilde{d}-1,k,z} > y_{\tilde{d}-1,k-1,z}$ must be true. If $y_{\tilde{d},k,z} > y_{\tilde{d}-1,k,z}$, then we can infinitesimally increase $(y_{\tilde{d},\tilde{d}+1,z}, \cdots, y_{\tilde{d},k-1,z})$ and infinitesimally decrease $y_{\tilde{d},k,z}$ without violating the order conditions, so $y_{\tilde{d},\tilde{d}+1,z}$ has not been maximized. If $y_{\tilde{d},k-1,z} > y_{\tilde{d}-1,k-1,z}$, then repeat the process until one finds some $\underline{d}$ such that $y_{\underline{d},k,z} > y_{\underline{d}-1,k,z}$, or $y_{d,k,z} > y_{d,k-1,z}$ for all d. In the first case, we can infinitesimally increase $(y_{d,\tilde{d}+1,z}, \cdots, y_{d,k-1,z})$ and infinitesimally decrease $y_{d,k,z}$ for all $\underline{d} \leq d \leq \tilde{d}$. In the second case, we can infinitesimally increase $(y_{d,\tilde{d}+1,z}, \cdots, y_{d,k-1,z})$ and infinitesimally decrease $y_{d,k,z}$ for all d. The increasing and decreasing margins are common for all corresponding d due to the common weights $p_{d,z}$. Since we have $y_{\tilde{d},\tilde{d}+1,z} = \cdots = y_{\tilde{d},k-1,z}$, this adjustment is always possible as long as $y_{\tilde{d},\tilde{d}+1,z} < y_{\tilde{d}+1,\tilde{d}+1,z}$.
- Step 3: Repeat Step 2-1 until we have either $y_{\tilde{d},\tilde{d}+1,z} = y_{\tilde{d}+1,\tilde{d}+1,z}$ or $y_{\tilde{d},\tilde{d}+1,z} = \cdots = y_{\tilde{d},\overline{d},z}$. If $y_{\tilde{d},\tilde{d}+1,z} = y_{\tilde{d}+1,\tilde{d}+1,z}$, then move onto the next row. If $y_{\tilde{d},\tilde{d}+1,z} = \cdots = y_{\tilde{d},\overline{d},z} < y_{\tilde{d}+1,\tilde{d}+1,z}$, then shrink all below-diagonal elements towards their lower bounds and inflate all above-diagonal elements towards their upper bounds. First, try replacing all below-diagonal elements $y_{d,k,z}$ with $y_{k,k,z}$ and all above-diagonal elements $y_{d,k,z}$ with $y_{d,k,z} + \varepsilon_d$ where ε_d are constants that hold the equations true. If $y_{\tilde{d},k,z} + \varepsilon_{\tilde{d}} > y_{\tilde{d}+1,\tilde{d}+1,z}$, then replace below-diagonal elements with $y_{d,k,z} + w(y_{d,k,z} - y_{k,k,z})$ for some common weight $w \in (0, 1]$ so that $y_{\tilde{d},k,z} + \varepsilon_{\tilde{d}} = y_{\tilde{d}+1,\tilde{d}+1,z}$. Note that ε_d is increasing in d in both cases since we replace less below-diagonal elements in higher rows and the margin by which the weighted sum of the below-diagonal elements decrease is increasing in the row number. This shows that the transformation preserves the vertical rank of elements. It also preserves the horizontal rank since below-diagonal elements shrink by the same proportion where the two endpoints

of the shrinkage, $y_{d,k,z}$ and $y_{k,k,z}$, horizontally increase, and above-diagonal elements inflate by the same margin. Therefore, we have condition (*) and the $(\tilde{d}+1)$th row can be filled in.

Proof of Proposition 2

The proof generally follows that of Proposition 1. We first prove the validity and sharpness of the bound for arm-specific means $\mathbb{E}[Y(d)]$. For validity, define $n_{\tilde{d},z} = \mathbb{E}[Y(\tilde{d}) \mid Z = z]$. Then,

$$\begin{aligned} n_{\tilde{d},z} &= \sum_{e} \mathbb{E}[Y(\tilde{d}) \mid D = e, Z = z] \cdot \mathrm{P}[D = e \mid Z] \\ &= \sum_{e \succ \tilde{d}} \mathbb{E}[Y(\tilde{d}) \mid D = e, Z = z] \cdot \mathrm{P}[D = e \mid Z] + \sum_{e \preceq \tilde{d}} \mathbb{E}[Y(\tilde{d}) \mid D = e, Z = z] \cdot \mathrm{P}[D = e \mid Z] \\ &\geq \sum_{e \succ \tilde{d}} \mathbb{E}[Y(\tilde{d}) \mid D = \tilde{d}, Z = z] \cdot \mathrm{P}[D = e \mid Z] + \sum_{e \preceq \tilde{d}} \mathbb{E}[Y(e) \mid D = e, Z = z] \cdot \mathrm{P}[D = e \mid Z] \\ &= m^-_{\tilde{d},z} \pi^-_{\tilde{d},z} + m_{\tilde{d},z}(1 - \pi_{\tilde{d},z}). \end{aligned}$$

