An Odd Couple: Monotone Instrumental Variables and Binary Treatments

(Forthcoming in *Econometric Reviews*)

Jeremiah Richey
jarichey80@gmail.com

Kyungpook National University

Abstract

This paper investigates Monotone Instrumental Variables (MIV) and their ability to aid in identifying treatment effects when the treatment is binary in a nonparametric bounding framework. I show that an MIV can only aid in identification beyond that of a Monotone Treatment Selection assumption if for some region of the instrument the observed conditional-on-received-treatment outcomes exhibit monotonicity in the instrument in the opposite direction as that assumed by the MIV in a Simpson's Paradox-like fashion. Furthermore, an MIV can only aid in identification beyond that of a Monotone Treatment Response assumption if for some region of the instrument either the above Simpson's Paradox-like relationship exists or the instrument's indirect effect on the outcome (as through its influence on treatment selection) is the opposite of its direct effect as assumed by the MIV. The implications of the main findings for empirical work are discussed and the results are highlighted with an application investigating the effect of criminal convictions on job match quality using data from the 1997 National Longitudinal Survey of the Youth. Though the main results are shown to hold only for the binary treatment case in general, they are shown to have important implications for the multi-valued treatment case as well.

JEL Classification: C14, J63, K40

Key Words: Instrumental variables, Nonparametric bounds, Partial identification, Criminal convictions

This work has benefited from comments by Brent Kreider, Joseph Herriges, David Frankel, Quinn Weninger, Gray Calhoun, participants at the 2013 Singapore Economic Review Conference, seminar participants at Kyungpook National University, and two anonymous referees. All errors are my own.

1 Introduction

Endogenous variables are often encountered in empirical work. In order to identify these variables' causal effects on outcomes, econometricians often rely on some form of Instrumental Variable (IV) assumption. In this vein, Manski and Pepper (2000) introduced the concept of a Monotone Instrumental Variable (MIV). Their MIV assumption assumes that mean responses vary weakly monotonically with an instrument as opposed to assuming that mean responses are constant across an instrument (as in the mean-independence form of the traditional IV assumption). For example, consider a variable measuring individuals' involvement in delinquent activities as a youth and their job tenure's response function to criminal conviction status. To use delinquency as an IV would be to assume individuals with different levels of delinquency rates have the same mean tenure functions, whereas to use delinquency as an MIV would be to assume individuals with higher delinquency rates have weakly lower mean tenure functions. The appeal of this weaker assumption has led to it becoming an increasingly popular identifying assumption (Gonzalez 2005, Gerfin and Schellhorn 2006, de Haan 2011, Kang 2011).

Though Manski and Pepper (2000) highlighted very general conditions under which an MIV may or may not have identifying power, this paper explores the implications of those general conditions for the relationships between the instrument, the treatment, and the outcome of interest when treatment is binary. Specifically, I show that, for binary treatments, an MIV cannot have identifying power on both the upper and lower bound of the treatment effect in a nonparametric bounding framework if the treatment is monotonic in the instrument (Condition TM) and the observed conditional-on-received-treatment outcomes have the same monotonicity (Condition CM) assumed by the MIV. More important for applied work, under similar circumstances, an MIV cannot aid in identification beyond the identifying power of Monotone Treatment Response or Monotone Treatment Selection. These findings have two implications. First, many seemingly promising MIVs will fail to yield identifying power, and understanding these conditions should aid researchers in their search for more fruitful MIVs. Second, MIVs that do aid in identification must derive their identifying power from a complex, and perhaps counterintuitive, relationship with the treatment and the outcome which should be considered. While an MIV assumption may at first appear trivially valid, it would seem a *useful* MIV demands greater scrutiny.

This paper proves the above propositions within a potential outcomes framework and explores the implications for empirical work. These results are highlighted with an application investigating the effect of criminal convictions on job tenure. Though the main results are shown to hold only for the binary treatment case in general, they are shown to have important implications for the multi-valued treatment case as well.

