Journal of Econometrics 250 (2025) 106009

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Pairwise valid instruments

Zhenting Sun [a,*], Kaspar Wüthrich [b]

[a] Department of Economics, University of Melbourne, Australia
[b] Department of Economics, University of Michigan, United States

ARTICLE INFO

Keywords:
Invalid instruments
Local average treatment effects
Identification
Instrumental variable estimation
Asymptotic bias reduction

ABSTRACT

Finding valid instruments is difficult. We propose Validity Set Instrumental Variable (VSIV) estimation, a method for estimating local average treatment effects (LATEs) in heterogeneous causal effect models when the instruments are partially invalid. We consider settings with pairwise valid instruments, that is, instruments that are valid for a subset of instrument value pairs. VSIV estimation exploits testable implications of instrument validity to remove invalid pairs and provides estimates of the LATEs for all remaining pairs, which can be aggregated into a single parameter of interest using researcher-specified weights. We show that the proposed VSIV estimators are asymptotically normal under weak conditions and remove or reduce the asymptotic bias relative to standard LATE estimators (that is, LATE estimators that do not use testable implications to remove invalid variation). We evaluate the finite sample properties of VSIV estimation in application-based simulations and apply our method to estimate the returns to college education using parental education as an instrument.

1. Introduction

Instrumental variable (IV) methods based on the local average treatment effect (LATE) framework (Imbens and Angrist, 1994; Angrist and Imbens, 1995; Angrist et al., 1996) rely on three assumptions:[1] (i) *exclusion* (the instrument does not have a direct effect on the outcome), (ii) *random assignment* (the instrument is independent of potential outcomes and treatments), and (iii) *monotonicity* (the instrument has a monotonic impact on treatment take-up).[2] In many applications, some of these assumptions are likely to be violated or at least questionable. This has motivated the derivation of testable restrictions and tests for IV validity in various settings (e.g., Balke and Pearl, 1997; Imbens and Rubin, 1997; Heckman and Vytlacil, 2005; Huber and Mellace, 2015; Kitagawa, 2015; Mourifié and Wan, 2017; Kédagni and Mourifié, 2020; Carr and Kitagawa, 2021; Farbmacher et al., 2022; Frandsen et al., 2023; Jiang and Sun, 2023; Sun, 2023).[3] The main contribution of this paper is to propose a method for exploiting the information available in the testable restrictions of IV validity to remove or reduce the asymptotic bias when estimating LATE parameters.[4]

We consider settings where the available instruments are partially invalid. A leading example of such a setting is when there is a multivalued instrument for which only some pairs of instrument values satisfy the IV assumptions. In Section 5.3, we revisit the analysis of the causal effect of college education on earnings using parental education as an instrument. This instrument is likely

* Correspondence to: Department of Economics, University of Melbourne, Grattan Street, Parkville Victoria 3010, Australia.
E-mail addresses: zhentingsun@gmail.com (Z. Sun), kasparwu@umich.edu (K. Wüthrich).

1 See, for example, Imbens (2014), Melly and Wüthrich (2017), and Huber and Wüthrich (2019) for recent reviews, and Angrist and Pischke (2009, 2014), and Imbens and Rubin (2015) for textbook treatments.

2 Some papers also include the instrument first-stage assumption as part of the LATE assumptions. We will maintain suitable first-stage assumptions.

3 There is a related literature on inference with invalid instruments in linear IV models (e.g., Conley et al., 2012; Nevo and Rosen, 2012; Armstrong and Kolesár, 2021; Goh and Yu, 2022).

4 We define the asymptotic bias as the probability limit of the ℓ^2 difference between an estimator and the true value.

https://doi.org/10.1016/j.jeconom.2025.106009
Received 21 July 2022; Received in revised form 11 October 2024; Accepted 4 November 2024
Available online 7 May 2025
0304-4076/

partially invalid due to parental education having a positive effect on future earnings, at least up to a certain education level (Li et al., 2024). Another example is the quarter of birth (QOB) instrument of Angrist and Krueger (1991). A potential concern with this instrument is that the seasonality in birth patterns renders the QOB instrument partially invalid (e.g., Bound et al., 1995; Buckles and Hungerman, 2013), motivating some studies to only consider a subset of QOBs as instruments (e.g., Dahl et al., 2023). Another empirically relevant setting where partially invalid instruments may arise is when there are multiple instruments.[5] In applications with multiple instruments, the validity of a subset of the instruments may be questionable, or the instruments may be partially invalid because the heterogeneity in individual choice behavior renders standard monotonicity assumptions invalid (Mogstad et al., 2021). As an example of the latter, consider the study by Thornton (2008), who estimates the causal effect of knowing HIV status on the likelihood of buying condoms using two randomly assigned instruments: Monetary incentives and distance to results centers. In this application, monotonicity is likely to fail due to differences in individual preferences over monetary incentives and distance (Mogstad et al., 2021, Online Appendix B.1).[6]

The proposed method, which we refer to as *Validity Set IV (VSIV) estimation*, uses testable implications of IV validity to remove invalid variation in the instruments and provides LATE estimates based on the remaining variation in the instruments. We establish the asymptotic normality of the proposed VSIV estimators and show that they always remove or reduce the asymptotic bias relative to standard LATE estimators, that is, LATE estimators that do not exploit testable implications of IV validity to remove invalid pairs. Thus, VSIV estimation constitutes a data-driven approach for removing or reducing the asymptotic bias of LATE estimators, given all the information about IV validity available in the data.

The use of the testable implications of IV validity in VSIV estimation is more constructive than the standard practice where researchers first test for IV validity, discard the instruments if they reject IV validity, and proceed with standard IV analyses if they do not reject IV validity. VSIV estimation uses the testable implications to remove invalid information in the instruments. Consequently, it can be used to estimate causal effects in settings where the instruments are only partially invalid so that existing tests reject the null of full IV validity.[7] VSIV estimation salvages falsified instruments by exploiting the variation in the instruments not refuted by the data and thereby contributes to the literature on salvaging falsified models (e.g., Masten and Poirier, 2021; Li et al., 2024).

Our goal is to estimate the causal effect of an endogenous treatment D on an outcome of interest Y, using a potentially vector-valued discrete instrument Z. We consider binary treatments in the main text and multivalued ordered or unordered treatments in the Appendix. In the ideal case, Z is fully valid, that is, the LATE assumptions hold for all instrument values (the instrument is valid for the whole population). However, full IV validity is questionable in many applications, especially when there are many instruments or instrument values. To this end, we introduce the notion of *pairwise valid instruments*. Pairwise valid instruments are only valid for a subset of all pairs of instrument values, which we refer to as *validity pair set*. Intuitively, the instruments are valid for some subpopulations but invalid for the others.

