GENERALIZED INSTRUMENTAL INEQUALITIES: TESTING THE IV INDEPENDENCE ASSUMPTION

DÉSIRÉ KÉDAGNI AND ISMAEL MOURIFIÉ

The Pennsylvania State University and University of Toronto

Abstract. This paper proposes a new set of testable implications of the instrumental variable (IV) independence assumption: *generalized instrumental inequalities*. In our leading case with a binary outcome, we show that the generalized instrumental inequalities are necessary and sufficient to detect all observable violations of the IV independence assumption. To test the generalized instrumental inequalities, we propose an approach combining a sample splitting procedure and intersection bounds inferential methods. This idea allows one to easily implement the test using the Stata package of Chernozhukov, Kim, Lee, and Rosen (2015).We apply our proposed strategy to assess the validity of the IV independence assumption for various instruments used in the returns to college literature.

Keywords: Instrumental variable, Independence assumption, Sharp inequalities, Intersection bounds.

JEL subject classification: C12, C15, C21, C24.

Date: The present version is of April 4, 2018. We have greatly benefited from insightful comments received from Marc Henry, Romuald Méango, and Yuanyuan Wan. We are grateful to from Marc Henry, Romuald Méango, Judea Pearl, and Yuanyuan Wan for valuable discussions and comments. We benefited from discussions with participants at University of Toronto, Penn State, University of Rochester, Syracuse University. All errors are ours. Mourifié gratefully acknowledges support from the Connaught Fund.
Correspondence address: Department of Economics, University of Toronto, 150 St. George Street, Toronto ON M5S 3G7, Canada, ismael.mourifie@utoronto.ca.

1. Introduction

An instrument is a random variable that is independent of certain (unobserved) latent terms, thereby facilitating the identification of the causal effect of an endogenous treatment on a particular outcome, or the identification of structural functions. Depending on the model specification, the function of interest, and the identification strategy a researcher wishes to use, different types of statistical independence may be invoked. To illustrate this idea, let us consider the simple potential outcome model:

$$Y = Y_1 D + Y_0(1 - D) \tag{1.1}$$

where $Y \in \mathcal{Y} \subset \mathbb{R}$ is the observed outcome, $D \in \{0, 1\}$ is the observed treatment indicator, and Y_1 and Y_0 are the potential outcomes. Similarly, let $\{D_z : z \in \mathcal{Z}\}$ denote the potential treatments.[1] One of the most commonly used instrumental variable (IV) independence assumptions in the causal inference literature is the so-called random assignment (RA) assumption, where the IV is statistically independent of the vector of potential outcomes and treatments, i.e., $Z \perp (Y_1, Y_0, D_z)_{z \in \mathcal{Z}}$. See for instance Imbens and Angrist (1994) and Heckman and Vytlacil (2005).

A weaker version of the IV independence assumption is $Z \perp (Y_1, Y_0)$, namely the joint statistical independence (JSI). It requires the IV to be statistically independent of the vector of potential outcomes only. The JSI has become popular with the recent development of the partial identification approach, where researchers want to learn about certain features of the joint distribution of the potential outcome variables without imposing any restrictions on the treatment or the selection mechanism. See for instance Manski (1990), Chernozhukov and Hansen (2005), and Mourifié et al. (2015). In this paper, we mainly focus on the testability of the JSI assumption.

There is a general belief in the applied-economics world that the IV independence assumptions are, in general, fundamentally non-testable (e.g., Imbens and Angrist, 1994; Deaton, 2009). Therefore, the validity of these assumptions have often been argued based only on economic intuition, creating a great deal of controversy among researchers; see for instance Deaton, Heckman, and Imbens's (2010) symposia in *The Journal of Economic Literature*. Because the IV independence assumptions are fundamental in both statistics and economics, it is crucial to develop a statistical method to identify variables that could not serve as valid instruments. Pearl (1994) appears to be the first to discuss the validity of the RA assumption. Pearl (1994) proposed a testable implication of the RA assumption, which he called the *instrumental inequality*. While Pearl's instrumental inequality is necessary for RA to hold, it is not sufficient to detect all observable violations of the RA assumption. Using an optimization program in a simple numerical example, Bonet (2001) showed the non-sufficiency of Pearl's inequality.

[1] Formal definitions of those quantities are given in the next section.

The first main contribution of this paper is to show that Pearl's instrumental inequality is not sufficient to detect all observable violations of the JSI assumption and to complement the Pearl instrumental inequality with a new set of inequalities we call *generalized instrumental inequalities* (GII). Second, we show that the GII are necessary and sufficient to detect all possible observable violations of the JSI assumption for the leading case of a binary outcome. In other words, the GII are necessary and sufficient conditions for the distribution of the observable variables (Y, D, Z) to rationalize the IV model entertained here, i.e., $Y = Y_1 D + Y_0(1 - D)$ and $Z \perp (Y_1, Y_0)$.

As a by-product contribution, the GII can also be used to assess the validity of the RA assumption since the JSI assumption is weaker than the RA assumption. Although we do not show here that the GII are sufficient to detect all the observable violations of the RA assumption, to the best of our knowledge, these are the most informative testable implications of the RA assumption proposed in the literature so far.

Third, we propose a sample splitting procedure that allows one to test the GII using intersection bounds inferential methods like those in Chernozhukov, Lee, and Rosen (2013, CLR) or Andrews and Shi (2013, AS). Therefore, our testing procedure can be implemented using the Stata packages provided by Chernozhukov et al (2015, CLKR) or Andrews et al (2016, AKS). It is worth noting that prior to this work, Ramsahai and Lauritzen (2011) and Wang, Robins and Richardson (2017) were also interested in the same testing question. Ramsahai and Lauritzen (2011) consider testing only Pearl's instrumental inequality using a likelihood ratio test in the special case where the instrument, the outcome, and the treatment are all binary. In addition, Wang, Robins and Richardson (2017) recently proposed another approach to test Pearl's instrumental inequality conditional on covariates but again in the binary IV model.[2]

Our work complements these papers by proposing a testing procedure for the GII which applies to continuous instruments, non-binary outcomes, and discrete treatment. In empirical application the IV validity is often invoked conditionally on covariates, see Baiocchi et al. (2014). An interesting feature of the procedure is that one can easily incorporate additional covariates and test conditional independence assumptions.

The last contribution is empirical. Considering the sample of white males from the National Longitudinal survey of Youth of 1979 (NSLY), we use our test to assess the validity of the JSI assumption of some instruments used in the returns to college literature such as parental education, college tuition fees at the age 17, and number of siblings.

[2]Notice that Wang, Robins and Richardson (2017) are aware that the Pearl's instrumental inequality are not, in general, sharp. They say: "In general, there are other observed data constraints implied by the discrete instrumental variable model (Bonet, 2001), the testing of which is an interesting topic for future research." This current paper fills this gap.

Outline. The remainder of the paper is organized as follows. Section 2 details the general framework of analysis. In Section 3, we derive the sharp testable implications of the JSI assumption when the outcome is binary. Section 4 generalizes the JSI testable implications to the non-binary outcome. Section 5 describes our testing procedure. Section 6 presents our empirical application. The last section concludes the paper. Proofs of the main results are collected in the appendix. Additional results, containing a generalization of the GII to the discrete multi-valued treatment case, simulations, and detailed help for the implementation of the tests using the CKLR Stata package are collected in an online appendix.

2. Analytical Framework

We consider the potential outcome model of Rubin (1974). We denote by $Y \in \mathcal{Y} \subseteq \mathbb{R}$ the observed outcome, by $D \in \{0, 1, ..., T\}$ the multi-valued discrete and observed treatment status, and by $Z \in \mathcal{Z} \subseteq \mathbb{R}$ the observed instrumental variable. For the sake of simplicity, we consider a single instrument; however all our results hold trivially for a vector of instruments, i.e., $Z \in \mathcal{Z} \subseteq \mathbb{R}^{d_z}$. Let $\{Y_{dz} \in \mathcal{Y} : d \in \{0, 1, ..., T\}, z \in \mathcal{Z}\}$ be the potential outcomes that would have been observed if the treatment D and the instrument Z had been externally set to d and z, respectively. Similarly, let $\{D_z \in \{0, 1, ..., T\} : z \in \mathcal{Z}\}$ be the potential treatments that would have been observed if the instrument Z had been externally set to z. The first assumption we use is an "exclusion restriction":

Assumption 1 (Exclusion restriction (ER))**.** *$Y_{dz} = Y_d$ for all $d \in \{0, 1, ..., T\}$ and $z \in \mathcal{Z}$.*

The ER assumption means that Z does not have a direct causal effect on Y, but rather has an effect only through the treatment D. ER is almost always used when invoking an instrument. In many papers, it is directly assumed from the beginning and not even clearly stated (e.g. Imbens and Angrist, 1994; Heckman and Vytlacil, 2005). In the rest of the paper, we will assume that the ER assumption is valid. Therefore, when the treatment is binary, we can write the observed outcome as follows: $Y = Y_1 D + Y_0(1 - D)$. We will mainly consider that the treatment is binary $D \in \{0, 1\}$, but we show in the online appendix how to extend the results to the multi-valued discrete treatment case. In this paper, we study the validity of the JSI assumption.

Assumption 2 (Joint Statistical Independence (JSI))**.** *The vector of potential outcomes (Y_1, Y_0) is statistically independent of Z, i.e., $(Y_1, Y_0) \perp Z$.*

We will refer to MSI when Z is statistically independent of each Y_1 and Y_0, i.e., $Y_0 \perp Z$ and $Y_1 \perp Z$. Notice that we have the following implications between these different IV statistical independence assumptions:

$$RA \Rightarrow JSI \Rightarrow MSI.$$

3. Sharp instrumental inequalities in Binary outcome Models

In this section, we consider the case where the observed outcome is binary, i.e., $\mathcal{Y} = \{0,1\}$. For the sake of simplicity, we will use the following notation: $q_{ij}(z) \equiv \mathbb{P}(Y = i, D = j|Z = z)$ and $p_{ij} \equiv \mathbb{P}(Y_0 = i, Y_1 = j)$ for $i, j \in \{0,1\}$. $q_{ij}(z)$ are identified, observable probabilities obtained using the vector (Y, D, Z), while $\mathbb{P}(Y_0 = i, Y_1 = j)$ are unobservable probabilities. For all $i, j \in \{0,1\}$ and $z \in \mathcal{Z}$ we have:

$$\begin{aligned} \mathbb{P}(Y = i, D = j|Z = z) &= \mathbb{P}(Y_j = i, D = j|Z = z), \\ &\leqslant \mathbb{P}(Y_j = i|Z = z), \\ &= \mathbb{P}(Y_j = i), \\ &= \mathbb{P}(Y_j = i, Y_{1-j} = i) + \mathbb{P}(Y_j = i, Y_{1-j} = 1 - i). \end{aligned} \tag{3.1}$$

The second equality holds under the JSI or MSI assumption. The last inequality implies that

$$\mathbb{P}(Y = i, D = j|Z = z) \leqslant \mathbb{P}(Y_j = i, Y_{1-j} = i) + \mathbb{P}(Y_j = i, Y_{1-j} = 1 - i) \text{ for all } z \in \mathcal{Z}. \tag{3.2}$$

Hereafter, we use the following notation: $\overline{q}_i \equiv \sup_z \{q_{i0}(z) + q_{i1}(z)\}$, $\underline{q}_i \equiv \inf_z \{q_{i0}(z) + q_{i1}(z)\}$, $\overline{q}_{ij} \equiv \sup_z q_{ij}(z)$, $\underline{q}_{ij} \equiv \inf_z q_{ij}(z)$, $\overline{q}_{10-01} \equiv \sup_z \{q_{10}(z) + q_{01}(z)\}$, $\underline{q}_{10-01} \equiv \inf_z \{q_{10}(z) + q_{01}(z)\}$, and $\underline{q}_{00-11} \equiv \inf_z \{q_{00}(z) + q_{11}(z)\}$, where we use $\inf_z$ ($\sup_z$) as a shorthand notation for $\inf_{z\in\mathcal{Z}}$ ($\sup_{z\in\mathcal{Z}}$). Therefore, Equation (3.2) provides the following inequalities:

$$\overline{q}_{11} \leqslant p_{01} + p_{11}, \tag{3.3}$$

$$\overline{q}_{01} \leqslant p_{00} + p_{10}, \tag{3.4}$$

$$\overline{q}_{10} \leqslant p_{11} + p_{10}, \tag{3.5}$$

$$\overline{q}_{00} \leqslant p_{00} + p_{01}. \tag{3.6}$$

We can easily see that by adding (3.3) to (3.4) and (3.5) to (3.6) we obtain the Pearl *instrumental inequality*:

$$\begin{aligned} \overline{q}_{11} + \overline{q}_{01} &\leqslant 1, \\ \overline{q}_{10} + \overline{q}_{00} &\leqslant 1. \end{aligned}$$

Although Pearl (1994) derived his inequality under the RA assumption, notice that it requires only that the MSI assumption holds. A violation of one of these two inequalities implies a violation of the MSI assumption, which in turn is a violation of the JSI and the RA assumptions. However, these inequalities are not sufficient to screen all violations of the JSI assumption.

Now, we are going to derive our new set of testable implications. For all $i, j \in \{0,1\}$ and $z \in \mathcal{Z}$ we have:

$$\begin{aligned} \mathbb{P}(Y_0 = i, Y_1 = j) &= \mathbb{P}(Y_0 = i, Y_1 = j | Z = z), \\ &= \mathbb{P}(Y_0 = i, Y_1 = j, D = 0 | Z = z) + \mathbb{P}(Y_0 = i, Y_1 = j, D = 1 | Z = z), \\ &\leqslant \mathbb{P}(Y_0 = i, D = 0 | Z = z) + \mathbb{P}(Y_1 = j, D = 1 | Z = z), \\ &= \mathbb{P}(Y = i, D = 0 | Z = z) + \mathbb{P}(Y = j, D = 1 | Z = z). \end{aligned} \tag{3.7}$$

The first equality holds under JSI. Equation (3.7) provides the following inequalities:

$$p_{11} \leqslant \underline{q}_1, \tag{3.8}$$

$$p_{01} \leqslant \underline{q}_{00-11}, \tag{3.9}$$

$$p_{10} \leqslant \underline{q}_{10-01}, \tag{3.10}$$

$$p_{00} \leqslant \underline{q}_0. \tag{3.11}$$

By adding (3.8) to (3.9) and combining it to (3.3) we have $\overline{q}_{11} \leqslant p_{01} + p_{11} \leqslant \underline{q}_{00-11} + \underline{q}_1$. Using this idea, we can therefore derive the following four testable implications of the JSI:

$$\begin{aligned} \overline{q}_{11} &\leqslant \underline{q}_1 + \underline{q}_{00-11}, \\ \overline{q}_{01} &\leqslant \underline{q}_0 + \underline{q}_{10-01}, \\ \overline{q}_{10} &\leqslant \underline{q}_1 + \underline{q}_{10-01}, \\ \overline{q}_{00} &\leqslant \underline{q}_0 + \underline{q}_{00-11}. \end{aligned}$$

Finally, by adding up all four inequalities (3.8)–(3.11) we obtain the last testable inequality:

$$1 \leqslant \underline{q}_1 + \underline{q}_{10-01} + \underline{q}_0 + \underline{q}_{00-11}.$$

To summarize, we define the *sharp instrumental inequalities* by the following set of inequalities:

$$\overline{q}_{10} + \overline{q}_{00} \leqslant 1, \tag{3.12}$$

$$\overline{q}_{11} + \overline{q}_{01} \leqslant 1, \tag{3.13}$$

$$\overline{q}_{10} \leqslant \underline{q}_1 + \underline{q}_{10-01}, \tag{3.14}$$

$$\overline{q}_{00} \leqslant \underline{q}_0 + \underline{q}_{00-11}, \tag{3.15}$$

$$\overline{q}_{11} \leqslant \underline{q}_1 + \underline{q}_{00-11}, \tag{3.16}$$

$$\overline{q}_{01} \leqslant \underline{q}_0 + \underline{q}_{10-01}, \tag{3.17}$$

$$1 \leqslant \underline{q}_1 + \underline{q}_{10-01} + \underline{q}_0 + \underline{q}_{00-11}. \tag{3.18}$$

Notice that the last five inequalities cannot in general be recovered from the first two, except when the instrument is binary. Indeed, whenever the instrument is binary, we will show that the last five inequalities are redundant.

