



UNIVERSITY OF
TECHNOLOGY SYDNEY

ECONOMICS DISCIPLINE GROUP

UTS BUSINESS SCHOOL

WORKING PAPER NO. 7

April 2013

Is Monotonicity in an IV and RD Design Testable? No, But You Can Still Check on it

Ben Edwards
Mario Fiorini
Katrien Stevens
Matthew Taylor

ISSN: 2200-6788

<http://www.business.uts.edu.au/economics/>

Is Monotonicity in an IV and RD design testable? No, but you can still check on it*

Ben Edwards[†], Mario Fiorini[‡], Katrien Stevens[§], Matthew Taylor[¶]

April 4, 2013

Abstract

Whenever treatment effects are heterogeneous and there is sorting into treatment based on the gain, monotonicity is a condition that both Instrumental Variable and fuzzy Regression Discontinuity designs have to satisfy for their estimand to be interpretable as a LATE. Angrist and Imbens (1995) argue that the monotonicity assumption is testable whenever the treatment is multivalued. We show that their test is informative if counterfactuals are observed. Yet applying the test without observing counterfactuals, as it is generally done, is not. Nevertheless, we argue that monotonicity can and should be investigated using a mix of economic intuition and data patterns, just like other untestable assumptions in an IV or RD design. We provide examples in a variety of settings as a guide to practice.

*We thank Jerome Adda, Michael Keane, Matthew Lindquist, Olena Stavrunova and seminar participants at the Swedish Institute for Social Research (SOFI) for helpful comments and suggestion. We also thank Lance Lochner for a discussion of their model.

[†]Australian Institute of Family Studies

[‡]Economics Discipline Group, University of Technology Sydney

[§]School of Economics, University of Sydney

[¶]Australian Institute of Family Studies

1 Introduction

In the early 90's, work by Imbens and Angrist (1994), Angrist and Imbens (1995) and Angrist, Imbens, and Rubin (1996) provided the theoretical foundation for the identification of the Local Average Treatment Effect (LATE): the treatment effect for those individuals who are influenced by the instrument. Their result is crucial in any context where (i) the gain from treatment is heterogeneous across the population and (ii) there is some degree of sorting into treatment based on the anticipated gain from treatment. Heckman, Urzua, and Vytlačil (2006) define any context where both (i) and (ii) occur as *essential heterogeneity*. They show that, under essential heterogeneity, the identification of the LATE requires the additional assumption of monotonicity: for a given change in the value of the instrument, it can not be that some individuals switch into treatment while others switch out of treatment. Hahn, Todd, and Van der Klaauw (2001) show that under essential heterogeneity, the assumption of monotonicity is also needed for the identification of a LATE in a fuzzy regression discontinuity approach. Whenever monotonicity does not hold, the IV and RD estimands are uninterpretable.

Angrist and Imbens (1995) go on to argue that the monotonicity assumption is testable whenever the treatment is multivalued: if monotonicity holds, the cumulative distribution functions (CDFs) of the endogenous variable should not cross for alternative values of the instrument. In other words, one CDF should first-order stochastically dominate the other. The justification for the Angrist and Imbens (1995) test is based on counterfactuals, as one should compare CDFs for the full sample under alternative values of the instrument. Nevertheless, counterfactuals are generally not observed. Angrist and Imbens (1995) thus apply their test to cross sectional data and claim that it is informative about the monotonicity assumption.

The aim of this paper is twofold. First, we reiterate that monotonicity is fundamentally untestable, whether the treatment is binary or multivalued. We argue that the monotonicity test proposed in Angrist and Imbens (1995) is informative only when applied to counterfactuals. However it is misleading to apply the test to any data where individuals are observed under only one of the possible values of the instrument. To clarify the distinction between the theoretical result of Angrist and Imbens (1995) based on counterfactual CDFs and its implementation using cross-section CDFs, we show plausible examples where:

1. the test rejects the monotonicity assumption even though monotonicity holds. Two recent studies, Barua and Lang (2009) and Aliprantis (2012), use cross-section CDFs to support their argument that monotonicity is violated. Our example shows that, since the test is uninformative, the monotonicity assumption cannot be rejected based on the test alone.
2. the test fails to reject the monotonicity assumption even though monotonicity is violated. Angrist and Imbens (1995) rely on the test to argue that the Angrist and Krueger (1991) quarter of birth instrument satisfies the monotonicity assumption because the cross-section CDFs do not cross. Our example clarifies that monotonicity might be violated even if the cross-section CDFs do not cross.

For clarity, we often provide an illustration using the school entry age setting, where the endogenous variable is school entry age, the instrument or the discontinuity is motivated

by school entry laws, and the outcome of interest is given by test scores or lifetime outcomes. We use this setting because this literature is among the first to discuss the monotonicity assumption but also to implement the Angrist and Imbens (1995) test. Nevertheless, our arguments can be easily generalized to different contexts.

The second aim of this paper is to argue that applied work should devote space to discussing monotonicity. Heckman, Urzua, and Vytlačil (2006) point out that there is an asymmetry in the way empirical researchers discuss heterogeneity: outcomes of choices are often allowed to be heterogeneous, choices themselves are not. But since choices are plausibly also heterogeneous in the sense that individuals might respond differently to changes in some variable of interest (such as the instrument), monotonicity cannot be assumed to hold a priori. While the assumption of monotonicity is not testable, common sense, economic intuition and data patterns can be used to argue in favour of or against the assumption. After all, this is what researchers ordinarily do when discussing the untestable independence assumption in an IV setting, or the untestable continuity assumption in an RD setting. We go through three different studies in a variety of settings that adopt either the IV or RD estimator. For each study, we try to make a case for or against monotonicity.

2 General Setting

To illustrate the notion of essential heterogeneity and monotonicity in the simplest way, let us define a model where both the endogenous variable (D) and the instrumental variable (Z) are binary:

$$\begin{aligned} Y_i &= \beta_{0i} + \beta_{1i}D_i && \text{(Outcome)} \\ D_i^* &= \alpha_{0i} + \alpha_{1i}Z_i && \text{(Latent Utility)} \end{aligned} \tag{1}$$

with

$$D_i = \begin{cases} 1 & \text{if } D_i^* > 0 \\ 0 & \text{if } D_i^* \leq 0 \end{cases} \quad \text{(Observed Treatment)}$$

where $\{\beta_{0i}, \beta_{1i}, \alpha_{0i}, \alpha_{1i}\}$ are random coefficients. Here β_{1i} is the heterogeneous gain from treatment and α_{1i} is the heterogeneous response to the instrument. The case of essential heterogeneity arises when $E(\beta_{1i}\alpha_{0i}) \neq 0$, that is when the treatment decision is partly driven by the gain from treatment.¹ Monotonicity is violated when α_{1i} has a positive sign for some individuals and a negative sign for some others.

2.1 Monotonicity in the IV design

Let $Y_i(0)$ be the response without treatment for individual i . $Y_i(1)$ is the response with treatment. Define $D_i(w)$ as the treatment when $Z_i = w$. Imbens and Angrist (1994) show that, provided a random variable Z_i satisfying the 3 following conditions is available, instrumental variable estimation identifies β_1 for those individuals who are influenced by the instrument.

IV1. $P(w) = E[D_i|Z_i = w]$ is a non trivial function of w (rank)

¹Essential heterogeneity is different from selection on levels: $E(\beta_{0i}\alpha_{0i}) \neq 0$.