Pairwise validity separates the instrument value pairs into two groups: Valid pairs for which all LATE assumptions hold and invalid pairs for which at least one of the LATE assumptions is violated. Pairwise validity does not require researchers to specify which LATE assumptions are violated for the invalid pairs and to which extent; it allows for failures of exclusion, random assignment, monotonicity, or combinations thereof. As a result, there is no information about the LATE for the invalid instrument value pairs absent additional restrictions (see Appendix B). Pairwise validity is motivated by the fact that it is often difficult to determine why exactly specific instrument pairs are invalid based on contextual knowledge (that is, which combinations of assumptions are violated and how), especially when there are many potentially invalid pairs. If additional information is available on which LATE assumptions fail and how, we can exploit such information for partial identification and sensitivity analysis (e.g., Huber, 2014; Noack, 2021; Kédagni, 2023; Cui et al., 2024) or focus on target parameters that are identified under relaxations of the LATE assumptions (e.g., De Chaisemartin, 2017; Frandsen et al., 2023). Even in settings where such information is available, pairwise validity provides a useful benchmark and starting point.

VSIV estimation provides estimates of the LATEs for all pairs of instrument values that satisfy the testable restrictions for IV validity. Specifically, we obtain an estimator, $\widehat{\mathscr{Z}}_0$, of the set of pairs of instrument values that satisfy the testable restrictions in Kitagawa (2015), Mourifié and Wan (2017), Kédagni and Mourifié (2020), and Sun (2023) and estimate LATEs for all pairs of instrument values in $\widehat{\mathscr{Z}}_0$. These LATEs can then be aggregated into a single parameter of interest based on user-specified weights.

We study the theoretical properties of VSIV estimation under two scenarios. First, we assume that the estimated validity pair set, $\widehat{\mathscr{Z}}_0$, is consistent for the largest validity pair set $\mathscr{Z}_{\bar{M}}$ (that is, the union of all validity pair sets) in the sense that $\mathbb{P}(\widehat{\mathscr{Z}}_0 = \mathscr{Z}_{\bar{M}}) \to 1$. In this case, VSIV estimation is asymptotically unbiased and normal under standard conditions. Second, since the estimator of the validity pair set, $\widehat{\mathscr{Z}}_0$, is typically constructed based on necessary (but not necessarily sufficient) conditions for IV validity, it could converge to a *pseudo-validity pair set* $\mathscr{Z}_0$ that is larger than $\mathscr{Z}_{\bar{M}}$, that is, $\mathbb{P}(\widehat{\mathscr{Z}}_0 = \mathscr{Z}_0) \to 1$.[8] Let $\mathscr{Z}_P$ be a presumed set of valid pairs of instrument values, incorporating prior information about instrument validity ($\mathscr{Z}_P$ is equal to the set of all pairs if no such prior information is available). We prove that VSIV estimation based on $\widehat{\mathscr{Z}}_0 \cap \mathscr{Z}_P$ leads to a smaller asymptotic bias than standard LATE

5 Settings with multiple instruments are common in empirical research (Mogstad et al., 2021, Section I).

6 Mogstad et al. (2021) propose a weaker version of monotonicity, referred to as partial monotonicity, that they argue is more plausible in this application. We discuss the connection between our assumptions and partial monotonicity in Section 2.2. Jiang and Sun (2023) develop formal tests for partial monotonicity and apply these tests in the context of the Thornton (2008) application.

7 See Appendix D.7 for a comparison between VSIV estimation and pairwise pretests based on existing tests for IV validity.

8 Kitagawa (2015, Proposition 1.1) shows that there exist no sufficient conditions for IV validity when D and Z are both binary.

estimators based on $\mathscr{Z}_P$. Taken together, our theoretical results show that, irrespective of whether the largest validity pair set can be estimated consistently or not, VSIV estimation leads to asymptotically normal LATE estimators with reduced asymptotic bias.

Finally, we use VSIV estimation to revisit the estimation of the causal effect of college education on earnings using parental education as an instrument. We evaluate the finite sample performance of VSIV estimation in a simulation study calibrated to this application and use these simulations to determine the choice of the tuning parameter required for VSIV estimation. Based on this choice of tuning parameter, VSIV estimation screens out the pairs of instrument values corresponding to low levels of parental education. This is consistent with the above discussion and the findings in Li et al. (2024) in that for low levels of parental education the exclusion restriction may fail. The LATEs for the pairs of instrument values that are not screened out are positive and significant.

Notation: We introduce some standard notation, following Sun (2023). All random elements are defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. For all $m \in \mathbb{N}$, $\mathcal{B}_{\mathbb{R}^m}$ is the Borel σ-algebra on $\mathbb{R}^m$. We denote by $\mathcal{P}$ the set of probability measures such that if the data $\{(Y_i, D_i, Z_i)\}_{i=1}^n$ are i.i.d. and distributed according to some probability measure $Q \in \mathcal{P}$, then $Q(G) = \mathbb{P}((Y_i, D_i, Z_i) \in G)$ for all measurable sets G. For every $Q \in \mathcal{P}$ and every measurable function v, with some abuse of notation, we define $Q(v) = \int v \, dQ$. The symbol $\rightsquigarrow$ denotes weak convergence in a metric space in the Hoffmann–Jørgensen sense. For every set B, let 1_B denote the indicator function for B. Finally, to simplify the exposition of the theoretical results, we adopt the convention (e.g., Folland, 1999, p. 45), that

$$0 \cdot \infty = 0. \tag{1.1}$$

2. Identification with pairwise valid instruments

2.1. Weakening instrument validity to pairwise validity

Consider a setting with an outcome variable $Y \in \mathbb{R}$, a treatment $D \in \mathcal{D}$, and an instrument (vector) $Z \in \mathcal{Z}$. In the main text, we focus on the leading case where the treatment is binary with $\mathcal{D} = \{0, 1\}$. The extensions to multivalued ordered and unordered treatments can be found in the Appendix. The instrument is discrete with $\mathcal{Z} = \{z_1, \dots, z_K\}$, and can be ordered or unordered. Let $Y_{dz} \in \mathbb{R}$ for $(d, z) \in \mathcal{D} \times \mathcal{Z}$ denote the potential outcomes and let D_z for $z \in \mathcal{Z}$ denote the potential treatments. The following assumption generalizes the standard LATE assumptions with binary instruments to multivalued instruments.