The non-sufficiency of Pearl's instrumental inequality in the binary treatment case has also been discussed in Bonet (2001). Using a convex analysis approach and an optimization program, Bonet (2001) pointed out the non-sufficiency of Pearl's instrumental inequality with the help of a numerical example in the case where $Card(\mathcal{Z}) = 3$.[3] Indeed, assuming $\mathcal{Z} = \{z_1, z_2, z_3\}$, Bonet (2001) found that the inequality[4]

$$q_1(z_2) + q_{10-01}(z_1) + q_{00}(z_3) \leqslant 2, \tag{3.19}$$

is necessary and cannot be recovered using Pearl's instrumental inequality. Therefore, Bonet (2001) strengthened Pearl's *instrumental inequality* using inequality (3.19). However, as recognized by Bonet (2001), it was difficult to derive a general form of the equation (3.19) using his computational method. Here, as a result of our constructive approach, it can be shown that inequality (3.15) is equivalent to

$$\overline{q}_1 + \overline{q}_{10-01} + \overline{q}_{00} \leqslant 2$$

which generalizes inequality (3.19). Notice that Bonet (2001) derived his testable implications under the RA assumption, which means that whenever the outcome and the treatment are binary and the instrument is binary or trivariate, the RA assumption does not impose more observable restrictions on the data than the JSI assumption.

In a numerical numerical illustration relegated in the online Appendix Section 1, we show that Pearl's instrumental inequality and Bonet's (2001) inequality are not sufficient in general to detect all observable violations of the JSI assumption. However, the following theorem shows that inequalities (3.13) to (3.18) are the most informative observable restrictions for assessing the validity of the JSI assumption.

Theorem 1 (Sharp characterization of the JSI assumption)**.** *Let Y, D, Y_1, Y_0, Z define a potential outcome model $Y = Y_1 D + Y_0(1 - D)$, where $\mathcal{Y} = \{0, 1\}$. (i) If JSI holds, then inequalities (3.12) to (3.18) hold. (ii) For any joint distribution of (Y, D, Z) such that inequalities (3.12) to (3.18) hold, there exists a joint distribution of $(\tilde{D}, \tilde{Y}_1, \tilde{Y}_0, Z)$ such that JSI holds, and $(\tilde{Y}, \tilde{D}, Z)$ has the same distribution as (Y, D, Z).*

Theorem 1 means that inequalities (3.12) to (3.18) are necessary and sufficient to detect all possible observable violations of the JSI assumption. In other words, inequalities (3.12) to (3.18) are necessary and sufficient conditions for the distribution of the observable variables (Y, D, Z) to rationalize the IV model entertained here, i.e., the potential outcome model and the JSI assumption. To the best of our knowledge, this is the first time a sufficiency result about the testable implication of the JSI assumption has been given in the literature.

[3]Bonet (2001) used the computer program called PORTA.

[4]See Equation (11), page 54 of Bonet (2001).

Theorem 2. *Let Y, D, Y_1, Y_0, Z define a potential outcome model $Y = Y_1 D + Y_0(1-D)$, where $\mathcal{Y} = \{0,1\}$ and $\mathcal{Z} = \{0,1\}$. (i) If JSI holds, then inequalities (3.13) and (3.12) hold. (ii) For any joint distribution of (Y, D, Z) such that inequalities (3.13) and (3.12) hold, there exists a joint distribution of $(\tilde{D}, \tilde{Y}_1, \tilde{Y}_0, Z)$ such that JSI holds, and $(\tilde{Y}, \tilde{D}, Z)$ has the same distribution as (Y, D, Z).*

Theorem 2 states that the Pearl instrumental inequality is sufficient to detect all observable violations of the JSI assumption whenever the instrument, the outcome, and the treatment are all binary. In the proof collected in Appendix 8.2, we show why (3.14) to (3.18) are redundant when the instrument is binary and explain why this is no longer the case when the instrument is non-binary.

4. Generalization of the JSI Testable Implications

In this section, we do not restrict the observed outcome to be binary. Therefore, we generalize the *sharp instrumental inequalities* to the case where the outcome Y is either discrete, continuous, or mixed continuous-discrete. Let us write $f_{Y,D}(y, d|z) \equiv f_{Y|D=d}(y|Z = z)\mathbb{P}(D = d|Z = z)$, where $f_{Y|D=d}(y|Z = z)$ is the conditional density function of Y given $\{D = d, Z = z\}$ that is absolutely continuous with respect to a known dominating measure μ on Y, $P_{\mathcal{Y}}$ denotes a partition of the set $\mathcal{Y}$, and A and B denote generic elements of $P_{\mathcal{Y}}$. Also to simplify the notation, we will use the following shorthand: $\mathbb{P}(A, d|z) \equiv \mathbb{P}(Y \in A, D = d|Z = z)$. The derivations made in the previous section can be accordingly adapted to obtain the following generalized instrumental inequalities (GII):

$$\max_{d\in\{0,1\}} \int_{\mathcal{Y}} \sup_z f_{Y,D}(y, d|z) d\mu(y) \leqslant 1. \tag{4.1}$$

$$\sup_{d\in\{0,1\}} \sup_{A\in P_{\mathcal{Y}}} \left(\int_A \sup_z f_{Y,D}(y, d|z) d\mu(y) - \sum_{B\in P_{\mathcal{Y}}} \inf_z \{\mathbb{P}(A, d|z) + \mathbb{P}(B, 1-d|z)\} \right) \leqslant 0. \tag{4.2}$$

$$1 \quad \leqslant \quad \sum_{A\in P_{\mathcal{Y}}} \sum_{B\in P_{\mathcal{Y}}} \inf_z \{\mathbb{P}(A, d|z) + \mathbb{P}(B, 1-d|z)\}. \tag{4.3}$$

An adequate generalization of inequality (3.1) to a non-binary outcome implies that $\sup_z f_{Y,D}(y, d|z) \leqslant f_{Y_d}(y)$, where $f_{Y_d}(y)$ is the density function of the potential outcome Y_d. From this latter inequality, we obtain inequality (4.1). Similarly, a generalization of inequality (3.7) leads to the following bound on the joint distribution

$$\mathbb{P}(Y_d \in A, Y_{1-d} \in B) \quad \leqslant \quad \inf_z \{\mathbb{P}(Y \in A, D = d|Z = z) + \mathbb{P}(Y \in B, D = 1-d|Z = z)\}. \tag{4.4}$$

This latter inequality allows us to derive inequalities (4.2) and (4.3).

Inequality (4.1) is the generalization of the discrete version of the instrumental inequality derived by Pearl (1994). To the best of our knowledge, this is the first time that inequalities (4.2) and (4.3) have been derived as testable implications of both the JSI and RA assumptions. Inequalities (4.2) and (4.3) can be derived with any arbitrary partition $P_{\mathcal{Y}}$. This result suggests that, we should take the supremum or the infinimum of those inequalities over the set of all possible partitions to produce stronger testable implications. However, this would make the testing procedure very involved. In unreported simulations, we notice that inequalities (4.2) and (4.3) produce more informative testable implications when the elements of $P_{\mathcal{Y}}$ is a finer partition. So far we have been unable to prove that this is always the case, but given the simulation evidence we would advise practitioners to start with the finest possible partition. For example, when Y is discrete, the practitioner can use the partition of singletons on $\mathcal{Y}$, i.e., $P_{\mathcal{Y}} = \{\{y\} : y \in \mathcal{Y}\}$. Using the shorthand notation $\mathbb{P}(y, d|z) \equiv \mathbb{P}(Y = y, D = d|Z = z)$ in the multi-valued discrete outcome case, we obtained the following inequalities

$$\max_{d\in\{0,1\}} \sum_{y\in\mathcal{Y}} \sup_z \mathbb{P}(y, d|z) \leqslant 1, \tag{4.5}$$

$$\max_{d\in\{0,1\}} \max_{y\in\mathcal{Y}} \left(\sup_z \mathbb{P}(y, d|z) - \sum_{y'\in\mathcal{Y}} \inf_z \left\{\mathbb{P}(y, d|z) + \mathbb{P}(y', 1-d|z)\right\} \right) \leqslant 0, \tag{4.6}$$

and

$$1 \quad \leqslant \quad \sum_{y\in\mathcal{Y}, y'\in\mathcal{Y}} \inf_z \left\{\mathbb{P}(y, d|z) + \mathbb{P}(y', 1-d|z)\right\}. \tag{4.7}$$

When the outcome is continuous, for example, with a bounded support, i.e., $\mathcal{Y} = [\underline{y}; \overline{y}]$. A partition of $\mathcal{Y}$ can be constructed using a union of bin intervals, i.e., $[\underline{y}; \overline{y}] = \cup_{l=1}^{L} [y_l; y_{l+1}]$, where $y_l = \underline{y} + lh$ and h is the bin interval with width $\frac{\overline{y}-\underline{y}}{L}$. The simulation evidence suggests that the ability of inequalities (4.2) and (4.3) to detect violations of the JSI assumption will increase as h goes to 0. Notice that even if a smaller value of h increases the chance to detect a violation of the JSI assumption, it could considerably complicate the testing procedure, as we will see in Section 2.1 in the online appendix.

4.1. **GII and emptiness of the identification region.** The testable implications of the JSI assumption could also be derived using the emptiness of the identification region as entertained in Balke and Pearl (1997) who focused only on the special case where the instrument, the outcome, and the treatment are all binary. In our context, the GII can be obtained using the emptiness of the following bounds derived on the potential outcome $\mathbb{P}(Y_d \in A)$, $d \in \{0, 1\}$, for all A in the partition $P_{\mathcal{Y}}$.

Proposition 1. *Under the JSI assumption, bounds on $\mathbb{P}(Y_d \in A)$ are given by*

$$\begin{aligned} & \max\left[\int_A \sup_z f_{Y,D}(y,d|z)d\mu(y), 1 - \sum_{B\in P_{\mathcal{Y}}\setminus A}\sum_{C\in P_{\mathcal{Y}}} \inf_z\left\{\mathbb{P}(B,d|z)+\mathbb{P}(C,1-d|z)\right\}\right] \\ \leqslant\ & \mathbb{P}(Y_d \in A) \leqslant \\ & \min\left[1-\int_{\overline{A}} \sup_z f_{Y,D}(y,d|z)d\mu(y), \sum_{B\in P_{\mathcal{Y}}} \inf_z\left\{\mathbb{P}(A,d|z)+\mathbb{P}(B,1-d|z)\right\}\right] \end{aligned} \tag{4.8}$$

for all $A \in P_{\mathcal{Y}}$ and $d \in \{0,1\}$.

Although we construct the bounds for all $A \in P_{\mathcal{Y}}$, we can easily adapt the procedure to derive the bounds for all $A \in B_{\mathcal{Y}}$, where $B_{\mathcal{Y}}$ is a collection of Borel sets generated from $\mathcal{Y}$. It is worth noting that the JSI assumption is violated whenever the bounds cross and the bounds cross if and only if the GII do not hold. Whenever the bounds cross the JSI assumption is violated. Remarkably, the bounds cross if and only if the GII do not hold.

5. Testing procedure

Leading case: Binary outcome. In this section, we propose a simple strategy to test the *sharp instrumental inequalities*, i.e., inequalities (3.12) to (3.18), using the intersection bounds or conditional moment inequalities framework of CLR or AS. When Z is discrete, the *sharp instrumental inequalities* can be written as a finite set of moment inequalities and be tested using, for example, the specification tests developed by Romano and Shaikh (2010), Andrews and Soares (2010) among others. However, in general, whenever the support of Z is large or infinite, inferential methods for intersection bounds or conditional moment inequalities like CLR, and AS are more appropriate. In practice, the researcher may just choose one based on computational feasibility. Hereafter, we will mainly refer to the CLR framework and use the CKLR Stata package, but one can similarly refer to the AS framework and use the AKS Stata package.

In fact, we can easily see that inequalities (3.12) and (3.13) are special cases of intersection bounds that can be tested using the framework of CLR and implemented using the Stata package of CKLR. Indeed, (3.13) and (3.12) can be equivalently written as

$$\sup_z \mathbb{E}[Y(1-D)|Z=z] \quad \leqslant \quad \inf_z \mathbb{E}[Y(1-D)+D|Z=z]. \tag{5.1}$$

$$\sup_z \mathbb{E}[YD|Z=z] \quad \leqslant \quad \inf_z \mathbb{E}[YD+(1-D)|Z=z], \tag{5.2}$$

However, the rest of the inequalities, i.e., (3.14) and (3.18), cannot be so easily carried out in the CLR framework. For instance, (3.14) is equivalent to

$$\sup_z \mathbb{E}[Y(1-D)|Z=z] \quad \leqslant \quad \inf_z \mathbb{E}[Y|Z=z] + \inf_z \mathbb{E}[Y(1-D)+(1-Y)D|Z=z]. \tag{5.3}$$

More precisely, we are not aware about any existing inferential methods that provide valid approach to test such an inequality when Z is continuous. Here, we present a simple idea that allows inequality (5.3) to be analyzed in the CLR framework. We first need to make the following assumption:

Assumption 3. *Assume that there exist vectors of random variables* $(Y^{(k)}, D^{(k)}, Z^{(k)})$ *for* $k \in \{1, ..., K\}$ *such that (i)* $(Y^{(k)}, D^{(k)}, Z^{(k)}) \perp (Y^{(l)}, D^{(l)}, Z^{(l)})$ *for* $k \neq l$ *with* $l \in \{1, ..., K\}$*, and (ii)* $(Y^{(k)}, D^{(k)}, Z^{(k)})$ *has the same distribution as* (Y, D, Z)*.*

Assume that $\left\{(Y^{(l)}, D^{(l)}, Z^{(l)})\right\}_{l=1}^{3}$ satisfy Assumption 3. We have

$$\begin{aligned}
& \sup\nolimits_{(z_1,z_2,z_3)} \Big\{\mathbb{E}[Y^{(1)}(1-D^{(1)}) - Y^{(2)} - Y^{(3)}(1-D^{(3)}) \\
- & \quad (1-Y^{(3)})D^{(3)}|Z^{(1)} = z_1, Z^{(2)} = z_2, Z^{(3)} = z_3]\Big\} \\
= & \quad \sup\nolimits_{(z_1,z_2,z_3)} \Big\{\mathbb{E}[Y^{(1)}(1-D^{(1)})|Z^{(1)} = z_1, Z^{(2)} = z_2, Z^{(3)} = z_3] \\
+ & \quad \mathbb{E}[-Y^{(2)}|Z^{(1)} = z_1, Z^{(2)} = z_2, Z^{(3)} = z_3] \\
+ & \quad \mathbb{E}[-Y^{(3)}(1-D^{(3)}) - (1-Y^{(3)})D^{(3)}|Z^{(1)} = z_1, Z^{(2)} = z_2, Z^{(3)} = z_3]\Big\} \\
= & \quad \sup\nolimits_{(z_1,z_2,z_3)} \Big\{\mathbb{E}[Y^{(1)}(1-D^{(1)})|Z^{(1)} = z_1] + \mathbb{E}[-Y^{(2)}|Z^{(2)} = z_2] \\
+ & \quad \mathbb{E}[-Y^{(3)}(1-D^{(3)}) - (1-Y^{(3)})D^{(3)}|Z^{(3)} = z_3]\Big\} \\
= & \quad \sup\nolimits_{z_1} \Big\{\mathbb{E}[Y^{(1)}(1-D^{(1)})|Z^{(1)} = z_1] + \sup\nolimits_{z_2} \mathbb{E}[-Y^{(2)}|Z^{(2)} = z_2] \\
+ & \quad \sup\nolimits_{z_3} \mathbb{E}[-Y^{(3)}(1-D^{(3)}) - (1-Y^{(3)})D^{(3)}|Z^{(3)} = z_3]\Big\} \\
= & \quad \sup\nolimits_{z_1} \Big\{\mathbb{E}[Y(1-D)|Z = z_1]\Big\} - \inf\nolimits_{z_2} \Big\{\mathbb{E}[Y^{(2)}|Z^{(2)} = z_2] \\
- & \quad \inf\nolimits_{z_3} \mathbb{E}[Y^{(3)}(1-D^{(3)}) + (1-Y^{(3)})D^{(3)}Z^{(3)} = z_3]\Big\} \\
= & \quad \sup\nolimits_{z} \Big\{\mathbb{E}[Y(1-D)|Z = z]\Big\} - \inf\nolimits_{z} \Big\{\mathbb{E}[Y|Z = z]\Big\} \\
- & \quad \inf\nolimits_{z} \Big\{\mathbb{E}[Y(1-D) + (1-Y)D|Z = z]\Big\}.
\end{aligned}$$