2 Set Up and Worst Case Bounds

Assume treatment is binary, say whether or not one has been convicted of a crime, and the outcome of interest is job tenure. Define d as a potential treatment and t as the realized or selected treatment. Throughout this paper any conditioning on additional covariates is suppressed, though all results are generalizable to the inclusion of additional regressors. Suppose the goal is to estimate the average treatment effect (ATE) of criminal convictions on job tenure:

$$ATE = E[y(1)] - E[y(0)] \tag{1}$$

where y(1) indicates job tenure under a conviction treatment and y(0) indicates tenure under a non-convicted treatment. There is reason to believe conviction status is endogenous in such an equation and so the ATE is not identified by the data alone. In such a case a researcher might aim to bound the treatment effect by bounding each of the unknown values in Equation 1. The following focuses on one value, $E[y(1)]$, as similar arguments will hold for the other.

By iterated expectations

$$E[y(1)] = E[y(1)|t = 1] \cdot P(t = 1) + E[y(1)|t = 0] \cdot P(t = 0). \tag{2}$$

The data identify all of the right hand side quantities except the counterfactual $E[y(1)|t = 0]$. But, if this counterfactual has a naturally bounded outcome space, defined by K_L and K_U, we can define Manski's (1990) worst case bounds on our unknown:

$$E[y(1)]_{lb} = E[y(1)|t=1] \cdot P(t=1) + K_L \cdot P(t=0)$$

$$\leq E[y(1)] \leq \tag{3}$$

$$E[y(1)]_{ub} = E[y(1)|t=1] \cdot P(t=1) + K_U \cdot P(t=0).$$

3 The MIV Assumption and Identification

3.1 MIV in Isolation

Manski and Pepper (2000) introduced the concept of an MIV, a weakened form of a mean independent IV assumption, that can aid in identification by tightening the bounds in Equation (3).

Assumption MIV: *Let Z be an ordered set. Covariate z is a monotone instrumental variable if, for each $d \in D$ and all $(z, z') \in (Z \times Z)$ such that $z_2 \geq z_1$,*

$$E[y(d)|z_2] \leq E[y(d)|z_1].$$

If the instrument z is delinquency rates[1], this assumption implies that individuals with higher delinquency rates have weakly lower job tenure functions.

Following Manski and Pepper, $E[y(1)]$ can be bounded under the MIV assumption by

$$\sum_{z \in Z} P(z) \left\{ \sup_{z' \geq z} \left[E[y(1)|z', t=1] \cdot P(t=1|z') + K_L \cdot P(t=0|z') \right\} \right.$$

$$\leq E[y(1)] \leq \tag{4}$$

[1]By delinquency rates I refer to the degree to which an individual participates in delinquent activities.

$$\sum_{z\in Z} P(z)\left\{ \inf_{z''\leq z} \left[E[y(1)|z'', t=1] \cdot P(t=1|z'') + K_U \cdot P(t=0|v=z'')\right)\right\}.$$

If the worst case bounds, conditional on the instrument, exhibit the same monotonicity in the instrument as assumed by the MIV assumption, then the MIV has no identifying power. For the MIV to have any 'bite,' there must exist a region of the instrument in which the bounds run counter to the monotonicity of the MIV assumption. This is the general condition noted by Manski and Pepper (2000) under which an MIV will have identifying power. What follows extends this general observation by investigating its implications for the relationships in the data. Define Treatment Monotonicity as follows:

Condition TM: Treatment Monotonicity[2] *Let Z be an ordered set. For all $(z, z') \in (Z \times Z)$ such that $z_2 \geq z_1$,*

$$P(t=1|z_2) \geq P(t=1|z_1).$$

TM would imply, in the current example, that higher delinquency rates weakly increase the probability of being convicted of a crime. Though a somewhat similar (yet statistically different) assumption is required in the standard IV literature (the rank condition), this is not explicitly needed in the general framework set up by Manski and Pepper. The only necessary (though not sufficient) condition for the MIV to have identifying power is that the pair (y, d) not be independent of the instrument (Manski 2003). Define conditional MIV as follows:

Condition CM: Conditional MIV *Let Z be an ordered set. Covariate z is a conditional monotone instrumental variable if, letting t be received treatment, for each $t \in T$ and all $(z, z') \in (Z \times Z)$ such that $z_2 \geq z_1$,*

$$E[y(d)|t=d, z_2] \leq E[y(d)|t=d, z_1].$$

This condition differs from Manski and Pepper's MIV assumption by proposing observed outcomes

