NONPARAMETRIC BOUNDS ON TREATMENT EFFECTS WITH IMPERFECT INSTRUMENTS

KYUNGHOON BAN AND DÉSIRÉ KÉDAGNI

Iowa State University - Department of Economics

Abstract. This paper extends the identification results in Nevo and Rosen (2012) to nonparametric models. We derive nonparametric bounds on the average treatment effect when an imperfect instrument is available. As in Nevo and Rosen (2012), we assume that the correlation between the imperfect instrument and the unobserved latent variables has the same sign as the correlation between the endogenous variable and the latent variables. We show that the monotone treatment selection and monotone instrumental variable restrictions, introduced by Manski and Pepper (2000, 2009), jointly imply this assumption. We introduce the concept of *comonotone instrumental variable*, which also satisfies this assumption. Moreover, we show how the assumption that the imperfect instrument is less endogenous than the treatment variable can help tighten the bounds. We also use the monotone treatment response assumption to get tighter bounds. The identified set can be written in the form of intersection bounds, which is more conducive to inference. We illustrate our methodology using the National Longitudinal Survey of Young Men data to estimate returns to schooling.

Keywords: Imperfect instrumental variables, comonotone IV, nonparametric bounds, average treatment effect, monotone treatment response.

JEL subject classification: C14, C21, C25, C26.

Date: The present version is as of December 17, 2020. We thank Santiago Acerenza, Otavio Bartalotti, Helle Bunzel, Vitor Possebom, and participants at the Iowa State econometrics workshop for helpful comments. All errors are ours. Corresponding address: 260 Heady Hall, 518 Farm House Lane, Ames, IA, 50011, USA. Email address: dkedagni@iastate.edu.

1. Introduction

The use of an instrumental variable (IV) is a popular solution to deal with endogeneity in social sciences. However, this approach may yield misleading conclusions when the instrument is invalid. A valid instrument must be uncorrelated with the unobservables in the model.[1] This requirement is often difficult to justify, and may call into question empirical findings. For this reason, Nevo and Rosen (2012) derived bounds on the parameters of interest (e.g., the average treatment effect) in parametric models under weaker conditions. They first assume that the sign of correlation between the imperfect IV (IIV)[2] and the unobserved latent variables is the same as that of the correlation between the endogenous variable and the latent variables. Second, they add the assumption that the correlation between the IIV and the latent variables is less than the correlation between the endogenous variable and the latent variables to tighten the bounds on the parameters of interest.

In this paper, we derive nonparametric bounds on the average treatment effect with an imperfect IV under the above assumptions when the outcome variable has bounded support. We introduce the concept of binarized MTS-MIV, which is implied by the monotone treatment selection (MTS) and monotone IV (MIV) assumptions developed by Manski and Pepper (2000, 2009). We show that the correlation between a binarized MTS-MIV and the unobserved latent variables has the same sign as the correlation between the endogenous variable and the latent variables. Hence, we link the Nevo and Rosen (2012) same direction of correlation assumption to the Manski and Pepper (2000) monotone treatment selection and monotone IV assumptions. We also introduce the concept of comonotone instrumental variable, which satisfies Nevo and Rosen's (2012) assumption. We believe these results are new in the literature. Furthermore, we show how additional restrictions such as the less endogenous instrument, the monotone treatment response, and the Roy selection can help tighten the bounds. As in Nevo and Rosen (2012), the bounds take the form of intersection bounds and can be implemented using the inferential methods developed by Chernozhukov, Lee, and Rosen (2013) or Andrews and Shi (2013). We illustrate our methodology using the National Longitudinal Survey of Young Men (NLSYM) data to estimate returns to schooling.

There is an increasing interest in the identification of causal effects with imperfect instrumental variables. Recently, Masten and Poirier (2020) developed a methodology that allows

[1]A stronger version of this condition is that the instrument is statistically (or mean) independent of the unobservables.

[2]An instrument that is potentially correlated with the unobservables.

researchers to consider continuous relaxations of the IV models when they are refuted by the data. Their approach is data driven as it exploits the extent of falsification of the model to construct the identified set for the parameter of interest. Although their method helps save some invalid IVs, their identifying assumptions may seem difficult to interpret. Our approach, as well as that of Nevo and Rosen (2012) and Manski and Pepper (2000, 2009), is not data driven and has clearer interpretations of the identifying assumptions. This paper also relates to the work of Kédagni, Li, and Mourifié (2018), who consider weaker version of the mean independence assumption for the IV model. They derived nonparametric bounds on the average treatment effect under unconditional moment restrictions for the IV. Several other papers have also studied identification of model parameters when the IV is invalid using a framework different from ours, see Hotz, Mullin, and Sanders (1997), Conley, Hansen, and Rossi (2012), among others.

The remainder of the paper is organized as follows. Section 2 presents the model, the assumptions and their link with the literature. In Section 3, we derive our main identification results. We discuss inference and implementation in Section 4. Section 5 presents an empirical illustration of our proposed methodology, while Section 6 concludes. Proofs and additional results are relegated to the appendix.

2. Analytical Framework

Consider the following potential outcome model (POM)

$$Y = \sum_{d=1}^{T} Y_d \mathbb{1}\{D = d\}, \tag{2.1}$$

where Y is the outcome variable taking values in $\mathcal{Y} \subset \mathbb{R}$, D is a discrete endogenous treatment variable taking values in $\mathcal{D} = \{1, 2, \ldots, T\}$, Y_d is the potential outcome that would have been observed if the treatment D had externally been set to d. Let $Z \in \mathcal{Z} \subseteq \mathbb{R}^+$ be an imperfect IV in the sense that it may be correlated with the potential outcome Y_d. In what follows, we assume that the random variable Y_d is integrable, i.e., $\mathbb{E}[Y_d] < \infty$. The objects of interest in this paper are the potential outcome means $\theta_d \equiv \mathbb{E}[Y_d]$, for all $d \in \mathcal{D}$, and some treatment effects $ATE(d, d') \equiv \theta_d - \theta_{d'}$, for $d, d' \in \mathcal{D}$. We allow for heterogeneous treatment effects, so that $ATE(d, d')$ may vary across (d, d'). The methodology that we develop in this paper can also be used to identify other commonly used parameters of interest such as the average treatment effect on the treated $ATT(d, d') \equiv \mathbb{E}[Y_d - Y_{d'} | D = d]$, and the average treatment effect on the untreated $ATU(d, d') \equiv \mathbb{E}[Y_d - Y_{d'}|D = d']$. But, for the sake of clarity of the exposition, we focus our attention on the ATE.

We observe a random sample of the vector (Y, D, Z). For simplicity, we drop exogenous covariates from the analysis. For example, Y could be earnings, D years of schooling, and Z parental education. In this example, Y_d is the potential earnings for an individual with d years of schooling. We now state our main identifying assumptions:

Assumption 1 (Bounded support (BoS))**.** *$Supp(Y_d|D \neq d) = Supp(Y_d|D = d) = \left[\underline{y}_d, \overline{y}_d\right]$.* □

Assumption BoS states that the support of the counterfactual outcome is the same as that of the factual. It is standard and similar to the usual bounded outcome assumption considered in Manski (1990, 1994), and many other papers. Like in Kédagni, Li, and Mourifié (2018), it does not require the support of the potential outcome Y_d to be uniform across all treatment levels d.

Assumption 2 (Same direction of correlation (SDC))**.** *$Cov\,(Y_d, D)\,Cov\,(Y_d, Z) \geq 0$.* □

Assumption SDC is equivalent to Assumption 3 in Nevo and Rosen (2012). It states that the correlation between the imperfect instrument Z and the potential outcome Y_d has weakly the same sign as the correlation between the endogenous treatment D and the potential outcome. For example, it is documented that parental education is not a valid instrument, see Kédagni and Mourifié (2020), Mourifié, Henry, and Méango (2020), among many others. However, one could assume that parental education has the same sign of correlation with the potential earnings as does the individual's education. Note that if either the treatment D or the instrument Z is exogenous, this assumption holds. When Assumption BoS holds (the researcher can check this), Assumption SDC has a testable implication. Indeed, when BoS holds and the identified set for θ_d is empty, then SDC is rejected.

Assumption SDC is a weaker version of the concepts of monotone IV (MIV: $\mathbb{E}\,[Y_d|Z = z]$ is monotone in z) and monotone treatment selection (MTS: $\mathbb{E}\,[Y_d|D = \ell]$ is monotone in ℓ) developed by Manski and Pepper (2000, 2009). To show this result, we introduce the concept of binarized MTS-MIV that is intermediate between SDC and MTS-MIV. We use the following notation.

Notation 1. *Denote $g_d^+(j) = \mathbb{E}[Y_d|D \geq j]$, $g_d^-(j) = \mathbb{E}[Y_d|D < j]$, $h_d^+(z) = \mathbb{E}[Y_d|Z \geq z]$, $h_d^-(z) = \mathbb{E}[Y_d|Z < z]$. ρ_{UV} denotes the coefficient of correlation between two random variables U and V.* □

Definition 1. *The variable Z is a binarized MTS-MIV for D if for each $d \in \mathcal{D}$,*

$$\left(g_d^+(j) - g_d^-(j)\right)\left(h_d^+(z) - h_d^-(z)\right) \geq 0 \quad \textit{for all } j,\ z. \tag{2.2}$$

□

In words, we say that Z is a binarized MTS-MIV for D if all binarized treatments $\mathbb{1}\{D \geq j\}$ satisfy the MTS restriction, and all binarized instruments $\mathbb{1}\{Z \geq z\}$ satisfy the MIV restriction.

Remark 1. *If Z is a binarized MTS-MIV for D then the functions g_d^+ and g_d^- do not overlap, nor do the functions h_d^+ and h_d^- for all d. Moreover, if $g_d^+ \geq g_d^-$ for some d then $h_d^+ \geq h_d^-$, and vice versa.* □

Lemma 1 shows that MTS-MIV is a sufficient condition for binarized MTS-MIV, while Lemma 2 shows that binarized MTS-MIV is a sufficient condition for SDC.

Lemma 1. *MTS-MIV in the same direction for D and Z implies that Z is a binarized MTS-MIV for D.* □

Lemma 2. *If Z is a binarized MTS-MIV for D, then Assumption SDC holds.* □

Remark 2. *From Lemmas 1 and 2, we conclude that MTS-MIV in the same direction implies Assumption SDC. Moreover, when both the treatment D and the imperfect instrument Z are binary, MTS-MIV in the same direction, binarized MTS-MIV and SDC are equivalent.* □

Definition 2. *Let $(\Omega, \mathcal{F})$ be a measurable space. Two random variables X_1 and X_2 defined on Ω are said to be comonotonic if*

$$\left(X_1(\omega) - X_1(\omega')\right)\left(X_2(\omega) - X_2(\omega')\right) \geq 0 \quad \textit{for all } \ \omega, \omega' \in \Omega. \tag{2.3}$$

□

Definition 3. *The variable Z is said to be a comonotone instrumental variable (CoMIV) for the treatment D if Z and D are comonotonic.* □

To better understand the CoMIV concept, we provide some sufficient conditions for it in the following lemma.

Lemma 3. *The following results hold.*

(1) *If D is a deterministic increasing function of Z (or vice versa), then Z is a CoMIV for D.*
(2) *Suppose $D = h(Z, V)$, where h is increasing in both of its arguments, and V represents unobserved heterogeneity. If Z and V are comonotonic, then Z is a CoMIV for D.*

□

For example, when $D = 2Z + V$ and $Z = e^V$, Z is a CoMIV for D.

The following lemma provides another sufficient condition for the SDC assumption. It shows that the SDC assumption can be satisfied as a form of comonotonicity between the treatment D and the imperfect instrument Z with respect to the potential outcome Y_d.

Lemma 4. *If Z is a CoMIV for D, then Assumption SDC holds.* □

At this point, we have shown that the Manski and Pepper (2000, 2009) MTS-MIV assumption and the comonotonicity of Z and D are two sufficient conditions for Nevo and Rosen's (2012) SDC assumption. However, comonotonicity of Z and D does not imply MTS-MIV. The following example shows a case where Z and D are comonotonic, but fail to satisfy the MTS-MIV restrictions. In Appendix D, we show by Example 3 that MTS-MIV (or binarized MTS-MIV) does not imply CoMIV either. Furthermore, Example 5 shows that CoMIV does not imply binarized MTS-MIV. See the appendix for more examples.