The second equality holds under Assumption 3 (i), and the last equality holds under Assumption 3 (ii). Therefore, if Assumption 3 holds, inequality (5.3) can be rewritten as follows:

$$\sup_{(z_1,z_2,z_3)} \Big\{\mathbb{E}[Y^{(1)}(1-D^{(1)}) - Y^{(2)} - Y^{(3)}(1-D^{(3)}) - (1-Y^{(3)})D^{(3)}|Z^{(1)} = z_1, Z^{(2)} = z_2, Z^{(3)} = z_3]\Big\} \leqslant 0,$$

where the latter inequality can be easily tested using the CLR framework. Inequalities (3.15) to (3.18) can be rewritten accordingly. To summarize, testing the *sharp instrumental inequalities* is equivalent to testing the following inequalities:

$$
\begin{aligned}
&\sup_{(z_1,z_2)} \mathbb{E}[Y^{(1)}(1-D^{(1)}) - Y^{(2)}(1-D^{(2)}) - D^{(2)}|Z^{(1)}=z_1, Z^{(2)}=z_2] && \leqslant 0,\\
&\sup_{(z_1,z_2)} \mathbb{E}[Y^{(1)}D^{(1)} - Y^{(2)}D^{(2)} - (1-D^{(2)})|Z^{(1)}=z_1, Z^{(2)}=z_2] && \leqslant 0,\\
&\sup_{(z_1,z_2,z_3)} \mathbb{E}[Y^{(1)}(1-D^{(1)}) - Y^{(2)} - Y^{(3)}(1-D^{(3)}) \\
&\qquad -(1-Y^{(3)})D^{(3)}|Z^{(1)}=z_1, Z^{(2)}=z_2, Z^{(3)}=z_3] && \leqslant 0,\\
&\sup_{(z_1,z_2,z_3)} \mathbb{E}[(1-Y^{(1)})(1-D^{(1)}) - (1-Y^{(2)}) - (1-Y^{(3)})(1-D^{(3)}) \\
&\qquad -Y^{(3)}D^{(3)}|Z^{(1)}=z_1, Z^{(2)}=z_2, Z^{(3)}=z_3] && \leqslant 0,\\
&\sup_{(z_1,z_2,z_3)} \mathbb{E}[Y^{(1)}D^{(1)} - (1-Y^{(2)}) - (1-Y^{(3)})(1-D^{(3)}) \\
&\qquad -Y^{(3)}D^{(3)}|Z^{(1)}=z_1, Z^{(2)}=z_2, Z^{(3)}=z_3] && \leqslant 0,\\
&\sup_{(z_1,z_2,z_3)} \mathbb{E}[(1-Y^{(1)})D^{(1)} - (1-Y^{(2)}) - Y^{(3)}(1-D^{(3)}) \\
&\qquad -(1-Y^{(3)})D^{(3)}|Z^{(1)}=z_1, Z^{(2)}=z_2, Z^{(3)}=z_3] && \leqslant 0,\\
&\sup_{(z_1,z_2,z_3,z_4)} \mathbb{E}[Y^{(1)} - Y^{(2)}(1-D^{(2)}) - (1-Y^{(2)})D^{(2)} - Y^{(3)} \\
&\qquad -(1-Y^{(4)})(1-D^{(4)}) - Y^{(4)}D^{(4)}|Z^{(1)}=z_1, Z^{(2)}=z_2, Z^{(3)}=z_3, Z^{(4)}=z_4] && \leqslant 0,
\end{aligned}
$$

where we use $\sup_{(z_1,\ldots,z_j)}$ as a shorthand notation for $\sup_{z_1,\ldots,z_j \in \mathcal{Z}\times\ldots\times\mathcal{Z}}$, $j=2,3,4$. In practice a vector of random variables $(Y^{(k)}, D^{(k)}, Z^{(k)})$, $k=1,2,3,4$, satisfying Assumption 3 can be obtained by *splitting the sample*.

Sample splitting. Let us assume that we have an independent and identically distributed (i.i.d.) sample $\{(Y_i,D_i,Z_i)\}_{i=1}^n$. Let us split the sample into four different samples $(Y^{(1)},D^{(1)},Z^{(1)})_{n_1} \equiv \{(Y_i,D_i,Z_i)\}_{i=1}^{n_1}$, $(Y^{(2)},D^{(2)},Z^{(2)})_{n_2} \equiv \{(Y_i,D_i,Z_i)\}_{i=n_1+1}^{n_2}$, $(Y^{(3)},D^{(3)},Z^{(3)})_{n_3} \equiv \{(Y_i,D_i,Z_i)\}_{i=n_2+1}^{n_3}$, and $(Y^{(4)},D^{(4)},Z^{(4)})_{n_4} \equiv \{(Y_i,D_i,Z_i)\}_{i=n_3+1}^{n_4}$ with $n_1 < n_2 < n_3 < n_4 = n$ where $(Y^{(k)},D^{(k)},Z^{(k)})_{n_k}$, $k=1,2,3,4$, represent data matrices. Therefore,

(1) $(Y^{(1)},D^{(1)},Z^{(1)})_{n_1}$, $(Y^{(2)},D^{(2)},Z^{(2)})_{n_2}$, $(Y^{(3)},D^{(3)},Z^{(3)})_{n_3}$, and $(Y^{(4)},D^{(4)},Z^{(4)})_{n_4}$ are jointly independent,
(2) $(Y^{(1)},D^{(1)},Z^{(1)})_{n_1}$, $(Y^{(2)},D^{(2)},Z^{(2)})_{n_2}$, $(Y^{(3)},D^{(3)},Z^{(3)})_{n_3}$, and $(Y^{(4)},D^{(4)},Z^{(4)})_{n_4}$ have the same distribution as $\{(Y_i,D_i,Z_i)\}_{i=1}^n$.

Then, we propose to test the above inequalities using CKLR's Stata command by splitting the sample. This idea has three main advantages. First, it allows the researcher to interchange supremum (or infimum) with some sums of expectations, without having to worry about the statistical dependence between terms. Second, we carry out all good properties of the testing procedure developed by CLR, ensuring the consistency of the test against any fixed alternatives as well as against the class of local alternatives defined in Theorems 1 to 4 of CLR. Third, the testing procedure can be used easily by any applied researchers with the help of the CKLR Stata command. A drawback of this method is that it is data demanding, and the sample splitting procedure can generate an efficiency loss. Moreover, it is worth noting that while the natural choice would be to split the sample into sub-samples with same cardinality as we entertained hereafter, this would not be necessarily the optimal choice.[5] To the best of our knowledge, this is the first testing procedure that can be easily used to assess the validity of different IV independence assumptions (i.e., MSI, JSI, and RA) proposed in the literature, and can deal with discrete, continuous, or mixed continuous-discrete outcomes and instruments.

6. Empirical results

Estimating the causal impact of college education on later earnings has always been troublesome for economists because of the endogeneity of the education level. To evaluate the returns to schooling, different approaches have been proposed, and most of them rely on the validity of instruments such as parents' education, number of siblings, tuition fees, quarter of birth, distance to college, etc. In this section, we will assess the validity of some of those IVs, more precisely "parents' education", "number of siblings", and "tuition fees".

Although these instruments are controversial, some of them are still widely used in applied works. Often used since the work of Willis and Rosen (1979), parental background variables such as parental education have been criticized as valid instruments for their potential correlation with unobserved child ability. Some applied researchers argue that they could be valid instruments in the presence of controls such as parental income or a measure of child ability, see for instance Lemke and Rischall (2003). In addition, Card (2001) provides economic reasons that warn researchers against instrumenting college education with variables that potentially affect educational institutions, such as tuition fees or distance to college, arguing that they may violate the independence assumptions.

6.1. **Data.** We consider the data used in Heckman, Tobias and Vytlacil (2001, HTV). The data consists of a sample of 1,230 white males taken from the National Longitudinal Survey of Youth of 1979 (NSLY). The data contains information on the log weekly wage labelled "lwage", college education "college", labor market experience "exper", college tuition at age 17 "tuit17", father's education "fatheduc", mother's education "motheduc", number of siblings "sibs", and a measure of ability "abil." As explained in HTV, the ability measure is constructed from the ten component tests of the Armed Services Vocational Aptitude Battery (ASVAB) provided in the NSLY. Following HTV, we consider the college enrolment indicator as the treatment: it is equal to 1 if the individual has completed at least 13 years of education and 0 otherwise. We dichotomize the outcome. Notice that by dichotomizing, we sacrifice asymptotic testing power. On the other hand, this will produce a reasonable number of inequalities to test. Some summary statistics are reported in Table 1. We

[5]However, this avoids running into other issues. Indeed, for example in the case of two sub-samples, if we do not split the sample in half, it is not obvious, how the regression $\mathbb{E}[(.)|Z^{(1)}, Z^{(2)}]$ should be run, except by dropping some observations.

are interested in testing the conditional JSI assumption using the testing procedure described in Section 5.

TABLE 1. Summary Statistics

	Total
Observations	1,230
lwage	2.4138 (0.5937)
college	0.4325 (0.4956)
tuit17	8.5686 (4.1277)
fatheduc	12.44715 (3.2638)
motheduc	12.1781 (2.2781)
sibs	2.9488 (1.8630)
abil	1.7966 (2.1844)
exper	10.7415 (3.1125)

Average and standard deviation (in the parentheses)

6.2. **Methodology and Results.** As we discussed in the previous section, using our sample splitting method allow us to test all 7 sharp inequalities simultaneously, but may generate an efficiency loss. Since our sample is not very large, we start by testing inequalities that do not need the sample to be split and to see if they are informative enough to reject some instruments in the dataset under analysis.

6.2.1. *Non-sharp test without sample splitting.* We test inequalities (5.2) and (5.1) for each instrument separately. We perform the tests using the "clr2bound" command and the "local linear" method, we use the default choices of bandwidth and kernel functions recommended in CLR and CKLR, that is, $K(u) = \frac{15}{16}(1-u^2)^2\mathbf{1}\{|u| \leqslant 1\}$ and $h_{ROT} \times \hat{s} \times n^{\frac{1}{5}} \times n^{-\frac{2}{7}}$, where h_{ROT} is the rule of thumb choice given by Fan and Gijbels (1996). To avoid boundary issues, we compute the maximum of the test statistics over the interval $[Q^z_\alpha, Q^z_{1-\alpha}]$, where Q^z_α is the α-quantile of the instrument under testing. As robustness checks we consider four different dichotomizations of the outcome.[6] Table 2 reports the results of the test using inequality (5.1) at the three usual significance levels. We find also similar results for (5.2).

As can be seen, the validity of the parental education and number of siblings instruments is rejected at all three significance levels while the validity of college tuition is rejected at 5%. If we view this as jointly testing 4 hypotheses, then we should be concerned about controlling the family-wise error rate (FWER). Given the results in Table 2, the multiple testing procedure of Holm (1979) concludes that the validity of the parents' education and number of siblings instruments is rejected

[6] Y^k, $k \in \{20, 40, 60, 80\}$ denotes the indicator than the outcome is less than or equal to its k^{th} percentile.

Table 2. Testing instruments using GII: local linear

	Y^{20}			Y^{40}			Y^{60}			Y^{80}		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
motheduc	R	R	R	R	R	R	R	R	R	R	R	R
fatheduc	R	R	R	R	R	R	R	R	R	R	R	R
tuit17	NR	R	R	NR	R	R	NR	R	R	NR	R	R
sibs	R	R	R	R	R	R	R	R	R	R	R	R

"R" stands for rejection and "NR" stands for no rejection.

with the FWER controlled at the nominal rate of 4%. Additional tests show that the highest p-values of all 4 hypotheses of the college tuition fees validity is lower than 2.5%, so that the validity of the college tuition fees instrument is rejected with the FWER controlled at the nominal rate of 10%.

In Table 3, we conducted Pearson correlation statistical tests to see if there is a possible statistical association between the instruments and the "exper" variable and also between the instruments and the "abil" variable.

Table 3. Pearson's correlation coefficient

	abil	tuit17	fatheduc	motheduc	sibs	exper
abil	1					
tuit17	0.0567	1				
	(0.0468)					
fatheduc	0.3786	-0.0012	1			
	(0.0000)	(0.9665)				
motheduc	0.3799	-0.0498	0.5991	1		
	(0.0000)	(0.0810)	(0.0000)			
sibs	-0.1684	-0.0614	-0.1369	-0.2116	1	
	(0.0000)	(0.0313)	(0.0000)	(0.0000)		
exper	-0.4456	-0.0404	-0.3506	-0.3312	0.2010	1
	(0.0000)	(0.1564)	(0.0000)	(0.0000)	(0.0000)	

Pvalues (in parentheses)

As can be seen in Table 3, the absence of correlation between the instruments "fatheduc," "motheduc," and "sibs," and the covariates "exper," and "abil" is rejected at 1%. The absence of correlation between the instrument "tuit17" and "abil" is rejected at 5%, while we do not find significant correlation between "tuit17" and "exper." These statistical associations between the instruments and the covariates may be the cause of the rejections we observe in Table 2 since the tests have not been performed conditionally to the experience and/or the measure of ability variables. Interestingly, recall that Lemke and Rischall (2003) suggest that after conditioning on a measure of students' ability, parental education might be used as a valid instrument. To investigate this claim, we will test the independence between parental education and the potential earnings conditional on the "abil" and "exper" variables.

6.2.2. *Sharp test with controls using sample splitting.* We perform the tests using the "clrbound" command but this time with the "parametric regression" method. In fact, notice that in the presence of multiple instruments and controls the "nonparametric regression" method of the CLR test cannot be performed, mainly due to the curse of dimensionality: see CKLR and AKS for discussion. Examples of our use of the command are clearly spelled out in Section 4 of the online appendix. With this method, we may be concerned that some rejections are driven by a potential parametric misspecification. Therefore, we suggest that the researcher uses different parametric specifications as a robustness check when using the "parametric regression" method. For example, one can apply the command using different polynomial transformations of the instruments when they are continuous or ordered discrete-multivalued.[7]

We investigate the violation of the independence assumption for each instrument conditional on two main covariates $X \equiv [X_1, X_2] \equiv [abil, exper]$. The conditional JSI is a stronger assumption and so it implies a larger number of inequalities. As we show in the online appendix Section 3, the GII have to hold conditional on $X = x$ for each x in the support of X. To summarize the information, we will proceed as follows: for each instrument Z_k we will test the GII associated with the following assumptions:

(1) $(Y_1, Y_0) \perp Z_k | X_d = x_d$, where $x_d \in \mathcal{X}_d$, d $\in \{0, 1\}$.
(2) $(Y_1, Y_0) \perp Z_k | X_1 = x_1, X_2 = x_2$, where $x_2 \in \mathcal{X}_2$.

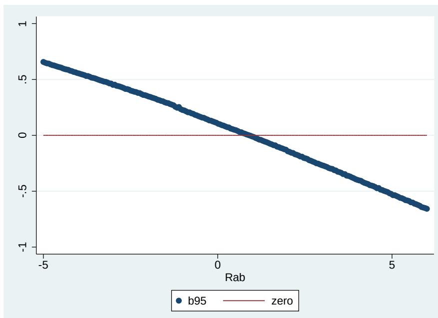
Here, $\mathcal{X}_d$ denotes the set of points on which the conditional JSI will be tested. Because the number of years of experience, *exper*, is a discrete random variable, $\mathcal{X}_d$ corresponds to its entire support. However, since *abil* has a continuous support, we discretize its support. As we discuss earlier we use different polynomial specifications for X as robustness checks. Detailed help for the implementation of these tests using the CKLR Stata packages can be found in Section 4.2 of the online appendix.