IV2. for all $w \in \mathfrak{R}$ the triple $[Y_i(0), Y_i(1), D_i(w)]$ is jointly independent of Z_i (independence)

IV3. for all $z, w \in \mathfrak{R}$, either $D_i(w) \geq D_i(z) \forall i$ OR $D_i(w) \leq D_i(z) \forall i$ (monotonicity)

Condition IV1 is the rank condition. Condition IV2 is stronger than the standard IV exclusion restriction: in model (1) it not only implies that Z_i is *independent* of β_{0i} but also that Z_i is independent of β_{1i} . Condition IV3, named monotonicity in Imbens and Angrist (1994), is needed to identify a LATE whenever there is essential heterogeneity. The monotonicity assumption requires that, for every individual, a change in the value of the instrument from z to w must either leave the treatment decision unchanged or change the treatment in the same direction. Monotonicity is violated if, because of the same change in the value of the instrument from z to w , some individuals respond by getting the treatment (“switching in”) while others stop getting it (“switching out”). Monotonicity is a condition on counterfactuals: it refers to an individual’s behaviour in 2 *alternative* states of the world $Z = w$ and $Z = z$.² To understand the importance of the monotonicity condition Angrist, Imbens, and Rubin (1996) rewrite the IV estimate of β_1 as

$$\beta_1^{IV} = (1 + \lambda) \times E[Y_i(1) - Y_i(0) | D_i(z) - D_i(w) = 1] - \lambda \times E[Y_i(1) - Y_i(0) | D_i(z) - D_i(w) = -1] \quad (2)$$

where

$$\lambda = \frac{P[D_i(z) - D_i(w) = -1]}{P[D_i(z) - D_i(w) = 1] - P[D_i(z) - D_i(w) = -1]}$$

Equation 2 is very informative. It says that

- IV estimation does not “use” individuals that do not respond to changes in the value of the instrument: $D_i(z) - D_i(w) = 0$.
- if monotonicity holds, such that for instance $D_i(z) - D_i(w) \geq 0 \forall i$, then $\lambda = 0$ and IV estimates $\beta_1^{IV} = E[Y_i(1) - Y_i(0) | D_i(z) - D_i(w) = 1]$: the effect for those individuals that are induced to take the treatment because of the instrument (LATE). Viceversa if $D_i(z) - D_i(w) \leq 0 \forall i$ then $\lambda = -1$, and $\beta_1^{IV} = E[Y_i(1) - Y_i(0) | D_i(z) - D_i(w) = -1]$: the effect for those individuals that stop getting the treatment because of the instrument (a different LATE).
- if monotonicity does not hold and individuals respond differently to a given change in the value of Z_i , such that for some $D_i(z) - D_i(w) = 1$ while for others $D_i(z) - D_i(w) = -1$, then the IV estimate is not a treatment effect for any specific individual or group of individuals. It is not necessarily a weighted average of treatment effects either: λ is unbounded, thus the IV estimand does not necessarily fall in between $E[Y_i(1) - Y_i(0) | D_i(z) - D_i(w) = 1]$ and $E[Y_i(1) - Y_i(0) | D_i(z) - D_i(w) = -1]$ but can be more extreme.

²Because every individual has to respond in the same direction, Heckman and Vytlacil (2005) and Heckman, Urzua, and Vytlacil (2006) rename monotonicity with the term “uniformity”.

- if the return to treatment is heterogeneous but there is no sorting on gain, then monotonicity is not required. Without sorting on gain $E[Y_i(1) - Y_i(0)|D_i(z) - D_i(w) = 1] = E[Y_i(1) - Y_i(0)|D_i(z) - D_i(w) = -1] = E[Y_i(1) - Y_i(0)]$: that is, the expected return from treatment is the same among those who switch in and out, and equal to the average treatment effect (ATE). In this case $\beta_1^{IV} = E[Y_i(1) - Y_i(0)|D_i(z) - D_i(w) = 1] = E[Y_i(1) - Y_i(0)]$, so the IV estimand is the ATE.

2.2 Monotonicity in the fuzzy RD design

Let the setting be the same as in equation (1). Also let

$$\lim_{z \downarrow z_c} P[D_i = 1|Z_i = z] \neq \lim_{z \uparrow z_c} P[D_i = 1|Z_i = z]$$

such that the probability of treatment jumps at the threshold z_c , without requiring the jump to be equal to 1. Hahn, Todd, and Van der Klaauw (2001) show that there is a very close analogy between the fuzzy Regression Discontinuity design and the IV estimator. Both the IV and the RD estimand can be expressed as a Wald coefficient with

$$\beta_1^{IV} = \frac{E[Y_i|Z = z] - E[Y_i|Z = w]}{E[D_i|Z = z] - E[D_i|Z = w]}$$

and

$$\beta_1^{RD} = \frac{\lim_{z \downarrow z_c} E[Y_i|Z = z] - \lim_{z \uparrow z_c} E[Y_i|Z = z]}{\lim_{z \downarrow z_c} E[D_i|Z = z] - \lim_{z \uparrow z_c} E[D_i|Z = z]}$$

Hahn, Todd, and Van der Klaauw (2001) show that under essential heterogeneity, regression discontinuity estimation identifies β_1 for those individuals with $Z_i = z_c$ and only for those individuals who are affected by the threshold (LATE at z_c) under the following conditions

RD1. $\lim_{z \downarrow z_c} P[D_i = 1|Z_i = z] \neq \lim_{z \uparrow z_c} P[D_i = 1|Z_i = z]$ (RD)

RD2. $E[Y(0)|Z = z]$ is continuous in z at z_c (continuity)

RD3. $(\beta_{1i}, D_i(z))$ is jointly independent of Z_i near z_c (independence)

RD4. There exist a small number $\xi > 0$ such that either $D_i(z_c + e) \geq D_i(z_c - e), \forall i$ OR $D_i(z_c + e) \leq D_i(z_c - e), \forall i$ and for all $0 < e < \xi$ (monotonicity)

Condition RD1 is the RD equivalent of the rank condition in the IV setting. Condition RD2 implies that in the absence of treatment, individuals close to the threshold z_c are similar. These first two conditions must hold whether there is essential heterogeneity or not. Like in IV estimation, whenever there is essential heterogeneity the independence between Z_i and β_{1i} together with monotonicity are needed to identify a LATE at z_c . Note again that monotonicity is a condition on counterfactuals: for every individual, crossing the threshold must either leave the treatment decision unchanged or change the treatment in the same direction. Invoking the reasoning in Angrist, Imbens, and Rubin (1996), we can rewrite the RD estimate of β_1 as

$$\beta_1^{RD} = (1 + \lambda) \times E[Y_i(1) - Y_i(0) | D_i(z_c + e) - D_i(z_c - e) = 1] - \lambda \times E[Y_i(1) - Y_i(0) | D_i(z_c + e) - D_i(z_c - e) = -1] \quad (3)$$

where

$$\lambda = \frac{P[D_i(z_c + e) - D_i(z_c - e) = -1]}{P[D_i(z_c + e) - D_i(z_c - e) = 1] - P[D_i(z_c + e) - D_i(z_c - e) = -1]}$$

Equation (3) is the equivalent of equation (2) in an RD setting.³ Thus, if monotonicity holds, λ is equal to either 0 or -1 , and the RD estimand can be interpreted as a LATE. When monotonicity is violated, λ is neither 0 nor -1 , and the RD estimand does not measure the treatment effect for any particular group of individuals.

2.3 Monotonicity when the treatment is multi-valued

Angrist and Imbens (1995) extend the setting in equation (1) to the case where the treatment D takes more than two values. They show that IV estimates a weighted average of causal responses to a unit change in treatment, for those whose treatment status is affected by the instrument. They call this parameter the average causal response (ACR). The ACR still requires monotonicity otherwise the same interpretation problem arises as in the binary treatment case. Similarly, Lee and Lemieux (2010) show that monotonicity is also required in the fuzzy RD setting with a multi-valued treatment, in which case the interpretation of the RD estimand is still the same as that of the IV estimand.

3 Can we test for Monotonicity?

Given the importance of the monotonicity condition in both IV and RD designs, it is remarkable that this condition is seldom investigated in applied studies. This is in stark contrast to the lengthy discussions dedicated to the IV independence and rank conditions, and to the RD discontinuity (in the probability of treatment) and continuity (in the conditional regression function) conditions.