Example 1. *Consider the following data generating process (DGP)*

$$\begin{cases} Y &= 2D + U \\ D &= 0 \cdot \mathbb{1}\left\{V \in [0,1]\right\} + 1 \cdot \mathbb{1}\left\{V \in (1, \tfrac{3}{2}]\right\} + 2 \cdot \mathbb{1}\left\{V \in (\tfrac{3}{2}, 5]\right\} \\ Z &= 2D \\ U &= 4V\mathbb{1}\left\{V \in [1,2]\right\} + V\mathbb{1}\left\{V \notin [1,2]\right\} \end{cases} \tag{2.4}$$

where $V \sim \mathcal{U}_{[0,5]}$. This example illustrates a case where MTS and MIV fail to hold, but binarized MTS-MIV holds, and D and Z are comonotonic, suggesting that MTS-MIV and CoMIV are two different sufficient conditions for the SDC assumption.

First, note that Z is a CoMIV for D by (1) of Lemma 3 since $Z = 2D$. However, it can be shown that the conditional expectation function $\mathbb{E}[Y_d|D = \ell]$ in the given DGP is not monotone in ℓ for each $d = 0, 1, 2$, implying that MTS fails. Likewise, the conditional expectation function $\mathbb{E}[Y_d|Z = z]$ is not monotone in z for each $d = 0, 1, 2$, which implies that MIV is violated. Figure 1 illustrates these facts for the case $d = 1$. On the other hand,

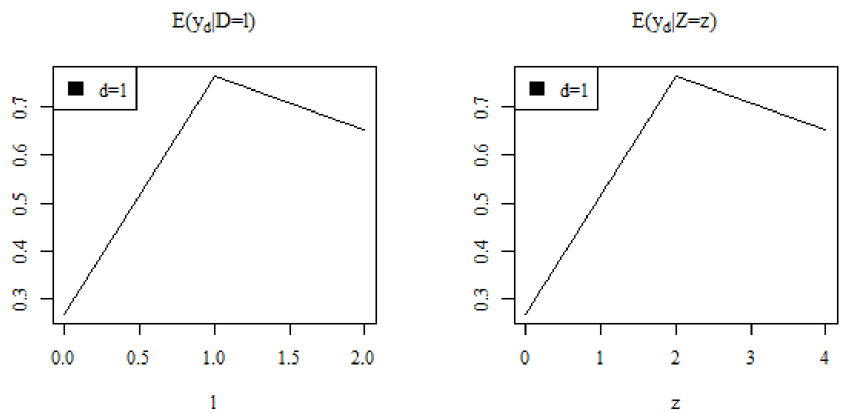

FIGURE 1. Numerical illustration of a violation of MTS and MIV

Figure 2 shows that the same DGP satisfies the binarized MTS-MIV restriction (2.2), and accordingly the SDC assumption by Lemma 1, as the function g_d^+ *is always greater than the function* g_d^-, *and the function* h_d^+ *is always greater than the function* h_d^- *for all* d. □

Assumption 3 (Less endogenous instrument (LEI))**.** $\mid \rho_{Y_d D} \mid \geq \mid \rho_{Y_d Z} \mid$. □

Assumption LEI is the same as Assumption 4 in Nevo and Rosen (2012). It states that the imperfect instrument Z is less correlated with the potential outcome than is the endogenous treatment D. In the context of our empirical example, it reasonable to assume that parental education is less correlated with the individual's potential wage than is the individual's own education.

Assumption 4 (Monotone treatment response (MTR))**.** $Y_d \geq Y_{d'}$ *for all* $d > d'$. □

Assumption MTR states that the potential outcome weakly increases with the level of the treatment. It was introduced by Manski (1997), and considered in Manski and Pepper (2000, 2009), among many others. For instance, in the returns to schooling example, it implies that the wage that a worker earns weakly increases as a function of the worker's years of schooling. We show how this assumption can help tighten the bounds derived under Assumptions BoS, SDC and LEI.

Assumption 5 (Roy Selection (RS))**.** $\{D = d\} \iff \{Y_d > Y_{d'}$ *for all* $d' \neq d\}$. □

Assumption RS states that agents choose the level of treatment that maximizes their potential outcome. This version of Roy selection implicitly assumes that agents have perfect

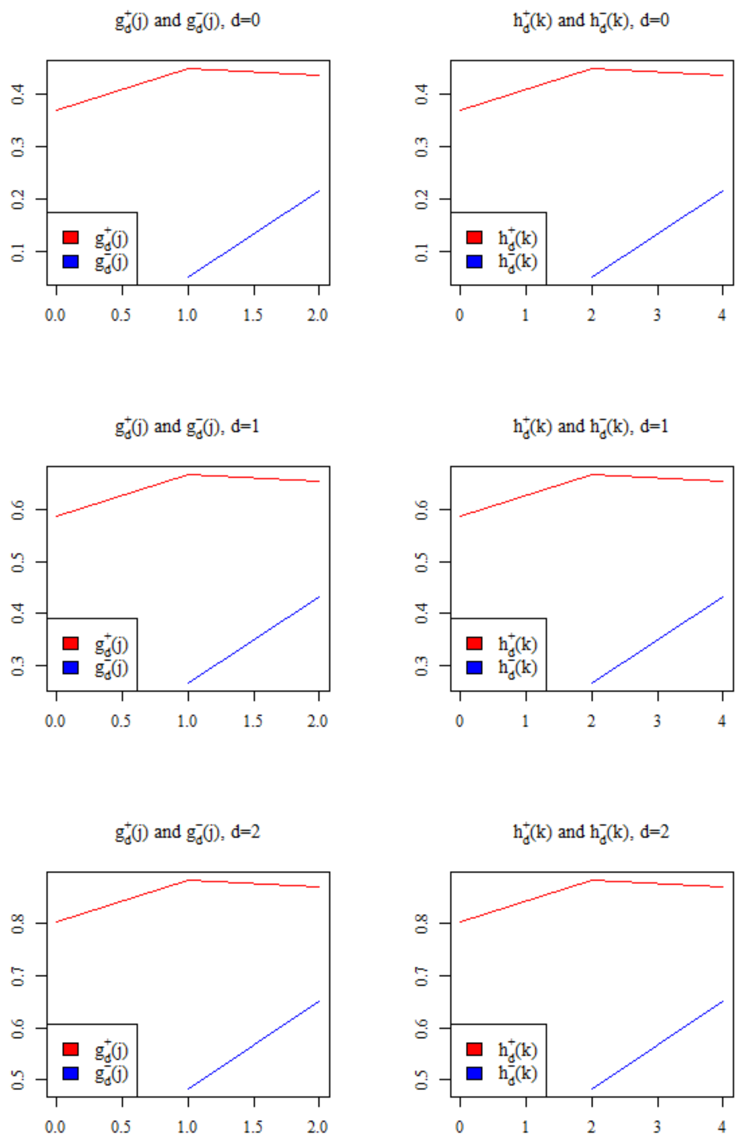

Figure 2. Numerical illustration of binarized MTS-MIV

foresight. This assumption is inspired from the seminal work of Roy (1951) and has been considered in many papers. Like the MTR assumption, Assumption RS also helps tighten the bounds derived under Assumptions BoS, SDC and LEI. Note that Assumption RS is not compatible with the MTS and MTR assumptions, while Assumption SDC is.

Now that we have discussed the model and our identifying assumptions, we are going to present our main identification results.

3. Identification results

3.1. **Identification under the same direction of correlation assumption.**

Assumption SDC is equivalent to $\mathbb{E}\left[Y_d\tilde{D}\right]\mathbb{E}\left[Y_d\tilde{Z}\right] \geq 0$, where $\tilde{D} \equiv D - \mathbb{E}[D]$ and $\tilde{Z} \equiv Z - \mathbb{E}[Z]$, which in turn is equivalent to: either

$$\mathbb{E}\left[Y_d\tilde{D}\right] \geq 0, \tag{3.1}$$

$$\mathbb{E}\left[Y_d\tilde{Z}\right] \geq 0, \tag{3.2}$$

or

$$\mathbb{E}\left[Y_d\tilde{D}\right] \leq 0, \tag{3.3}$$

$$\mathbb{E}\left[Y_d\tilde{Z}\right] \leq 0. \tag{3.4}$$

We first derive bounds on the potential outcome mean θ_d using inequality (3.1). Similarly, we can derive the bounds implied by inequalities (3.2), (3.3), and (3.4). Inequality (3.1) implies that, for any $\lambda \geq 0$, we have the following inequalities

$$\mathbb{E}\left[Y_d\lambda\tilde{D}\right] \geq 0 \text{ and } \mathbb{E}\left[-Y_d\lambda\tilde{D}\right] \leq 0,$$

which respectively imply

$$\mathbb{E}\left[Y_d\alpha\tilde{D}\right] \geq 0 \text{ and } \mathbb{E}\left[-Y_d\alpha\tilde{D}\right] \leq 0,$$

where $\alpha = \frac{\lambda}{1+\lambda} \in [0,1)$. The latter inequalities are respectively equivalent to:[3]

$$\mathbb{E}\left[Y_d\left(1+\alpha\tilde{D}\right)\right] \geq \mathbb{E}[Y_d] \equiv \theta_d \text{ and } \mathbb{E}\left[Y_d\left(1-\alpha\tilde{D}\right)\right] \leq \mathbb{E}[Y_d] \equiv \theta_d,$$

which we rewrite using the identity $\mathbb{1}\{D=d\} + \mathbb{1}\{D \neq d\} = 1$ as

$$\mathbb{E}\left[Y\left(1+\alpha\tilde{D}\right)\mathbb{1}\{D=d\} + Y_d\left(1+\alpha\tilde{D}\right)\mathbb{1}\{D\neq d\}\right] \geq \theta_d, \tag{3.5}$$

$$\mathbb{E}\left[Y\left(1-\alpha\tilde{D}\right)\mathbb{1}\{D=d\} + Y_d\left(1-\alpha\tilde{D}\right)\mathbb{1}\{D\neq d\}\right] \leq \theta_d, \tag{3.6}$$

[3] In the case where the ATT/ATU is our parameter of interest, we would bound $\theta_{d|d'} \equiv \mathbb{E}[Y_d|D=d'] = \frac{\mathbb{E}[Y_d\mathbb{1}\{D=d'\}]}{\mathbb{E}[\mathbb{1}\{D=d'\}]}$. In such a case, we will equivalently write these inequalities as:

$$\mathbb{E}\left[Y_d\left(\mathbb{1}\{D=d'\}+\alpha\tilde{D}\right)\right] \geq \mathbb{E}[Y_d\mathbb{1}\{D=d'\}] \text{ and } \mathbb{E}\left[Y_d\left(\mathbb{1}\{D=d'\}-\alpha\tilde{D}\right)\right] \leq \mathbb{E}[Y_d\mathbb{1}\{D=d'\}],$$

and use the same technique we develop in this paper.

respectively, given that $Y = Y_d$ when $D = d$. Now, using Assumption BoS, we can bound the counterfactuals $Y_d\left(1+\alpha\tilde{D}\right)\mathbb{1}\{D \neq d\}$ and $Y_d\left(1-\alpha\tilde{D}\right)\mathbb{1}\{D \neq d\}$ as follows:

$$\begin{aligned} Y_d\left(1+\alpha\tilde{D}\right)\mathbb{1}\{D \neq d\} &\leq \max\left\{\underline{y}_d\left(1+\alpha\tilde{D}\right), \overline{y}_d\left(1+\alpha\tilde{D}\right)\right\}\mathbb{1}\{D \neq d\} \\ Y_d\left(1-\alpha\tilde{D}\right)\mathbb{1}\{D \neq d\} &\geq \min\left\{\underline{y}_d\left(1-\alpha\tilde{D}\right), \overline{y}_d\left(1-\alpha\tilde{D}\right)\right\}\mathbb{1}\{D \neq d\}. \end{aligned}$$