Assumption 2.1. IV validity for LATEs with binary treatments and multivalued instruments:

(i) Exclusion: For each $d \in \{0, 1\}$, $Y_{dz_1} = \cdots = Y_{dz_K}$ almost surely (a.s.).
(ii) Random Assignment: Z is jointly independent of $(Y_{0z_1}, \dots, Y_{0z_K}, Y_{1z_1}, \dots, Y_{1z_K})$ and $(D_{z_1}, \dots, D_{z_K})$.
(iii) Monotonicity: For all $k \in \{1, \dots, K-1\}$, $D_{z_{k+1}} \geq D_{z_k}$ a.s.

Assumption 2.1 is similar to the LATE assumptions in, for example, Imbens and Angrist (1994), Angrist and Imbens (1995), Frölich (2007), Kitagawa (2015), and Sun (2023). It imposes the IV validity assumptions with respect to all possible values of the instrument $z \in \mathcal{Z}$. This assumption has a lot of identifying power: It identifies LATEs with respect to every pair of IV values (z_k, z_{k+1}) with $\mathbb{P}(D_{z_{k+1}} > D_{z_k}) > 0$. However, Assumption 2.1 can be restrictive in applications. We therefore introduce the notion of ***pairwise instrument validity***, which weakens the conditions in Assumption 2.1. Define the set of all possible pairs of values of Z as

$$\mathscr{Z} = \left\{ \left(z_1, z_2\right), \dots, \left(z_1, z_K\right), \left(z_2, z_3\right), \dots, \left(z_2, z_K\right), \dots, \left(z_{K-1}, z_K\right), \left(z_2, z_1\right), \dots, \left(z_K, z_{K-1}\right) \right\}.$$

The number of the elements in $\mathscr{Z}$ is $K \cdot (K-1)$. We use $\mathcal{Z}_{(k,k')}$ to denote a pair $(z_k, z_{k'}) \in \mathscr{Z}$. Note that we include both $(z_k, z_{k'})$ and $(z_{k'}, z_k)$ in $\mathscr{Z}$ so that we do not restrict the direction of the monotonicity (Assumption 2.1(iii)) a priori.

Definition 2.1. The instrument Z is **pairwise valid** for the treatment $D \in \{0, 1\}$ if there is a set $\mathscr{Z}_M = \{(z_{k_1}, z_{k'_1}), \dots, (z_{k_M}, z_{k'_M})\} \subseteq \mathscr{Z}$ such that the following conditions hold for every $(z, z') \in \mathscr{Z}_M$:

(i) Exclusion: For each $d \in \{0, 1\}$, $Y_{dz} = Y_{dz'}$ a.s.
(ii) Random Assignment: Z is jointly independent of $(Y_{0z}, Y_{0z'}, Y_{1z}, Y_{1z'}, D_z, D_{z'})$.[9]
(iii) Monotonicity: $D_{z'} \geq D_z$ a.s.

The set $\mathscr{Z}_M$ is called a **validity pair set** of Z.[10] The union of all validity pair sets is the largest validity pair set, denoted by $\mathscr{Z}_{\bar{M}}$. A pair of instrument values (z, z') is called a **valid pair** if $(z, z') \in \mathscr{Z}_{\bar{M}}$. A pair (z, z') is called an **invalid pair** if $(z, z') \notin \mathscr{Z}_{\bar{M}}$.

[9] This condition can be further weakened: The conditional distribution of $(Y_{0z}, Y_{0z'}, Y_{1z}, Y_{1z'}, D_z, D_{z'})$ given $Z = z$ or $Z = z'$ is the same as the unconditional distribution.

[10] We use $\mathscr{Z}_M$ to denote an arbitrary validity pair set throughout the paper. To simplify the notation, we therefore only index $\mathscr{Z}$ by M and not by the full index set $\{(k_1, k'_1), \dots, (k_M, k'_M)\}$.

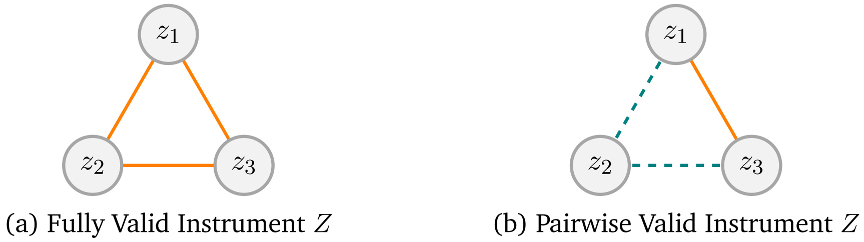

Fig. 2.1. Full IV Validity vs. Pairwise IV Validity.

Definition 2.1 separates the instrument value pairs into two groups: Valid pairs for which all the LATE assumptions hold and invalid pairs for which the LATE assumptions fail due to failures of exclusion, independence, monotonicity, or combinations thereof. We show in Appendix B that absent additional restrictions on how Definition 2.1 can be violated, the sharp identified set for the LATEs for the invalid pairs is the entire real line; that is, there is no information about the LATEs for the invalid pairs in the data.

Pairwise validity does not require researchers to specify which LATE assumptions are violated and to which extent for the invalid pairs. The motivation for this is that it can be difficult to determine why exactly instrument pairs are invalid in applications. While pairwise validity does not require researchers to impose additional assumptions for the invalid pairs, it allows for the possibility that valid pairs restrict which LATE assumptions are violated for invalid pairs.[11]

Definition 2.1 complements the existing approaches for relaxing the LATE assumptions. These approaches typically impose more structure on which LATE assumptions are violated and how exactly the LATE assumptions are violated. They provide partial identification results and methods for performing sensitivity analyses (e.g., Huber, 2014; Noack, 2021; Kédagni, 2023; Cui et al., 2024) or focus on target parameters that are identified under weaker assumptions (e.g., De Chaisemartin, 2017; Frandsen et al., 2023).

To illustrate Definition 2.1, consider a simple example where $Z \in \mathcal{Z} = \{z_1, z_2, z_3\}$. If Z is fully valid as in Assumption 2.1 such that $D_{z_3} \geq D_{z_2} \geq D_{z_1}$ a.s., then $\mathscr{Z}_{\bar{M}} = \{(z_1, z_2), (z_1, z_3), (z_2, z_3)\}$. The orange solid lines in Fig. 2.1(a) indicate that two instrument values, $\{z_k, z_{k'}\}$, form a valid pair: Either $(z_k, z_{k'})$ or $(z_{k'}, z_k)$ satisfies the conditions in Definition 2.1. The full validity Assumption 2.1 requires that every pair of instrument values forms a valid pair. Definition 2.1 relaxes Assumption 2.1 as it does not require every pair to form a valid pair. For example, it could be that only (z_1, z_3) satisfies the conditions in Definition 2.1. The teal dashed lines in Fig. 2.1(b) indicate that $\{z_1, z_2\}$ and $\{z_2, z_3\}$ do not form valid pairs. In this case, the instrument Z is pairwise but not fully valid.