Angrist and Imbens (1995) argue that the monotonicity assumption has a “testable” implication whenever the treatment D_i takes more than two values. Let $D \in \{0, \dots, K\}$ with $K > 1$, and let monotonicity hold such that $D_i(z) \geq D_i(w)$, $\forall i$. Since $D_i(z) \geq D_i(w)$ with probability 1, then $P[D_i(z) \geq k] \geq P[D_i(w) \geq k]$ for every individual i and for every value k . That is, since the multivalued treatment is always (weakly) larger under $Z = z$ than under $Z = w$, the CDF for $D(z)$, $F_z(D)$, dominates the CDF for $D(w)$, $F_w(D)$:

$$F_z(D) \leq F_w(D) \quad (4)$$

Angrist and Imbens (1995) use this result based on *counterfactuals* to argue that, under monotonicity, the *empirical* CDF for $Z = z$ and $Z = w$ will not cross.⁴ They then apply the test in the Angrist and Krueger (1991) context, where quarter of birth is used as an

³Proof in the Appendix.

⁴If D is binary the CDFs cannot cross by construction.

instrument (Z) to estimate the return to years of schooling (D). Figure 1 is extracted from Angrist and Imbens (1995) and shows that the two CDFs do not cross: the schooling CDF for individuals born in the fourth quarter first-order stochastically dominates the CDF for individuals born in the first quarter. Angrist and Imbens (1995) then state that “This is important evidence in favor of the monotonicity assumption in this example.” Moreover, since in their data first-order stochastic dominance holds for any adjacent pair of quarters, they also state that “This is evidence that any adjacent pair of quarters can be used to define a binary instrumental variable that satisfies the monotonicity assumption.” Thus, they conclude that the IV estimates can be interpreted as the return to schooling for those individuals induced to take more schooling because they were born in the fourth quarter (LATE).⁵

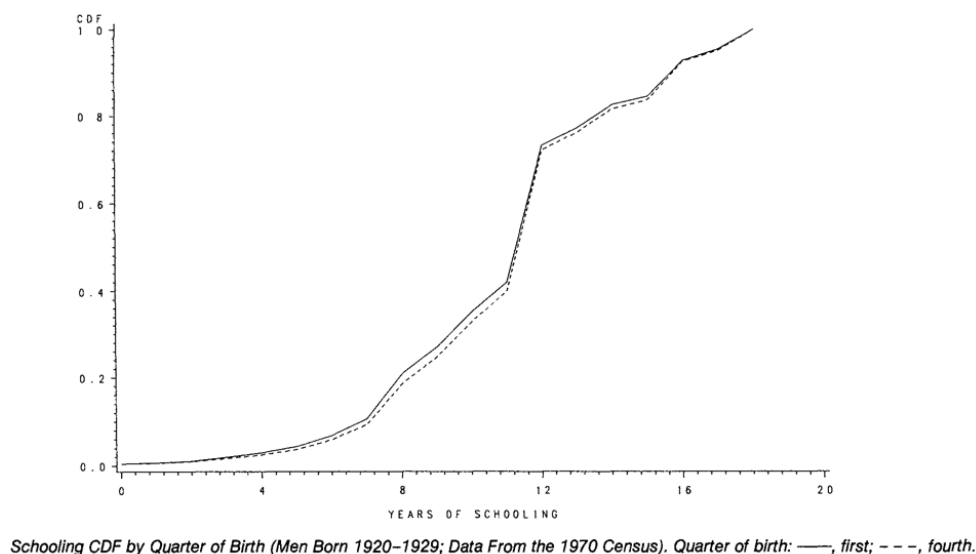


Figure 1: Angrist and Imbens (1995), page 438

3.1 Testing Monotonicity in the Age at School Entry setting

Two recent studies investigate monotonicity according to the Angrist and Imbens (1995) procedure: Barua and Lang (2009) and Aliprantis (2012). Both papers study the effect of school entry age on educational attainment.

There has been a recent explosion of interest in school entry age. Following evidence that in any grade, older children tend to perform better academically than younger children, it has become increasingly common to keep children out of kindergarten or first grade even when they are legally eligible to attend (see Deming and Dynarski (2008) for a discussion). This practice is often called “red-shirting”, a term originally used to describe the practice of holding college athletes out of play to add size and strength

⁵Angrist and Krueger (1991) explain that, for men in the 1980 US Census who were born in the 1930s and 1940s, school districts typically required a student to have turned age six by January 1 of the year in which he or she entered school. Therefore, students born earlier in the year entered school at an older age and attained the legal dropout age at an earlier point in their educational careers than students born later in the year.

prior to participating. Estimating the causal effect of school entry age is problematic because of potential endogeneity. Affluent parents can afford child-care costs associated with delaying their child’s school entry and are therefore more likely to do so. On the other hand, children who are less precocious intellectually and/or emotionally are also more likely to be delayed. These children may also perform poorly on cognitive tests. For both reasons OLS estimates could be biased, albeit in different directions. Bedard and Dhuey (2006) suggest solving the endogeneity problem by using the legal entry age (LEA) as an instrument for entry age. LEA is the age at which a child could start school given his/her birth date and given the country-specific school entry cut-off date. Figure 2 illustrates the relationship between legal entry age and date of birth in a context where the school year starts on September 1 and the law stipulates that children should be 5 years old when they start school. The LEA then jumps at the cut-off date: children born August 31st can start school in September of the year they turn 5, while children born September 1st (or later) are expected to start school a year later. Unless month of birth is related to parental background characteristics, it is reasonable to assume that the IV independence condition is satisfied. Even if one believes that month of birth and parental background are correlated, the RD continuity and independence conditions are still satisfied for a narrow window around the cut-off. Bedard and Dhuey (2006) also show that actual entry age is strongly correlated with the LEA in their data.

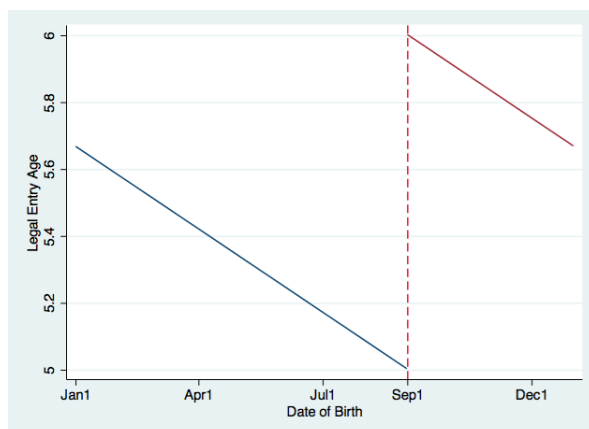


Figure 2: Legal Entry Age as a function of Date of Birth (September 1 cutoff date and school start)

The idea of using LEA as an instrument for school entry age is very similar to the Angrist and Krueger (1991) idea of using quarter of birth as an instrument for schooling. In both cases the date of birth provides the variation in the instrument. A variety of studies then use the same intuition in an IV or RD setting to estimate the causal effect of school entry age for different countries and/or outcomes: Datar (2006), Puhani and Weber (2007), McEwan and Shapiro (2008), Black, Devereux, and Salvanes (2011), Elder and Lubotsky (2009), Fredriksson and Ockert (2009), Muhlenweg and Puhani (2010), Muhlenweg, Blomeyer, Stichnoth, and Laucht (2012).