Therefore, using inequalities (3.5) and (3.6), it follows that

$$\mathbb{E}\left[\overline{f}_d\left(Y, D, 1+\alpha\tilde{D}\right)\right] \geq \theta_d \text{ and } \mathbb{E}\left[\underline{f}_d\left(Y, D, 1-\alpha\tilde{D}\right)\right] \leq \theta_d$$

for any $\alpha \in [0, 1)$, where we define the function $\underline{f}_d$ and $\overline{f}_d$ as

$$\begin{aligned} \underline{f}_d(Y, D, \delta) &\equiv \mathbb{1}\{D = d\}\delta Y + \mathbb{1}\{D \neq d\}\min\left\{\delta\underline{y}_d, \delta\overline{y}_d\right\} \\ \overline{f}_d(Y, D, \delta) &\equiv \mathbb{1}\{D = d\}\delta Y + \mathbb{1}\{D \neq d\}\max\left\{\delta\underline{y}_d, \delta\overline{y}_d\right\}. \end{aligned}$$

We can then take the supremum and the infimum of the lower and upper bounds over α, respectively, to obtain the following bounds for θ_d:

$$I^d_{SDC1} \equiv \left[\sup_{\alpha\in[0,1)} \mathbb{E}\left[\underline{f}_d\left(Y, D, 1-\alpha\tilde{D}\right)\right], \inf_{\alpha\in[0,1)} \mathbb{E}\left[\overline{f}_d\left(Y, D, 1+\alpha\tilde{D}\right)\right]\right].$$

Similarly, using inequalities (3.2), (3.3), and (3.4), we derive the following bounds for θ_d:

$$\begin{aligned} I^d_{SDC2} &\equiv \left[\sup_{\alpha\in[0,1)} \mathbb{E}\left[\underline{f}_d\left(Y, D, 1-\alpha\tilde{Z}\right)\right], \inf_{\alpha\in[0,1)} \mathbb{E}\left[\overline{f}_d\left(Y, D, 1+\alpha\tilde{Z}\right)\right]\right], \\ I^d_{SDC3} &\equiv \left[\sup_{\alpha\in[0,1)} \mathbb{E}\left[\underline{f}_d\left(Y, D, 1+\alpha\tilde{D}\right)\right], \inf_{\alpha\in[0,1)} \mathbb{E}\left[\overline{f}_d\left(Y, D, 1-\alpha\tilde{D}\right)\right]\right], \\ I^d_{SDC4} &\equiv \left[\sup_{\alpha\in[0,1)} \mathbb{E}\left[\underline{f}_d\left(Y, D, 1+\alpha\tilde{Z}\right)\right], \inf_{\alpha\in[0,1)} \mathbb{E}\left[\overline{f}_d\left(Y, D, 1-\alpha\tilde{Z}\right)\right]\right], \end{aligned}$$

respectively. All these results are summarized in the following proposition.

Proposition 1. *Under Assumptions BoS and SDC, the identification region for the parameter θ_d is:*

$$I^d_{SDC} \equiv \left(I^d_{SDC1} \cap I^d_{SDC2}\right) \cup \left(I^d_{SDC3} \cap I^d_{SDC4}\right).$$

□

Proposition 1 provides two-sided bounds on the potential outcome means, and then on the average treatment effects, which mainly relies on the bounded outcome assumption. While Nevo and Rosen (2012) can provide two-sided bounds under some circumstances, they usually get one-sided bounds. We relax the parametric linear assumption at the expense of the bounded support assumption. In light of the following statement from the fourth paragraph of Section VI in Nevo and Rosen (2012) "*. . . However, with a nonparametric functional, it is doubtful that our assumptions on the correlations of endogenous regressors and imperfect instruments with econometric errors would prove anywhere near as fruitful*," we believe that the result of Proposition 1 makes a positive contribution to the literature.

Remark 3. *The bounds derived in Proposition 1 may not be sharp. For example, Equations (3.1) and (3.2) imply: for all* $(\lambda, \gamma) \in \mathbb{R}^2_+$*, we have*

$$\mathbb{E}\left[Y_d\left(\lambda\tilde{D}+\gamma\tilde{Z}\right)\right] \geq 0, \quad \textit{and} \quad \mathbb{E}\left[Y_d\left(-\lambda\tilde{D}-\gamma\tilde{Z}\right)\right] \leq 0.$$

Likewise, Equations (3.3) and (3.4) imply: for all $(\lambda, \gamma) \in \mathbb{R}^2_+$*, we have*

$$\mathbb{E}\left[Y_d\left(\lambda\tilde{D}+\gamma\tilde{Z}\right)\right] \leq 0, \quad \textit{and} \quad \mathbb{E}\left[Y_d\left(-\lambda\tilde{D}-\gamma\tilde{Z}\right)\right] \geq 0.$$

The bounds derived from these conditions could be tighter than the identification region in Proposition 1, but our simulation and empirical results suggest that they are generally wider. □

3.2. **Adding the less endogenous instrument assumption.** In this subsection, we combine Assumptions SDC and LEI in order to get tighter bounds on the parameter θ_d. Assumption LEI is equivalent to $|\frac{\mathbb{E}[Y_d\tilde{D}]}{\sigma_D}| \geq |\frac{\mathbb{E}[Y_d\tilde{Z}]}{\sigma_Z}|$. Hence, Assumptions LEI and SDC imply that either of the followings is always true:

$$\frac{\mathbb{E}\left[Y_d\tilde{D}\right]}{\sigma_D} \geq \frac{\mathbb{E}\left[Y_d\tilde{Z}\right]}{\sigma_Z} \geq 0 \;\; \text{or} \;\; \frac{\mathbb{E}\left[Y_d\tilde{D}\right]}{\sigma_D} \leq \frac{\mathbb{E}\left[Y_d\tilde{Z}\right]}{\sigma_Z} \leq 0.$$

Differently, we can rewrite these inequalities as either

$$\mathbb{E}\left[Y_d\left(\tilde{D}\sigma_Z-\tilde{Z}\sigma_D\right)\right] \geq 0 \qquad \text{and} \qquad \mathbb{E}\left[Y_d\tilde{Z}\right] \geq 0 \tag{3.7}$$

or

$$\mathbb{E}\left[Y_d\left(\tilde{D}\sigma_Z-\tilde{Z}\sigma_D\right)\right] \leq 0 \qquad \text{and} \qquad \mathbb{E}\left[Y_d\tilde{Z}\right] \leq 0. \tag{3.8}$$

We first note that each of the second inequalities from (3.7) and (3.8) is the same as (3.2) and (3.4), implying the bounds I^d_{SDC2} and I^d_{SDC4} for θ_d, respectively. Second, using a similar reasoning as in the previous subsection, and replacing $\tilde{Z}$ by $\left(\tilde{D}\sigma_Z-\tilde{Z}\sigma_D\right)$, we obtain the following bounds for θ_d from each of the first inequalities of (3.7) and (3.8), respectively:

$$\theta_d \in \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{f}_d\left(Y, D, 1-\alpha\left(\tilde{D}\sigma_Z - \tilde{Z}\sigma_D\right)\right)\Big], \right.$$
$$\left. \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{f}_d\left(Y, D, 1+\alpha\left(\tilde{D}\sigma_Z - \tilde{Z}\sigma_D\right)\right)\Big] \right] \equiv I^d_{LEI1}$$
$$\theta_d \in \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{f}_d\left(Y, D, 1+\alpha\left(\tilde{D}\sigma_Z - \tilde{Z}\sigma_D\right)\right)\Big], \right.$$
$$\left. \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{f}_d\left(Y, D, 1-\alpha\left(\tilde{D}\sigma_Z - \tilde{Z}\sigma_D\right)\right)\Big] \right] \equiv I^d_{LEI2}$$

Thus, the following proposition holds.

Proposition 2. *Under Assumptions BoS, SDC and LEI, the identification region for the parameter θ_d is:*

$$I^d_{LEI} \equiv \left(I^d_{LEI1} \cap I^d_{SDC2}\right) \cup \left(I^d_{LEI2} \cap I^d_{SDC4}\right).$$

□

Remark 4. *When the less endogenous instrument assumption LEI is redundant, the identified sets I^d_{SDC} and I^d_{LEI} will be identical. Otherwise, I^d_{LEI} will be tighter than I^d_{SDC}.* □

3.3. **Adding the monotone treatment response assumption.** In this subsection, we derive bounds on the potential outcome mean θ_d under Assumptions BoS, SDC, LEI, and MTR . Under Assumption MTR we have

$$\begin{aligned} Y_d \mathbb{1}\{D > d\} &= Y_d \sum_{j=d+1}^{T} \mathbb{1}\{D = j\} = \sum_{j=d+1}^{T} Y_d \mathbb{1}\{D = j\} \\ &\leq \sum_{j=d+1}^{T} Y \mathbb{1}\{D = j\} = Y \sum_{j=d+1}^{T} \mathbb{1}\{D = j\} = Y \mathbb{1}\{D > d\}, \end{aligned}$$

where the inequality holds as $Y_d \mathbb{1}\{D = j\} \leq Y_j \mathbb{1}\{D = j\} = Y \mathbb{1}\{D = j\}$ for all $j > d$ for each d. This result shows that under the MTR assumption, the counterfactual random variable $Y_d \mathbb{1}\{D > d\}$ is bounded from above by the observed variable $Y \mathbb{1}\{D > d\}$. As we can see, this assumption considerably shrinks the upper bound on $Y_d \mathbb{1}\{D > d\}$, which would be $\overline{y}_d \mathbb{1}\{D > d\}$ otherwise under Assumption BoS. Similarly, we also have

$$Y_d \mathbb{1}\{D < d\} \quad \geq \quad Y \mathbb{1}\{D < d\}$$

for each d. Without the MTR assumption, the lower bound on $Y_d \mathbb{1}\{D < d\}$ would be $\underline{y}_d \mathbb{1}\{D < d\}$ under Assumption BoS. Combining these results together, the following inequalities hold under Assumptions BoS and MTR:

$$\begin{array}{rcccl} \underline{y}_d \mathbb{1}\{D > d\} & \leq & Y_d \mathbb{1}\{D > d\} & \leq & Y \mathbb{1}\{D > d\}, \\ Y \mathbb{1}\{D < d\} & \leq & Y_d \mathbb{1}\{D < d\} & \leq & \overline{y}_d \mathbb{1}\{D < d\}. \end{array}$$

Thus, for any $\delta \in \mathbb{R}$, we have the following bounds

$$\begin{array}{rcccl} \min\{\delta \underline{y}_d, \delta Y\} \mathbb{1}\{D > d\} & \leq & \delta Y_d \mathbb{1}\{D > d\} & \leq & \max\{\delta \underline{y}_d, \delta Y\} \mathbb{1}\{D > d\}, \\ \min\{\delta \overline{y}_d, \delta Y\} \mathbb{1}\{D < d\} & \leq & \delta Y_d \mathbb{1}\{D < d\} & \leq & \max\{\delta \overline{y}_d, \delta Y\} \mathbb{1}\{D < d\}. \end{array} \tag{3.9}$$