In applications where the instrument Z is randomly assigned (e.g., in experiments with imperfect compliance), joint independence (Assumption 2.1(ii)) holds by design. In such applications, Definition 2.1 captures violations of exclusion and monotonicity. Such violations are easy to interpret. The pairwise exclusion assumption in Definition 2.1(i) requires that Y_{dz}, viewed as a function of z, is constant over some regions of $\mathcal{Z}$ and varies over others. This nests, for example, Condition E.3 in Li et al. (2024), which requires that $Y_{dt} \leq Y_{dt'}$ for all $t \leq t'$ and $Y_{dt} = Y_{dz}$ for all $t \geq z$. The pairwise monotonicity assumption (Definition 2.1(iii)) requires that D_z, viewed as a function of z, is monotonic over some regions of $\mathcal{Z}$ and non-monotonic over others. We discuss the relationship to existing relaxations of LATE monotonicity in more detail in Section 2.2.

In many quasi-experimental applications, the instrument Z is not randomly assigned, and joint independence (Assumption 2.1(ii)) may fail. A leading and practically relevant case where joint independence fails but pairwise independence (Definition 2.1(ii)) holds is when there are multiple instruments, and some of them are not independent of all potential variables.[12] To illustrate, consider the following example based on Mogstad et al. (2021, Section II.C) and the empirical application in Carneiro et al. (2011). Let D be an indicator for college attendance. There are two binary instruments, $Z = (Z_1, Z_2)$, where Z_1 is an indicator for college proximity (e.g., Card, 1993; Kane and Rouse, 1993) and Z_2 is an indicator for tuition subsidy.[13] Individuals decide whether to attend college based on the following selection mechanism,

$$D_z = 1\left\{B_0 + B_1 z_1 + z_2 \geq 0\right\}, \tag{2.1}$$

where B_0 and B_1 are random coefficients. For simplicity, we assume that $Z \perp\!\!\!\perp B_0$. The coefficient B_1 measures the "taste" for proximity relative to tuition subsidy. If the taste for proximity, B_1, is correlated with actual proximity, Z_1, for example, due to spatial sorting, then the pair $(D_{(0,0)}, D_{(0,1)})$ is independent of Z but the pair $(D_{(1,0)}, D_{(1,1)})$ is not.[14]

---

[11] For example, suppose that $\mathcal{Z} = \{z_1, z_2, z_3\}$, where the pairs (z_1, z_3) and (z_2, z_3) are valid. This configuration implies that exclusion (Definition 2.1(i)) cannot be violated for the pair (z_1, z_2).

[12] It is possible that joint independence fails but pairwise independence holds even if there is only one original instrument. To illustrate, let D indicate college enrollment, and let $Z \in \{1, 2, 3\}$ measure distance to the closest college (e.g., Kane and Rouse, 1993), where $Z = 1$ indicates `close`, $Z = 2$ indicates `far`, and $Z = 3$ indicates `very far`. Consider the selection mechanism $D_z = 1\{B_0 + B_1 1\{z = 1\} + f(z) 1\{z > 1\} \leq 0\}$, where B_0 and B_1 are random coefficients, $f(2) < f(3)$, and $B_0 \perp\!\!\!\perp Z$. The coefficient B_1 captures the taste for living `close` to college relative to living farther away. If B_1 is correlated with actual distance Z, then $Z \not\perp\!\!\!\perp D_1$ and $Z \perp\!\!\!\perp (D_2, D_3)$.

[13] We swap the order of the instruments relative to Mogstad et al. (2021, Section II.C) for the purpose of illustration.

[14] See, for example, Card (1993), Carneiro et al. (2011), Slichter (2014), and Kitagawa (2015) for discussions of the validity of the college proximity instrument.

Remark 2.1 (*Weakening Definition 2.1 with Multiple Instruments*). In Appendix D.2, we introduce a weaker notion of pairwise validity (Definition 2.1) for settings where Z contains multiple instruments: $Z = (Z_1, \ldots, Z_L)^T$, where Z_l is a scalar instrument for $l \in \{1, \ldots, L\}$.

2.2. Relationship to other variants and relaxations of monotonicity

Here we discuss the connection between our pairwise monotonicity assumption (Definition 2.1(iii)) and three recently proposed variants and relaxations of the LATE monotonicity assumption.

First, Mogstad et al. (2021) propose a partial monotonicity (PM) condition for settings with multiple instruments, which is a special case of Condition (iii) in Definition 2.1; see also Goff (2024) for vector monotonicity assumption.[15] For example, suppose that $Z = (Z_1, Z_2) \in \mathbb{R}^2$ and each element of Z is binary so that $\mathcal{Z} = \{(0,0),(0,1),(1,0),(1,1)\}$. Suppose that Assumption PM of Mogstad et al. (2021) holds with $D_{(0,0)} \geq D_{(0,1)}$, $D_{(0,0)} \geq D_{(1,0)}$, $D_{(1,1)} \geq D_{(0,1)}$, and $D_{(1,1)} \geq D_{(1,0)}$ a.s. (the sex composition instrument in Angrist and Evans (1998) discussed in Mogstad et al. (2021)), and that Conditions (i) and (ii) of Definition 2.1 hold. Then a validity pair set is

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

Second, Frandsen et al. (2023, Section IV) study the interpretation of 2SLS under relaxations of monotonicity and exclusion. The relaxation of monotonicity, referred to as average monotonicity, requires D_z to be positively correlated with the instrument propensity. Average monotonicity is fundamentally different from pairwise monotonicity (Definition 2.1(iii)). Pairwise monotonicity operates at the level of pairs of instrument values, whereas average monotonicity implies restrictions across all instrument values. Also, Frandsen et al. (2023) show that average monotonicity can be used to identify averages of treatment effects. By contrast, pairwise validity identifies the LATE for $(z, z') \in \mathscr{Z}_{\bar{M}}$, but does not identify the LATE for $(z, z') \notin \mathscr{Z}_{\bar{M}}$ (Corollary B.1 in Appendix B).

Finally, Noack (2021) considers a continuous relaxation of monotonicity when Z is binary, parameterized by the fraction of defiers. We do not consider continuous relaxations and make no assumptions on the degree of violation. Combining VSIV estimation with continuous relaxations as in Noack (2021) is an interesting direction for future research, as we discuss in Section 6.