Barua and Lang (2009) and Aliprantis (2012) independently argue that the results in this literature on school entry age should be interpreted with caution because the monotonicity condition is most likely violated. Barua and Lang (2009) use 1960 US census data, for individuals born between 1949-1953, to show that many parents do

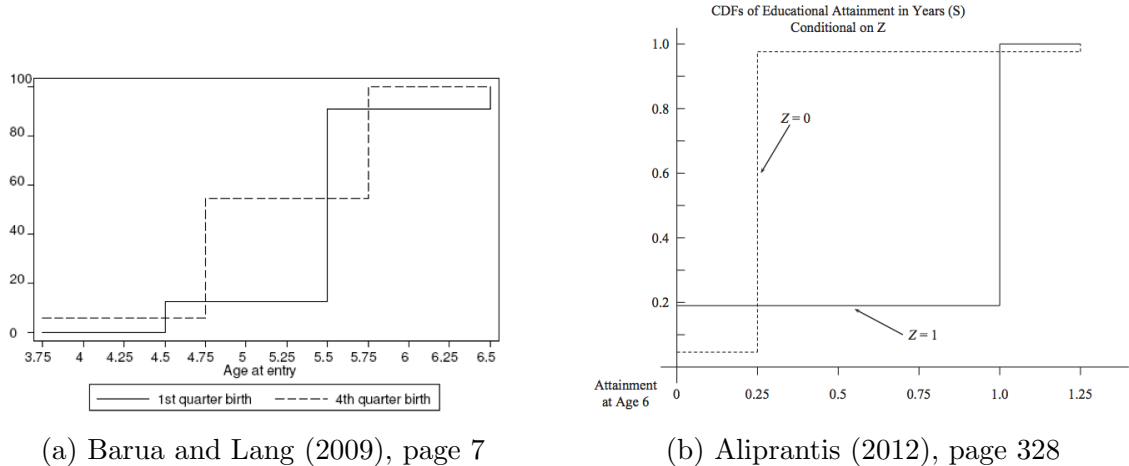


Figure 3: Applications of the monotonicity test

not enroll their children when eligible, with many parents delaying school start and others enrolling them earlier than is formally allowed. For instance, in states with a September cutoff date, Barua and Lang (2009) show that almost all students born in the first quarter enter kindergarten in September following their fifth birthday. In contrast, some children born in the fourth quarter enter before their fifth birthday, when they are younger than those born in the first quarter, while others enter the following year when they are older than entrants born in the first quarter. Therefore, the authors believe that quarter of birth (QoB) (and thus legal entry age, LEA) is not monotonically related to school entry age (EA): $EA(QoB = 1) < EA(QoB = 4)$ for some individuals but $EA(QoB = 1) > EA(QoB = 4)$ for some others. Barua and Lang (2009) complement their discussion with the Angrist and Imbens (1995) stochastic dominance test. Figure 3a is extracted from Barua and Lang (2009) and shows the CDFs for those individuals born in the first and last quarters of 1952 using reported age and grade at the time of the 1960 US Census. The CDFs clearly cross multiple times.⁶ Barua and Lang (2009) conclude that figure 3a gives clear evidence that monotonicity is not satisfied when quarter of birth or legal entry age are used as an instrument for entry age.

Aliprantis (2012) similarly argues that redshirting makes it all but impossible to interpret estimates of the effects of educational attainment when date or quarter of birth is used as an instrument for educational attainment. He compares the schooling of six year old children in the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) data set. Figure 3b is extracted from Aliprantis (2012). It shows the schooling CDFs for children born the quarter before ($Z = 1$) and the quarter after ($Z = 0$) the entrance cutoff date: the CDFs of school entry age cross.⁷ Aliprantis (2012) uses figure 3b to strengthen his argument that monotonicity fails in this instance as well.

⁶Barua and Lang (2009) do not observe the exact date of birth, but only the quarter. Thus, all individuals born in the first quarter are categorized as being born 1 March, while individuals born in the fourth quarter are categorized as being born 1 December. The Age at Entry values in figure 3a are derived accordingly.

⁷In the figure, schooling attainment at age 6 is measured in quarters and thus takes discrete values in steps of 0.25 years.

3.2 Why monotonicity is not testable

In this section we use simple examples to show that the Angrist and Imbens (1995) test applies to counterfactuals only: a situation where individuals are observed under different values of the instrument (or on both sides of the discontinuity).

If counterfactuals were observed, the test would detect a violation of the monotonicity assumption: if the CDFs cross, then monotonicity is violated. However, when counterfactuals are not observed and a cross-section of individuals is used instead, the test is not informative: *even if the CDFs cross, monotonicity might still hold*.

We then clarify that the test cannot assure that monotonicity holds even if counterfactuals are observed: *if the CDFs do not cross, monotonicity might still be violated*. Therefore, first order stochastic dominance of the counterfactual CDFs can only be interpreted as weak evidence in favour of monotonicity. Moreover, when a cross-section is used instead of counterfactuals, the test is again is not at all informative. Table 1 summarizes our main points.

Table 1: Monotonicity and CDFs

	Monotonicity	
	Holds	Violated
Counterfactual CDFs	cannot cross	can cross or not
Observed Cross-Section CDFs	can cross or not	can cross or not

3.2.1 How the test can reject monotonicity even though monotonicity holds

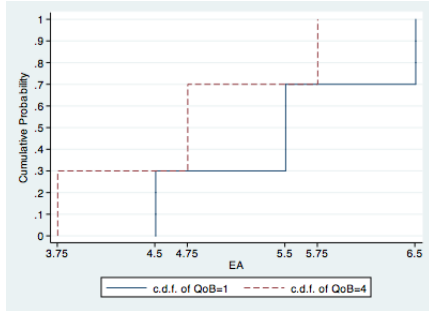
In table 2 we construct a simple example in the school entry age setting: we think of counterfactuals, where each of 10 children could be born either in the first or fourth quarter. Depending on the quarter of birth the parents might make a different decision as to when the child will start school. We assume a context where the school year starts on 1 September and the law stipulates that children should be 5 years old when they start school. Thus the cut-off is the same as the school start date. Nevertheless, parents can send their children to school early, on-time or might redshirt them.

Barua and Lang (2009) find that there was a large variance in the school starting age of children born in 1952 with the youngest children entering when 3.75 years old and the oldest entering when 6.5 years old (see figure 3a). Thus for many of those children the parents opted for an early entry. Our example in table 2 tries to replicate this variance and the entry age pattern. For instance, John’s parents would send him to school early, aged 4.5 years, had he been born in the first quarter, and send him to school in the same year, aged 3.75 years, had he been born in the fourth quarter. Importantly, in our example all children would go to school younger had they been born in the fourth quarter. Thus, if quarter of birth is used as instrument for entry age, monotonicity holds since $EA(QoB = 1) \geq EA(QoB = 4)$ for every child. It is easy to see that the entry age CDFs derived using all counterfactuals do not cross. This is illustrated in figure 4a. Nevertheless, when Barua and Lang (2009) derive the empirical CDFs, they cannot observe the counterfactuals but have to use a cross-section: children were either born in the first or fourth quarter. Let’s now consider a cross-section of children in our example,

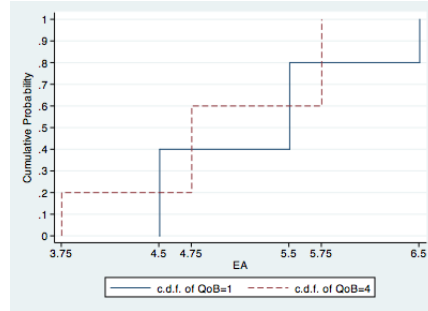
Table 2: Test rejects monotonicity even though monotonicity holds

Child	Date of Birth			
	01/03 (QoB=1)		01/12 (QoB=4)	
	Choice	Entry Age	Choice	Entry Age
John	Early (1)	4.5	Early (2)	3.75
Mike	Early (1)	4.5	Early (2)	3.75
Pete	On-Time	5.5	Early (1)	4.75
Nick	On-Time	5.5	Early (1)	4.75
Rob	Redshirt	6.5	On-Time	5.75
Laura	Early (1)	4.5	Early (2)	3.75
Sam	On-Time	5.5	Early (1)	4.75
Jeanie	On-Time	5.5	Early (1)	4.75
Emily	Redshirt	6.5	On-Time	5.75
Kirsten	Redshirt	6.5	On-Time	5.75

Values in bold indicate observed QoB and Entry Age (Cross-Section)
 Early (1) indicates that the child enters 1 school-year prior to when he/she is supposed to start
 Early (2) indicates that the child enters 2 school-years prior to when he/she is supposed to start



(a) Table 2: Counterfactuals



(b) Table 2: Observed Cross-Section

Figure 4: Monotonicity holds but the observed cross-section CDFs cross

indicated by the observations in bold in table 2: we *observe* all the boys, John-Rob being born in $QoB = 1$ and all the girls, Laura-Kirsten, being born in $QoB = 4$. Figure 4b plots the CDFs using the observations in bold only. Importantly, even though monotonicity is satisfied, the CDFs now cross with the pattern of crossings replicating the one in Barua and Lang (2009). Note that we could extract a different cross-section selecting other observations in the table, and possibly ensure that the cross-section CDFs do not cross. The point we want to emphasize here is that with a cross-section any pattern is possible, even if the monotonicity assumption is satisfied in the underlying counterfactuals.⁸

⁸To obtain a crossing of cross-section CDFs (while monotonicity holds) in the example in table 2, there needs to be at least one child among the December-borns that makes different school entry decisions than in the January-born group (e.g. Mike versus Kirsten). When comparing all children on both sides of the threshold, as done in the Angrist and Imbens (1995) test, it is this discrepancy in behaviour that creates a type of “observed” rise in EA and thus a crossing of CDFs. This difference in EA behaviour across the threshold can be driven by aspects that do not have a direct impact on the outcome, such as monetary costs of redshirting a child or parental preferences with respect to school EA.