Figure 1 shows the results of the tests described in the two assumptions listed above for the parental education instruments. To be parsimonious, we analyze both mother and father education together as parental education. In (1), we control only on one covariate when implementing the GII test. Figure 2(a) depicts the result for *abil* as a control, and Figure 2(b) for *exper*. The dotted line represents the 95% lower bound of the one-sided confidence interval obtained at different values of x_d. When the dot associated with a value x_d^* is above the zero line, the assumption $(Y_1, Y_0) \perp Z_k | X_d = x_d^*$ is rejected. As robustness check of (1), we also consider a second-order polynomial instead of the linear parametric specification used in (1). The results are depicted in Figure 2(c) and Figure 2(d), where X_d represents both the ability and the experience. In general,

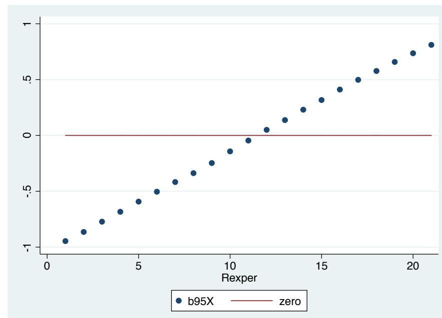
[7]This is explained in more detailed in Section 4 of the online appendix.

applied researchers control simultaneously on the vector of covariates $[X_1, X_2]$. In (2), we therefore fix *exper* at the value 4, and vary the ability, and afterwards fix *abil* at the value -3 and vary the experience.

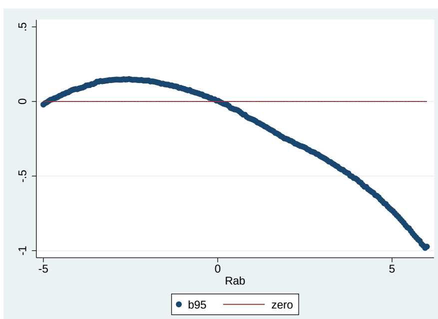
We use again a second order polynomial parametrization. The results are shown in Figures 2(e) and 2(f), respectively. Similarly, Figure 2 shows the results of the tests for the college tuition fees instrument.



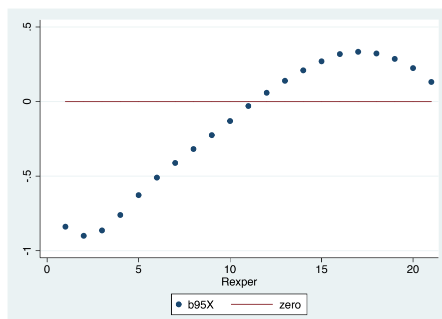
(a) *abil*



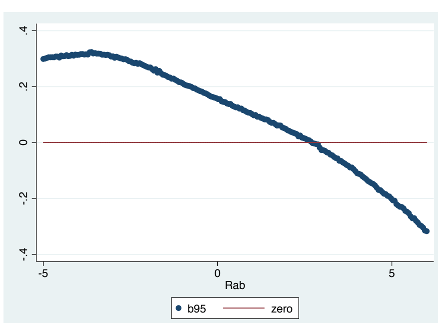
(b) *exper*



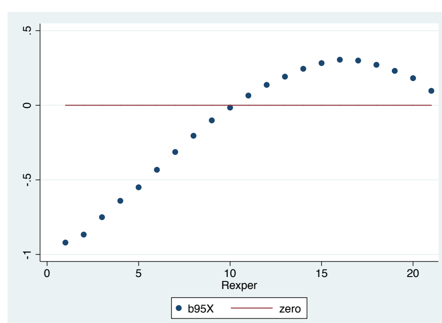
(c) *abil*, $abil^2$



(d) *exper*, $exper^2$

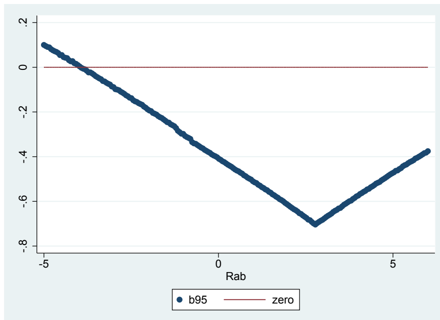


(e) *abil*, $abil^2$, *exper*, $exper^2$

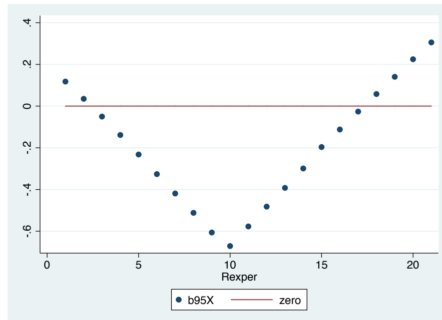


(f) *abil*, $abil^2$, *exper*, $exper^2$

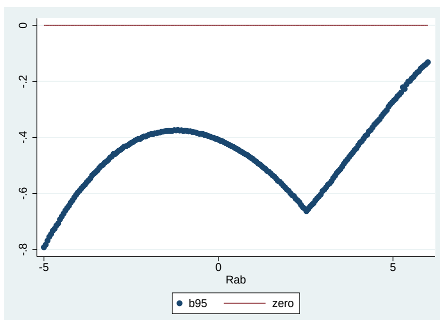
FIGURE 1. Testing GII for parents' education with controls.



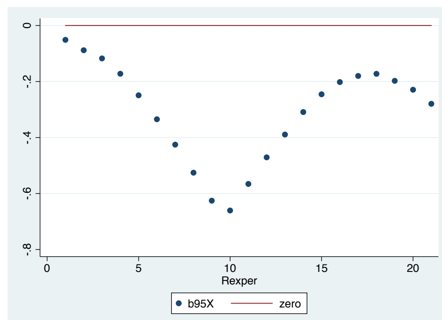
(a) *abil*



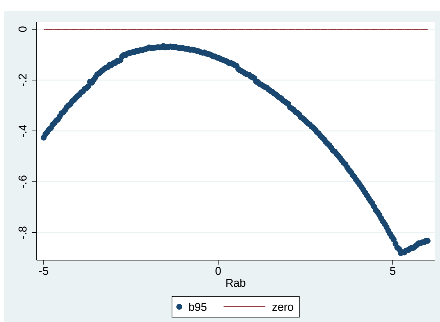
(b) *exper*



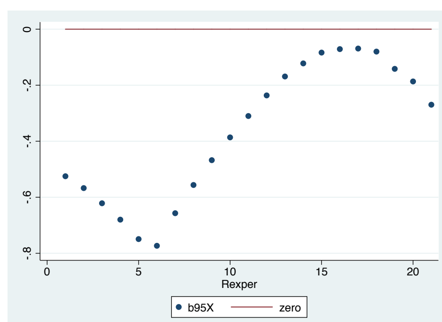
(c) $abil$, $abil^2$



(d) $exper$, $exper^2$



(e) $abil$, $abil^2$, $exper$, $exper^2$



(f) $abil$, $abil^2$, $exper$, $exper^2$

Figure 2. Testing GII for college tuition fees with controls.

6.3. **Summary and interpretation of the results.** The first main results obtained from our various tests and robustness checks suggest that the independence assumption between individual's potential earnings and parental education and number of siblings is rejected by the data, even conditional on experience and our measure of ability. Regarding the parents' education, the result seems to reject the idea suggested by Lemke and Rischall (2003), i.e., that controlling for some measure of ability might make parental education a valid instrument.

The second important result is the rejection of the independence assumption between individual's potential earnings and college tuition fees. However, while we still find rejections when controlling for experience only, we do not find evidence of violations of the independence assumption when controlling for ability or both ability and experience. The latter results suggests that, either the college tuition fees instrument is valid when controlling on measures of ability, or our test does not have enough power to reject this conditional independence assumption.

Overall, our test appears to be an interesting tool for applied researchers to verify the validity of any potential instruments, given the controls they have in hand.

7. Conclusion

In this paper, we propose a new set of testable implications of the instrumental variable (IV) independence assumption. When the outcome is binary, we show that the derived testable implications (GII) are the most informative for detecting all observable violations of a version of the JSI assumption. The GII also appear to be the strongest available in the literature so far for detecting the observable violations of the RA assumption.[8] Additionally, we propose a simple strategy for testing the generalized instrumental inequalities using intersection bounds inferential methods.

References

Andrews, D. W. K., W. Kim, and X. Shi (2016): "Stata Commands for Testing Conditional Moment Inequalities/Equalities," *Unpublished manuscript.*

Andrews, D. W. K., and X. Shi (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81, 609–666.

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

Baiocchi, M., J. Cheng, and D. S. Small (2014): "Instrumental variable methods for causal inference," *Statistics in Medicine*, 33, 2297–2340.

Balke, A., and J. Pearl (1997): "Bounds on Treatment Effects from Studies with Imperfect Compliance," *Journal of the American Statistical Association*, 92(439), 1171–1176.

Bonet, B. (2001): "Instrumentality Tests Revisited," *Proc. 17th Conf. on Uncertainty in Artificial Intelligence (UAI). Seattle, WA. Morgan Kaufmann.*, pp. 48–55.

Card, D. (2001): "Estimating the Return to Schooling: Progress on some Persistent Econometric Problems," *Econometrica*, 69, 1127–1160.

Chernozhukov, V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73(1), 245–261.

[8]However, at this point, we are not able to show that they are sufficient to detect all observable violations of the RA assumption. We are actively working on this question.

Chernozhukov, V., W. Kim, S. Lee, and A. M. Rosen (2015): "Implementing Intersection Bounds in Stata," *Stata Journal*, 15(1), 21–44.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

Deaton, A. S. (2009): "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development," *NBER Working Paper*, 14690.

Deaton, A. S., J. J. Heckman, and G. W. Imbens (2010): "Forum on the Estimation of Treatment Effects," *The Journal of Economic Literature*, 48, 356–455.

Fan, J., and I. Gijbels (1996): "Local Polynomial Modelling and Its Applications," *London, U.K.: Chapman & Hall*, 712.

Galichon, A., and M. Henry (2011): "Set identification in models with multiple equilibria," *The Review of Economic Studies*, 78(4), 1264–1298.

Heckman, J. J., J. L. Tobias, and E. J. Vytlacil (2001): "Four Parameters of Interest in the Evaluation of Social Programs," *Southern Economic Journal*, 68(2), 210–23.

Heckman, J. J., and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73(3), 669–738.

Holm, S. (1979): "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, 6(2), 65–70.

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–475.

Lemke, R. J., and I. C. Rischall (2003): "Skill, Parental Income, and IV Estimation of the Returns to Schooling," *Applied Economics Letters*, 10(5), 281–286.

Manski, C. F. (1990): "Nonparametric Bounds on Treatment Effects," *American Economic Reviews, Papers and Proceedings of the Hundred and Second Annual Meeting of the American Economic Association*, 80(2), 319–323.

Mourifié, I., M. Henry, and R. Méango (2015): "Sharp Bounds for the Roy Model," *Unpublished manuscript*.

Pearl, J. (1994): "On the Testability of Causal Models with Latent and Instrumental Variables," *Uncertainty in Articial Intelligence*, 11, 435–443.

Ramsahai, R. R., and S. L. Lauritzen (2011): "Likelihood Analysis of the Binary Instrumental Variable Model," *Biometrika*, 98(4), 987–994.

Romano, J. P., and A. M. Shaikh (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78(1), 169–211.

Rubin, D. B. (1974): "Estimating causal effects of treatments in randomized and nonrandomized studies.," *Journal of Educational Psychology*, 66(5), 688.

Wang, L., J. M. Robins, and T. S. Richarson (2017): "On falsification of the Binary Instrumental Variable Model," *Biometrika*, 104(1), 229–236.

Willis, R., and S. Rosen (1979): "Education and self-selection," *Journal of Political Economy*, 87(5), Pt2:S7–36.

8. Appendix

8.1. **Proof of Theorem 1.** (i) is proven in the main text.

(ii) We propose the following joint distribution $\tilde{p}_{ijk|z} \equiv \mathbb{P}(\tilde{Y_0} = i, \tilde{Y_1} = j, \tilde{D} = k | Z = z)$:

$$
\begin{aligned}
\tilde{p}_{010|z} &= \min(A, q_{00}(z), L - q_{11}(z), U + A - q_1(z)), \\
\tilde{p}_{011|z} &= A - \min(A, q_{00}(z), L - q_{11}(z), U + A - q_1(z)) = A - \tilde{p}_{010|z}, \\
\tilde{p}_{110|z} &= L - q_{11}(z) - \min(A, q_{00}(z), L - q_{11}(z), U + A - q_1(z)) = L - q_{11}(z) - \tilde{p}_{010|z}, \\
\tilde{p}_{111|z} &= q_{11}(z) - A + \min(A, q_{00}(z), L - q_{11}(z), U + A - q_1(z)) = q_{11}(z) - A + \tilde{p}_{010|z}, \\
\tilde{p}_{100|z} &= q_1(z) - L + \min(A, q_{00}(z), L - q_{11}(z), U + A - q_1(z)) = q_1(z) - L + \tilde{p}_{010|z}, \\
\tilde{p}_{101|z} &= U + A - q_1(z) - \min(A, q_{00}(z), L - q_{11}(z), U + A - q_1(z)) = U + A - q_1(z) - \tilde{p}_{010|z}, \\
\tilde{p}_{000|z} &= q_{00}(z) - \min(A, q_{00}(z), L - q_{11}(z), U + A - q_1(z)) = q_{00}(z) - \tilde{p}_{010|z}, \\
\tilde{p}_{001|z} &= 1 - U - A - q_{00}(z) + \min(A, q_{00}(z), L - q_{11}(z), U + A - q_1(z)) = 1 - U - A - q_{00}(z) + \tilde{p}_{010|z},
\end{aligned}
$$

where $A \equiv \max(0, 1 - \underline{q}_1 - \underline{q}_{10-01} - \underline{q}_0, \overline{q}_{00} - \underline{q}_0, \overline{q}_{11} - \underline{q}_1)$, $U \equiv \min(1 - \overline{q}_{00}, \underline{q}_1 + \underline{q}_{10-01})$ and $L \equiv \max(\overline{q}_{11}, 1 - \underline{q}_0 - \underline{q}_{10-01})$.

Intuitively, the distribution we propose follows the idea from the copula theory where the joint distribution can be decomposed in terms of marginals and the copula that describes the dependence structure. Specializing Proposition 1 to the binary outcome case provides bounds on the marginals $\mathbb{P}(Y_d = 1)$. We consider a joint distribution $\tilde{p}_{ij} \equiv \mathbb{P}(\tilde{Y_0} = i, \tilde{Y_1} = j)$ such that $\tilde{p}_{10} + \tilde{p}_{11}$ is equal to the derived upper bound of $\mathbb{P}(Y_0 = 1)$; $\tilde{p}_{01} + \tilde{p}_{11}$ is equal to the lower bounds of $\mathbb{P}(Y_1 = 1)$, and the inequalities on the joint distribution derived in Section 3 are verified, i.e., $\tilde{p}_{ij} \leqslant \underline{q}_{i0-j1}$. Moreover, because the potential outcome model we entertain here does not impose direct observable restrictions on $\tilde{p}_{ijk|z}$, we construct these probabilities such that $\tilde{p}_{000|z} + \tilde{p}_{010|z} = q_{00}(z)$, $\tilde{p}_{001|z} + \tilde{p}_{101|z} = q_{01}(z)$, $\tilde{p}_{100|z} + \tilde{p}_{110|z} = q_{10}(z)$, and $\tilde{p}_{011|z} + \tilde{p}_{111|z} = q_{11}(z)$.

In what follows, we use repeatedly the following: $a \geqslant b$ and $c \geqslant b$ implies $\max(a, c) \geqslant \max(b, d)$.

(1) well-defined probability. Notice that all defined quantities are positive and sum to one whenever equation (3.13) to (3.18) hold. So it is a well-defined probability distribution. Indeed, it is easy to see that $\tilde{p}_{011|z}, \tilde{p}_{101|z}, \tilde{p}_{000|z}$, and $\tilde{p}_{110|z}$ are non-negative since $a \geqslant \min(a, b)$.

$\tilde{p}_{010|z} \geqslant 0$, since $A \geqslant 0$, $q_{00}(z) \geqslant 0$, $L - q_{11}(z) = \max(\overline{q}_{11} - q_{11}(z), 1 - \underline{q}_0 - \underline{q}_{10-01} - q_{11}(z)) \geqslant 0$, $U + A - q_1(z) \geqslant 0$. The latter inequality holds because $A = \max(0, 1 - \underline{q}_1 - \underline{q}_{10-01} - \underline{q}_0, \overline{q}_{00} - \underline{q}_0, \overline{q}_{11} - \underline{q}_1) \geqslant \overline{q}_1 - U = \max(\overline{q}_1 - 1 + \overline{q}_{00}, \overline{q}_1 - \underline{q}_1 - \underline{q}_{10-01}) = \max(\overline{q}_{00} - \underline{q}_0, 1 - \underline{q}_0 - \underline{q}_1 - \underline{q}_{10-01})$.