[2]This monotonicity is not the same monotonicity assumption of Imbens and Angrist (1994). The monotonicity of Imbens and Angrist assumes every individual responds in the same way to exposure of the instrument. Their's is a non-verifiable identifying assumption while the TM condition here can be verified with data.

conditional on received treatment are monotonic in the instrument. This should not be confused with Manski and Pepper's Monotone Treatment Response (MTS) assumption, which while also conditions on received treatment, uses received treatment as the instrument. In the current example this condition implies that, within the subpopulation of individuals observed to have been convicted of a crime (as well as within the subpopulation of individuals observed to not have been convicted of a crime), those with higher delinquency rates have lower mean job tenures. Whereas the MIV assumption is unverifiable, Condition CM is observable in the data.

A key finding, for which a simple proof is provided in the appendix, is the following proposition.

Proposition 1.1 *An MIV assumption cannot provide identifying power on both sides of the treatment effect when the treatment is binary if Conditions TM and CM hold.*

Before discussing the implications of Proposition 1.1, I briefly note two extensions more relevant for applied work.

3.2 MIV with MTR and MTS

MIVs are rarely invoked in isolation as the MIV assumption alone generally leads to bounds too wide to be very informative. Rather, it is routinely invoked along with other assumptions. Two common assumptions are Monotone Treatment Response (MTR) and Monotone Treatment Selection (MTS).

MTR Assumption: *Let D be ordered. For each $j \in J$*

$$d_1 \leq d_0 \Rightarrow y_j(d_1) \leq y_j(d_0). \tag{5}$$

This implies that being convicted of a crime will not increase any individuals' job tenure. The MTR assumption assumes the (weak) sign of the treatment effect and results in an upper bound of zero.[3] The MTS assumption (Manski and Pepper 2000) assumes the direction of the selection

[3] The distinction between lower and upper bounds is a bit arbitrary and depends on how one defines the treatment.

mechanism and implies individuals who selected into the conviction treatment have weakly lower mean responses. The MTS assumption aids in bounding the lower bound of the treatment effect.

MTS Assumption: *Let T be ordered. For each $d \in D$ and all $(u_0, u_1) \in T \times T$ such that $u_1 \leq u_0$,*

$$E[y(d)|t = u_1] \leq E[y(d)|t = u_0]. \tag{6}$$

Proposition 1.1 referred to MIV in isolation. Similar results hold when MIV is assumed along with MTR and MTS. Proofs of the following extensions are simple variations of the proof for Proposition 1.1 and are also available in the appendix.

Proposition 1.2 *An MIV cannot provide identifying power beyond MTS on the treatment effect if condition CM holds.*

and

Proposition 1.3 *An MIV cannot provide identifying power beyond MTR on the treatment effect if conditions CM and TM hold if 'weak monotonicity' - $E[y(1)|t = 1, v] \geq E[y(0)|t = 0, v]$ - also holds.*

Proposition 1.2 implies that an MIV can only aid in identification beyond MTS if the instrument, treatment, and outcome have a Simpson's Paradox-like relationship. It must be that observed mean outcomes, conditional on received treatment, vary in one direction with the instrument while mean mean response functions vary in the opposite direction with the instrument. An example of this in the literature can be found in de Haan (2011). Her paper investigates (in part) the impact a parental college degree has on children's education attainment and uses grandparents' education as an MIV. This MIV implies children's mean education function is weakly increasing in grandparents' education levels. So in such a setting her MIV can only have identifying power beyond MTS (which it does) if, for some portion of the instrument, observed children's mean education levels, conditional on realized parents' education, are *decreasing* in grandparents' education levels. It is not that the

If one were to define treatment as 'not convicted' then the opposite would hold in this example.

failure of Condition CM invalidates the MIV assumption, rather it complicates the simple nature of its credibility and should be considered, especially since this is precisely the source of identification.

Proposition 1.3 implies an MIV assumption can only aid in identification beyond MTR if one of two situations occur: either condition CM must fail and the above Simpson's Paradox-like relationship must exist or Condition TM must fail. The failure of Condition TM, under assumptions MTR and MIV, implies that though the instrument and treatment affect the outcome in the same direction, the instrument makes the treatment *less* likely. A similar situation and its identifying power in a different parametric setting is discussed by Nevo and Rosen (2012). It is worth noting that in the literature it seems common for researchers to assume Condition TM either implicitly or explicitly (see Gonzalez 2005, de Haan 2011, Kang 2011). This is perhaps a result of MIVs being viewed as 'not quite valid' IVs and would imply identification comes at the expense of Condition CM.