3.2.2 How the test can fail to reject monotonicity even though monotonicity is violated

In this section we construct a simple example using the Angrist and Krueger (1991) setting, where quarter of birth partly determines years of schooling through a combination of school entry cutoff date and legal drop out age. Compulsory school attendance age varies by state but the minimum is generally age 16. This context is used by Angrist and Imbens (1995) to apply their monotonicity test. Again, we think of counterfactuals, where each of 10 children could be born either in the 1st or 4th quarter.

Case A In our example in table 3 most children have less schooling when born in the first quarter. For instance, John would drop out before completing High School if born in the first quarter, but completes High School if born in the fourth quarter. Similarly, Nick also stays in schooling longer, attaining college education, if born in the fourth quarter. However, this is not true for *all* children: Mike, Pete and Laura have less schooling when born in the fourth quarter. Thus, monotonicity does not hold. Figure 5a plots the counterfactual CDFs: the CDFs do not cross while monotonicity is violated. This confirms that the first-order stochastic dominance of the counterfactual CDFs is a necessary but not sufficient condition for monotonicity to hold. Figure 5b is obtained using the observations in bold only: also the cross-section CDFs do not cross.

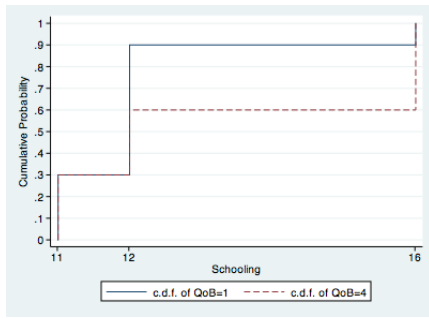
Table 3: Test fails to reject monotonicity even though monotonicity is violated: Case A

Child	Date of Birth			
	01/03 (QoB=1)		01/12 (QoB=4)	
	Choice	Schooling	Choice	Schooling
John	Dropout	11	High-School	12
Mike	High-School	12	Dropout	11
Pete	High School	12	Dropout	11
Nick	High School	12	College	16
Rob	College	16	College	16
Laura	High School	12	Dropout	11
Sam	Dropout	11	High School	12
Jeanie	Dropout	11	High School	12
Emily	High School	12	College	16
Kirsten	High School	12	College	16

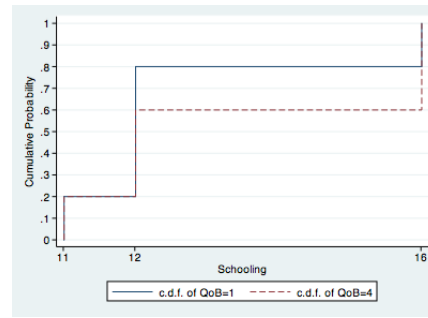
Values in bold indicate actual observed students/choices

Case B Table 4 is identical to table 3 but for Sam’s choice: she would now attain high school irrespectively of when she is born. Monotonicity is still violated because of Mike, Pete and Laura’s choices. Figure 6a plots the counterfactual CDF from table 4: the counterfactual CDFs now cross and monotonicity is violated. The cross-section CDFs in figure 6d do not cross and are identical to those in figure 5b: that’s because the observed (bold) choices have not changed.

The examples in section 3.2.1 and 3.2.2 could be altered to derive different scenarios. The key point is that the stochastic dominance test suggested by Angrist and Imbens (1995) only applies when counterfactuals are observed. Since this is generally impossible, monotonicity remains fundamentally untestable. The stochastic dominance test applied



(a) Table 3: Counterfactuals



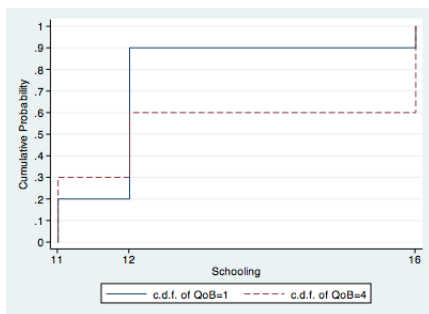
(b) Table 4: Cross-Section

Figure 5: Monotonicity does not hold, and both the counterfactual and observed CDFs do not cross

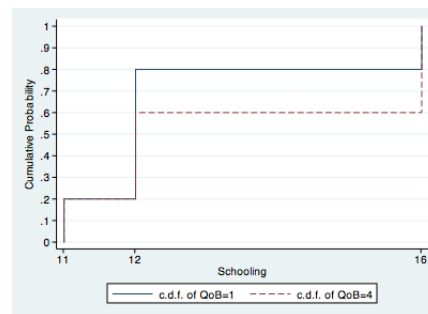
Table 4: Test fails to reject monotonicity even though monotonicity is violated: Case B

Child	Date of Birth			
	01/03 (QoB=1)		01/12 (QoB=4)	
	Choice	Schooling	Choice	Schooling
John	Dropout	11	High-School	12
Mike	High-School	12	Dropout	11
Pete	High-School	12	Dropout	11
Nick	High-School	12	College	16
Rob	College	16	College	16
Laura	High-School	12	Dropout	11
<i>Sam</i>	High-School	12	High-School	12
Jeanie	Dropout	11	High School	12
Emily	High School	12	College	16
Kirsten	High School	12	College	16

Values in bold indicate actual observed students/choices



(a) Table 4: Counterfactuals



(b) Table 4: Cross-Section

Figure 6: Monotonicity does not hold, but the counterfactuals CDFs cross while the observed cross-section CDFs do not

to an observed cross section of data is not informative. Whether stochastic dominance is satisfied in the cross section will depend on which choices we observe, and the result of the test could go either way irrespective of whether monotonicity is satisfied.

Thus when Barua and Lang (2009) show the CDFs crossing in figure 3a and conclude that monotonicity is violated we can not be sure, because monotonicity might still be satisfied. Similarly for Aliprantis (2012) and the evidence in figure 3b. When Angrist and Imbens (1995) show first order stochastic dominance in figure 1 and conclude that monotonicity is likely to hold, we are again skeptical: while stochastic dominance in the counterfactual CDFs is necessary but not sufficient for monotonicity to hold, the cross section CDFs can exhibit any kind of stochastic dominance pattern (whether monotonicity holds or not).