To show $\tilde{p}_{111|z} \geqslant 0$, we have to show that $\min(A, q_{00}(z), L - q_{11}(z), U + A - q_1(z)) \geqslant A - q_{11}(z)$.
(i) $q_{00}(z) \geqslant A - q_{11}(z) \Leftrightarrow A \leqslant q_{00}(z) + q_{11}(z)$. The latter inequality holds because $A - \underline{q}_{00-11} = \max(-\underline{q}_{00-11}, 1 - \underline{q}_1 - \underline{q}_{10-01} - \underline{q}_0 - \underline{q}_{00-11}, \overline{q}_{00} - \underline{q}_0 - \underline{q}_{00-11}, \overline{q}_{11} - \underline{q}_1 - \underline{q}_{00-11}) \leqslant 0$ under **(3.18) (3.15) and (3.16)**.
(ii) $L - q_{11}(z) \geqslant A - q_{11}(z) \Leftrightarrow L \geqslant A$.
(iii) $U + A - q_1(z) \geqslant A - q_{11}(z) \Leftrightarrow U \geqslant q_{10}(z)$, and the latter inequality holds under **(3.12) and (3.14)**.

To show $\tilde{p}_{100|z} \geqslant 0$, we have to show that $\min(A + q_1(z), q_{00}(z) + q_1(z), L + q_1(z) - q_{11}(z), U + A) \geqslant L \Leftrightarrow \min(A + q_1(z), 1 - q_{01}(z), L + q_{10}(z), U + A) \geqslant L$. (i) $A + q_1(z) \geqslant L \Leftrightarrow A \geqslant L - q_1(z) \Leftrightarrow A \geqslant \max(\overline{q}_{11} - q_1(z), 1 - \underline{q}_0 - \underline{q}_{10-01} - q_1(z))$.
(ii) $1 - q_{01}(z) \geqslant L$ holds since **(3.13) and (3.17)**.
(iii) To show $U + A \geqslant L$ we will show that $U + A \geqslant \overline{q}_{11}$ and $U + A \geqslant 1 - \underline{q}_0 - \underline{q}_{10-01}$
(iii.1)$U + A \geqslant \overline{q}_{11} \Leftrightarrow A \geqslant \max(-1 + \overline{q}_{00} + \overline{q}_{11}, \overline{q}_{11} - \underline{q}_1 - \underline{q}_{10-01})$. The latter inequality holds since $\overline{q}_{00} - \underline{q}_0 = \overline{q}_{00} - 1 + \overline{q}_1 \geqslant -1 + \overline{q}_{00} + \overline{q}_{11}$ and $\overline{q}_{11} - \underline{q}_1 \geqslant \overline{q}_{11} - \underline{q}_1 - \underline{q}_{10-01}$.
(iii.2)$U + A \geqslant 1 - \underline{q}_0 - \underline{q}_{10-01} \Leftrightarrow A \geqslant \max(\overline{q}_{00} - \underline{q}_0 - \underline{q}_{10-01}, 1 - \underline{q}_0 - \underline{q}_{10-01} - \underline{q}_1 - \underline{q}_{10-01})$, which follows from $\overline{q}_{00} - \underline{q}_0 \geqslant \overline{q}_{00} - \underline{q}_0 - \underline{q}_{10-01}$ and $1 - \underline{q}_0 - \underline{q}_{10-01} - \underline{q}_1 \geqslant 1 - \underline{q}_0 - \underline{q}_{10-01} - \underline{q}_1 - \underline{q}_{10-01}$.

To show $\tilde{p}_{001|z} \geqslant 0$ we have to show that $1 - U - A - q_{00}(z) + \min(A, q_{00}(z), L - q_{11}(z), U + A - q_1(z)) \geqslant 0 \Leftrightarrow 1 - U \geqslant \max(q_{00}(z), A, A + q_{11-00}(z) - L, q_{00}(z) + q_1(z) - U)$. The latter inequality holds because:
(i) $1 - U \geqslant q_{00}(z) \Leftrightarrow \max(\overline{q}_{00}, 1 - \underline{q}_1 - \underline{q}_{10-01}) \geqslant q_{00}(z)$, where the latter trivially holds.
(ii) $1 - U \geqslant A \Leftrightarrow \max(\overline{q}_{00}, 1 - \underline{q}_1 - \underline{q}_{10-01}) \geqslant \max(0, 1 - \underline{q}_1 - \underline{q}_{10-01} - \underline{q}_0, \overline{q}_{00} - \underline{q}_0, \overline{q}_{11} - \underline{q}_1)$. The latter inequality holds because (1) $\overline{q}_{00} \geqslant \max(0, \overline{q}_{00} - \underline{q}_0)$ and (2) $\overline{q}_{00-11} \geqslant \overline{q}_{11}$ implies $1 - \underline{q}_1 - \underline{q}_{10-01} \geqslant \max(1 - \underline{q}_1 - \underline{q}_{10-01} - \underline{q}_0, \overline{q}_{11} - \underline{q}_1)$.
(iii) $1 - U \geqslant A + q_{11-00}(z) - L \Leftrightarrow L + 1 - U - A \geqslant q_{00-11}(z)$. The latter holds since $L + 1 - U \geqslant A + \overline{q}_{00-11}$. In fact, $L + 1 - U \geqslant A + \overline{q}_{00-11} \Leftrightarrow \max(\overline{q}_{00} + \overline{q}_{11}, 1 - \underline{q}_0 - \underline{q}_{10-01} + \overline{q}_{00}, 1 - \underline{q}_1 - \underline{q}_{10-01} + \overline{q}_{11}, 2 - \underline{q}_1 - 2\underline{q}_{10-01} - \underline{q}_0) \geqslant \max(\overline{q}_{00-11}, 1 - \underline{q}_1 - \underline{q}_{10-01} - \underline{q}_0 + \overline{q}_{00-11}, \overline{q}_{00} - \underline{q}_0 + \overline{q}_{00-11}, \overline{q}_{11} - \underline{q}_1 + \overline{q}_{00-11})$ where the latter inequality holds whenever we remark that $\overline{q}_{00} + \overline{q}_{11} \geqslant \overline{q}_{00-11}$, $1 - \underline{q}_0 - \underline{q}_{10-01} + \overline{q}_{00} = \overline{q}_{00} - \underline{q}_0 + \overline{q}_{00-11}$, $1 - \underline{q}_1 - \underline{q}_{10-01} + \overline{q}_{11} = \overline{q}_{11} - \underline{q}_1 + \overline{q}_{00-11}$) and $2 - \underline{q}_1 - 2\underline{q}_{10-01} - \underline{q}_0 = 1 - \underline{q}_1 - \underline{q}_{10-01} - \underline{q}_0 + \overline{q}_{00-11}$. Finally,
(iv) $1 - U \geqslant q_{00}(z) + q_1(z) - U \Leftrightarrow q_0(z) \geqslant q_{00}(z)$.

(2) Verification of the JSI assumption. We can easily verify that:

$$\begin{aligned}
\tilde{p}_{01|z} &= \max(0, 1-\underline{q}_1 - \underline{q}_{10-01} - \underline{q}_0, \overline{q}_{00} - \underline{q}_0, \overline{q}_{11} - \underline{q}_1) = A, \\
\tilde{p}_{11|z} &= L - A, \\
\tilde{p}_{10|z} &= U - L + A, \\
\tilde{p}_{00|z} &= 1 - U - A.
\end{aligned}$$

Notice that all the right-side quantities are independent of z. Therefore, $(Y_1, Y_0) \perp Z$, meaning that JSI holds by construction. For the sake of simplicity, let us denote $\tilde{p}_{ij} \equiv \tilde{p}_{ij|z}$.

(3) $(\tilde{Y}, \tilde{D}, Z)$ **has the same distribution as** (Y, D, Z)**.** We can easily see that $\mathbb{P}(\tilde{Y} = j, \tilde{D} = 1|Z = z) = \mathbb{P}(\tilde{Y}_1 = j, \tilde{D} = 1|Z = z) = \tilde{p}_{0j1|z} + \tilde{p}_{1j1|z} = q_{j1}(z) = \mathbb{P}(Y = j, D = 1|Z = z)$ for $j \in \{0, 1\}$. Similarly, we have $\mathbb{P}(\tilde{Y} = i, \tilde{D} = 0|Z = z) = \mathbb{P}(\tilde{Y}_0 = i, \tilde{D} = 0|Z = z) = \tilde{p}_{i10|z} + \tilde{p}_{i00|z} = q_{i0}(z) = \mathbb{P}(Y = i, D = 0|Z = z)$ for $i \in \{0, 1\}$.

(4) Compatibility between the observable and the unobservable distributions. The potential outcome model and the JSI assumption impose a set of restrictions between the unobservables (Y_0, Y_1) and the observables (Y, D, Z). Equations (3.2) and (3.7) are some examples of those restrictions. We have to show that our constructed distribution respects all those restrictions. Theorem 1 of Galichon and Henry (2011) allows to characterize all the restrictions imposed by the potential outcome and the JSI assumption on the distribution of unobservables (Y_0, Y_1) and the distribution of the observables (Y, D, Z). Indeed, the potential outcome model can be defined equivalently as a correspondence G between values of unobservables $(y_0, y_1) \in \mathcal{A} = \{(0,0), (0,1), (1,0), (1,1)\}$ and values of observables $(y, d) \in \mathcal{A}$. G is defined by

$$G(0,0) = \{(0,1), (0,0)\}, G(0,1) = \{(0,0), (1,1)\}, G(1,1) = \{(1,1), (1,0)\}, G(1,0) = \{(1,0), (0,1)\}.$$

By Theorem 1 of Galichon and Henry (2011), we have:

$$\begin{aligned}
\forall\ A \subset \mathcal{A}, \qquad \mathbb{P}((Y, D) \in A|Z = z) &\leqslant \mathbb{P}(G(Y_0, Y_1) \cap A \neq \varnothing|Z = z) \\
&= \mathbb{P}(G(Y_0, Y_1) \cap A \neq \varnothing), \ \forall z \in \mathcal{Z},
\end{aligned}$$

that is:
for singletons,

$$\begin{aligned}
q_{11}(z) &\leqslant p_{01} + p_{11}, && (8.1) \\
q_{01}(z) &\leqslant p_{00} + p_{10}, && (8.2) \\
q_{10}(z) &\leqslant p_{11} + p_{10}, && (8.3) \\
q_{00}(z) &\leqslant p_{00} + p_{01}; && (8.4)
\end{aligned}$$

for pairs,

$$q_{11}(z)+q_{10}(z) \leqslant p_{11}+p_{01}+p_{10}, \tag{8.5}$$

$$q_{11}(z)+q_{01}(z) \leqslant 1, \tag{8.6}$$

$$q_{11}(z)+q_{00}(z) \leqslant p_{11}+p_{00}+p_{01}, \tag{8.7}$$

$$q_{10}(z)+q_{01}(z) \leqslant p_{00}+p_{11}+p_{10}, \tag{8.8}$$

$$q_{10}(z)+q_{00}(z) \leqslant 1, \tag{8.9}$$

$$q_{01}(z)+q_{00}(z) \leqslant p_{00}+p_{01}+p_{10}; \tag{8.10}$$

for triplets,

$$q_{11}(z)+q_{10}(z)+q_{01}(z) \leqslant 1, \tag{8.11}$$

$$q_{11}(z)+q_{10}(z)+q_{00}(z) \leqslant 1, \tag{8.12}$$

$$q_{11}(z)+q_{01}(z)+q_{00}(z) \leqslant 1, \tag{8.13}$$

$$q_{10}(z)+q_{01}(z)+q_{00}(z) \leqslant 1. \tag{8.14}$$

From the previous enumerates, we know that $\tilde{p} \equiv (\tilde{p}_{01}, \tilde{p}_{00}, \tilde{p}_{11}, \tilde{p}_{10})$ is a well-defined probability distribution, that is, $\tilde{p}_{01} \geqslant 0,\ \tilde{p}_{00} \geqslant 0,\ \tilde{p}_{11} \geqslant 0,\ \tilde{p}_{10} \geqslant 0$ and $\tilde{p}_{01}+\tilde{p}_{00}+\tilde{p}_{11}+\tilde{p}_{10}$=1.

It remains to show that the probability distribution $\tilde{p} \equiv (\tilde{p}_{01}, \tilde{p}_{00}, \tilde{p}_{11}, \tilde{p}_{10})$ satisfies all inequalities (8.1) - (8.14). Note that $\tilde{p}$ trivially satisfies inequalities (8.6), (8.9), and (8.11) to (8.14).

(a) Singletons: By construction, we have $\tilde{p}_{01}+\tilde{p}_{11} = L = \max(\overline{q}_{11}, 1-\underline{q}_0-\underline{q}_{10-01}) \geqslant \overline{q}_{11}$, then (8.1) is satisfied. Similarly, $\tilde{p}_{00}+\tilde{p}_{10} = 1-L = \min(1-\overline{q}_{11}, \underline{q}_0+\underline{q}_{10-01}) \geqslant \overline{q}_{01}$ since **(3.13) and (3.17)** hold, then (8.2) is satisfied. $\tilde{p}_{11}+\tilde{p}_{10} = U = \min(1-\overline{q}_{00}, \underline{q}_1+\underline{q}_{10-01}) \geqslant \overline{q}_{10}$ since **(3.12) and (3.14)** hold, then (8.3) is satisfied. $\tilde{p}_{00}+\tilde{p}_{01} = 1-U = \max(\overline{q}_{00}, 1-\underline{q}_1-\underline{q}_{10-01}) \geqslant \overline{q}_{00}$ trivially holds then (8.4) is satisfied.

(b) Pairs: By construction, we have $\tilde{p}_{11}+\tilde{p}_{01}+\tilde{p}_{10} = U+\tilde{p}_{01} = \min(1-\overline{q}_{00}+\tilde{p}_{01}, \underline{q}_1+\underline{q}_{10-01}+\tilde{p}_{01}) \geqslant \overline{q}_1$, because $1-\overline{q}_{00}+\tilde{p}_{01} \geqslant \overline{q}_1$ and $\underline{q}_1+\underline{q}_{10-01}+\tilde{p}_{01} \geqslant \overline{q}_1$. Indeed, we have $1-\overline{q}_{00}+\tilde{p}_{01} = \max(1-\overline{q}_{00}, 1-\overline{q}_{00}+1-\underline{q}_1-\underline{q}_{10-01}-\underline{q}_0, 1-\underline{q}_0, 1-\overline{q}_{00}+\overline{q}_{11}-\underline{q}_1) \geqslant 1-\underline{q}_0 = \overline{q}_1$ and $\underline{q}_1+\underline{q}_{10-01}+\tilde{p}_{01} = \max(\underline{q}_1+\underline{q}_{10-01}, 1-\underline{q}_0, \underline{q}_1+\underline{q}_{10-01}+\overline{q}_{00}-\underline{q}_0, \underline{q}_1+\underline{q}_{10-01}+\overline{q}_{11}-\underline{q}_1) \geqslant 1-\underline{q}_0 = \overline{q}_1$. Therefore, (8.5) is satisfied. Now, let us show that $\tilde{p}_{11}+\tilde{p}_{00}+\tilde{p}_{01} \geqslant \overline{q}_{00-11} \Leftrightarrow \tilde{p}_{10} \leqslant 1-\overline{q}_{00-11} = \underline{q}_{10-01}$, where $\tilde{p}_{10} = U-L+\tilde{p}_{01} = \max(U-L, U-L+1-\underline{q}_1-\underline{q}_{10-01}-\underline{q}_0, U-L+\overline{q}_{00}-\underline{q}_0, U-L+\overline{q}_{11}-\underline{q}_1)$. (8.7) is satisfied since each argument of the max function is less than or equal to $\underline{q}_{10-01}$. Indeed, we have:

(i) $U-L = \min(-\overline{q}_{11}+U, -1+\underline{q}_0+\underline{q}_{10-01}+U) \leqslant -\overline{q}_{11}+U = \min(1-\overline{q}_{00}-\overline{q}_{11}, -\overline{q}_{11}+\underline{q}_{10-01}) \leqslant 1-\overline{q}_{00}-\overline{q}_{11} \leqslant \underline{q}_{10-01}$ since $\overline{q}_{00}+\overline{q}_{11} \geqslant \overline{q}_{00-11} = 1-\underline{q}_{10-01}$.