Now, we note that $Y_d \delta \mathbb{1}\{D > d\} + Y_d \delta \mathbb{1}\{D < d\} = Y_d \delta \mathbb{1}\{D \neq d\}$. Recall that inequality (3.2) implied by the second inequality of (3.7) yields

$$\begin{array}{rcl} \mathbb{E}\left[Y \left(1 + \alpha \tilde{Z}\right) \mathbb{1}\{D = d\} + Y_d \left(1 + \alpha \tilde{Z}\right) \mathbb{1}\{D \neq d\}\right] & \geq & \theta_d, \\ \mathbb{E}\left[Y \left(1 - \alpha \tilde{Z}\right) \mathbb{1}\{D = d\} + Y_d \left(1 - \alpha \tilde{Z}\right) \mathbb{1}\{D \neq d\}\right] & \leq & \theta_d, \end{array}$$

for every $\alpha \in [0, 1)$. Combining this with (3.9), and replacing δ by $1 + \alpha \tilde{Z}$ or $1 - \alpha \tilde{Z}$ respectively, implies

$$\begin{array}{rcl} Y_d \left(1 + \alpha \tilde{Z}\right) \mathbb{1}\{D \neq d\} & \leq & \max\left\{\underline{y}_d \left(1 + \alpha \tilde{Z}\right), Y \left(1 + \alpha \tilde{Z}\right)\right\} \mathbb{1}\{D > d\} \\ & & + \max\left\{\overline{y}_d \left(1 + \alpha \tilde{Z}\right), Y \left(1 + \alpha \tilde{Z}\right)\right\} \mathbb{1}\{D < d\}, \\ Y_d \left(1 - \alpha \tilde{Z}\right) \mathbb{1}\{D \neq d\} & \geq & \min\left\{\underline{y}_d \left(1 - \alpha \tilde{Z}\right), Y \left(1 - \alpha \tilde{Z}\right)\right\} \mathbb{1}\{D > d\} \\ & & + \min\left\{\overline{y}_d \left(1 - \alpha \tilde{Z}\right), Y \left(1 - \alpha \tilde{Z}\right)\right\} \mathbb{1}\{D < d\}. \end{array}$$

Thus, we have

$$\begin{array}{c} \mathbb{E}\left[\overline{m}_d \left(Y, D, 1 + \alpha \tilde{Z}\right)\right] \geq \theta_d, \\ \mathbb{E}\left[\underline{m}_d \left(Y, D, 1 - \alpha \tilde{Z}\right)\right] \leq \theta_d \end{array}$$

for any $\alpha \in [0, 1)$, where we define the functions $\underline{m}_d$ and $\overline{m}_d$ as

$$\begin{array}{rcl} \underline{m}_d (Y, D, \delta) & \equiv & \mathbb{1}\{D = d\} \delta Y + \mathbb{1}\{D > d\} \min\{\delta \underline{y}_d, \delta Y\} + \mathbb{1}\{D < d\} \min\{\delta \overline{y}_d, \delta Y\}, \\ \overline{m}_d (Y, D, \delta) & \equiv & \mathbb{1}\{D = d\} \delta Y + \mathbb{1}\{D > d\} \max\{\delta \underline{y}_d, \delta Y\} + \mathbb{1}\{D < d\} \max\{\delta \overline{y}_d, \delta Y\}. \end{array}$$

Likewise, inequality (3.4) implied by the second inequality of (3.8) yields the following implications for θ_d:

$$\begin{aligned}\mathbb{E}\Big[\overline{m}_d\left(Y, D, 1-\alpha\tilde{Z}\right)\Big] &\geq \theta_d,\\ \mathbb{E}\Big[\underline{m}_d\left(Y, D, 1+\alpha\tilde{Z}\right)\Big] &\leq \theta_d.\end{aligned}$$

for any $\alpha \in [0,1)$. Hence, we derive new bounds for θ_d implied by the second inequalities of (3.7) and (3.8) from adding the MTR assumption:

$$\begin{aligned}\theta_d &\in \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{m}_d\left(Y, D, 1-\alpha\tilde{Z}\right)\Big], \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{m}_d\left(Y, D, 1+\alpha\tilde{Z}\right)\Big]\right] \equiv I^d_{MTR2},\\ \theta_d &\in \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{m}_d\left(Y, D, 1+\alpha\tilde{Z}\right)\Big], \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{m}_d\left(Y, D, 1-\alpha\tilde{Z}\right)\Big]\right] \equiv I^d_{MTR4}.\end{aligned}$$

Similarly, by replacing $\delta = 1-\alpha\left(\tilde{D}\sigma_Z - \tilde{Z}\sigma_D\right)$ and $\delta = 1+\alpha\left(\tilde{D}\sigma_Z - \tilde{Z}\sigma_D\right)$, respectively, in (3.9), we derive the following bounds

$$\begin{aligned}\theta_d \in \Bigg[&\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{m}_d\left(Y, D, 1-\alpha\left(\tilde{D}\sigma_Z - \tilde{Z}\sigma_D\right)\right)\Big],\\ &\inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{m}_d\left(Y, D, 1+\alpha\left(\tilde{D}\sigma_Z - \tilde{Z}\sigma_D\right)\right)\Big]\Bigg] \equiv I^d_{MTR1}\end{aligned}$$

implied by the first inequality of (3.7) using a similar reasoning from the previous subsections. In the same way, the first inequality of (3.8) now implies

$$\begin{aligned}\theta_d \in \Bigg[&\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{m}_d\left(Y, D, 1+\alpha\left(\tilde{D}\sigma_Z - \tilde{Z}\sigma_D\right)\right)\Big],\\ &\inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{m}_d\left(Y, D, 1-\alpha\left(\tilde{D}\sigma_Z - \tilde{Z}\sigma_D\right)\right)\Big]\Bigg] \equiv I^d_{MTR3}\end{aligned}$$

after adding MTR assumption. Therefore, we conclude that Assumptions BoS, SDC, LEI, and MTR together imply

$$\theta_d \in \left(I^d_{MTR1} \cap I^d_{MTR2}\right) \cup \left(I^d_{MTR3} \cap I^d_{MTR4}\right) \equiv I^d_{MTR}.$$

3.4. **Adding the Roy selection assumption.** In this subsection, we derive bounds on θ_d under Assumptions BoS, SDC, LEI, and RS. Under Assumption RS, we have

$$\begin{aligned} Y_d \mathbb{1}\{D = d'\} &= Y_d \mathbb{1}\left\{Y_{d'} > \max\{Y_d : d \neq d'\}\right\} \\ &\leq Y_{d'} \mathbb{1}\left\{Y_{d'} > \max\{Y_d : d \neq d'\}\right\} \\ &= Y_{d'} \mathbb{1}\{D = d'\} = Y \mathbb{1}\left\{D = d'\right\}, \end{aligned}$$

where the inequality holds from the fact that $Y_d \leq Y_{d'}$ for all d when $D = d'$. By taking the summation over all $d' \neq d$, the derived inequality implies

$$\sum_{d' \neq d} Y_d \mathbb{1}\{D = d'\} \leq \sum_{d' \neq d} Y \mathbb{1}\left\{D = d'\right\}, \text{ which is the same as } \; Y_d \mathbb{1}\{D \neq d\} \leq Y \mathbb{1}\{D \neq d\}.$$

By adding $Y_d \mathbb{1}\{D = d\}$ to each side of this latter inequality, it implies that $Y_d \leq Y$ for all d. This last inequality shows that the observed outcome Y is an upper bound for the potential outcome Y_d under Assumption RS. This result together with Assumption BoS imply $\underline{y}_d \leq Y_d \leq Y \leq \overline{y}_d$ for all d. Thus, we note that the second inequality of (3.7), which is equivalent to (3.2), now implies

$$\begin{aligned} \mathbb{E}\Big[\overline{r}_d\left(Y, D, 1 + \alpha \tilde{Z}\right)\Big] &\geq \theta_d, \\ \mathbb{E}\Big[\underline{r}_d\left(Y, D, 1 - \alpha \tilde{Z}\right)\Big] &\leq \theta_d, \end{aligned}$$

for any $\alpha \in [0, 1)$, where we define the function $\underline{r}_d$ and $\overline{r}_d$ as

$$\begin{aligned} \underline{r}_d\left(Y, D, \delta\right) &\equiv \mathbb{1}\left\{D = d\right\} \delta Y + \mathbb{1}\left\{D \neq d\right\} \min\left\{\delta \underline{y}_d, \delta Y\right\}, \\ \overline{r}_d\left(Y, D, \delta\right) &\equiv \mathbb{1}\left\{D = d\right\} \delta Y + \mathbb{1}\left\{D \neq d\right\} \max\left\{\delta \underline{y}_d, \delta Y\right\}, \end{aligned}$$

since we have

$$\begin{aligned} Y_d \left(1 + \alpha \tilde{Z}\,\right) &\leq \max\left\{\underline{y}_d \left(1 + \alpha \tilde{Z}\right), Y \left(1 + \alpha \tilde{Z}\right)\right\}, \\ Y_d \left(1 - \alpha \tilde{Z}\,\right) &\geq \min\left\{\underline{y}_d \left(1 - \alpha \tilde{Z}\right), Y \left(1 - \alpha \tilde{Z}\right)\right\}. \end{aligned}$$

In a way similar to the derivations in the previous sections, we can show that Assumptions BoS, SDC, LEI, and RS yield the following bounds for θ_d:

$$I^d_{RS} \equiv \left(I^d_{RS1} \cap I^d_{RS2}\right) \cup \left(I^d_{RS3} \cap I^d_{RS4}\right),$$

where

$$
\begin{aligned}
I_{RS1}^{d} &\equiv \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{r}_d\left(Y,D,1-\alpha\left(\tilde{D}\sigma_Z-\tilde{Z}\sigma_D\right)\right)\Big], \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{r}_d\left(Y,D,1+\alpha\left(\tilde{D}\sigma_Z-\tilde{Z}\sigma_D\right)\right)\Big]\right],\\
I_{RS2}^{d} &\equiv \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{r}_d\left(Y,D,1-\alpha\tilde{Z}\right)\Big], \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{r}_d\left(Y,D,1+\alpha\tilde{Z}\right)\Big]\right],\\
I_{RS3}^{d} &\equiv \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{r}_d\left(Y,D,1+\alpha\left(\tilde{D}\sigma_Z-\tilde{Z}\sigma_D\right)\right)\Big], \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{r}_d\left(Y,D,1-\alpha\left(\tilde{D}\sigma_Z-\tilde{Z}\sigma_D\right)\right)\Big]\right],\\
I_{RS4}^{d} &\equiv \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{r}_d\left(Y,D,1+\alpha\tilde{Z}\right)\Big], \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{r}_d\left(Y,D,1-\alpha\tilde{Z}\right)\Big]\right].
\end{aligned}
$$

4. Inference

We want to construct confidence bounds for the set $I_{SDC}^d = \left(I_{SDC1}^d \cap I_{SDC2}^d\right) \cup \left(I_{SDC3}^d \cap I_{SDC4}^d\right)$, where

$$
\begin{aligned}
I_{SDC1}^{d} &= \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{f}_d\left(Y,D,1-\alpha\tilde{D}\right)\Big], \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{f}_d\left(Y,D,1+\alpha\tilde{D}\right)\Big]\right]\\
I_{SDC2}^{d} &= \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{f}_d\left(Y,D,1-\alpha\tilde{Z}\right)\Big], \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{f}_d\left(Y,D,1+\alpha\tilde{Z}\right)\Big]\right]\\
I_{SDC3}^{d} &= \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{f}_d\left(Y,D,1+\alpha\tilde{D}\right)\Big], \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{f}_d\left(Y,D,1-\alpha\tilde{D}\right)\Big]\right]\\
I_{SDC4}^{d} &= \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{f}_d\left(Y,D,1+\alpha\tilde{Z}\right)\Big], \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{f}_d\left(Y,D,1-\alpha\tilde{Z}\right)\Big]\right].
\end{aligned}
$$

This is an intersection-union test as described in Berger (1982). We are going to construct confidence regions for the sets $I_{SDC1}^d \cap I_{SDC2}^d$ and $I_{SDC3}^d \cap I_{SDC4}^d$ using the intersection bounds framework of Chernozhukov, Lee, and Rosen (2013) or Andrews and Shi (2013), and then take the union of the two confidence regions. Berger and Hsu (1996) showed that the union of the confidence regions has at least the same coverage rate as each confidence region.