If assumptions MIV, MTS and MTR are maintained, the failure of either Condition CM or TM implies a somewhat counterintuitive data generating processes. Whether either of these circumstances makes sense or could be expected is a case-by-case question and should be determined by the selection model at hand. Nonetheless this implies that, though an MIV assumption may appear more credible than an IV assumption, for an MIV to provide any identifying power requires the selection process to have a unique relationship with the instrument. Knowledge of this necessity should help researchers search for MIVs that will more likely aid in identification and caution them to be more scrutinizing of MIVs that provide identifying power. The following application highlights this latter point.

4 An Application: The Effect of Criminal Convictions on Match Quality

4.1 Background and Assumptions

The effects of criminal convictions on labor market outcomes is an important area of study for labor economists as well as social scientists in general. In 2009, nearly 7.2 million adults, or 3.1% of the adult population, were incarcerated, on parole, or on probation (U.S. Department of Justice 2010). Studies of the demand side of the labor market reveal that many employers are averse to hiring individuals with criminal records (Holzer 2007). Such an aversion among employers, when set in a equilibrium search model of employment, leads to several predictions. One of these predictions is that individuals with criminal records can be expected to have lower match qualities (Black 1995). Though unobserved, match quality is routinely measured as job tenure.

In order to infer the causal effects of criminal convictions on labor market outcomes, one could assume that treatment assignment is exogenous, but a criminal conviction is likely to be endogenous due to unobserved characteristics correlated with both convictions and labor market outcomes. In light of this, a researcher might aim to bound the treatment effect with a few weak, yet plausible, assumptions. One might simply assume treatment selection and response are monotone, and that mean job tenure functions vary monotonically with delinquency rates. Delinquency rates would seem to surely affect the 'first stage' (convictions), yet there is a good chance they also affect the 'second stage' (tenure) beyond their effect through convictions. Thus, though it is not a viable IV, it seems a prime candidate for an MIV.

4.2 Data

The data used in this application come from white male respondents in the 1997 NLSY with at most a high school diploma who are not enrolled in school. This yields a population of 892 individuals without criminal convictions and 146 individuals with criminal convictions. The conviction variable is based on criminal convictions not settled in juvenile court prior to the year of the reported outcome variable. Job tenure is measured as the average length in weeks of employment for all jobs begun no earlier than 2003 or year after first conviction, whichever is most recent. The delinquency variable is a measure of the degree to which the respondent participated in delinquent activities as a youth and is construction from a list of youth delinquency questions within the NLSY questionnaire.

4.3 Empirical Findings

Though job tenure is not a naturally bounded outcome, imposing bounds on expected job tenure of zero and 100 leads to the worst case bounds in left column of Table 1. These bounds are quite wide and not very informative. However, imposing the assumptions of monotonicity in treatment response and selection, MTR and MTS, yield fairly informative bounds on the ATE and can be found in the middle column of Table 1. The MTR/MTS bounds imply that a criminal conviction decreases one's expected tenure by at most 6.63 weeks and will not increase tenure. Imposing delinquency rates as an MIV along with the MTR/MTS assumptions can then lead to bounds on the ATE that tighten both the upper and lower bounds.[4] When paired with MTR, the MIV leads to an upper bound on the effect on tenure of -0.31 weeks - lower than the upper bound found under MTR alone. Similarly, combining the MIV with the MTS assumption yields a lower bound of -3.89 weeks - higher than the lower bound found under MTS alone[5].

The MIV has aided in identifying both the lower and upper bound beyond the MTS/MTR bounds. So where did this identifying power come from? Panel A of Figure 1 plots the probability of being convicted of a crime as a function of delinquency rates. The graph shows that, in general, condition TM holds for this instrument. This implies the identifying power of the MIV comes from condition CM not holding. Panel B of Figure 1 plots convicted and non-convicted mens' job tenure as a function of their delinquency rates. It is clear that in general 'weak monotonicity' holds and that

[4]It is well documented that MIV bounds have a finite sample bias that lead bounds to be tighter than they should. The bounds here have already been corrected for this bias using a bias correction recommended by Kreider and Pepper (2007).