(ii) $U-L+1-\underline{q}_1-\underline{q}_{10-01}-\underline{q}_0 = \min(1-\underline{q}_1-\underline{q}_{10-01}-\underline{q}_0-\overline{q}_{11}+U, -\underline{q}_1+U) \leqslant U-\underline{q}_1 = \min(1-\overline{q}_{00}-\underline{q}_1, \underline{q}_{10-01}) \leqslant \underline{q}_{10-01}$.

(iii) $U - L + \overline{q}_{00} - \underline{q}_0 = \min(\overline{q}_{00} - \underline{q}_0 - \overline{q}_{11} + U, \overline{q}_{00} - 1 + \underline{q}_{10-01} + U) \leqslant \overline{q}_{00} - 1 + \underline{q}_{10-01} + U = \min(\underline{q}_{10-01}, \overline{q}_{00} - 1 + \underline{q}_{10-01} + \underline{q}_1 + \underline{q}_{10-01}) \leqslant \underline{q}_{10-01}$.
(iv) $U - L + \overline{q}_{11} - \underline{q}_1 = \min(-\underline{q}_1 + U, \overline{q}_{11} - \underline{q}_1 - 1 + \underline{q}_0 + \underline{q}_{10-01} + U) \leqslant -\underline{q}_1 + U = \min(1 - \overline{q}_{00} - \underline{q}_1, \underline{q}_{10-01}) \leqslant \underline{q}_{10-01}$. Furthermore, we have $\tilde{p}_{00} + \tilde{p}_{11} + \tilde{p}_{10} = 1 - \tilde{p}_{01} \geqslant \overline{q}_{10-01} \Leftrightarrow \tilde{p}_{01} \leqslant 1 - \overline{q}_{10-01} = \underline{q}_{00-11} \Leftrightarrow \max(0, 1 - \underline{q}_1 - \underline{q}_{10-01} - \underline{q}_0, \overline{q}_{00} - \underline{q}_0, \overline{q}_{11} - \underline{q}_1) \leqslant \underline{q}_{00-11}$, where the latter inequality holds because **(3.15), (3.18) and (3.16)** hold. Then (8.8) is satisfied.

Finally, we have $\tilde{p}_{00} + \tilde{p}_{01} + \tilde{p}_{10} = 1 - \tilde{p}_{11} \geqslant \overline{q}_0 \Leftrightarrow \tilde{p}_{11} \leqslant 1 - \overline{q}_0 = \underline{q}_1$, where the latter inequality holds since we remark that $\tilde{p}_{11} = L - \tilde{p}_{01}$ and $\tilde{p}_{01} + \underline{q}_1 = \max(\underline{q}_1, 1 - \underline{q}_{10-01} - \underline{q}_0, \overline{q}_{00} - \underline{q}_0 + \underline{q}_1, \overline{q}_{11}) \geqslant \max(\overline{q}_{11}, 1 - \underline{q}_{10-01} - \underline{q}_0) = L$. Therefore, (8.10) is satisfied.

8.2. **Proof of Theorem 2.** It is enough to show that (3.14) to (3.18) always hold whenever (3.13) and (3.12) hold. We show that it is the case for (3.14), inequalities (3.15) to (3.17) can be derived accordingly. Because $\mathcal{Z} = \{0, 1\}$, $\overline{q}_{10} \leqslant \underline{q}_1 + \underline{q}_{10-01}$ holds if

$$q_{10}(z) \leqslant q_1(z) + q_{10-01}(z), \ z = 0, 1; \tag{8.15}$$

$$q_{10}(z) \leqslant q_1(1-z) + q_{10-01}(z), \ z = 0, 1; \tag{8.16}$$

$$q_{10}(z) \leqslant q_1(z) + q_{10-01}(1-z), \ z = 0, 1; \tag{8.17}$$

$$q_{10}(z) \leqslant q_1(1-z) + q_{10-01}(1-z), \ z = 0, 1. \tag{8.18}$$

The first three inequalities hold trivially. The last inequality holds since (3.12) holds. In fact (3.12) implies that $q_{10}(z) \leqslant q_{11}(1-z) + q_{01}(1-z) + q_{10}(1-z)$.

Notice that if z can take more than two values i.e., z_1, z_2 and z_3, $\overline{q}_{10} \leqslant \underline{q}_1 + \underline{q}_{10-01}$ would encompass the following case

$$q_{10}(z_1) \leqslant q_1(z_2) + q_{10-01}(z_3)$$

which is not implied by (3.12).

Now, let us show that inequality (3.18) holds, i.e., $1 \leqslant \underline{q}_1 + \underline{q}_{10-01} + \underline{q}_0 + \underline{q}_{00-11}$. The nontrivial cases where the latter inequality may not hold are those for which $\arg\min_z q_0(z) \neq \arg\min_z q_1(z)$ and $\arg\min_z q_{10-01}(z) \neq \arg\min_z q_{00-11}(z)$. Without loss of generality, let us assume that

$$\underline{q}_1 + \underline{q}_{10-01} + \underline{q}_0 + \underline{q}_{00-11} = q_1(z) + q_{10-01}(z) + q_0(1-z) + q_{00-11}(1-z).$$

We have

$$\begin{aligned} q_1(z) + q_{10-01}(z) + q_0(1-z) + q_{00-11}(1-z) &= 1 + q_{10}(z) - q_{00}(z) + 1 + q_{00}(1-z) - q_{10}(1-z) \\ &= 1 + q_{10}(z) + q_{00}(1-z) + (1 - q_{00}(z) - q_{10}(1-z)) \\ &\geqslant 1 + q_{10}(z) + q_{00}(1-z), \end{aligned}$$

where the last inequality follows from (3.12). This completes the proof.

8.3. **Proof of Proposition 1.** Under the MSI we have:

$$\begin{aligned} f_{Y_d}(y) &= f_{Y_d|Z=z}(y), \\ &= f_{Y,D}(y, D = d|z) + f_{Y_d,D}(y, D = 1-d|z), \\ &\geqslant f_{Y,D}(y, D = d|z). \end{aligned} \tag{8.19}$$

Then,

$$\sup_z f_{Y,D}(y, d|z) \leqslant f_{Y_d}(y). \tag{8.20}$$

So, for any arbitrary $A \in P_{\mathcal{Y}}$, we have

$$\begin{aligned} \mathbb{P}(Y_d \in A) &= 1 - \mathbb{P}(Y_d \in \overline{A}), \\ &\leqslant 1 - \int_{\overline{A}} \sup_z f_{Y,D}(y, d|z) d\mu(y), \end{aligned} \tag{8.21}$$

where $\overline{A}$ denotes the complement of the event A. The last inequality holds due to inequality (8.20).

For any arbitrary $A \in P_{\mathcal{Y}}$ and $B \in P_{\mathcal{Y}}$, we have:

$$\begin{aligned} \mathbb{P}(Y_d \in A, Y_{1-d} \in B) &= \mathbb{P}(Y_d \in A, Y_{1-d} \in B|Z = z), \\ &= \mathbb{P}(Y_d \in A, Y_{1-d} \in B, D = d|Z = z) + \mathbb{P}(Y_d \in A, Y_{1-d} \in B, D = 1-d|Z = z), \\ &\leqslant \mathbb{P}(Y_d \in A, D = d|Z = z) + \mathbb{P}(Y_{1-d} \in B, D = 1-d|Z = z), \end{aligned}$$

where the first equality holds under the JSI assumption. Therefore,

$$\mathbb{P}(Y_d \in A, Y_{1-d} \in B) \leqslant \inf_z \{\mathbb{P}(Y \in A, D = d|Z = z) + \mathbb{P}(Y \in B, D = 1-d|Z = z)\}. \tag{8.22}$$

From eq (8.20) and eq (8.22), we have

$$\int_A \sup_z f_{Y,D}(y, d|z) d\mu(y) \leqslant \sum_{B \in P_{\mathcal{Y}}} \mathbb{P}(Y_d \in A, Y_{1-d} \in B) \leqslant \sum_{B \in P_{\mathcal{Y}}} \inf_z \{\mathbb{P}(A, d|z) + \mathbb{P}(B, 1-d|z)\}. \tag{8.23}$$

Combining eq (8.21) with eq (8.23), we have the following upper bound:

$$\mathbb{P}(Y_d \in A) \leqslant \min\left[1 - \int_{\overline{A}} \sup_z f_{Y,D}(y, d|z) d\mu(y), \sum_{B \in P_{\mathcal{Y}}} \inf_z \left\{\mathbb{P}(A, d|z) + \mathbb{P}(B, 1-d|z)\right\}\right].$$

Notice that the first term of the upper bound can be derived just under the MSI assumption while the second term requires the JSI assumption. For the lower bounds, we have

$$\mathbb{P}(Y_d \in A) = 1 - \sum_{B \in P_{\mathcal{Y}} \setminus A} \mathbb{P}(Y_d \in B) \geqslant 1 - \sum_{B \in P_{\mathcal{Y}} \setminus A} \sum_{C \in P_{\mathcal{Y}}} \inf_z \left\{\mathbb{P}(B, d|z) + \mathbb{P}(C, 1-d|z)\right\}. \tag{8.24}$$

Combining eq (8.23) with eq (8.24), we have the following lower bound:

$$\mathbb{P}(Y_d \in A) \geqslant \max\left[\int_A \sup_z f_{Y,D}(y, d|z) d\mu(y), 1 - \sum_{B \in P_{\mathcal{Y}} \setminus A} \sum_{C \in P_{\mathcal{Y}}} \inf_z \left\{\mathbb{P}(B, d|z) + \mathbb{P}(C, 1-d|z)\right\}\right].$$

This completes the proof.

GENERALIZED INSTRUMENTAL INEQUALITIES: TESTING THE IV INDEPENDENCE ASSUMPTION

DÉSIRÉ KÉDAGNI AND ISMAEL MOURIFIÉ

The Pennsylvania State University and University of Toronto

Abstract. This is an online appendix that contains complementary results of the main paper and detailed information of our proposed testing procedure using the Chernozhukov et al. (2015, CKLR) Stata package.

Keywords: Instrumental variable, Independence assumption, Sharp inequalities, Intersection bounds.

JEL subject classification: C12, C15, C21, C24.

Date: The present version is of April 4, 2018. Correspondence address: Department of Economics, University of Toronto, 150 St. George Street, Toronto ON M5S 3G7, Canada, ismael.mourifie@utoronto.ca.

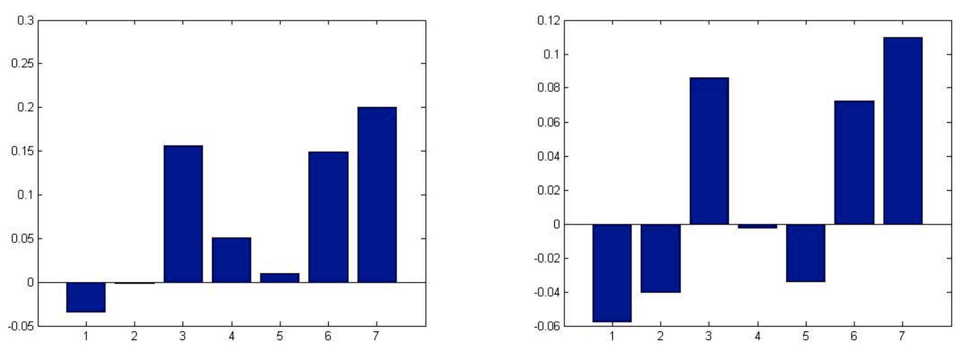
1. Numerical illustration.

We provide a numerical illustration, which shows that Pearl's instrumental inequality and Bonet's (2001) inequality are not sufficient in general to detect all observable violations of the JSI assumption. We consider the following data-generating process (DGP):

$$\begin{aligned} Y &= 1\{|D+U| > \delta\} \\ D &= 1\left\{|\frac{Z}{100} - V| > 1\right\} \\ Z &= \sum_{k=1}^{100} k * 1\left\{\Phi^{-1}\left(\frac{k-1}{100}\right) < Z^* \leqslant \Phi^{-1}\left(\frac{k}{100}\right)\right\}, \end{aligned}$$

where $U = Z^* + V$ and $\begin{pmatrix} Z^* \\ V \end{pmatrix} \sim N(0, \Sigma)$ with $\Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$.

The probabilities $q_{ij}(z)$ are approximated using their empirical analogous over an i.i.d. sample of size of 1,000,000. Figure 1 represents respectively inequalities (3.12) to (3.18) of the main text in a sequence 1 to 7. For instance, the number 1 on the x-axis represents the quantity $\overline{q}_{11} + \overline{q}_{01} - 1$. When the plot is below 0, we do not have a violation, otherwise the corresponding inequality is violated.



(a) $\delta = 2.8$ (b) $\delta = 3.1$

Figure 1. Violations of the sharp instrumental inequalities.

Figure 1(a) shows a case where all inequalities (3.14) to (3.18) are violated while Pearl's instrumental inequalities, (3.12) and (3.13), are not. Figure 1(b) shows a case where Pearl and Bonet's (2001) instrumental inequalities are simultaneously not violated while 3 of the new derived inequalities are violated. Therefore, Figure 1(a) proves that Pearl's instrumental inequalities are not sufficient to detect all observable violations of the JSI assumption while Figure 1(b) proves that Bonet's inequalities are also not sufficient for that.

2. Continuous/multi-valued discrete outcome and multi-valued discrete treatment.

In this section, we extend the model to the multi-valued discrete treatment case. The extended formulation of the JSI can be written as: the vector of potential outcomes $(Y_0, Y_1, ..., Y_T)$ is statistically independent of Z, i.e. $(Y_0, Y_1, ..., Y_T) \perp Z$. The Pearl inequality becomes:

$$\max_{d \in \{0,1,\ldots,T\}} \int_{\mathcal{Y}} \sup_z f_{Y,D}(y, d|z) d\mu(y) \leqslant 1. \tag{2.1}$$

The new instrumental inequalities can be extended as follows: for any arbitrary vector $(A_0, ..., A_t) \in P_{\mathcal{Y}}^{t+1} \equiv \underbrace{P_{\mathcal{Y}} \times \ldots \times P_{\mathcal{Y}}}_{t+1 \ times}$, $t \in \{2, \ldots, T\}$, we have:

$$\mathbb{P}(Y_0 \in A_1, ..., Y_t \in A_t) \quad \leqslant \quad \inf_z \Big\{ \mathbb{P}(A_d, d|z) + \sum_{l \neq d}^{t} \mathbb{P}(A_l, l|z) \Big\}.$$

Then

$$\mathbb{P}(Y_d \in A_d) \quad \leqslant \quad \sum_{(A_0,\ldots,A_{d-1},A_{d+1},\ldots,A_t) \in P_{\mathcal{Y}}^t} \inf_z \Big\{ \mathbb{P}(A_d, d|z) + \sum_{l \neq d}^{t} \mathbb{P}(A_l, l|z) \Big\}.$$

Using also $\mathbb{P}(Y_d \in A_d) \geqslant \sup_{A_d \in P_{\mathcal{Y}}} \int_{A_d} \sup_z f_{Y,D}(y, d|z) d\mu(y)$, we obtain the following inequality:

$$\sup_{d \in \{0,\ldots,t\}} \sup_{A_d \in P_{\mathcal{Y}}} \left(\int_{A_d} \sup_z f_{Y,D}(y, d|z) d\mu(y) - \sum_{(A_0,\ldots,A_{d-1},A_{d+1},\ldots,A_t) \in P_{\mathcal{Y}}^t} \inf_z \Big\{ \mathbb{P}(A_d, d|z) + \sum_{l \neq d}^{t} \mathbb{P}(A_l, l|z) \Big\} \right) \leqslant 0. \tag{2.2}$$

For the latter inequality, we have:

$$1 \quad = \quad \sum_{(A_0,\ldots,A_t) \in P_{\mathcal{Y}}^{t+1}} \mathbb{P}(Y_0 \in A_1, ..., Y_t \in A_t) \quad \leqslant \quad \sum_{(A_0,\ldots,A_t) \in P_{\mathcal{Y}}^{t+1}} \inf_z \Big\{ \sum_{l=0}^{t} \mathbb{P}(A_l, l|z) \Big\}.$$

Therefore, we finally have:

$$1 \quad \leqslant \quad \sum_{(A_0,\ldots,A_t) \in P_{\mathcal{Y}}^{t+1}} \inf_z \Big\{ \sum_{l=0}^{t} \mathbb{P}(A_l, l|z) \Big\}. \tag{2.3}$$

To summarize, inequalities (2.1), (2.2), and (2.3) represent the most general form of testable implications, obtained so far, for both the JSI and RA assumptions. They encompass cases where the outcome and the instrument can be discrete, continuous, or mixed. The only restriction is imposed on the treatment, which is restricted to be discrete. Pearl (1994) already pointed out the impossibility of extending his instrumental inequality to the continuous treatment case. Based on some additional remarks, Pearl conjectured that the RA assumption has no testable implications when the treatment is continuous. In other words, the RA assumption has no testable implications when the treatment is continuous. Subsequently, Bonet (2001) provided additional results, which appears to confirm Pearl's (1994) conjecture. However, both analyses did not fully take into account the new set of testable implications derived here. Using this new set of inequalities, we reopen the question of the non-testability of the IV independence assumption in models with fully unrestricted variables. We are currently working on this open conjecture.