4 Monotonically Hopeless?

Since Angrist and Imbens (1995) test is not informative, should we just ignore the monotonicity assumption as is common practice? We believe the answer to this question is no. In an IV setting, the exclusion restriction (or independence) is not testable. Yet, the tools of common sense, economic intuition and data patterns are ordinarily used to argue in favour of or against that assumption. Similarly in an RD setting the assumption that individuals close to the threshold are similar is not testable. Nevertheless, one can at least check whether the observable characteristics have the same distribution just above and just below the cutoff. In that case there are reasons to believe that the same is true for unobservable characteristics. In this section the goal is to show that the plausibility of the monotonicity assumption can and should be investigated like any other assumption. We go through three different studies in various settings that adopt either the IV or RD estimator. For each study, we try to make a case for or against monotonicity.⁹

4.1 Angrist (1990): Vietnam lottery draft

Angrist (1990) uses Vietnam era draft lottery numbers (Z) as an instrument to estimate the effect of veteran status (D) on subsequent earnings (Y). These numbers were randomly assigned and influenced the probability of military service. Individuals with a number below a certain threshold were draft eligible ($Z = 1$). This is the simplest setting where both the treatment and the instrument are binary and can be modelled as in equation (1). If we assume that $D_i^* = \alpha_{0i} + \alpha_{1i}Z_i$ and that α_{1i} is homogeneous across eligible individuals then monotonicity is clearly satisfied. Thus, with $\alpha_{1i} = \alpha_1 > 0$ eligible individuals ($Z = 1$) would always be more likely to serve in Vietnam than non-eligible ones ($Z = 0$). In this model, for monotonicity to be violated we need α_{1i} to be heterogeneous. One could imagine that there is a group of individuals (group A) that would like to serve in the army independently of the lottery number: for them $\alpha_1^A = 0$. There was certainly a large group of individuals (group B) that would go to Vietnam only

⁹As explained earlier, monotonicity is needed only under essential heterogeneity. We do not investigate or test essential heterogeneity. This would lengthen the discussion considerably and it is not our focus. Yet, at least on intuitive grounds we cannot see how essential heterogeneity could be ruled out a priori in any of these studies. Heckman, Urzua, and Vytlacil (2006) and Heckman, Schmierer, and Urzua (2010) provide a thorough discussion of essential heterogeneity and describe a way to test it.

if drafted: for them $\alpha_1^B > 0$. If every individual belongs to either group A or B, there is heterogeneity in the way individuals respond to the instrument but monotonicity still holds since $D_i(Z = 1) \geq D_i(Z = 0)$, $\forall i$. By contrast, monotonicity fails if there is also a group of individuals (group C) who would go to Vietnam only if not eligible ($Z = 0$): for them $\alpha_1^C < 0$. While this is possible it is very unlikely. Thus, in the Angrist (1990) setting one could argue that monotonicity is not a major concern.

Keane (2010) provides a hypothetical example, similar to Angrist (1990), where monotonicity is more likely to fail. Suppose we use draft eligibility as an IV for completed schooling in an earnings equation. This seems as sensible as using it as an IV for military service, since draft eligibility presumably affects schooling while being uncorrelated with unobserved characteristics. Thus $D = 1$ if the individual went to college and $D = 0$ otherwise; Y and Z are the same as above. Amongst the draft eligible group, some (group B) stay in school longer than they otherwise would have, as a draft avoidance strategy: $\alpha_1^B > 0$. Others (group C) get less schooling than they otherwise would have, either because their school attendance is directly interrupted by the draft, or because the threat of school interruption lowers the option value of continuing school: $\alpha_1^C < 0$. This behaviour is more plausible than in Angrist (1990). It implies that monotonicity is violated since the instrument, draft eligibility, lowers school for some and raises it for others.

4.2 Black, Devereux, and Salvanes (2011): school entry age effects on test scores and adult outcomes

Black, Devereux, and Salvanes (2011) investigate the effect of school entry age ($D = EA$) on military IQ test scores and adult outcomes (Y) in Norway. To deal with endogeneity, they use Legal Entry Age as an instrument ($Z = LEA$). In Norway, school starts in August and children are expected to enter school in the calendar year they turn 7 (implying a January 1st cut-off date). Note that Black, Devereux, and Salvanes (2011) use both an IV and RD approach, with the latter relying only on children born one month either side of the cut-off date. The RD approach is implemented to account for potential manipulation of the date of birth by parents or the seasonality of births. In this section we focus on the monotonicity assumption in the RD approach by centring our attention around the December-January discontinuity. Nevertheless, the reasoning equally applies to all months of birth in a more general IV setting. Figure 7b plots the LEA by month of birth. Variation in the LEA is fully determined by the date of birth.¹⁰ In this context, one case where monotonicity holds is when all children start school on-time ($EA = LEA$): all December-borns would enter school older had they been born in January, and vice versa.

In a next step, we exploit data patterns to scrutinise the monotonicity assumption. Although Black, Devereux, and Salvanes (2011) do not discuss monotonicity, they provide useful information. Table 7a is taken from their paper and shows the proportion of children who enter school on-time, before and after the expected school entry age. Throughout the year, almost all children start school in the year they turn 7 (On Time). However, about 15% of December-borns are redshirted (Late), while 10% of January-borns start school before the year they turn 7 (Early). Overall, the youngest children in

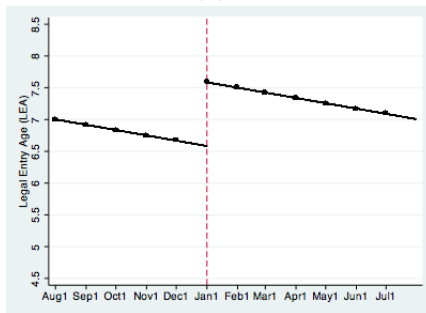
¹⁰LEA is defined as $7.7 - \frac{(\text{month of birth} - 1)}{12}$.

TABLE 1.—COMPLIANCE RATES BY MONTH OF BIRTH

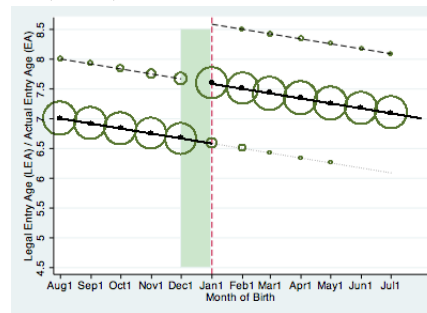
	Early	On Time	Late
January	.10	.90	0.0
February	.04	.96	.01
March	.02	.97	.01
April	.01	.98	.01
May	.01	.98	.01
June	0.0	.98	.01
July	0.0	.98	.01
August	0.0	.98	.02
September	0.0	.97	.02
October	0.0	.96	.04
November	0.0	.93	.07
December	0.0	.85	.15

Each number in the "Early" column refers to the percentage of children in each birth month who started school before the year they turned 7 years old. Each number in the "On Time" column refers to the percentage of children in each birth month who started school the year they turned 7 years old. Each number in the "Late" column refers to the percentage of children in each birth month who started school after the year they turned 7 years old.

(a) Black, Devereux, and Salvanes (2011) page 458.



(b) LEA by month of birth



(c) LEA and observed school Entry Age

Figure 7: Monotonicity in Black, Devereux, and Salvanes (2011)

an eligible school entry cohort (Oct-Dec borns) are the most likely to be redshirted, while the oldest ones (Jan-Feb borns) are the most likely to enter school early. This entry age behaviour is consistent with parents/educators making school-entry decisions based on either a child’s absolute or relative age. Figure 7c replicates figure 7b but we now add the observed EA patterns as shown in the table.¹¹ The size of the circles mirrors the proportions by month of birth. The largest circles are found along the LEA line, reflecting the high on-time entry rates. The smaller circles on the dotted line reflect early school entry ($EA < LEA$), which occurs mainly among children born in January-February. Instead, the smaller circles on the dashed lines reflect delayed school entry ($EA > LEA$), which is most common among October-December borns.

We then consider counterfactuals to discuss monotonicity. It is possible to distinguish 9 groups based on actual and counterfactual EA behaviour, as shown in table 5. The sign in each cell indicates the change in EA if a child is born in January rather than December. For instance, Group E represents December-born on-time school entrants who would also enter school on-time had they been born in January. This implies an increase in EA for members of group E: $EA_i(Dec) < EA_i(Jan)$, $\forall i \in E$. The (+) sign reflects the associated increase in EA. Figures 7a and 7c suggest that group E is most likely the largest group.