We now explain how to rewrite the intersection bounds I_{SDC1}^d in such a way that it can be easily implemented using the Chernozhukov et al. (2015) or Andrews, Kim, and Shi (2017) Stata packages. Suppose that we draw a random variable U from the uniform distribution over $[0,1)$, independently of the data (Y,D,Z). We have

$$\mathbb{E}\Big[\underline{f}_d\left(Y,D,1-U\tilde{D}\right)|U=\alpha\Big]=\mathbb{E}\Big[\underline{f}_d\left(Y,D,1-\alpha\tilde{D}\right)\Big],$$

since U is independent of (Y,D,Z). Therefore,

$$I^d_{SDC1}=\left[\sup_{\alpha\in[0,1)}\mathbb{E}\Big[\underline{f}_d\left(Y,D,1-U\tilde{D}\right)|U=\alpha\Big],\inf_{\alpha\in[0,1)}\mathbb{E}\Big[\overline{f}_d\left(Y,D,1+U\tilde{D}\right)|U=\alpha\Big]\right].$$

Hence, these bounds take the form of conditional moment inequalities, which can be implemented using existing inferential methods like Chernozhukov et al. (2015) or Andrews, Kim, and Shi (2017). A similar technique has been proposed in Kédagni, Li, and Mourifié (2018) where they construct confidence sets for the potential outcome means under the IV zero-covariance assumption.

5. Empirical illustration

In this application, we use a data set drawn from the NLSYM. This data includes 3,010 young men who were ages 24-34 in 1976. It is the same data used in Card (1995). In our analysis, the outcome variable is log hourly wage in cents ($lwage$), and the treatment variable is education ($educ$) grouped in 4 categories: less than high school ($educ < 12\, years$), high school ($12 \leq educ < 16$), college degree ($16 \leq educ < 18$), and graduate ($educ \geq 18$). Our imperfect IV is parental education. Since the work of Willis and Rosen (1979), parental education has been used as an IV. However, an individual's ability can be dependent on her parents' ability, which is correlated with parental education. For this reason, parents' education will not be a valid instrument. This fact is documented in Kédagni and Mourifié (2020), who provided evidence that even after controlling for a measure of ability, parental education is not a good instrument. This result is contrary to the Lemke and Rischall (2003) idea that controlling for some measure of child ability could make parental education a valid IV. Nonetheless, it is reasonable to assume that parental education has the same sign of correlation with the individual's potential wage as the correlation between the person's potential wage and her own education. It is also likely that parental education be less endogenous than is the person's own education. Finally, as in Manski and Pepper (2000), we use the monotone treatment response assumption to tighten the bounds on the average returns to education.

In theory, the outcome variable $lwage$ is unbounded. For practical reasons, we follow Ginther (2000) to trim the log wage. The outcome variable that we use is defined as $Y = \tau$-quantile of $lwage$ if $lwage$ is less than or equal to its τ-quantile, $Y = (1-\tau)$-quantile of *lwage* if *lwage* is greater than or equal to its $(1-\tau)$-quantile, and $Y = lwage$ otherwise. In our empirical illustration, we set $\tau = 0.05$. We construct two-sided confidence bounds on the potential average log wages using the `clr2bound` command of Chernozhukov et al. (2015) in the Stata software. We estimate the conditional expectations using local linear methods. See Appendix B for more details on the implementation.

We present the results with mother's education as an IIV. The results for father's education are in Appendix C. Table 1 displays the 95% confidence bounds for the potential wage means and average returns to schooling under the SDC assumption, while Table 2 shows the confidence set under both the SDC and LEI assumptions. The results are similar in both tables. This suggests that the constraints imposed by Assumption LEI are not binding. The bounds seem wide and uninformative.

Table 1. Confidence sets for parameters under SDC

Parameters	95% conf. LB	95% conf. UB
θ_0 (< high)	5.53	6.86
θ_1 (high)	5.89	6.66
θ_2 (college)	5.65	6.88
θ_3 (graduate)	5.55	6.94
$\theta_0 - \theta_1$	-1.13	0.97
$\theta_2 - \theta_1$	-1.01	0.98
$\theta_3 - \theta_1$	-1.11	1.05

conf. LB: confidence lower bound; conf. UB: confidence upper bound.

Table 2. Confidence sets for parameters under SDC and LEI

Parameters	95% conf. LB	95% conf. UB
θ_0 (< high)	5.53	6.86
θ_1 (high)	5.89	6.66
θ_2 (college)	5.65	6.86
θ_3 (graduate)	5.55	6.94
$\theta_0 - \theta_1$	-1.13	0.97
$\theta_2 - \theta_1$	-1.01	0.97
$\theta_3 - \theta_1$	-1.11	1.05

conf. LB: confidence lower bound; conf. UB: confidence upper bound.

However, the confidence regions for the parameters considerably shrink and become informative when we add the MTR assumption (see Table 3). We assume that the lower bound of θ_d is equal to the upper bound of θ_{d-1} as Y_{d-1} cannot exceed Y_d for each $d = 1, 2, 3$ under the MTR assumption. Individuals with less than high school education could earn up to 93% less than high school graduates. College graduates could earn up to 52% more than high school graduates, while individuals with a graduate degree earn between 36% and 64% higher wages than high school graduates (which approximately represents an annual return between 6.0% and 10.7%).

Table 3. Confidence sets for parameters under SDC, LEI, and MTR

Parameters	95% conf. LB	95% conf. UB
θ_0 (< high)	5.53	6.30
θ_1 (high)	6.30	6.46
θ_2 (college)	6.46	6.82
θ_3 (graduate)	6.82	6.94
$\theta_0 - \theta_1$	-0.93	0.00
$\theta_2 - \theta_1$	0.00	0.52
$\theta_3 - \theta_1$	0.36	0.64

conf. LB: confidence lower bound; conf. UB: confidence upper bound.

6. Conclusion

In this paper, we derive nonparametric bounds on the average treatment effect when an imperfect instrument is available. We extend Nevo and Rosen's (2012) identification results to nonparametric models. We first assume that the sign of correlation between the imperfect instrument and the unobserved latent variables is the same as the correlation between the endogenous variable and the latent variables. We show that the MTS-MIV restrictions introduced by Manski and Pepper (2000, 2009), jointly imply this assumption. We introduce the concept of comonotone IV, which also satisfies this assumption. Second, we show how the assumption that the imperfect instrument is less endogenous than the treatment variable can help tighten the bounds. We also use the monotone treatment response assumption to get tighter bounds. The identified set takes the form of intersection bounds, which can be implemented using Nevo and Rosen's (2012)Chernozhukov, Lee, and Rosen's (2013) inferential method. Finally, we illustrate our methodology using the National Longitudinal Survey of Young Men data to estimate returns to schooling.

Appendix A. Proofs

A.1. **Proof of Lemma 1.**

Proof. We have

$$g_d^+(j) - g_d^-(j) \quad = \quad \sum_{\ell=j}^{T} \mathbb{E}[Y_d|D=\ell]\frac{\mathbb{P}(D=\ell)}{\mathbb{P}(D\geq j)} - \sum_{\ell=1}^{j-1} \mathbb{E}[Y_d|D=\ell]\frac{\mathbb{P}(D=\ell)}{\mathbb{P}(D<j)}$$

Suppose without loss of generality that $\mathbb{E}[Y_d|D=\ell]$ is increasing in ℓ. Then

$$\sum_{\ell=j}^{T} \mathbb{E}[Y_d|D=\ell]\frac{\mathbb{P}(D=\ell)}{\mathbb{P}(D\geq j)} \geq \mathbb{E}[Y_d|D=j] \text{ and } \sum_{\ell=1}^{j-1} \mathbb{E}[Y_d|D=\ell]\frac{\mathbb{P}(D=\ell)}{\mathbb{P}(D<j)} \leq \mathbb{E}[Y_d|D=j-1].$$

Therefore

$$g_d^+(j) - g_d^-(j) \quad \geq \quad \mathbb{E}[Y_d|D=j] - \mathbb{E}[Y_d|D=j-1] \geq 0.$$

On the other hand, we have

$$\begin{aligned} h_d^+(z) &= \mathbb{E}\left[Y_d|Z\geq z\right] = \int_z^{\infty} \mathbb{E}\left[Y_d|Z=v\right] \frac{f_Z(v)}{\mathbb{P}(Z\geq z)} dv, \\ &\geq \int_z^{\infty} \mathbb{E}\left[Y_d|Z=z\right] \frac{f_Z(v)}{\mathbb{P}(Z\geq z)} dv = \mathbb{E}\left[Y_d|Z=z\right], \end{aligned}$$

where f_Z is the density (or probability mass) of Z, and the inequality holds because $\mathbb{E}\left[Y_d|Z=v\right]$ is increasing in v. Similarly, we have

$$\begin{aligned} h_d^-(z) &= \mathbb{E}\left[Y_d|Z\geq z\right] = \int_0^{z} \mathbb{E}\left[Y_d|Z=v\right] \frac{f_Z(v)}{\mathbb{P}(Z< z)} dv, \\ &\leq \int_0^{z} \mathbb{E}\left[Y_d|Z=z\right] \frac{f_Z(v)}{\mathbb{P}(Z< z)} dv = \mathbb{E}\left[Y_d|Z=z\right], \end{aligned}$$

Therefore

$$h_d^+(z) - h_d^-(z) \quad \geq \quad \mathbb{E}\left[Y_d|Z=z\right] - \mathbb{E}\left[Y_d|Z=z\right] = 0.$$

□

A.2. **Proof of Lemma 2.**

Proof. We first notice that

$$D = \sum_{j=1}^{T} j\mathbb{1}\{D=j\} = \sum_{j=1}^{T} j\left(\mathbb{1}\{D\geq j\} - \mathbb{1}\{D>j\}\right) = \sum_{j=1}^{T} \mathbb{1}\{D\geq j\},$$

and

$$Z = \int_0^Z dz = \int_0^\infty \mathbb{1}\{Z \geq z\}\, dz \text{ (layer cake representation).}$$

Then

$$Cov(Y_d, D) = \sum_{j=1}^{T} Cov(Y_d, \mathbb{1}\{D \geq j\}),$$

and

$$Cov(Y_d, Z) = \int_0^\infty Cov(Y_d, \mathbb{1}\{Z \geq z\})dz \text{ by the Fubini-Tonelli theorem.}$$

We also show that

$$\begin{aligned} Cov(Y_d, \mathbb{1}\{D \geq j\}) &= \mathbb{P}(D \geq j)\mathbb{P}(D < j)\left(\mathbb{E}[Y_d|D \geq j] - \mathbb{E}[Y_d|D < j]\right) \\ Cov(Y_d, \mathbb{1}\{Z \geq z\}) &= \mathbb{P}(Z \geq z)\mathbb{P}(Z < z)\left(\mathbb{E}[Y_d|Z \geq z] - \mathbb{E}[Y_d|Z < z]\right). \end{aligned}$$

From these results, it is straightforward to verify that if Z is binarized MTS-MIV for D, then

$$\begin{aligned} Cov(Y_d, D)Cov(Y_d, Z) &= \sum_{j=1}^{T} Cov(Y_d, \mathbb{1}\{D \geq j\}) \int_0^\infty Cov(Y_d, \mathbb{1}\{Z \geq z\})dz, \\ &= \sum_{j=1}^{T} \int_0^\infty Cov(Y_d, \mathbb{1}\{D \geq j\})Cov(Y_d, \mathbb{1}\{Z \geq z\})dz, \\ &= \sum_{j=1}^{T} \int_0^\infty P(j,z)\left(g_d^+(j) - g_d^-(j)\right)\left(h_d^+(z) - h_d^-(z)\right)dz \geq 0, \end{aligned}$$

where $P(j,z) = \mathbb{P}(D \geq j)\mathbb{P}(D < j)\mathbb{P}(Z \geq z)\mathbb{P}(Z < z)$.

Note that Z does not have to be continuous, and the result still holds when Z is discrete.