Throughout the text, we enumerate three different types of IV independence assumption. The validity of the MSI assumption should be tested using only (2.1), while the validity of the JSI and RA assumption should be tested using inequalities (2.1), (2.2), and (2.3). Now, let us propose the generalization of our testing procedure to a multi-valued discrete outcome.

2.1. **Testing procedure for a multi-valued discrete outcome.** We have

$$\max_{d\in\{0,1\}} \sum_{y\in\mathcal{Y}} \sup_{z} \mathbb{P}(y,d|z) \leqslant 1, \tag{2.4}$$

$$\max_{d\in\{0,1\}} \max_{y\in\mathcal{Y}} \left(\sup_{z} \mathbb{P}(y,d|z) - \sum_{y'\in\mathcal{Y}} \inf_{z} \left\{\mathbb{P}(y,d|z) + \mathbb{P}(y',1-d|z)\right\} \right) \leqslant 0, \tag{2.5}$$

and

$$1 \quad \leqslant \sum_{y\in\mathcal{Y},y'\in\mathcal{Y}} \inf_{z} \left\{\mathbb{P}(y,d|z) + \mathbb{P}(y',1-d|z)\right\} \tag{2.6}$$

The testing procedure suggested in Section 5 of the main text can be extended to a case with a multi-valued discrete outcome, i.e., $\mathcal{Y} = \{y_1, ..., y_K\}$. In this case, (2.4) is equivalent to

$$\sum_{k=1}^{K} \sup_{z} \mathbb{P}(Y = y_k, D = d|Z = z) \leqslant 1, \ d = 0, 1. \tag{2.7}$$

Assumption 1. *Assume that there exist vectors of random variable* $(Y^{(k)}, D^{(k)}, Z^{(k)})$ *for* $k \in \{1, ..., K\}$ *such that (i)* $(Y^{(k)}, D^{(k)}, Z^{(k)}) \perp (Y^{(l)}, D^{(l)}, Z^{(l)})$ *for* $k \neq l$ *with* $l \in \{1, ..., K\}$*, and (ii)* $(Y^{(k)}, D^{(k)}, Z^{(k)})$ *has the same distribution as* (Y, D, Z)*.*

Let us assume that there exists a sequence of vectors $(Y^{(k)}, D^{(k)}, Z^{(k)})$, $k = 1, ..., K$ that satisfies Assumption 1. Then, the latter inequality can be equivalently written as

$$\sup_{(z_1,...,z_K)} \mathbb{E}\Big[\sum_{k=1}^{K} 1\{Y^{(k)} = y_k, D^{(k)} = d\} | Z^{(1)} = z_1, ..., Z^{(K)} = z_K \Big] \leqslant 1, \ d = 0, 1. \tag{2.8}$$

By splitting the sample, the latter equation can be tested using the Chernozhukov, Lee and Rosen (2013, CLR) method. Notice that this formulation is data demanding since we need to split the sample in at least $K = card(\mathcal{Y})$ sub-samples.

For the last inequality, we will also require a sequence of vectors satisfying Assumption 1. However, we will need a longer sequence, so we denote this sequence by $(Y^{(k,l)}, D^{(k,l)}, Z^{(k,l)})$ $k, l = 1, ..., K$. Inequality (2.6) can be written as follows:

$$1 \leqslant \inf_{(z_1,...,z_K)} \mathbb{E}\Big[\sum_{k=1}^{K} \sum_{l=1}^{K} \Big\{ 1\{Y^{(k,l)} = y_k, D^{(k,l)} = d\} + 1\{Y^{(k,l)} = y_l, D^{(k,l)} = 1-d\} \Big\} | Z^{(1)} = z_1, ..., Z^{(K^2)} = z_{K^2} \Big] \tag{2.9}$$

In practice, this latter inequality requires the sample to be split into K^2 sub-samples. Since every sub-sample requires reasonable number of observations, this strategy would be data demanding. One way to deal with the large data demand is to consider a partition where its elements are not just singletons but have a bigger size. If the size of each element of the partition is n_g, the testing procedure will require only $(\frac{K}{n_g})^2$ sub-samples if K is a multiple of n_g or $([\frac{K}{n_g}] + 1)^2$ if not. However, by using fewer sub-samples, the testing procedure would lose power, since it would detect less observable violations, as discussed in Section 4 of the main text. More sample splitting could increase power but may in such a case increase the efficiency loss, whereas less sample splitting could decrease power but may in such a case reduce the efficiency loss. Therefore, it could exist a trade-off here. Basically, we do think that a formal discussion regarding the optimal way of splitting the sample is beyond the scope of this paper, and we leave this question for future research.

Remark 1. *Notice that when the outcome is continuous, we can apply the latter test using a suitable discretization of the outcome.*

3. Introducing Covariates in the inference procedure

In empirical studies, the IV independence assumptions are most often invoked conditionally on covariates. Researchers often introduce covariates to reduce the level of endogeneity of the treatment variable D. Indeed, introducing covariates allows us to control for potential observable confounders. Let X denote the vector of covariates and $\mathcal{X}$ its support. Depending on the identification question, one of the following IV independence assumptions is used:

Assumption 2 (Conditional IV independence assumptions)**.** *Let us assume the following IV independence assumptions:*

(2.1): $Y_1 \perp Z|X$ *and* $Y_0 \perp Z|X$.
(2.2): $(Y_1, Y_0) \perp Z|X$.
(2.3): $(Y_1, Y_0, D_z)_{z\in\mathcal{Z}} \perp Z|X$.

An interesting feature of the CLR method is the fact that one can easily incorporate additional covariates into the conditional moment inequalities as additional conditioning variables without additional work. Therefore, the conditional JSI assumption, i.e., $(Y_1, Y_0) \perp Z|X$ can be tested using the sample splitting procedure and the CKLR Stata package.

Assumption 3. *Assume that there exist vectors of random variable* $(Y^{(k)}, D^{(k)}, Z^{(k)}, X^{(k)})$ *for* $k \in \{1, ..., K\}$ *such that (i)* $(Y^{(k)}, D^{(k)}, Z^{(k)}, X^{(k)}) \perp (Y^{(l)}, D^{(l)}, Z^{(l)}, X^{(l)})$ *for* $k \neq l$ *with* $l \in \{1, ..., K\}$*, and (ii)* $(Y^{(k)}, D^{(k)}, Z^{(k)}, X^{(k)})$ *has the same distribution as* (Y, D, Z, X)*.*

For the sake of simplicity, we consider the leading binary outcome case. Let us consider the following inequality:

$$\sup_z \mathbb{E}[YD|Z = z, X = x] + \sup_z \mathbb{E}[-YD - (1 - D)|Z = z, X = x] \quad \leqslant \quad 0, \text{ for all } x \in \mathcal{X}. \quad (3.1)$$

Assume that $(Y^{(1)}, D^{(1)}, Z^{(1)}, X^{(1)})$ and $(Y^{(2)}, D^{(2)}, Z^{(2)}, X^{(2)})$ satisfy Assumption 1. We have

$$\begin{aligned}
&\sup_{z_1, z_2} \Big\{\mathbb{E}[Y^{(1)}D^{(1)} - Y^{(2)}D^{(2)} - (1 - D^{(2)})|Z^{(1)} = z_1, Z^{(2)} = z_2, X^{(1)} = x, X^{(2)} = x]\Big\} \\
&= \sup_{z_1, z_2} \Big\{\mathbb{E}[Y^{(1)}D^{(1)}|Z^{(1)} = z_1, Z^{(2)} = z_2, X^{(1)} = x, X^{(2)} = x] \\
&\qquad\qquad + \mathbb{E}[-Y^{(2)}D^{(2)} - (1 - D^{(2)})|Z^{(1)} = z_1, Z^{(2)} = z_2, X^{(1)} = x, X^{(2)} = x]\Big\} \\
&= \sup_{z_1, z_2} \Big\{\mathbb{E}[Y^{(1)}D^{(1)}|Z^{(1)} = z_1, X^{(1)} = x] \\
&\qquad\qquad + \mathbb{E}[-Y^{(2)}D^{(2)} - (1 - D^{(2)})|Z^{(2)} = z_2, X^{(2)} = x]\Big\} \\
&= \sup_{z_1} \Big\{\mathbb{E}[Y^{(1)}D^{(1)}|Z^{(1)} = z_1, X^{(1)} = x] \\
&\qquad\qquad + \sup_{z_2} \mathbb{E}[-Y^{(2)}D^{(2)} - (1 - D^{(2)})|Z^{(2)} = z_2, X^{(2)} = x]\Big\} \\
&= \sup_z \mathbb{E}[YD|Z = z, X = x] + \sup_z \mathbb{E}[-YD - (1 - D)|Z = z, X = x].
\end{aligned}$$

The second equality holds under Assumption 3 (i), and the last holds under Assumption 3 (ii). Therefore, if Assumption 3 holds, inequality (3.1) can be rewritten as follows:

$$\sup_{z_1, z_2} \Big\{\mathbb{E}[Y^{(1)}D^{(1)} - Y^{(2)}D^{(2)} - (1 - D^{(2)})|Z^{(1)} = z_1, Z^{(2)} = z_2, X^{(1)} = x, X^{(2)} = x]\Big\} \leqslant 0.$$

The rest of the inequalities can be rewritten in a similar way. To summarize, testing the *sharp instrumental inequalities* is equivalent to testing the following inequalities:

$$\sup_{(z_1,z_2)} \mathbb{E}[Y^{(1)}D^{(1)} - Y^{(2)}D^{(2)} - (1-D^{(2)})|Z^{(1)} = z_1, Z^{(2)} = z_2, X^{(1)} = x, X^{(2)} = x] \leqslant 0, \tag{3.2}$$

$$\sup_{(z_1,z_2)} \mathbb{E}[Y^{(1)}(1-D^{(1)}) - Y^{(2)}(1-D^{(2)}) - D^{(2)}|Z^{(1)} = z_1, Z^{(2)} = z_2, X^{(1)} = x, X^{(2)} = x] \leqslant 0, \tag{3.3}$$

$$\begin{aligned}\sup_{(z_1,z_2,z_3)} \; & \mathbb{E}[Y^{(1)}(1-D^{(1)}) - Y^{(2)} - Y^{(3)}(1-D^{(3)}) \\ & -(1-Y^{(3)})D^{(3)}|Z^{(1)} = z_1, Z^{(2)} = z_2, Z^{(3)} = z_3, X^{(1)} = x, X^{(2)} = x, X^{(3)} = x] \leqslant 0,\end{aligned} \tag{3.4}$$

$$\begin{aligned}\sup_{(z_1,z_2,z_3)} \; & \mathbb{E}[(1-Y^{(1)})(1-D^{(1)}) - (1-Y^{(2)}) - (1-Y^{(3)})(1-D^{(3)}) \\ & -Y^{(3)}D^{(3)}|Z^{(1)} = z_1, Z^{(2)} = z_2, Z^{(3)} = z_3, X^{(1)} = x, X^{(2)} = x, X^{(3)} = x] \leqslant 0,\end{aligned} \tag{3.5}$$

$$\begin{aligned}\sup_{(z_1,z_2,z_3)} \; & \mathbb{E}[Y^{(1)}D^{(1)} - (1-Y^{(2)}) - (1-Y^{(3)})(1-D^{(3)}) \\ & -Y^{(3)}D^{(3)}|Z^{(1)} = z_1, Z^{(2)} = z_2, Z^{(3)} = z_3, X^{(1)} = x, X^{(2)} = x, X^{(3)} = x] \leqslant 0,\end{aligned} \tag{3.6}$$

$$\begin{aligned}\sup_{(z_1,z_2,z_3)} \; & \mathbb{E}[(1-Y^{(1)})D^{(1)} - (1-Y^{(2)}) - Y^{(3)}(1-D^{(3)}) \\ & -(1-Y^{(3)})D^{(3)}|Z^{(1)} = z_1, Z^{(2)} = z_2, Z^{(3)} = z_3, X^{(1)} = x, X^{(2)} = x, X^{(3)} = x] \leqslant 0,\end{aligned} \tag{3.7}$$

$$\begin{aligned}\sup_{(z_1,z_2,z_3,z_4)} \; & \mathbb{E}[Y^{(1)} - Y^{(2)}(1-D^{(2)}) - (1-Y^{(2)})D^{(2)} - Y^{(3)} - (1-Y^{(4)})(1-D^{(4)}) - Y^{(4)}D^{(4)} \\ & |Z^{(1)} = z_1, Z^{(2)} = z_2, Z^{(3)} = z_3, Z^{(4)} = z_4, X^{(1)} = x, X^{(2)} = x, X^{(3)} = x, X^{(4)} = x] \leqslant 0.\end{aligned} \tag{3.8}$$

4. Using CKLR'S Stata package to test the JSI assumption.

In the case we would like to test each of the Pearl instrumental inequality separately we do not need to split the sample. In such a context, we only need to directly use the clr2bound command as exemplified below:

Step 0: CKLR-package installation.

We need to install the CLKR package using the following commands in Stata. So, first type

```
ssc install clrbound
```

then type

```
ssc install moremata, replace
```

Step 1: Using non-parametric clrbound command.

```
centile(Ztuit17), centile(1.5 98.5)
scalar UB = r(c2)
scalar LB = r(c1)
gen Rtuit17 = LB + 0.1 * (n - 1)
replace Rtuit17 = . if Rtuit17 > UB
```

```
clr2bound ((ldepen1 Ztuit17 Rtuit17)) ((udepen1 Ztuit17 Rtuit17)), met("local") level(0.5
0.9 0.95 0.99) norseed rnd(20000)
```

Now, we will present how to use the CKLR's Stata command with the sample splitting, but before we will briefly visit and discuss the conditions under which the CLR approach can be used in this context.