Similarly, $n_{\tilde{d},z} \leq m^+_{\tilde{d},z} \pi^+_{\tilde{d},z} + m_{\tilde{d},z}(1 - \pi^+_{\tilde{d},z})$. MIV implies that $n_{\tilde{d},z}$ increases in z. Therefore, $\sup_{z' \leq z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + m_{\tilde{d},z'}(1 - \pi^-_{\tilde{d},z'})\} \leq n_{\tilde{d},z} \leq \inf_{z' \geq z}\{m^+_{\tilde{d},z'}\pi^+_{\tilde{d},z'} + m_{\tilde{d},z'}(1 - \pi^+_{\tilde{d},z'})\}$, and $\mathbb{E}[Y(d)] = \sum_z \mathrm{P}[Z = z] \cdot n_{\tilde{d},z} \in [L_d, U_d]$. An implication is

$$\sup_{z' \leq z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + m_{\tilde{d},z'}(1 - \pi^-_{\tilde{d},z'})\} \leq \inf_{z' \geq z}\{m^+_{\tilde{d},z'}\pi^+_{\tilde{d},z'} + m_{\tilde{d},z'}(1 - \pi^+_{\tilde{d},z'})\}. \tag{3}$$

For sharpness of the lower bound, define $n_{\tilde{d},z} = \sup_{z' \leq z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + m_{\tilde{d},z'}(1 - \pi^-_{\tilde{d},z'})\}$. $n_{\tilde{d},z}$ increases in $\tilde{d}$ since $n_{\tilde{d},z}$ replaces less observations with the respective lower bound $m_{\tilde{d},z'}$ as $\tilde{d}$ increases, and the lower bound itself increases in $\tilde{d}$ as well, which must be true under MRT and cMTS. Furthermore,

$$\begin{aligned} & m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + m_{\tilde{d},z}(1 - \pi^-_{\tilde{d},z'}) \\ \leq & \sup_{z' \leq z}\{m^-_{\tilde{d},z'}\pi^-_{\tilde{d},z'} + m_{\tilde{d},z'}(1 - \pi^-_{\tilde{d},z'})\} = n_{\tilde{d},z} \\ \leq & \inf_{z' \geq z}\{m^+_{\tilde{d},z'}\pi^+_{\tilde{d},z'} + m_{\tilde{d},z'}(1 - \pi^+_{\tilde{d},z'})\} \\ \leq & m^+_{\tilde{d},z'}\pi^+_{\tilde{d},z'} + m_{\tilde{d},z}(1 - \pi^+_{\tilde{d},z'}). \end{aligned}$$

The second inequality is from expression (3), which is again from the fact that the true joint data distribution satisfies MTR and MIV. As the conditions of Lemma 2 are met, we can find a population that satisfies $\mathbb{E}[Y(\tilde{d}) \mid Z = z] = n_{\tilde{d},z}$ for all $\tilde{d}$ and z. Meanwhile, the population satisfies MIV since $n_{\tilde{d},z}$ increases in z. The arm-specific mean recovers the lower bound. This concludes the proof. The sharpness for the upper bound can be analogously shown.

Next, we prove the validity and sharpness of the bound for average treatment effects $\mathbb{E}[Y(d) - Y(d')]$. The derivation of validity is exactly same as in Proposition 1. For sharpenss, note that the above sharpness proof for arm-specific means replaces $m_{e,z}$ for $\overline{y}$ and $\underline{y}$ in the proof of Proposition 1. We can do the same for the sharpness of the average treatment effect bounds. Formally, the sharpness argument in Proposition 1 depends on the monotonicity of $m^-_{\tilde{d},z}\pi^-_{\tilde{d},z} + \underline{y}(1 - \pi^-_{\tilde{d},z})$ and $m^+_{\tilde{d},z}\pi^+_{\tilde{d},z} + \underline{y}(1 - \pi^+_{\tilde{d},z})$ with respect to $\tilde{d}$. If we replace them with $m^-_{\tilde{d},z}\pi^-_{\tilde{d},z} + m_{\tilde{d},z}(1 - \pi^-_{\tilde{d},z})$ and $m^+_{\tilde{d},z}\pi^+_{\tilde{d},z} + m_{\tilde{d},z}(1 - \pi^+_{\tilde{d},z})$, which are also increasing in $\tilde{d}$, the construction of $n_{\tilde{d},z}$ will be still valid. Validity means that (i) $n_{\tilde{d},z}$ increases in z, and (ii) they satisfy the conditions for Lemma 2. Lemma 2 then guarantees the existence of a population that satisfies $\mathbb{E}[Y(\tilde{d}) \mid Z = z] = n_{\tilde{d},z}$ and

Assumptions 1, 2, 4 and 6. The population also satisfies Assumption 5 by construction of $n_{\tilde{d},z}$. Finally, it achieves the bounds for the average treatment effect.

Data construction

The log hourly wage was calculated as the natural log of `earnings` (yearly salary) divided by `weeks` (weeks worked) and `hours` (usual hours worked per week).

Year of arrival records one of the following choices: 0. 1987 to 1990; 1. 1985 to 1986; 2. 1982 to 1984; 3. 1980 to 1981; 4. 1975 to 1979; 5. 1970 to 1974; 6. 1965 to 1969; 7. 1960 to 1964; 8. 1950 to 1959; 9. Before 1950; 10. Born in the U.S. The midpoint for 9 was taken at 1945.5 considering the midpoint of choice 8.