□

A.3. **Proof of Lemma 3.**

Proof. (1) Suppose $D = \nu(Z)$ and ν is increasing. Then

$$\left(D(\omega) - D(\omega')\right)\left(Z(\omega) - Z(\omega')\right) = \left(\nu(Z(\omega)) - \nu(Z(\omega'))\right)\left(Z(\omega) - Z(\omega')\right).$$

Since ν is increasing, $\nu(Z(\omega)) - \nu(Z(\omega'))$ and $Z(\omega) - Z(\omega')$ have the same sign. Therefore,

$$\left(D(\omega) - D(\omega')\right)\left(Z(\omega) - Z(\omega')\right) \geq 0.$$

(2) Suppose $D = h(Z, V)$, where h is increasing in both of its arguments. Suppose that Z and V are comonotonic. Then

$$\left(D(\omega) - D(\omega')\right)\left(Z(\omega) - Z(\omega')\right) = \left[h(Z(\omega), V(\omega)) - h(Z(\omega'), V(\omega'))\right]\left(Z(\omega) - Z(\omega')\right),$$
$$= \left[\left(h(Z(\omega), V(\omega)) - h(Z(\omega), V(\omega'))\right) + \left(h(Z(\omega), V(\omega')) - h(Z(\omega'), V(\omega'))\right)\right]\left(Z(\omega) - Z(\omega')\right).$$

$h(Z(\omega), V(\omega)) - h(Z(\omega), V(\omega'))$ has the same sign as $V(\omega) - V(\omega')$ since h is increasing in its second argument, while $h(Z(\omega), V(\omega')) - h(Z(\omega'), V(\omega'))$ has the same sign as $Z(\omega) - Z(\omega')$ since h is increasing in its first argument. Since Z and V are comonotonic, it follows that $V(\omega) - V(\omega')$ and $Z(\omega) - Z(\omega')$ have the same sign. Therefore,

$$\left(h(Z(\omega), V(\omega)) - h(Z(\omega), V(\omega'))\right)\left(Z(\omega) - Z(\omega')\right) \geq 0,$$

and

$$\left(h(Z(\omega), V(\omega')) - h(Z(\omega'), V(\omega'))\right)\left(Z(\omega) - Z(\omega')\right) \geq 0.$$

Hence,

$$\left(D(\omega) - D(\omega')\right)\left(Z(\omega) - Z(\omega')\right) \geq 0.$$

□

A.4. **Proof of Lemma 4.**

Proof. Suppose Z and D are comonotonic. Then by definiton, we have for all $\omega, \omega' \in \Omega$,

$$\left(D(\omega) - D(\omega')\right)\left(Z(\omega) - Z(\omega')\right) \geq 0,$$

which implies

$$\left[\left(Y_d(\omega) - Y_d(\omega')\right)\left(D(\omega) - D(\omega')\right)\right]\left[\left(Y_d(\omega) - Y_d(\omega')\right)\left(Z(\omega) - Z(\omega')\right)\right] \geq 0$$

as $(Y_d(\omega) - Y_d(\omega'))^2 \geq 0$. This latter inequality is equivalent to: either

$$\left(Y_d(\omega) - Y_d(\omega')\right)\left(D(\omega) - D(\omega')\right) \geq 0, \tag{A.1}$$
$$\left(Y_d(\omega) - Y_d(\omega')\right)\left(Z(\omega) - Z(\omega')\right) \geq 0, \tag{A.2}$$

or

$$\left(Y_d(\omega) - Y_d(\omega')\right)\left(D(\omega) - D(\omega')\right) \leq 0, \tag{A.3}$$
$$\left(Y_d(\omega) - Y_d(\omega')\right)\left(Z(\omega) - Z(\omega')\right) \leq 0. \tag{A.4}$$

Using the results (i)-(ii) in Section 2.2 and (S1) in Section 4.3 from Wang and Zitikis (2020), we conclude that inequality (A.1) implies that $Cov(Y_d, D) \geq 0$. Similarly, inequalities (A.2), (A.3) and (A.4), respectively, imply $Cov(Y_d, Z) \geq 0$, $Cov(Y_d, D) \leq 0$, and $Cov(Y_d, Z) \leq 0$. Therefore, we have $Cov(Y_d, D)Cov(Y_d, Z) \geq 0$ in all cases. This concludes the proof. $\square$

Appendix B. Implementation of the bounds

In this section, we show the different steps for implementing the method developed in the paper. We want to construct confidence bounds on the parameter $\theta_d \in I^d_{SDC1}$, where

$$I^d_{SDC1} = \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{f}_d\left(Y, D, 1 - U\tilde{D}\right)|U = \alpha\Big], \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{f}_d\left(Y, D, 1 + U\tilde{D}\right)|U = \alpha\Big]\right].$$

As explained in the main text, the implementation can be done using the Stata package developed by Chernozhukov et al. (2015). Assuming that we have an i.i.d. data with finite moments, the conditions for the implementation of the Chernozhukov, Lee, and Rosen (2013) method are likely to hold in our framework. We provide below the code for the implementation.

```
set more off
gen Y = lwage
centile(Y), centile(5 95)
replace Y = r(c_1) if Y < r(c_1)
replace Y = r(c_2) if Y > r(c_2)
replace educ = 0 if educ < 12
replace educ = 1 if educ >= 12 & educ < 16
replace educ = 2 if educ >= 16 & educ < 18
replace educ = 3 if educ >= 18
replace motheduc = 0 if motheduc < 12
replace motheduc = 1 if motheduc >= 12 & motheduc < 16
replace motheduc = 2 if motheduc >= 16 & motheduc < 18
replace motheduc = 3 if motheduc >= 18 & motheduc != .
gen D = (educ==0)
gen Z = motheduc
sum Y if D == 1
scalar Y1up = r(max)
scalar Y1lo = r(min)
sum Y if D == 0
scalar Y0up = r(max)
scalar Y0lo = r(min)
sum Z
```

```
gen RZ = _n-1
replace RZ = . if RZ[_n]>3
set seed 12345
gen Lambda = runiform(0,1)
sum Z
scalar EZ = r(mean)
sum educ
scalar ED = r(mean)
gen ldepen1 = Y*D*(1-Lambda*(Z-EZ))
        + min(Y1lo*(1-D)*(1-Lambda*(Z-EZ)),Y1up*(1-D)*(1-Lambda*(Z-EZ)))
gen ldepen2 = Y*D*(1-Lambda*(educ-ED))
        + min(Y1lo*(1-D)*(1-Lambda*(educ-ED)),Y1up*(1-D)*(1-Lambda*(educ-ED)))
gen udepen1 = Y*D*(1+Lambda*(Z-EZ))
        + max(Y1lo*(1-D)*(1+Lambda*(Z-EZ)),Y1up*(1-D)*(1+Lambda*(Z-EZ)))
gen udepen2 = Y*D*(1+Lambda*(educ-ED))
        + max(Y1lo*(1-D)*(1+Lambda*(educ-ED)),Y1up*(1-D)*(1+Lambda*(educ-ED)))
gen RLambda = 0+0.01*_n
replace RLambda = . if RLambda>1
clr2bound ((ldepen1 Lambda RLambda) (ldepen2 Lambda RLambda))
        ((udepen1 Lambda RLambda) (udepen2 Lambda RLambda))
        , met("local") level(0.5 0.9 0.95 0.99) norseed rnd(20000)
gen ldepen12 = Y*D*(1+Lambda*(Z-EZ))
        + min(Y1lo*(1-D)*(1+Lambda*(Z-EZ)),Y1up*(1-D)*(1+Lambda*(Z-EZ)))
gen ldepen22 = Y*D*(1+Lambda*(educ-ED))
        + min(Y1lo*(1-D)*(1+Lambda*(educ-ED)),Y1up*(1-D)*(1+Lambda*(educ-ED)))
gen udepen12 = Y*D*(1-Lambda*(Z-EZ))
        + max(Y1lo*(1-D)*(1-Lambda*(Z-EZ)),Y1up*(1-D)*(1-Lambda*(Z-EZ)))
gen udepen22 = Y*D*(1-Lambda*(educ-ED))
        + max(Y1lo*(1-D)*(1-Lambda*(educ-ED)),Y1up*(1-D)*(1-Lambda*(educ-ED)))
clr2bound ((ldepen12 Lambda RLambda) (ldepen22 Lambda RLambda))
        ((udepen12 Lambda RLambda) (udepen22 Lambda RLambda))
        , met("local") level(0.5 0.9 0.95 0.99) norseed rnd(20000)
```

Appendix C. Additional empirical results

C.1. **Additional results using mother's education as an IIV.** Tables 4 and 5 show additional results using mother's education as an imperfect IV.

Table 4. Confidence sets for parameters under SDC, LEI, and RS

Parameters	95% conf. LB	95% conf. UB
θ_0 (< high)	5.53	6.30
θ_1 (high)	5.89	6.30
θ_2 (college)	5.65	6.30
θ_3 (graduate)	5.55	6.30
$\theta_0 - \theta_1$	-0.77	0.41
$\theta_2 - \theta_1$	-0.65	0.41
$\theta_3 - \theta_1$	-0.75	0.41

conf. LB: confidence lower bound; conf. UB: confidence upper bound.

Table 5. Confidence sets for parameters under SDC and MTR

Parameters	95% conf. LB	95% conf. UB
θ_0 (< high)	5.53	6.30
θ_1 (high)	6.30	6.47
θ_2 (college)	6.47	6.84
θ_3 (graduate)	6.84	6.94
$\theta_0 - \theta_1$	-0.94	0.00
$\theta_2 - \theta_1$	0.00	0.54
$\theta_3 - \theta_1$	0.38	0.64

conf. LB: confidence lower bound; conf. UB: confidence upper bound.

Note that the bounds in Table 5 are obtained under Assumptions BoS, SDC, and MTR without Assumption LEI:

$$\theta_d \in \left(\tilde{I}^d_{MTR1} \cap I^d_{MTR2}\right) \cup \left(\tilde{I}^d_{MTR3} \cap I^d_{MTR4}\right),$$

where

$$\tilde{I}^d_{MTR1} \equiv \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{m}_d\left(Y, D, 1-\alpha\tilde{D}\right)\Big], \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{m}_d\left(Y, D, 1+\alpha\tilde{D}\right)\Big]\right],$$

$$\tilde{I}^d_{MTR3} \equiv \left[\sup_{\alpha\in[0,1)} \mathbb{E}\Big[\underline{m}_d\left(Y, D, 1+\alpha\tilde{D}\right)\Big], \inf_{\alpha\in[0,1)} \mathbb{E}\Big[\overline{m}_d\left(Y, D, 1-\alpha\tilde{D}\right)\Big]\right].$$

C.2. **Results using father's education as an IIV.** Tables 6 and 7 present the results using father's education as an imperfect IV. The results in Table 6 are similar to those we obtain using mother's education as an IIV. Table 7 displays the results with race as a control variable.

Table 6. Confidence sets under SDC and MTR using father's education as IIV

Parameters	95% conf. LB	95% conf. UB
θ_0 (< high)	5.51	6.36
θ_1 (high)	6.36	6.48
θ_2 (college)	6.48	6.82
θ_3 (graduate)	6.82	6.96
$\theta_0 - \theta_1$	-0.97	0.00
$\theta_2 - \theta_1$	0.00	0.45
$\theta_3 - \theta_1$	0.34	0.59

conf. LB: confidence lower bound; conf. UB: confidence upper bound.

Table 7. Confidence sets under SDC and MTR using father's education as IIV with race as control

Parameters	95% conf. LB black =1	95% conf. UB black =1	95% conf. LB black =0	95% conf. UB black =0
θ_0 (< high)	5.52	6.42	5.48	6.15
θ_1 (high)	6.42	6.47	6.15	6.56
θ_2 (college)	6.47	6.81	6.56	6.95
θ_3 (graduate)	6.81	6.96	6.95	7.04
$\theta_0 - \theta_1$	-0.96	0.00	-1.08	0.00
$\theta_2 - \theta_1$	0.00	0.39	0.00	0.81
$\theta_3 - \theta_1$	0.34	0.54	0.40	0.89

conf. LB: confidence lower bound; conf. UB: confidence upper bound.

Appendix D. Supplementary Examples

Example 2. *MTS-MIV, binarized MTS-MIV, and comonotonicity hold.*

$$\begin{cases} Y &= U\mathbb{1}\{D=0\} + 2U\mathbb{1}\{D=1\} + \frac{1}{2}U\mathbb{1}\{D=2\} \\ D &= 0\cdot\mathbb{1}\left\{U\in[0,\frac{1}{2}]\right\} + 1\cdot\mathbb{1}\left\{U\in(\frac{1}{2},\frac{3}{2}]\right\} + 2\cdot\mathbb{1}\left\{U\in(\frac{3}{2},2]\right\} \\ Z &= 2D \end{cases} \tag{D.1}$$

where $U \sim \mathcal{U}_{[0,2]}$.