4.1. **Conditions for validity of the CLR inferential procedure.** We will focus mainly on the parametric approach proposed by CLR because of the relative high number of instruments and covariates we may have. As we explained earlier, our use of the CLR approach requires to split the sample and construct a new set of auxiliary variables. First, for the sake of simplicity, let us write the auxiliary dependent variables as follows:

$dep1 \equiv Y^{(1)} * D^{(1)} - (Y^{(2)} * D^{(2)} + (1 - D^{(2)}))$

$dep2 \equiv Y^{(1)} * (1 - D^{(1)}) - (Y^{(2)} * (1 - D^{(2)}) + D^{(2)})$

$dep3 \equiv Y^{(1)} * (1 - D^{(1)}) - (Y^{(3)} + Y^{(4)} * (1 - D^{(4)}) + (1 - Y^{(4)}) * D^{(4)})$

$dep4 \equiv (1 - Y^{(1)}) * (1 - D^{(1)}) - ((1 - Y^{(3)}) + (1 - Y^{(4)}) * (1 - D^{(4)}) + Y^{(4)} * D^{(4)})$

$dep5 \equiv Y^{(1)} * D^{(1)} - (Y^{(3)} + (1 - Y^{(4)}) * (1 - D^{(4)}) + Y^{(4)} * D^{(4)})$

$dep6 \equiv (1 - Y^{(1)}) * D^{(1)} - ((1 - Y^{(3)}) + Y^{(4)} * (1 - D^{(4)}) + (1 - Y^{(4)}) * D^{(4)})$

$dep7 \equiv Y^{(1)} - Y^{(2)} - (Y^{(3)} * (1 - D^{(3)}) + (1 - Y^{(3)}) * D^{(3)} + (1 - Y^{(4)}) * (1 - D^{(4)}) + Y^{(4)} * D^{(4)})$

Second, from each single instrument Z and single covariate X we construct four auxiliary instruments and covariates by splitting the sample. Let us refer to them as $Z_k \equiv Z^{(k)}$ and $X_k \equiv X^{(k)}$ for $k = 1, ..., 4$. To fix ideas, notice that the moment inequality (3.2) is conditioned on two of those auxiliary instruments while (3.8) is conditioned on the four auxiliary instruments. To be clearer let's denote by $\tilde{Z}_j$ the vector of all conditioning variables present in the moment inequality j, where $j \in \mathcal{J} \equiv \{1, ..., 7\}$, i.e. $\tilde{Z}_1 = [Z_1, Z_2, X_1, X_2]$ and $\tilde{Z}_7 \equiv [Z_1, Z_2, Z_3, Z_4, X_1, X_2, X_3, X_4]'$; let $\tilde{\mathcal{Z}}_j$ the support of $\tilde{Z}_j$, and $\tilde{\mathcal{Z}}_j^x$ denotes the joint support of the instruments used in the moment inequality j when the vector of covariates is fixed at a value $x \in \mathcal{X}$. For instance, $\mathcal{Z}_1^x = \mathcal{Z}_1 \times \mathcal{Z}_2 \times \{x\} \times \{x\}$, where $\mathcal{Z}_k$ denotes the support of Z_k. Let $v = (\tilde{z}, j)$, suppose $\mathcal{V}_j^x \subseteq \mathcal{Z}_j^x$ Therefore, our conditional sharp moment inequalities could be written as follows:

$$\min_{v \in \mathcal{V}^x} \theta(v) \equiv \min_{v \in \mathcal{V}^x} \theta(\tilde{z}, j) \equiv \min_{j \in \mathcal{J}} \min_{z_j^x \in \mathcal{V}_j^x} \mathbb{E}\left[-depj | \tilde{Z}_j = \tilde{z}\right] \geqslant 0. \tag{4.1}$$

The use of the CLR parametric approach consist in assuming that $\mathbb{E}\left[-depj | \tilde{Z}_j = \tilde{z}\right] = b_j(\tilde{z})'\beta(j)$ where $b_j : \tilde{\mathcal{Z}}_j \mapsto \mathbb{R}^m$ denotes some transformation of $\tilde{z}$, but with m independent of n (where n is considered here to be the sample size of the resulting sub-sample after the splitting) and where $\beta(j)$ is the vector of regression coefficients of $-depj$ on $b_j(\tilde{z})$. The CKLR's Stata package allow to use different transformation of $\tilde{Z}_j$, we therefore strongly recommend the empirical researchers to try different choice for $b_j(\tilde{z})$ as robustness checks. For example, in the empirical application, we considered different parametrization such that (i) a linear specification $b_1(\tilde{z}) = [1, z_1, z_2, x_1, x_2]$, (ii) a second order polynomial parametrization for the instruments $b_1(\tilde{z}) = [1, z_1, z_2, z_1^2, z_2^2, x_1, x_2]$, or (iii) $b_1(\tilde{z}) = [1, z_1, z_2, z_1^2, z_2^2, x_1, x_2, x_1^2, x_2^2]$. This parametric model we present here is a special case of the section 4 of CLR and especially the Example 3. The CLR's estimation and inference approach in this setting is valid under certain conditions enumerated in page 696-697 of CLR; conditions that we found worthwhile to recall here and discuss their interpretations in our context. Some conditions should be verified by the applied researcher while others are usual regularity conditions.

(s) The support of each instrument is compact.

(a) $b_j(\tilde{z})$ includes constant 1;
This condition has been already taken into in the CKLR package.

(b) $\tilde{z} \mapsto b_j(\tilde{z})$ has Lipschitz coefficient bounded above by L, where L is a finite constant.
Since from (a), we consider only instrument with bounded support, this condition is verified as soon as $b_j(.)$ is continuously differentiable.

(c) For $\tilde{Z}_i = (\tilde{Z}_{ji}, j \in \mathcal{J})'$, let $\epsilon_i = [-dep1_i \ldots - dep7_i]' - \mathbb{E}\Big[[-dep1_i \ldots - dep7_i]'\Big|\tilde{Z}_i\Big]$ be the vector of errors, then the eigenvalues of $\mathbb{E}_n[\epsilon_i\epsilon_i'|\tilde{Z}_i = \tilde{z}]$ are bounded away from zero and from above, uniformly in $\tilde{z} \in \tilde{\mathcal{Z}}$ and n, where $\mathbb{E}_n$ denotes the expectation with respect to the empirical distribution of the sub-sample.

(d) $Q = \mathbb{E}_n[b_j(\tilde{Z}_{ji})b_j(\tilde{Z}_{ji})']$ has eigenvalues bounded away from zero and from above, uniformly in n. This condition generalized the absence of multicolinearity that is often assumed in linear regression models;

(e) $\mathbb{E}_n \left\|b_j(\tilde{Z}_{ji})\right\|^4$ and $\mathbb{E}_n \left\|\epsilon_i\right\|^4$ are bounded from above uniformly in n. Notice that if $b_j(.)$ is a continuous function and the outcome, the instruments and the covariates have all compact supports this condition readily holds.

As explained in CLR Example 3 on page 696-697 the above conditions are sufficient to ensure the validity of the CLR parametric approach. The properties of the CLR estimation and inference methods in this context are summarized in their Theorem 4.

4.2. **CKLR's implementation.** In this subsection, we explain how the CKLR Stata package can be used to test the IV independence assumption and show some examples of the command used. We will consider two examples for illustration: (i) binary outcome, single continuous instrument with no covariates, i.e., $(Y_1, Y_0) \perp Z$, (ii) binary outcome, single continuous instrument with one covariate, i.e., $(Y_1, Y_0) \perp Z_k|X = x$.

Step 0: CKLR-package installation.

We need to install the CLKR package using the following commands in Stata. So, first type

```
ssc install clrbound
```

then type

```
ssc install moremata, replace
```

Step 1: Sample splitting. Assuming that you have an independent and identically distributed (i.i.d.) sample $\{(Y_i, D_i, Z_i, X_i)\}_{i=1}^n$. Split the sample into four different sub-samples and construct data matrices representing the sub-samples, i.e., $(Y^{(k)}, D^{(k)}, Z^{(k)}, X^{(k)})$, $k = 1, 2, 3, 4$.[1]

Step 2: Construction of new auxiliary variables. We generate 7 new dependent variables as follows:

gen $dep1 = Y^{(1)} * D^{(1)} - (Y^{(2)} * D^{(2)} + (1 - D^{(2)}))$
gen $dep2 = Y^{(1)} * (1 - D^{(1)}) - (Y^{(2)} * (1 - D^{(2)}) + D^{(2)})$
gen $dep3 = Y^{(1)} * (1 - D^{(1)}) - (Y^{(3)} + Y^{(4)} * (1 - D^{(4)}) + (1 - Y^{(4)}) * D^{(4)})$
gen $dep4 = (1 - Y^{(1)}) * (1 - D^{(1)}) - ((1 - Y^{(3)}) + (1 - Y^{(4)}) * (1 - D^{(4)}) + Y^{(4)} * D^{(4)})$
gen $dep5 = Y^{(1)} * D^{(1)} - (Y^{(3)} + (1 - Y^{(4)}) * (1 - D^{(4)}) + Y^{(4)} * D^{(4)})$
gen $dep6 = (1 - Y^{(1)}) * D^{(1)} - ((1 - Y^{(3)}) + Y^{(4)} * (1 - D^{(4)}) + (1 - Y^{(4)}) * D^{(4)})$
gen $dep7 = Y^{(1)} - Y^{(2)} - (Y^{(3)} * (1 - D^{(3)}) + (1 - Y^{(3)}) * D^{(3)} + (1 - Y^{(4)}) * (1 - D^{(4)}) + Y^{(4)} * D^{(4)})$

These are the dependent variables that are the arguments of the 7 expectations characterizing the sharp instrumental inequalities, i.e. eq (3.2)-(3.8).

Step 3: Specifying a *grid* of values for the instruments.

In the sharp instrumental inequalities, the supremums are taken over the support of Z, i.e., $\mathcal{Z}$. The CKLR pakage allows the researcher to define this support. When the cardinality of $\mathcal{Z}$ is finite, we can generate a grid to enumerate all the elements of $\mathcal{Z}$. For a continuous Z, we should discretize the support and generate only a finite number of points. This grid will be represented by the m-dimensional vector *Rinst* whose elements are evenly-spaced points in the interval $[Z_{\min}, Z_{\max}]$, and where m represents the number of points that comprise the grid. In this example, we analyze the tuition fees which is a continuous instrument with values varying between 0 and 18.

We therefore generate the range of the instrument Z (tuition at age 17), i.e., Rtuit17 as follows:

```
set more off
gen Rtuit171=.
gen Rtuit172=.
gen Rtuit173=.
gen Rtuit174=.
forvalues i=1/6
forvalues j=1/6
forvalues k=1/6
```

[1]The reader can implement his own Stata code to split the sample. However, if interested the reader can look over our code available online to see how we proceed.

```
forvalues l=1/6
replace Rtuit171=`i'*0.1+17.7 if n==(`i'-1)*6*6*6+(`j'-1)*6*6+(`k'-1)*6+`l'
replace Rtuit172=`j'*0.1+17.7 if n==(`i'-1)*6*6*6+(`j'-1)*6*6+(`k'-1)*6+`l'
replace Rtuit173=`k'*0.1+17.7 if n==(`i'-1)*6*6*6+(`j'-1)*6*6+(`k'-1)*6+`l'
replace Rtuit174=`l'*0.1+17.7 if n==(`i'-1)*6*6*6+(`j'-1)*6*6+(`k'-1)*6+`l'
```

In presence of covariate, we can test the conditional JSI at a particular fixed value $X = x^*$. For example, let us consider the experience as the covariate and $x^* = 16$. In such a case, we need to set the range of X in the code at the value $x^* = 16$. And this can be done as follows:

```
gen RX1=.
replace RX1=16 if n<=216
```

Step 4: Using clrbound command. We just use the clrbound command as follows:

Case 1: Without controls.

```
clrbound (ldepen1 Z1 Z2 Rtuit171 Rtuit172) (ldepen2 Z1 Z2 Rtuit171 Rtuit172) (ldepen3
Z1 Z3 Z4 Rtuit171 Rtuit172 Rtuit173) (ldepen4 Z1 Z3 Z4 Rtuit171 Rtuit172 Rtuit173) (ldepen5
Z1 Z3 Z4 Rtuit171 Rtuit172 Rtuit173) (ldepen6 Z1 Z3 Z4 Rtuit171 Rtuit172 Rtuit173) (ldepen7
Z1 Z2 Z3 Z4 Rtuit171 Rtuit172 Rtuit173 Rtuit174), low met("parametric") level(0.5 0.9
0.95 0.99) norseed rnd(20000)
```

Case 2: With one covariate.

```
clrbound (ldepen1 Z1 Z2 X1 X2 Rtuit171 Rtuit172 RX RX) (ldepen2 Z1 Z2 X1 X2 Rtuit171
Rtuit172 RX RX) (ldepen3 Z1 Z3 Z4 X1 X3 X4 Rtuit171 Rtuit172 Rtuit173 RX RX RX) (ldepen4
Z1 Z3 Z4 X1 X3 X4 Rtuit171 Rtuit172 Rtuit173 RX RX RX) (ldepen5 Z1 Z3 Z4 X1 X3 X4 Rtuit171
Rtuit172 Rtuit173 RX RX RX) (ldepen6 Z1 Z3 Z4 X1 X3 X4 Rtuit171 Rtuit172 Rtuit173 RX
RX RX) (ldepen7 Z1 Z2 Z3 Z4 X1 X2 X3 X4 Rtuit171 Rtuit172 Rtuit173 Rtuit174 RX RX RX
RX), low met("parametric") level(0.5 0.9 0.95 0.99) norseed rnd(20000)
```

The detailed code to replicate our results is available online.[2]

5. Simulation

In this section, we consider three different DGPs. In all three designs, the instrument Z is continuous and uniformly distributed. The outcome Y is binary. In the first design, DGP1 belongs to the "strict interior" of the null hypothesis, meaning that the lower bounds are strictly smaller than their respective upper bounds. In the second design, the Pearl instrumental inequalities are nearly binding, i.e., $\int_{\{0,1\}} \sup_z f_{Y,D}(y,d|z)d\mu(y) \approx 1, d \in \{0,1\}$, but still in the interior, i.e., $1 - \int_{\{0,1\}} \sup_z f_{Y,D}(y,d|z)d\mu(y) \approx 0.01$, while the other inequalities are in the strict interior. Finally,

[2]Please visit https://sites.google.com/site/ismaelymourifie/research-papers or email us.

in the last design, the null hypothesis is rejected. For all inequalities, the lower bounds are strictly bigger than their respective upper bounds, and therefore the JSI assumption is rejected. Let $\alpha \in (0,1)$ be a pre-specified significance level; we then can expect that the rejection frequencies in those three DGPs will be close to 0, α, and 1, respectively.

DGP1:

$$\begin{aligned} Y &= 1\{D - U > 0\} \\ D &= 1\{V > Z\}, \end{aligned}$$

where $Z \sim$ unif(0,1) and $\begin{pmatrix} U \\ V \end{pmatrix} \sim N(0, \Sigma)$ with $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ s.t. $Z \perp (U, V)$.

DGP2:

$$\begin{aligned} Y &= 1\{D - U > 0\} \\ D &= 1\{V > Z\}, \end{aligned}$$

where $Z \sim$ unif(-2.3,2.3) and $\begin{pmatrix} U \\ V \end{pmatrix} \sim N(0, \Sigma)$ with $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ s.t. $Z \perp (U, V)$.

DGP3:

$$\begin{aligned} Y &= 1\{D - U > 0\} \\ D &= 1\{V > 3Z\}, \end{aligned}$$

where $Z \sim$ unif(-3,3) and $V \sim N(0,1)$ and $U = Z + V$.

To test the null hypothesis (i.e., H_0: JSI holds), different commands specified in CKLR can be used. The "clrtest" and "clrbound" are the two obvious command to use when aiming to test the hypothesis that the maximum of the intersection bounds is nonpositive as we presented in Section 5 of the main text. In the following, we will use the "clrbound" command.

Table 1 lists the simulation results of our procedure based on the "clrbound" command under different choices of sample size and DGPs. We investigate the rejection frequency using the "parametric regression" option. For detailed descriptions of the Stata package, please see CKLR. All results are computed based on 1000 replications. For DGP3 where the JSI fails to hold, the null hypothesis is rejected with high probability, even when the sample size is small, for example, when $n = 200$. For DGP2, when considered very close to the least favorable null, the rejection rate gets closer to the target levels when the size increases. It is not surprising to see that the test does not reject DGP1 since it is in the "strict interior" of H_0.

References

Bonet, B. (2001): "Instrumentality Tests Revisited," *Proc. 17th Conf. on Uncertainty in Artificial Intelligence (UAI). Seattle, WA. Morgan Kaufmann.*, pp. 48–55.

Table 1. Rejection Frequency (clrbound)

	Parametric		
Sig. level	10%	5%	1%
DGP1			
$n = 200$	0.1%	0%	0%
$n = 400$	0%	0%	0%
$n = 800$	0%	0%	0%
DGP2			
$n = 200$	4.7%	3.2%	1.0%
$n = 400$	5.0%	2.0%	0.0%
$n = 800$	9.3%	4.2%	1.6%
DGP3			
$n = 200$	74.8%	63.6%	43.0%
$n = 400$	95.5%	92.2%	78.9%
$n = 800$	100.0%	99.9%	98.4%

Based on $1,000$ replications.

Chernozhukov, V., W. Kim, S. Lee, and A. M. Rosen (2015): "Implementing Intersection Bounds in Stata," *Stata Journal*, 15(1), 21–44.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

Heckman, J. J., J. L. Tobias, and E. J. Vytlacil (2001): "Four Parameters of Interest in the Evaluation of Social Programs," *Southern Economic Journal*, 68(2), 210–23.

Pearl, J. (1994): "On the Testability of Causal Models with Latent and Instrumental Variables," *Uncertainty in Articial Intelligence*, 11, 435–443.