2.3. Identification under pairwise validity

The following lemma establishes identification under pairwise validity.

Lemma 2.1. *Suppose that the instrument Z is pairwise valid according to Definition 2.1 with a known validity pair set $\mathscr{Z}_M = \{(z_{k_1}, z_{k'_1}), \ldots, (z_{k_M}, z_{k'_M})\}$.*[16] *Then we can define a random variable $Y_d(z_{k_m}, z_{k'_m}) = Y_{dz_{k_m}} = Y_{dz_{k'_m}}$ a.s. for each $d \in \{0,1\}$ and every $(z_{k_m}, z_{k'_m}) \in \mathscr{Z}_M$, and the following quantity can be identified for every $(z_{k_m}, z_{k'_m}) \in \mathscr{Z}_M$:*

$$\beta_{k'_m,k_m} \equiv E\left[Y_1(z_{k_m}, z_{k'_m}) - Y_0(z_{k_m}, z_{k'_m}) \middle| D_{z_{k'_m}} > D_{z_{k_m}}\right] = \frac{E\left[Y|Z = z_{k'_m}\right] - E\left[Y|Z = z_{k_m}\right]}{E\left[D|Z = z_{k'_m}\right] - E\left[D|Z = z_{k_m}\right]}. \tag{2.2}$$

Lemma 2.1 is a direct extension of Theorem 1 of Imbens and Angrist (1994) for the case where Z is pairwise valid. We follow Imbens and Angrist (1994) and refer to $\beta_{k'_m,k_m}$ as a LATE. Lemma 2.1 shows that if a validity pair set $\mathscr{Z}_M$ is known, we can identify every $\beta_{k'_m,k_m}$ with $(z_{k_m}, z_{k'_m}) \in \mathscr{Z}_M$.[17] In practice, however, $\mathscr{Z}_M$ is usually unknown. In this paper, we use testable implications of IV validity to estimate a pseudo-validity pair set $\mathscr{Z}_0$ containing $\mathscr{Z}_M$, and show how to use this estimated set to reduce the asymptotic bias in LATE estimation.

We focus on the vector of LATEs $\{\beta_{k'_m,k_m}\}$ as our object of interest. Traditional IV estimators estimate weighted averages of LATEs (e.g., Imbens and Angrist, 1994) and, thus, are strictly less informative (we can always compute linear IV estimands based on the LATEs). Moreover, VSIV estimation estimates LATEs that do not enter such weighted averages (Theorem 2 of Imbens and Angrist (1994)). To illustrate, suppose $\mathcal{Z} = \{z_1, z_2, z_3\}$ and $\mathscr{Z}_{\bar{M}} = \{(z_1, z_2), (z_1, z_3), (z_2, z_3)\}$. The traditional IV estimator estimates a weighted average of $\beta_{2,1}$ and $\beta_{3,2}$, whereas our method estimates $(\beta_{2,1}, \beta_{3,1}, \beta_{3,2})^T$.

Importantly, VSIV estimation allows for assigning researcher-specified weights to $\{\beta_{k'_m,k_m}\}$. In the traditional IV estimation, the weights are determined by the estimation procedure. Mogstad et al. (2021) show that the weights assigned to the LATEs by 2SLS could be negative under partial monotonicity. Negative weights are not an issue for VSIV estimation because the weights can be

[15] Mogstad et al. (2021) motivate the PM condition by showing that full monotonicity imposes strong restrictions on the heterogeneity in individual choice behavior and is therefore likely violated in many applications.

[16] Note that mathematically we do not need to impose a first-stage assumption here due to the convention (1.1).

[17] Note that if $(z_{k_m}, z_{k'_m}) \in \mathscr{Z}_M$ with $D_{z_{k_m}} = D_{z_{k'_m}}$ a.s., then $\beta_{k'_m,k_m} = 0$ by (1.1). Moreover, if $(z_{k_m}, z_{k'_m}) \in \mathscr{Z}_M$ and $(z_{k'_m}, z_{k_m}) \in \mathscr{Z}_M$, then by Definition 2.1, $D_{z_{k_m}} = D_{z_{k'_m}}$ a.s.

chosen by researchers, instead of being determined by the estimation procedure. For example, we may define the weighted average as

$$\beta_w = \frac{p_{12}}{p_{123}}\beta_{2,1} + \frac{p_{13}}{p_{123}}\beta_{3,1} + \frac{p_{23}}{p_{123}}\beta_{3,2}, \tag{2.3}$$

where $p_{ij} = \mathbb{P}(Z \in \{z_i, z_j\})$ and $p_{123} = \mathbb{P}(Z \in \{z_1, z_2\}) + \mathbb{P}(Z \in \{z_2, z_3\}) + \mathbb{P}(Z \in \{z_1, z_3\})$. The asymptotic properties of the estimated weighted averages of LATEs follow straightforwardly from the asymptotic theory in Section 3. See Corollary 3.1.

Remark 2.2 (*Extrapolation*)**.** The focus of VSIV estimation is on estimating the LATE parameters $\beta_{k',k}$ for all pairs of instrument values $(z_k, z_{k'})$ satisfying the testable restrictions of IV validity. This is because absent additional restrictions, there is no information in the data about the LATE for the invalid pairs (see Appendix B).

The local and DGP-dependent nature of LATE parameters has motivated the development of a variety of methods for assessing and restoring external validity (e.g., Heckman et al., 2003; Angrist and Fernández-Val, 2013; Brinch et al., 2017; Mogstad et al., 2018; Wüthrich, 2020; Kowalski, 2023). The use of invalid instrument pairs will result in these approaches being biased and inconsistent. VSIV estimation constitutes a natural complement to the existing approaches to external validity. For settings where researchers are interested in externally valid effects, we recommend a two-step procedure: (i) Use VSIV estimation to eliminate invalid pairs. (ii) Apply a suitable approach to external validity based on the estimated validity pair set. In step (i), we recommend also reporting the VSIV LATE estimates because they summarize the available information about pairwise LATEs and are important inputs for approaches to external validity.