[5]Note no form of statistical inference is presented. This paper's focus is identification and the example here is used to highlight the main findings.

condition CM does not. Moreover, it is precisely the regions where condition CM does not hold that yields identifying power. If one notices that in the range of delinquency rates 6-8, non-convicted mens' observed mean job tenures are *increasing* with increased delinquency rates, while the MIV assumption imposes mean job tenure functions are *decreasing* in delinquency rates. It is exactly this conflict that aids in identification. How to interpret this hinges on the selection model proposed, and so, in general, is unclear. However, it certainly would seem to imply a researcher should at a minimum address how this might affect the credibility of the MIV.

5 Multi-Valued Treatment

All of the above has focused on binary treatments, and in general the findings only hold in such settings. However, the main findings resurface when one considers the bounds on the ATE between the 'least' and 'greatest' treatments in the multi-valued treatment case. That is, assume treatment can be ordered ($t = 1, 2, 3...T$) and one is interested in the treatment effect $E[y(T)] - E[y(1)]$. Propositions 1.1-1.3 hold in this case as well[6]. Though at first glance this might appear a minor point, it has far greater implications for empirical work when one considers the non-sharpness of the joint MIV/MTR bounds (Manski and Pepper 2000). Because of this fact, joint MIV/MTR bounds on 'internal' treatment effects (those defined by $t \neq 1, T$) tend to rarely exclude zero. Thus, some practitioners have come to report bounds on the ATE of going from lowest treatment (t = 1) to highest treatment (t = T) (see Gonzalez (2004), Gerfin and Schellhorn (2006), de Haan (2011)). In all three of these applications the authors obtain identifying power from their MIVs on these bounds, highlighting the relevance of the findings here for empirical work even for multi-valued treatment cases.

[6]Note that when $t = 1$ treatment can be seen as binary (either $= 1$ or > 1) and also when $t = T$ treatment can be seen as binary (either $= T$ or $< T$).

6 Conclusions

This paper has highlighted conditions necessary for an MIV to help identify treatment effects when treatment is binary in a nonparametric bounds setting. If assumptions MTR, MTS and MIV are maintained, either of the conditions discussed - failure of Condition CM or TM - imply a complex and perhaps counterintuitive data generating process. The main findings have two implications for applied work. First, many valid MIVs will fail to yield identification power and an understanding of these underlying necessary conditions should aid researchers in finding helpful MIVs. Second, the implicit complex relationships necessary for an MIV to aid in identification should caution researchers to be more scrutinizing of instruments that provide identifying power. In general, researchers should be aware of these underlying conditions and consider their implications.

References

[1] Black, Dan (1995). "Discrimination in an Equilibrium Search Model" *Journal of Labor Economics*, 13, 309-334.

[2] 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.

[3] Gerfin, Michael, Martin Schellhorn (2006). "Nonparametric Bounds on the Effect of Deductibles in Health Care Insurance on Doctor Visits - Swiss Evidence" *Health Economics*, 15, 1011 - 1020.

[4] Gonzalez, Libertad (2005). "Nonparametric Bounds on the Returns to Language Skills" *Journal of Applied Econometrics*, 20, 771-795.

[5] Holzer, Harry J. (2007). "Collateral Costs: The Effects of Incarceration on the Employment and Earnings of Young Workers" IZA Discussion Paper No. 3118

[6] Imbens, Guido, Joshua Angrist (1994) "Identification and Estimation of Local Average Treatment Effects" *Econometrica*, 62, 467-475.

[7] Kang, Changhui (2011) "Family Size and Educational Investments in Children: Evidence from Private Tutoring Expenditures in South Korea" *Oxford Bulletin of Economics and Statistics*, 73.

[8] Kreider, Brent and John Pepper (2007). "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors" *Journal of the American Statistical Association*, 102, 432-441.

[9] Manski, Charles (1990). "Nonparametric Bounds on Treatment Effects" *American Economic Review, Papers and Procedings*, 80, 319-23.

[10] Manski, Charles (1997). "Monotone Treatment Response" *Econometrica*, 65, 1311-1334.