This example illustrates a case where the three aforementioned conditions hold; MTS-MIV, binarized MTS-MIV, and comonotonicity between D *and* Z. *Hence, we note that MTS-MIV and CoMIV are two different sufficient conditions for the SDC assumption, but not exclusive to each other.*

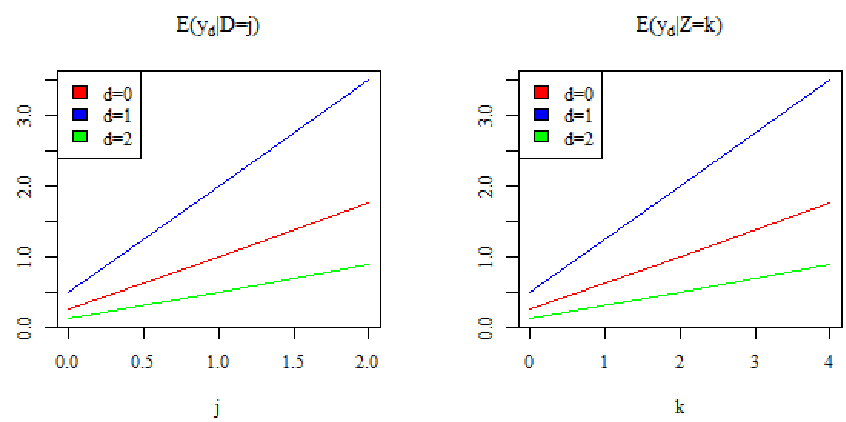

Figure 3. Numerical illustration of MTS-MIV

First, note that Z *is a CoMIV for* D *by (1) of Lemma 3 since* $Z = 2D$. *Moreover, it can be shown that the conditional expectation functions* $\mathbb{E}[Y_d|D=\ell]$ *and* $\mathbb{E}[Y_d|Z=z]$ *are monotone in* ℓ *and* z, *respectively, for all* $d = 0, 1, 2$ *(see Figure 3). Furthermore, Figure 4 shows that the DGP satisfies the binarized MTS-MIV restriction (2.2), and accordingly the SDC assumption by Lemma 1, as the function* g_d^+ *is always greater than the function* g_d^-, *and the function* h_d^+ *is always greater than the function* h_d^- *for all* d. $\square$

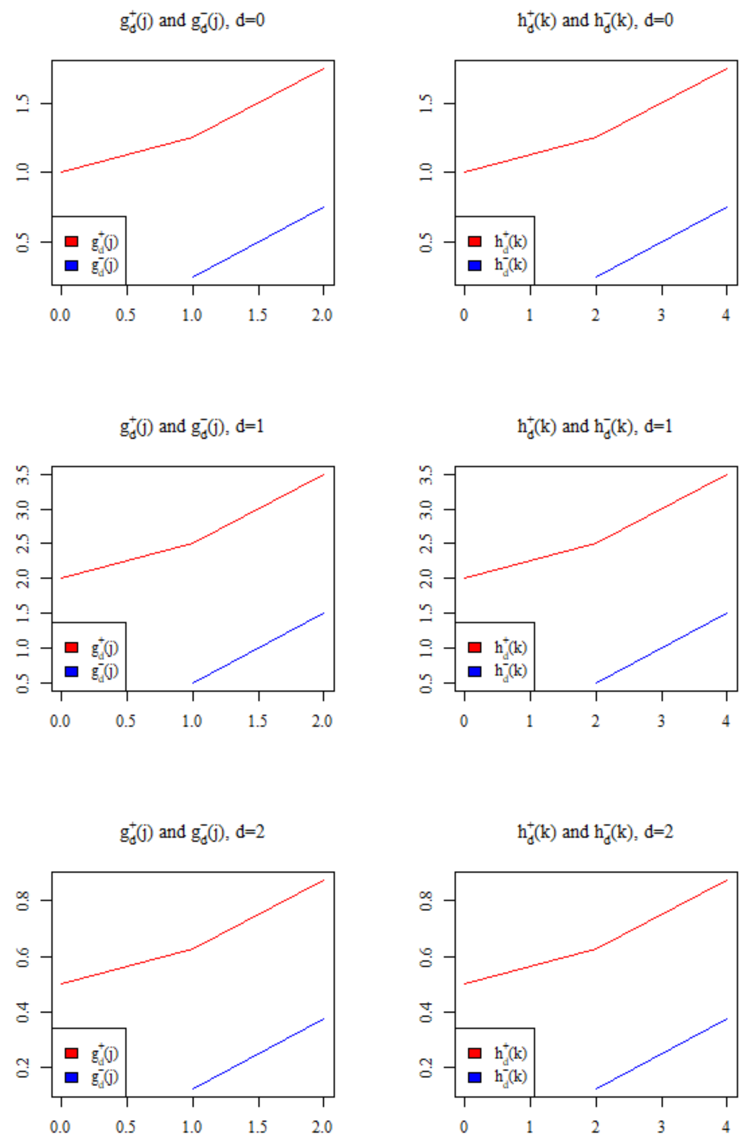

FIGURE 4. Numerical illustration of binarized MTS-MIV

Example 3. *MTS-MIV and binarized MTS-MIV hold, but CoMIV fails.*

$$\begin{cases} Y &= 2D + U \\ D &= 0 \cdot \mathbb{1}\{U \in [0,2]\} + 1 \cdot \mathbb{1}\{U \in (2,3]\} + 2 \cdot \mathbb{1}\{U \in (3,5]\} \\ Z &= 0 \cdot \mathbb{1}\{U \in [0,1] \cup (2,3]\} + 1 \cdot \mathbb{1}\{U \in (1,2] \cup (3,4]\} + 2 \cdot \mathbb{1}\{U \in (4,5]\} \end{cases} \tag{D.2}$$

where $U \sim \mathcal{U}_{[0,5]}$.

This example illustrates a case where MTS-MIV and binarized MTS-MIV hold, but CoMIV fails. Thus, we note that MTS-MIV does not imply CoMIV. Recall that Example 1 shows the case where MTS-MIV fails, but CoMIV holds. Hence, we conclude that neither of MTS-MIV nor CoMIV imply each other.

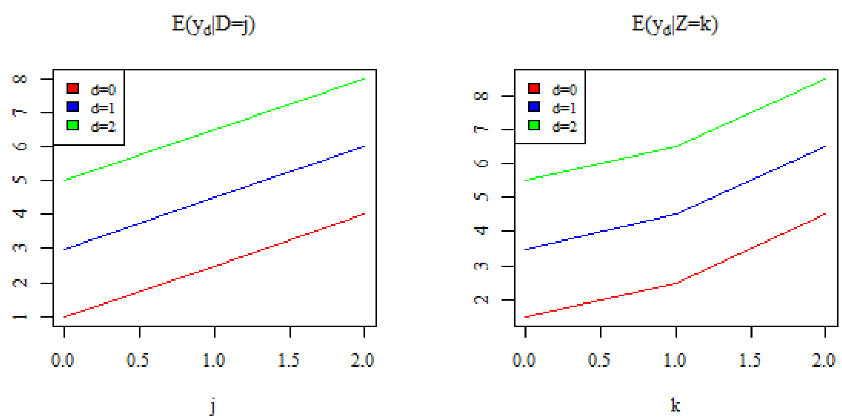

FIGURE 5. Numerical Illustration of MTS-MIV

It can be shown that the given DGP satisfies MTS-MIV and a numerical illustration is shown in Figure 5. Moreover, by Lemma 1, the same DGP also satisfies binarized MTS-MIV, which is shown in Figure 6; the conditional expectation functions $\mathbb{E}[Y_d|D=\ell]$ *and* $\mathbb{E}[Y_d|Z=z]$ *are monotone in* ℓ *and* z*, respectively, for all* $d = 0, 1, 2$*. However,* D *and* Z *are not comonotonic. Indeed, we have* $(D, Z) = (1, 0)$ *if* $U = 2.5$*, but* $(D, Z) = (0, 1)$ *if* $U = 1.5$*. This implies that*

$$[D(2.5) - D(1.5)]\,[Z(2.5) - Z(1.5)] = -1 < 0.$$

□

Example 4. *MTS and CoMIV fail, but binarized MTS-MIV holds.*

$$\begin{cases} Y &= \Phi\left(-\left(\alpha D + U - \frac{3}{2}\right)^2 + U\right) \\ D &= \mathbb{1}\{\delta Z > 1\} + \varepsilon \\ Z &= \sum_{k=1}^{K+1} (k-1)\mathbb{1}\{\frac{k-1}{K+1} < \Phi(V) \leq \frac{k}{K+1}\} \\ \varepsilon &= \sum_{k=1}^{K} (k-1)\mathbb{1}\{\frac{k-1}{K} < \Phi(U) \leq \frac{k}{K}\} \end{cases} \tag{D.3}$$

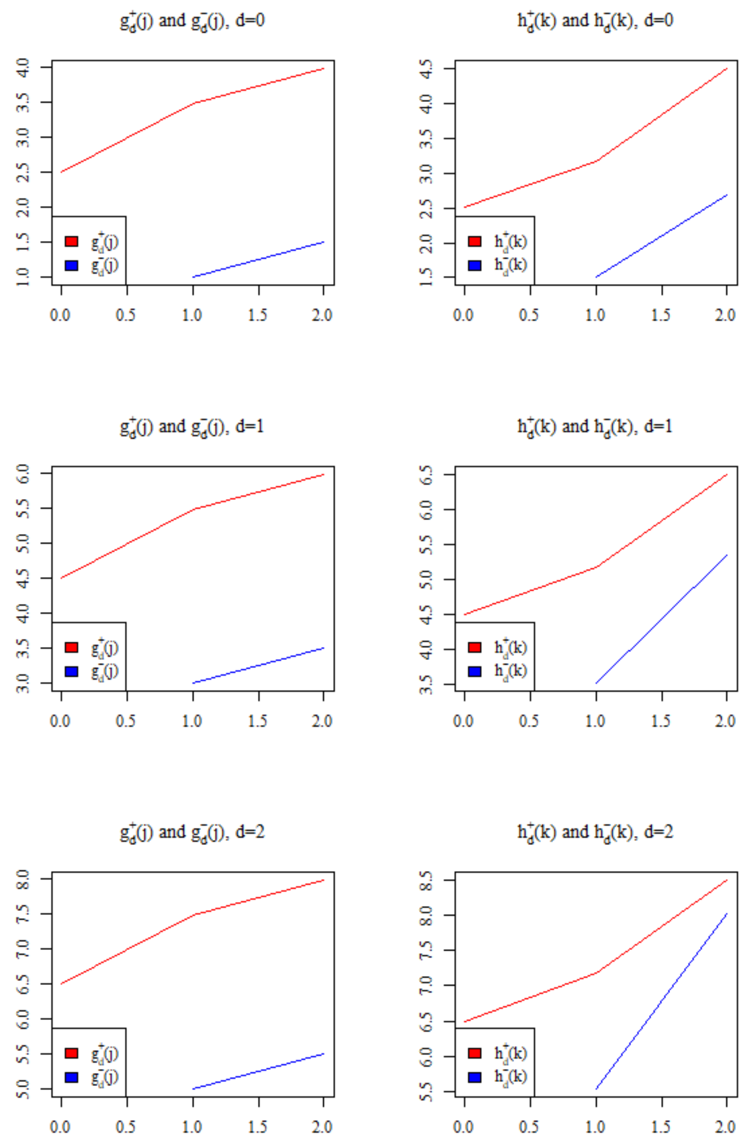

FIGURE 6. Numerical illustration of binarized MTS-MIV

where $\alpha = 0.5$, $K = 3$, $\delta \sim \mathcal{U}_{[-1,1]}$ *and* $\begin{pmatrix} U \\ V \end{pmatrix} \sim \mathcal{N}(\mu, \Sigma)$, *with* $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, *and* $\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$.