3. Validity set IV estimation

3.1. Overview

The goal of VSIV estimation is to exclude invalid instrument pairs. Specifically, we seek to exclude $(z_k, z_{k'}) \notin \mathscr{Z}_{\bar{M}}$ from $\mathscr{Z}$, since if $(z_k, z_{k'}) \notin \mathscr{Z}_{\bar{M}}$, then $\beta_{k',k}$ in (2.2) is not identified absent additional restrictions (Appendix B). Suppose that there is a set $\mathscr{Z}_0 \subseteq \mathscr{Z}$ that satisfies the testable implications in Kitagawa (2015), Mourifié and Wan (2017), Kédagni and Mourifié (2020), and Sun (2023). Then we construct an estimator $\widehat{\mathscr{Z}_0}$ for $\mathscr{Z}_0$ and construct IV estimators based on $(z_k, z_{k'}) \in \widehat{\mathscr{Z}_0}$. We refer to these estimators as *VSIV estimators*. In the following, we assume that we have access to an estimator $\widehat{\mathscr{Z}_0}$, which is consistent for $\mathscr{Z}_0$ in the sense that $\mathbb{P}(\widehat{\mathscr{Z}_0} = \mathscr{Z}_0) \to 1$. We describe the testable implications and the construction of the proposed estimator satisfying $\mathbb{P}(\widehat{\mathscr{Z}_0} = \mathscr{Z}_0) \to 1$ in detail in Section 4.

In Section 3.2, we study VSIV estimation under the assumption that $\mathscr{Z}_0 = \mathscr{Z}_{\bar{M}}$ so that $\mathbb{P}(\widehat{\mathscr{Z}_0} = \mathscr{Z}_{\bar{M}}) \to 1$. In this case, the proposed VSIV estimators are asymptotically unbiased and normal under standard weak regularity conditions. Since $\mathscr{Z}_0$ is constructed based on the necessary (but not necessarily sufficient) conditions for the pairwise IV validity, $\mathscr{Z}_0$ could be larger than $\mathscr{Z}_{\bar{M}}$. (There exist no sufficient testable conditions for IV validity in general (Kitagawa, 2015).) In Section 3.3, we show that even if $\mathscr{Z}_0$ is larger than $\mathscr{Z}_{\bar{M}}$, VSIV estimators yield asymptotic bias reductions relative to standard LATE estimators that do not exploit testable implications to remove invalid instrument value pairs.

Note that if $\mathscr{Z}_0 = \emptyset$, VSIV estimation is trivial asymptotically since $\mathbb{P}(\widehat{\mathscr{Z}_0} = \emptyset) \to 1$. All the VSIV estimators converge to 0 by the convention in (1.1). In this case, we do not report any IV estimates in practice.

3.2. VSIV estimation under consistent estimation of validity pair set

Suppose that $\mathscr{Z}_0 = \mathscr{Z}_{\bar{M}}$ so that $\mathbb{P}(\widehat{\mathscr{Z}_0} = \mathscr{Z}_{\bar{M}}) \to 1$, and we use $\widehat{\mathscr{Z}_0}$ to construct VSIV estimators for the LATEs. We impose the following standard regularity conditions. Let g be a prespecified function that maps the value of Z to $\mathbb{R}$. For example, we can simply set $g(z) = z$ for all z if Z is a scalar instrument.[18]

Assumption 3.1. $\{(Y_i, D_i, Z_i)\}_{i=1}^n$ is an i.i.d. sample from a population such that all relevant moments exist.

Assumption 3.2. For every $\mathcal{Z}_{(k,k')} \in \mathscr{Z}_{\bar{M}}$,

$$E[g(Z_i)D_i | Z_i \in \mathcal{Z}_{(k,k')}] - E[D_i | Z_i \in \mathcal{Z}_{(k,k')}] \cdot E[g(Z_i) | Z_i \in \mathcal{Z}_{(k,k')}] \neq 0. \tag{3.1}$$

Assumption 3.1 assumes an i.i.d. data set and requires the existence of the relevant moments. Assumption 3.2 imposes a first-stage condition for every $\mathcal{Z}_{(k,k')} \in \mathscr{Z}_{\bar{M}}$. Note that (3.1) may not hold for $\mathcal{Z}_{(k,k')} \notin \mathscr{Z}_{\bar{M}}$. This creates additional technical difficulties when establishing the asymptotic normality of the VSIV estimators, which we discuss below. Assumption 3.2 also implies that if $\mathcal{Z}_{(k,k')} \in \mathscr{Z}_{\bar{M}}$, then $\mathcal{Z}_{(k',k)} \notin \mathscr{Z}_{\bar{M}}$. Otherwise, by Definition 2.1, $D_{z_k} = D_{z_{k'}}$ a.s., and (3.1) does not hold. For every scalar random sample $\{\xi_i\}_{i=1}^n$ and every $\mathcal{A} \in \mathscr{Z}$, we define

$$\mathcal{E}_n\left(\xi_i, \mathcal{A}\right) = \frac{\frac{1}{n}\sum_{i=1}^n \xi_i 1\left\{Z_i \in \mathcal{A}\right\}}{\frac{1}{n}\sum_{i=1}^n 1\left\{Z_i \in \mathcal{A}\right\}} \text{ and } \mathcal{E}\left(\xi_i, \mathcal{A}\right) = \frac{E\left[\xi_i 1\left\{Z_i \in \mathcal{A}\right\}\right]}{E\left[1\left\{Z_i \in \mathcal{A}\right\}\right]}.$$

[18] The choice of g may affect the efficiency of the VSIV estimators. We leave the formal analysis of the optimal choice of g for future study.

We define the VSIV estimators using regression-based IV estimators, following Imbens and Angrist (1994). For every $\mathcal{Z}_{(k,k')} \in \mathscr{Z}$, we run the IV regression

$$Y_i 1\left\{Z_i \in \mathcal{Z}_{(k,k')}\right\} = \gamma^0_{(k,k')} 1\left\{Z_i \in \mathcal{Z}_{(k,k')}\right\} + \gamma^1_{(k,k')} D_i 1\left\{Z_i \in \mathcal{Z}_{(k,k')}\right\} + \epsilon_i 1\left\{Z_i \in \mathcal{Z}_{(k,k')}\right\}, \tag{3.2}$$

using $g(Z_i)1\{Z_i \in \mathcal{Z}_{(k,k')}\}$ as the instrument for $D_i 1\{Z_i \in \mathcal{Z}_{(k,k')}\}$. Given the estimated validity set $\widehat{\mathscr{Z}}_0$, we set the VSIV estimator for each $\mathcal{Z}_{(k,k')}$ as