[11] Manski, Charles *Partial Identification of Probability Distributions* (New York, NY: Springer-Verlag, 2003)

[12] Manski, Charles and John Pepper (2000). "Monotone Instrumental Variables: With an Application to the Returns to Schooling" *Econometrica*, 68, 997-1010.

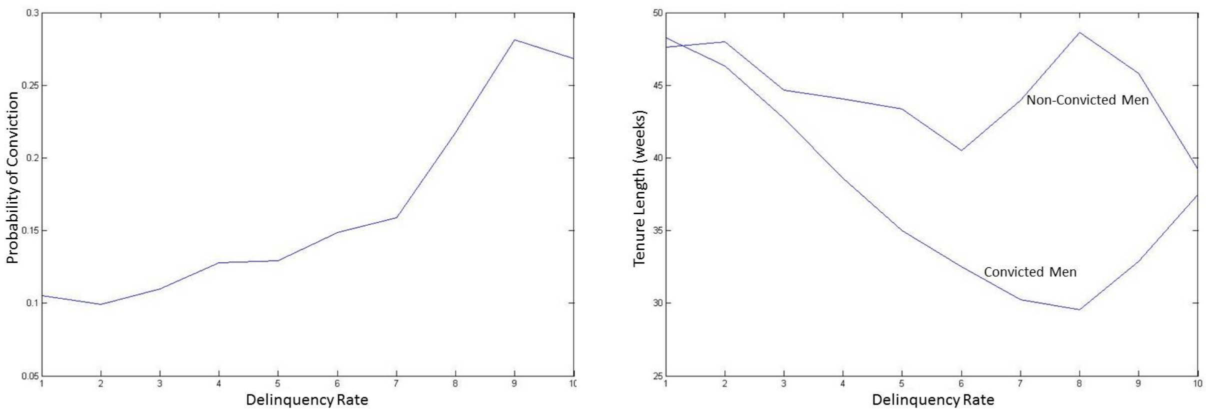
[13] Nevo, Aviv and Adam M. Rosen (2012). "Identification With Imperfect Instruments" *Review of Economics and Statistics*, 94, 659-671.

[14] U.S. Department of Justice. Office of Justice Programs. Bureau of Justice Statistics. 2010. *Bulletin: Correctional Population in the United States, 2009*. Washington, D.C.: U.S. Government Printing Office.

Table 1: Bounds on the ATE of criminal convictions on job tenure in weeks.

Population	Worst Case	MTR/MTS	MIV + MTR/MTS
White Men	(-47.4, 52.6)	(-6.63, 0)	(-3.89, -0.31)

Figure 1: Panel A: Probability of conviction as a function of delinquency rates. Panel B: Convicted and non-convicted mens' job tenure as a function of delinquency rates.



(a) Condition TM

(b) Condition CM

Appendix

Proposition 1.1

Which side of the treatment effect Proposition 1.1 applies to depends on how one defines the ATE, Condition TM, and Assumption MIV. Define $ATE = E[y(1)] - E[y(0)]$ and call $t = 1$ the 'treatment' and thus $t = 0$ the 'control.' Furthermore, define A to be positive if Condition TM is defined such that a 'higher' instrument ($u_2 > u_1$) implies an increased probability of receiving treatment and A to be negative otherwise. Also define B to be positive if Assumption MIV (and thus also Condition CM) is defined such that a 'higher' instrument ($u_2 > u_1$) implies a higher mean response and B to be negative otherwise.

Under the assumptions of Proposition 1.1 an MIV has no identifying power on the upper bound of the ATE if $A \times B$ is negative and no identifying power on the lower bound of the ATE if $A \times B$ is positive. In what follows I prove Proposition 1.1 for the case which coincides with the running example in the paper ($A = (+)$ and $B = (-)$) and focus on $E[y(1)]$ as a parallel argument holds for $E[y(0)]$. For $E[y(1)]$ it is the upper bound where MIV has no identifying power, and for $E[y(0)]$ it is the lower bound. Together these imply the lack of identification power on the upper bound of the treatment effect.