Table 5: Monotonicity

	January Born		
December born	Early	On Time	Late
Early	A(+)	B(+)	C(+)
On Time	D(-)	E(+)	F(+)
Late	G(-)	H(-)	I(+)

The (+) term indicates that children enter school older when born in January. Viceversa, the (-) term indicates that children enter school younger when born in January.

Since late school entry among January-borns does not occur, we choose to ignore groups C, F and I. Similarly, since early school entry among December-borns does not occur, we choose to ignore groups A, B and C.¹² Any other December-born on-time entrants would enter school early had they been born in January (group D). Crossing the cut-off date implies a drop in EA for this group: $EA_i(Dec) > EA_i(Jan)$, $\forall i \in D$. Since 10% of January-born children enter school early, existence of this group cannot be ruled out. Similarly, groups G and H represent December-borns entering school late, but who would enter school early or on-time if they had been born in January. This again implies a drop in EA: $EA_i(Dec) > EA_i(Jan)$, $\forall i \in G, H$. Figures 7a and 7c again suggest that the existence of these groups cannot be ruled out. Crucially, the existence of either of the (-) groups D, G or H alongside the most numerous group E (+) creates a violation of monotonicity.¹³

¹¹Since table 7a reports entry age by month, figure 7c is derived assuming children are born the first day of the corresponding month.

¹²We choose to ignore these groups for simplicity, but including them in the discussion does not change any of the intuition provided.

¹³The above reasoning equally applies to children born in January, had they been born in December

Since counterfactuals are fundamentally unobserved, this discussion does not provide definite evidence that monotonicity is violated. However, the data patterns should raise serious concerns.

4.3 Heckman, Lochner, and Taber (1998): General-Equilibrium treatment effects

Heckman, Lochner, and Taber (1998) consider a setting where changes in tuition are used as an instrument (Z) to measure the impact of having a college degree (D) on earnings (Y). Their focus is on illustrating general equilibrium effects in skill prices. The idea is that a quantitatively large tuition-reduction policy financed by a proportional tax can have two-way effects. On the one hand, it can increase enrollment by directly reducing the cost of attending college. This is a standard partial equilibrium effect. On the other hand, it can reduce enrollment by reducing the wage premium of going to college, because of the increased supply of college-educated individuals. This is a general equilibrium (GE) effect.

Extending their discussion, let Y_i be earnings and D_i be a binary college attainment indicator. Suppose a tuition subsidy had been introduced across states at different points in time. One could then consider using the state \times time interaction as an instrument for college attainment. Assuming the instrument satisfies the independence and rank conditions, monotonicity holds if being in a state/time where the subsidy is available either leaves individuals' enrollment decisions unchanged (because one would enroll even without the subsidy or because the subsidy was too small to change the decision) or leads to additional enrollment: $D_i(\text{Subsidy}) \geq D_i(\text{NoSubsidy}), \forall i$. Monotonicity would be violated if, given the GE effect of reduced college returns, someone in a state/time where the subsidy is available would no longer enroll. This can happen if for a group of individuals the partial equilibrium effect dominates and pushes them into taking the treatment, while for another group of individuals the general equilibrium effect outweighs the partial equilibrium effect and stops them from taking the treatment. When simulating the impact of a tuition subsidy, Heckman, Lochner, and Taber (1998) find evidence that both groups are present.¹⁴ Hence, in the case of policies that cause general equilibrium effects, using that policy as a source of identification in an IV or RD setting requires a thorough discussion of the monotonicity assumption.

instead. For example, group E also includes January-born on-time entrants who would also start school on-time had they been born in December. The (+) sign still indicates an increase in EA when moving from a counterfactual birth in December to an actual birth in January.

¹⁴Let V_j be the discounted present value of earnings from choice j , with $j = C$ (College) or $j = H$ (High-School). In their model V_j is proportional to the price of skill j , so these prices enter multiplicatively, while tuition enters additively. Individuals with a larger V_C will lose more from a decrease in the price of skill C . Those with a high V_H will benefit more from an increase in the price of skill H . A change in tuition affects everyone equally. This is what generates the two-way flow.

4.4 Special Case: violations of monotonicity occurring at random

In this section we discuss a special case. Klein (2010) shows how to recover the LATE for violations of monotonicity occurring at random. Let us rewrite the IV estimand in equation (2) as

$$\beta_1^{IV} = E[Y_i(1) - Y_i(0)|D_i(z) - D_i(w) = 1] + \lambda \underbrace{(E[Y_i(1) - Y_i(0)|D_i(z) - D_i(w) = 1] - E[Y_i(1) - Y_i(0)|D_i(z) - D_i(w) = -1])}_{bias} \quad (5)$$

When $P[D_i(z) - D_i(w) = -1] \neq 0$ such that $\lambda \neq 0$, one can see equation (5) as the Local Average Treatment Effect for those individuals taking the treatment because of the instrument, plus a bias term. Klein (2010) provides an approximation of this bias when violations of monotonicity occur at random. In the general setting of equation (1), this implies that there is some random variable U_i causing individuals to respond to the instrument differently, with U_i being independent of the instrumental variable Z_i and $(Y_i(0), Y_i(1), \alpha_{0i})$. In other words, U_i is a random variable that causes α_{1i} to have a different sign for different individuals.

Klein (2010) provides an example using schooling choices which can be applied to our discussion of the Black, Devereux, and Salvanes (2011) context. There we discussed how the existence of 2 groups of children would lead to a violation of the monotonicity assumption: group E who enters school on-time irrespective of their month of birth, and group H who is redshirted when born in December.¹⁵ For instance, suppose that a child belongs to either group based on parental preferences for absolute and relative age. December-born children who enter school on-time would be the youngest in class, and group H parents are sensitive to this issue while group E parents are not. Then, the bias correction proposed by Klein (2010) requires that these preferences are independent of the child's ability, the gain from delayed school entry, parental background or anything else that is related to the child's outcomes or affects the school entry decision. While these conditions seem restrictive in this context, in other settings violations of monotonicity might be more likely to occur at random.

5 Conclusion

Several studies have emphasised that IV and fuzzy RD estimates can be interpreted as a local average treatment effect only if monotonicity is satisfied, see Imbens and Angrist (1994), Angrist and Imbens (1995), Angrist, Imbens, and Rubin (1996), Hahn, Todd, and Van der Klaauw (2001), Heckman, Urzua, and Vytlacil (2006). This requirement applies in a context with essential heterogeneity, a situation with heterogeneous treatment effects and sorting into treatment based on the gain. Monotonicity is a restriction on the impact of the instrument on the treatment. It implies that a change in the instrument from value z to w affects the treatment of all individuals in the same direction: all individuals are either unaffected or switch into treatment, either all are unaffected or switch out

¹⁵In our discussion monotonicity would also be violated by the existence of group D or G alongside group E.

of treatment. If monotonicity does not hold then the IV and RD estimands are not interpretable.

Surprisingly, very few of the applied studies that rely on IV and fuzzy RD designs discuss the monotonicity assumption. This is in stark contrast to the lengthy discussions dedicated to the IV independence and rank conditions, and to the RD discontinuity (in the probability of treatment) and continuity (in the conditional regression function) conditions. But can we test for monotonicity or at least reasonably argue in favour of or against it? We believe that the answer to the former question is no, while the answer to the latter question is yes.

We show that a test of monotonicity designed by Angrist and Imbens (1995) only applies when counterfactuals are observed. Whenever counterfactuals are not observed, the test cannot provide conclusive evidence about monotonicity. Thus the test is generally uninformative. Unfortunately, we are not aware of any other test or rule of thumb that allows for a quick assessment. Nevertheless, we argue that the applied literature is not “monotonically hopeless”. IV and RD designs are incredibly powerful in dealing with endogeneity problems. A discussion of the monotonicity assumption is just an extra step that should be included to validate the results. We provide examples of such a discussion. In each case, we use a mix of common sense, economic intuition and data patterns to argue in favour of or against the monotonicity assumption. These are the same tools that are commonly used to justify, on a case by case basis, other untestable assumptions in IV and RD designs.