This example illustrates a case where MTS and CoMIV fail but binarized MTS-MIV holds, which provides the following implications. First, together with Lemma 1, binarized

MTS-MIV is a strictly weaker assumption than MTS-MIV (i.e., the converse of Lemma 1 does not hold). Second, binarized MTS-MIV is not a sufficient condition for CoMIV.

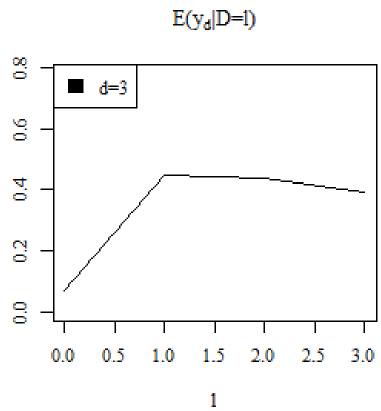

FIGURE 7. Numerical illustration of a violation of MTS

As shown in Figure 7, the conditional expectation function $\mathbb{E}\left[Y_d|D=\ell\right]$ *of the given DGP is not monotone in* ℓ *for* $d=3$. *Moreover,* D *and* Z *are not comonotonic because we have* $(D,Z)=(1,0)$ *if* $(U,V,\delta)=(0,-0.9,-0.7)$, *but* $(D,Z)=(0,1)$ *if* $(U,V,\delta)=(-0.8,-0.1,-0.5)$. *Indeed, we have:*

$$[D(0,-0.9,-0.7)-D(-0.8,-0.1,-0.5)]\,[Z(0,-0.9,-0.7)-Z(-0.8,-0.1,-0.5)]=-1<0.$$

On the other hand, Figure 8 shows that the same DGP satisfies binarized MTS-MIV, where the functions g_d^+ *and* g_d^- *do not overlap, nor do the functions* h_d^+ *and* h_d^- *for all* d.

□

Example 5. *CoMIV holds, but neither of MTS-MIV and binarized MTS-MIV holds.*

$$\left\{\begin{array}{lcl} Y &=& (2D+U-5)^2 \\ D &=& 0\cdot\mathbb{1}\left\{V\in[0,1]\right\}+1\cdot\mathbb{1}\left\{V\in(1,\tfrac{3}{2}]\right\}+2\cdot\mathbb{1}\left\{V\in(\tfrac{3}{2},5]\right\} \\ Z &=& 2D \\ U &=& 8V\mathbb{1}\left\{V\in[1,2]\right\}+V\mathbb{1}\left\{V\notin[1,2]\right\} \end{array}\right. \tag{D.4}$$

where $V\sim\mathcal{U}_{[0,5]}$.

This example illustrates a case where CoMIV holds, but neither of MTS-MIV and binarized MTS-MIV holds. Hence, we note that CoMIV is not a sufficient condition for binarized MTS-MIV. Together with Examples 1 and 4, it should be noted that neither of CoMIV and

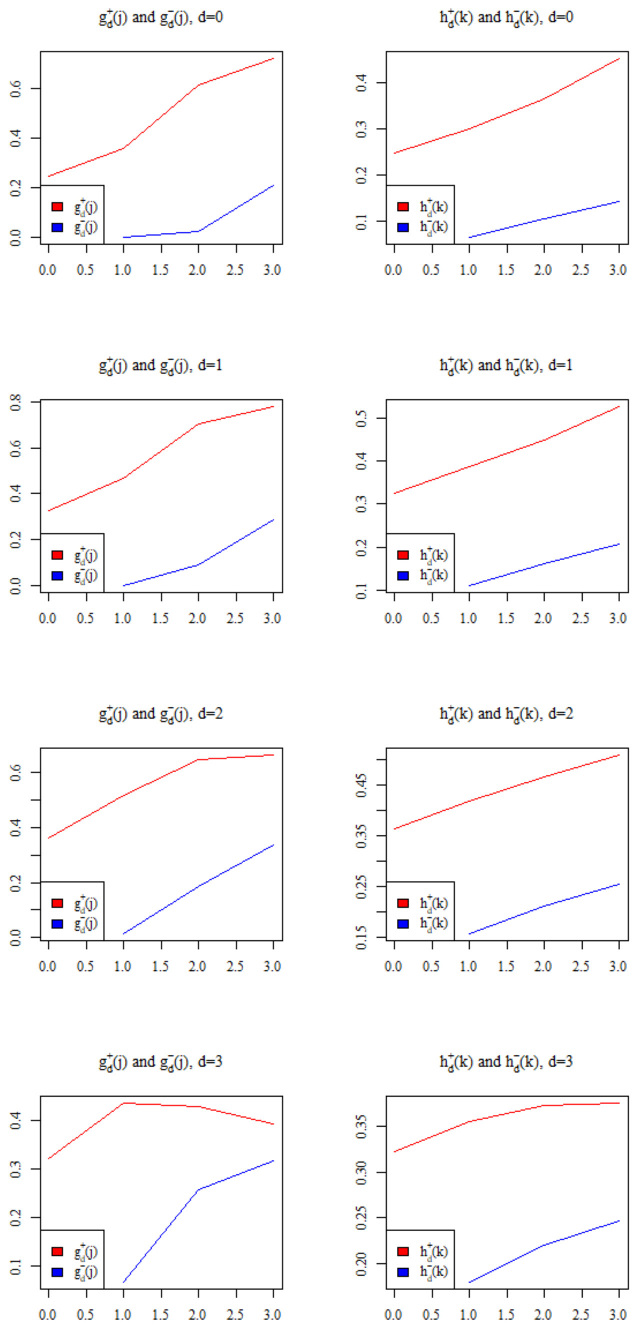

FIGURE 8. Numerical illustration of binarized MTS-MIV

binarized MTS-MIV implies each other, even though both are sufficient conditions for SDC assumption by Lemmas 2 and 4, respectively.

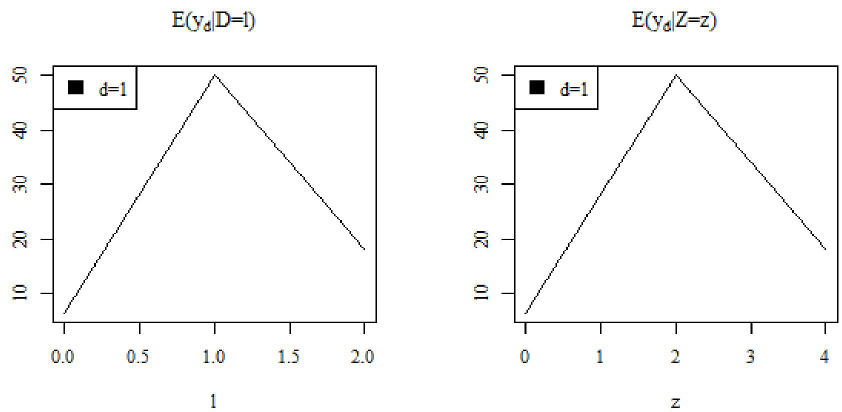

FIGURE 9. Numerical illustration of a Violation of MTS-MIV

Under the given DGP, Z is a CoMIV for Z by (1) of Lemma 3 since $Z = 2D$. However, it can be shown that the conditional expectation functions $\mathbb{E}\left[Y_d|D=\ell\right]$ and $\mathbb{E}\left[Y_d|Z=z\right]$ are not monotone in ℓ and z, respectively, for all $d = 0, 1, 2$ (see Figure 9 for $d = 1$). Furthermore,

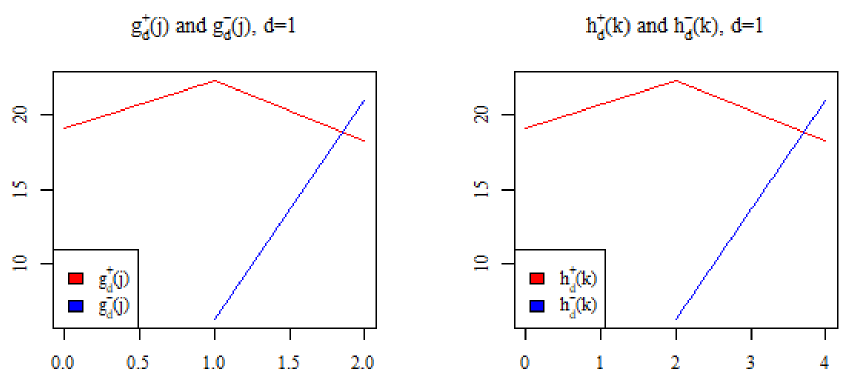

FIGURE 10. Numerical illustration of a Violation of Binarized MTS-MIV

Figure 10 shows that the DGP does not satisfy the binarized MTS-MIV restriction (2.2) either, as the function g_1^+ crosses the function g_1^-, and the function h_1^+ crosses the function h_1^-. $\square$

References

Andrews, D. W. K., W. Kim, and X. Shi. 2017. "Stata Commands for Testing Conditional Moment Inequalities/Equalities." Stata Journal 17 (1):56–72.

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

Berger, R. L. 1982. "Multiparameter Hypothesis Testing and Acceptance Sampling." Technometrics 24:295–300.

Berger, R. L. and J. C. Hsu. 1996. "Bioequivalence Trials, Intersection-Union Tests and Equivalence Confidence Sets." Statistical Science 11 (4):283–319.

Card, David. 1995. "Using Geographic Variation in College Proximity to Estimate the Return to Schooling." in Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp, ed, by Louis N. Christofides, E. Kenneth Grant, and Robert Swidinsky. Toronto: University of Toronto Press :201–222.

Chernozhukov, Victor, Wooyoung Kim, Sokbae Lee, and Adam M. Rosen. 2015. "Implementing Intersection Bounds in Stata." Stata Journal 15 (1):21–44.

Chernozhukov, Victor, Sokbae Lee, and Adam M. Rosen. 2013. "Intersection Bounds: Estimation and Inference." Econometrica 81 (2):667–737. URL `http://dx.doi.org/10.3982/ECTA8718`.

Conley, Timothy G., Christian B. Hansen, and Peter E. Rossi. 2012. "Plausibly Exogenous." The Review of Economics and Statistics 94 (1):260–272.

Ginther, D. K. 2000. "Alternative Estimates of the Effect of Schooling on Earnings." The Review of Economics and Statistics 82 (1):103–116.

Hotz, V. Joseph, Charles Mullin, and Seth Sanders. 1997. "Bounding Causal Effects Using Data from a Contaminated Natural Experiment: Analyzing the Effects of Teenage Childbearing." Review of Economic Studies 64 (4):575–603.

Kédagni, D., L. Li, and I. Mourifié. 2018. "Bounding Average Returns to Schooling using Unconditional Moment Restrictions." Economics Working Papers: Department of Economics, Iowa State University 18022.

Kédagni, D. and I. Mourifié. 2020. "Generalized Instrumental Inequalities: Testing the Instrumental Variable Independence Assumption." Biometrika 107 (3):661–675.

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. 1994. "The Selection Problem." in Advances Economics, Sixth World Congress, C. Sims (ed.), Cambridge University Press 1:143–170.

Manski, C. F. and J. Pepper. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." Econometrica 68:997–1010.

———. 2009. "More on Monotone Instrumental Variables." Econometrics Journal 12:S200–S216.

Manski, Charles 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.

———. 1997. "Monotone Treatment Response." Econometrica 65 (6):1311–1334.

Masten, M. A. and A. Poirier. 2020. "Salvaging Falsified Instrumental Variable Models." Working Paper .

Mourifié, I., M. Henry, and R. Méango. 2020. "Sharp Bounds and Testability of a Roy Model of STEM Major Choices." Journal of Political Economy 8 (128):3220–3283.

Nevo, Aviv and Adam Rosen. 2012. "Identification with Imperfect Instruments." The Review of Economics and Statistics 94 (3):659–671.

Roy, A. D. 1951. "Some Thoughts on the Distribution of Earnings." Oxford Economic Papers, New Series 3 (2):135–146.

Wang, R. and R. Zitikis. 2020. "Weak comonotonicity." European Journal of Operational Research 1 (282):386–397.

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