Assume Assumption MIV, Conditions TM and CM hold, and treatment is binary. For an MIV to aid in identification on the upper bound implies that there exists a pair $u_2, u_1 \in V$ st $u_2 > u_1$ and

$$\begin{gathered} E[y(1)|v = u_2, t = 1] \cdot P(t = 1|v = u_2) + K_u \cdot P(t = 0|v = u_2) \\ > \\ E[y(1)|v = u_1, t = 1] \cdot P(t = 1|v = u_1) + K_u \cdot P(t = 0|v = u_1). \end{gathered} \tag{7}$$

Note that

$$E[y(1)|v = u_2, t = 1] \leq E[y(1)|v = u_1, t = 1] \leq K_u. \tag{8}$$

The first inequality in Equation 8 holds by Condition CM and the second holds by the definition of the upper bound (K_u). For the inequality in Equation 7 to hold, this implies it must be that $P(t=0|v=u_2) > P(t=0|v=u_1)$.[7] But this contradicts Condition TM. ■

Proposition 1.2

In the set up throughout the paper it is the lower bound of the treatment effect that Assumption MTS applies to and thus the lower bound of $E[y(1)]$ and the upper bound of $E[y(0)]$ that Proposition 1.2 relates to. The following focuses on the lower bound of $E[y(1)]$. Similar arguments hold for the upper bound of $E[y(0)]$.

Assume Assumption MIV, Assumption MTS and Condition CM hold and treatment is binary.

For the MIV to have any identifying power on the lower bound implies that there exists a pair $u_2, u_1 \in V$ st $u_2 > u_1$ and

$$E[y(1)|v=u_2,t=1]\cdot P(t=1|v=u_2) + \underbrace{E[y(1)|v=u_2,t=1]}_{Replaces\ \ K_l\ \ under\ \ MTS}\cdot P(t=0|v=u_2)$$
$$> \tag{9}$$
$$E[y(1)|v=u_1,t=1]\cdot P(t=1|v=u_1) + \underbrace{E[y(1)|v=u_1,t=1]}_{Replaces\ \ K_l\ \ under\ \ MTS}\cdot P(t=0|v=u_1).$$

Note that this simplifies to

$$E[y(1)|v=u_2,t=1] > E[y(1)|v=u_1,t=1] \tag{10}$$

which contradicts Condition CM. ■

[7]Since $E[y(1)|v=u_2,t=1] \leq P[y(1)|v=u_1,t=1]$, for Eq.7 to hold $P(t=0|v=u_2)$ must be *greater* than $P(t=0|v=u_1)$ in order to put more weight on the upper bound K_u. This implies, due to the dichotomy of the treatment, that $P(t=1|v=u_2) < P(t=1|v=u_1)$.

Proposition 1.3

In the set up throughout the paper it is the upper bound of the treatment effect that Assumption MTR applies to and thus the upper bound of $E[y(1)]$ and the lower bound of $E[y(0)]$ that Proposition 1.3 relates to. The following focuses on the upper bound of $E[y(1)]$. Similar arguments hold for the lower bound of $E[y(0)]$.

Assume Assumption I (MIV), Conditions TM and CM as well as 'weak monotonicity' hold and treatment is binary.

For the MIV to have any identifying power on the upper bound implies that there exists a pair $u_2, u_1 \in V$ st $u_2 > u_1$ and

$$\begin{aligned} E[y(1)|v = u_2, t = 1] \cdot P(t = 1|v = u_2) + \underbrace{E[y(0)|v = u_2, t = 0]}_{Replaces\ \ K_u\ \ under\ \ MTR} \cdot P(t = 0|v = u_2) \\ > \\ E[y(1)|v = u_1, t = 1] \cdot P(t = 1|v = u_1) + \underbrace{E[y(0)|v = u_1, t = 0]}_{Replaces\ \ K_u\ \ under\ \ MTR} \cdot P(t = 0|v = u_1). \end{aligned} \tag{11}$$

Note both $E[y(1)|v = u_1, t = 1] > E[y(1)|v = u_2, t = 1]$ and $E[y(0)|v = u_1, t = 0] > E[y(0)|v = u_2, t = 0]$ hold by Condition CM. Also, $E[y(1)|v = u_2, t = 1] < E[y(0)|v = u_2, t = 0]$ and $E[y(1)|v = u_1, t = 1] < E[y(0)|v = u_1, t = 0]$ hold by weak monotonicity. This implies for Equation 11 to hold it must be that $P(t = 1|v = u_2) < P(t = 1|v = u_1)$ which contradicts Condition TM. ■