$$\widehat{\beta}^1_{(k,k')} = 1\left\{\mathcal{Z}_{(k,k')} \in \widehat{\mathscr{Z}}_0\right\} \cdot \frac{\mathcal{E}_n\left(g\left(Z_i\right)Y_i, \mathcal{Z}_{(k,k')}\right) - \mathcal{E}_n\left(g\left(Z_i\right), \mathcal{Z}_{(k,k')}\right)\mathcal{E}_n\left(Y_i, \mathcal{Z}_{(k,k')}\right)}{\mathcal{E}_n\left(g\left(Z_i\right)D_i, \mathcal{Z}_{(k,k')}\right) - \mathcal{E}_n\left(g\left(Z_i\right), \mathcal{Z}_{(k,k')}\right)\mathcal{E}_n\left(D_i, \mathcal{Z}_{(k,k')}\right)}, \tag{3.3}$$

which is the IV estimator of $\gamma^1_{(k,k')}$ in (3.2) multiplied by $1\{\mathcal{Z}_{(k,k')} \in \widehat{\mathscr{Z}}_0\}$. For every $\mathcal{Z}_{(k,k')} \in \widehat{\mathscr{Z}}_0$, this IV estimation is equivalent to a conventional IV regression in the subsample of $\{(Y_i, D_i, Z_i)\}_{i=1}^n$ with $Z_i \in \mathcal{Z}_{(k,k')}$. Note that $\widehat{\beta}^1_{(k,k')} = 0$ if $\mathcal{Z}_{(k,k')} \notin \widehat{\mathscr{Z}}_0$. We discuss this convention further below.

Define the vector of VSIV estimators as

$$\widehat{\beta}_1 = \left(\widehat{\beta}^1_{(1,2)}, \ldots, \widehat{\beta}^1_{(1,K)}, \ldots, \widehat{\beta}^1_{(K,1)}, \ldots, \widehat{\beta}^1_{(K,K-1)}\right)^T.$$

We also define

$$\beta^1_{(k,k')} = 1\left\{\mathcal{Z}_{(k,k')} \in \mathscr{Z}_{\bar{M}}\right\} \cdot \frac{\mathcal{E}\left(g\left(Z_i\right)Y_i, \mathcal{Z}_{(k,k')}\right) - \mathcal{E}\left(g\left(Z_i\right), \mathcal{Z}_{(k,k')}\right)\mathcal{E}\left(Y_i, \mathcal{Z}_{(k,k')}\right)}{\mathcal{E}\left(g\left(Z_i\right)D_i, \mathcal{Z}_{(k,k')}\right) - \mathcal{E}\left(g\left(Z_i\right), \mathcal{Z}_{(k,k')}\right)\mathcal{E}\left(D_i, \mathcal{Z}_{(k,k')}\right)} \tag{3.4}$$

and

$$\beta_1 = \left(\beta^1_{(1,2)}, \ldots, \beta^1_{(1,K)}, \ldots, \beta^1_{(K,1)}, \ldots, \beta^1_{(K,K-1)}\right)^T. \tag{3.5}$$

As we show formally in Theorem 3.1 below, $\beta^1_{(k,k')} = \beta_{k',k}$ as defined in (2.2) for every $(z_k, z_{k'}) \in \mathscr{Z}_{\bar{M}}$. If $\mathcal{Z}_{(k,k')} \notin \mathscr{Z}_{\bar{M}}$, we set $\beta^1_{(k,k')} = 0$ by (1.1) and (3.4). Similarly, if $\mathcal{Z}_{(k,k')} \notin \widehat{\mathscr{Z}}_0$, $\widehat{\beta}^1_{(k,k')} = 0$ by (1.1) and (3.3). Letting them be equal to 0 facilitates the description of the theoretical results in Theorem 3.1, and this will not affect the estimation of the weighted average of LATEs.[19] In practice, we recommend leaving $\widehat{\beta}^1_{(k,k')}$ to be blank if $1\{\mathcal{Z}_{(k,k')} \in \widehat{\mathscr{Z}}_0\} = 0$, as in the application in Section 5.3. We interpret the LATE corresponding to $\mathcal{Z}_{(k,k')} \in \mathscr{Z}_{\bar{M}}$ in the usual way as the average treatment effects for compliers in the corresponding subgroup. We do not report the estimates for LATEs corresponding to invalid pairs since they are not identified and there is no information about them in the data absent additional restrictions (Appendix B).

The next theorem establishes the asymptotic distribution of the VSIV estimator $\widehat{\beta}_1$, obtained based on the estimator of the instrument validity pair set $\widehat{\mathscr{Z}}_0$.

Theorem 3.1. *Suppose that the instrument Z is pairwise valid for the treatment D according to Definition 2.1 with the largest validity pair set $\mathscr{Z}_{\bar{M}} = \{(z_{k_1}, z_{k'_1}), \ldots, (z_{k_{\bar{M}}}, z_{k'_{\bar{M}}})\}$, that the estimator $\widehat{\mathscr{Z}}_0$ satisfies $\mathbb{P}(\widehat{\mathscr{Z}}_0 = \mathscr{Z}_{\bar{M}}) \to 1$, and that Assumptions 3.1 and 3.2 hold. Then*

$$\sqrt{n}(\widehat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, \Sigma), \tag{3.6}$$

where Σ is defined in (D.5) in the Appendix. In addition, $\beta^1_{(k,k')} = \beta_{k',k}$ as defined in (2.2) for every $(z_k, z_{k'}) \in \mathscr{Z}_{\bar{M}}$.

Theorem 3.1 establishes the joint asymptotic normality of the VSIV estimator of the LATEs. Establishing the asymptotic distribution in (3.6) requires a careful treatment of the case where the first-stage Assumption 3.2 does not hold for some pairs of instrument values $\mathcal{Z}_{(k,k')}$ that are not in the largest validity pair set $\mathscr{Z}_{\bar{M}}$, that is, $\mathcal{Z}_{(k,k')} \notin \mathscr{Z}_{\bar{M}}$. Specifically, we show that in this case, $\mathbb{P}(\widehat{\mathscr{Z}}_0 = \mathscr{Z}_{\bar{M}}) \to 1$ implies that, if $\mathcal{Z}_{(k,k')} \notin \mathscr{Z}_{\bar{M}}$, then for every $\rho > 0$, $n^\rho 1\{\mathcal{Z}_{(k,k')} \in \widehat{\mathscr{Z}}_0\} = o_p(1)$. This guarantees the convergence in (3.6) even when (3.1) does not hold for $\mathcal{Z}_{(k,k')} \notin \mathscr{Z}_{\bar{M}}$. The asymptotic covariance matrix Σ defined in the Appendix can be consistently estimated under standard conditions. Importantly, the estimation of the instrument validity pair set does not affect the asymptotic covariance matrix such that standard inference methods can be applied.