Appendix

In this section we want to illustrate what the Wald estimator in (3) measures if the monotonicity assumption RD4 does not hold. Let $e > 0$ denote an arbitrary small number.

Consider first the numerator. The mean difference in outcomes for individuals above and below the discontinuity point is

$$\begin{aligned} & E[Y_i|z_i = z_c + e] - E[Y_i|z_i = z_c - e] = \\ & E[Y_i(0) + (Y_i(1) - Y_i(0))D_i|z_i = z_c + e] - E[Y_i(0) + (Y_i(1) - Y_i(0))D_i|z_i = z_c - e] = \\ & E[Y_i(0)|z_i = z_c + e] - E[Y_i(0)|z_i = z_c - e] + \\ & E[(Y_i(1) - Y_i(0))D_i|z_i = z_c + e] - E[(Y_i(1) - Y_i(0))D_i|z_i = z_c - e] \end{aligned}$$

By assumption RD2, the first two terms cancel out at the limit $e \rightarrow 0$. By assumption RD3, the remaining terms can be written as

$$\begin{aligned} & E[(Y_i(1) - Y_i(0))D_i(z_c + e)] - E[(Y_i(1) - Y_i(0))D_i(z_c - e)] = \\ & E[(Y_i(1) - Y_i(0))(D_i(z_c + e) - D_i(z_c - e))] = \\ & 1 \times E[Y_i(1) - Y_i(0)|D_i(z_c + e) - D_i(z_c - e) = 1] \times P[D_i(z_c + e) - D_i(z_c - e) = 1] - \\ & 1 \times E[Y_i(1) - Y_i(0)|D_i(z_c + e) - D_i(z_c - e) = -1] \times P[D_i(z_c + e) - D_i(z_c - e) = -1] \end{aligned}$$

The denominator can be rewritten:

$$\begin{aligned} & E[D_i|z_i = z_c + e] - E[D_i|z_i = z_c - e] = \\ & E[D_i(z_c + e)] - E[D_i(z_c - e)] = \\ & E[D_i(z_c + e) - D_i(z_c - e)] = \\ & P[D_i(z_c + e) - D_i(z_c - e) = 1] - P[D_i(z_c + e) - D_i(z_c - e) = -1] \end{aligned}$$

The Wald estimator in (3) can then be rewritten as follows:

$$\begin{aligned} & \frac{\lim_{e \downarrow 0} E[Y_i|z_c + e] - \lim_{e \uparrow 0} E[Y_i|z_c - e]}{\lim_{e \downarrow 0} E[D_i|z_c + e] - \lim_{e \uparrow 0} E[D_i|z_c - e]} = \\ & \frac{\lim_{e \rightarrow 0} \{(1 + \lambda) \times E[Y_i(1) - Y_i(0)|D_i(z_c + e) - D_i(z_c - e) = 1] - \lambda \times E[Y_i(1) - Y_i(0)|D_i(z_c + e) - D_i(z_c - e) = -1]\}}{P[D_i(z_c + e) - D_i(z_c - e) = 1] - P[D_i(z_c + e) - D_i(z_c - e) = -1]} \end{aligned}$$

where

$$\lambda = \frac{P[D_i(z_c + e) - D_i(z_c - e) = -1]}{P[D_i(z_c + e) - D_i(z_c - e) = 1] - P[D_i(z_c + e) - D_i(z_c - e) = -1]}$$

References

- ALIPRANTIS, D. (2012): “Redshirting, Compulsory Schooling Laws, and Educational Attainment,” *Journal of Educational and Behavioral Statistics*, 37(2), 316–338.
- ANGRIST, J. D. (1990): “Lifetime earnings and the Vietnam era draft,” *American Economic Review*, 80(3), 313–336.
- ANGRIST, J. D., AND G. W. IMBENS (1995): “Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity,” *Journal of the American Statistical Association*, 90(430), 431–442.
- ANGRIST, J. D., G. W. IMBENS, AND D. B. RUBIN (1996): “Identification of Causal Effects Using Instrumental Variables,” *Journal of the American Statistical Association*, 91(434), 444–455.
- ANGRIST, J. D., AND A. B. KRUEGER (1991): “Does Compulsory School Attendance Affect Schooling and Earnings?,” *Quarterly Journal of Economics*, 106(4), 979–1014.
- BARUA, R., AND K. LANG (2009): “School Entry, Educational Attainment and Quarter of Birth: A Cautionary Tale of LATE,” Working Paper 15236, National Bureau of Economic Research.
- BEDARD, K., AND E. DHUEY (2006): “The Persistence Of Early Childhood Maturity: International Evidence Of Long-Run Age Effects,” *Quarterly Journal of Economics*, 121(4), 1437–1472.
- BLACK, S. E., P. J. DEVEREUX, AND K. G. SALVANES (2011): “Too Young to Leave the Nest: The Effects of School Starting Age,” *Review of Economics and Statistics*, 93(2), 455–467.
- DATAR, A. (2006): “Does delaying kindergarten entrance give children a head start?,” *Economics of Education Review*, 25(1), 43–62.
- DEMING, D., AND S. DYNARSKI (2008): “The Lengthening of Childhood,” *Journal of Economic Perspectives*, 22(3), 71–92.
- ELDER, T. E., AND D. H. LUBOTSKY (2009): “Kindergarten Entrance Age and Children’s Achievement: Impacts of State Policies, Family Background, and Peers,” *Journal of Human Resources*, 44(3), 641–683.
- FREDRIKSSON, P., AND B. OCKERT (2009): “The effect of school starting age on school and labor market performance,” Stockholm University, Department of Economics.
- HAHN, J., P. TODD, AND W. VAN DER KLAUW (2001): “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design,” *Econometrica*, 69(1), 201–09.
- HECKMAN, J. J., L. LOCHNER, AND C. TABER (1998): “General-Equilibrium Treatment Effects: A Study of Tuition Policy,” *American Economic Review*, 88(2), 381–386.

- HECKMAN, J. J., D. SCHMIERER, AND S. URZUA (2010): “Testing the Correlated Random Coefficient Model,” *Journal of Econometrics*, 158(2), 177 – 203, Specification Analysis in Honor of Phoebus J. Dhrymes.
- HECKMAN, J. J., S. URZUA, AND E. VYTLACIL (2006): “Understanding Instrumental Variables In Models With Essential Heterogeneity,” *Review of Economics and Statistics*, 88(3), 389–432.
- HECKMAN, J. J., AND E. VYTLACIL (2005): “Structural Equations, Treatment Effects, and Econometric Policy Evaluation,” *Econometrica*, 73(3), 669–738.
- IMBENS, G. W., AND J. D. ANGRIST (1994): “Identification and Estimation of Local Average Treatment Effects,” *Econometrica*, 62(2), 467–75.
- KEANE, M. P. (2010): “Structural vs. atheoretic approaches to econometrics,” *Journal of Econometrics*, 156(1), 3 – 20.
- KLEIN, T. J. (2010): “Heterogeneous treatment effects: Instrumental variables without monotonicity?,” *Journal of Econometrics*, 155(2), 99 – 116.
- LEE, D. S., AND T. LEMIEUX (2010): “Regression Discontinuity Designs in Economics.,” *Journal of Economic Literature*, 48(2), 281 – 355.
- MCEWAN, P. J., AND J. S. SHAPIRO (2008): “The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates,” *Journal of Human Resources*, 43(1), 1–29.
- MUHLENWEG, A., D. BLOMEYER, H. STICHNOTH, AND M. LAUCHT (2012): “Effects of age at school entry (ASE) on the development of non-cognitive skills: Evidence from psychometric data,” *Economics of Education Review*, 31(3), 68 – 76.
- MUHLENWEG, A. M., AND P. A. PUHANI (2010): “The Evolution of the School-Entry Age Effect in a School Tracking System,” *Journal of Human Resources*, 45(2), 407 – 438.
- PUHANI, P., AND A. WEBER (2007): “Does the early bird catch the worm?..” *Empirical Economics*, 32(2-3), 359 – 386.