Under Theorem 3.1, it is straightforward to obtain a consistent estimator of a weighted average of all $\beta^1_{(k,k')}$. Let $\widehat{W}$ be some estimated weights with

$$\widehat{W} = (\widehat{w}_{(1,2)}, \ldots, \widehat{w}_{(1,K)}, \ldots, \widehat{w}_{(K,1)}, \ldots, \widehat{w}_{(K,K-1)})^T$$

such that $\widehat{W} \xrightarrow{p} W$ for some W with

$$W = (w_{(1,2)}, \ldots, w_{(1,K)}, \ldots, w_{(K,1)}, \ldots, w_{(K,K-1)})^T.$$

Theorem 3.1 implies that $\widehat{W}^T\widehat{\beta}_1 \xrightarrow{p} W^T\beta_1$. To establish the asymptotic distribution of the estimated weighted average of all $\beta^1_{(k,k')}$, we assume that $(\widehat{W}^T, \widehat{\beta}_1^T)^T$ is asymptotically normal. This is a weak condition in practice. It holds, for example, if the weights are defined as in the weighted average in (2.3). The following corollary summarizes the result.

[19] It is equivalent to not including them into the averages.

Corollary 3.1. *Suppose* $\sqrt{n}\{(\widehat{W}^T,\widehat{\beta}_1^T)^T-(W^T,\beta_1^T)^T\}\xrightarrow{d}N(0,\Sigma_W)$ *for some matrix* Σ_W*. Then it follows that*

$$\sqrt{n}(\widehat{W}^T\widehat{\beta}_1-W^T\beta_1)\xrightarrow{d}N\left(0,(\beta_1^T,W^T)\Sigma_W(\beta_1^T,W^T)^T\right). \tag{3.7}$$

The LATE $\beta_{k',k}$ is not identified if $\mathcal{Z}_{(k,k')}\notin\mathscr{Z}_{\bar{M}}$. Let $\beta_{1S}=(\beta^1_{(\kappa_1,\kappa_1')},\dots,\beta^1_{(\kappa_S,\kappa_S')})^T$ for some $S>0$. In our context, it is interesting to test hypotheses about $\beta^1_{(k,k')}$ with $\mathcal{Z}_{(k,k')}\in\mathscr{Z}_{\bar{M}}$ ($\beta^1_{(k,k')}=\beta_{k',k}$ by Theorem 3.1):

$$\mathrm{H}_0:\mathcal{Z}_{(\kappa_1,\kappa_1')}\in\mathscr{Z}_{\bar{M}},\dots,\mathcal{Z}_{(\kappa_S,\kappa_S')}\in\mathscr{Z}_{\bar{M}},\ R\left(\beta_{1S}\right)=0, \tag{3.8}$$

where R is a (possibly nonlinear) smooth r-dimensional function. Let $R'(\beta_S)$ be the $r\times S$ matrix of the continuous first derivative functions of R at an arbitrary value β_S, that is, $R'(\beta_S)=\partial R(\beta_S)/\partial\beta_S^T$. Let $\mathcal{I}_S$ be a $S\times(K-1)K$ matrix such that

$$\mathcal{I}_S\beta=(\beta_{(\kappa_1,\kappa_1')},\dots,\beta_{(\kappa_S,\kappa_S')})^T$$

for every $\beta=(\beta_{(1,2)},\dots,\beta_{(1,K)},\dots,\beta_{(K,1)},\dots,\beta_{(K,K-1)})^T$. Theorem 3.1 implies that

$$\sqrt{n}(\widehat{\beta}_{1S}-\beta_{1S})=\sqrt{n}\mathcal{I}_S(\widehat{\beta}_1-\beta_1)\xrightarrow{d}N\left(0,\Sigma_S\right),$$

where $\Sigma_S=\mathcal{I}_S\Sigma\mathcal{I}_S^T$, so that by the delta method, we obtain

$$\sqrt{n}\left\{R\left(\widehat{\beta}_{1S}\right)-R\left(\beta_{1S}\right)\right\}\xrightarrow{d}N\left(0,R'\left(\beta_{1S}\right)\Sigma_S R'\left(\beta_{1S}\right)^T\right).$$

We construct the test statistics as

$$TS_{1n}=\prod_{s=1}^{S}1\left\{\mathcal{Z}_{(\kappa_s,\kappa_s')}\in\widehat{\mathscr{Z}}_0\right\}$$

and

$$TS_{2n}=\sqrt{n}R(\widehat{\beta}_{1S})^T\left\{R'(\widehat{\beta}_{1S})\mathcal{I}_S\hat{\Sigma}\mathcal{I}_S^T R'(\widehat{\beta}_{1S})^T\right\}^{-1}\sqrt{n}R(\widehat{\beta}_{1S}), \tag{3.9}$$

where $\hat{\Sigma}$ is a consistent estimator of Σ, which can be constructed based on the formula in (D.5). Suppose that Assumptions 3.1 and 3.2 hold and $\mathbb{P}(\widehat{\mathscr{Z}}_0=\mathscr{Z}_{\bar{M}})\to 1$. If H_0 is true and $R'(\beta_{1S})$ is of full row rank, then it follows from standard arguments that $TS_{2n}\xrightarrow{d}\chi^2_r$, where χ^2_r denotes the chi-square distribution with r degrees of freedom. The decision rule of the test is to reject H_0 if $TS_{1n}=0$ or $TS_{2n}>c_r(\alpha)$, where $c_r(\alpha)$ satisfies $\mathbb{P}(\chi^2_r>c_r(\alpha))=\alpha$ for some predetermined $\alpha\in(0,1)$. The following proposition establishes the formal properties of the proposed test.

Proposition 3.1. *Suppose that Assumptions 3.1 and 3.2 hold and* $\mathbb{P}(\widehat{\mathscr{Z}}_0=\mathscr{Z}_{\bar{M}})\to 1$.

(i) If H_0 *is true,* $\mathbb{P}(\{TS_{1n}=0\}\cup\{TS_{2n}>c_r(\alpha)\})\to\alpha$.
(ii) If H_0 *is false,* $\mathbb{P}(\{TS_{1n}=0\}\cup\{TS_{2n}>c_r(\alpha)\})\to 1$.

3.3. Asymptotic bias reduction under VSIV estimation

In Section 3.2, we show that if the estimator of the largest validity pair set is consistent, $\mathbb{P}(\widehat{\mathscr{Z}}_0=\mathscr{Z}_{\bar{M}})\to 1$, the VSIV estimators are consistent and asymptotically normal under weak conditions. However, since $\mathscr{Z}_0$ is constructed based on necessary (but not necessarily sufficient) conditions for IV validity, we have $\mathbb{P}(\widehat{\mathscr{Z}}_0=\mathscr{Z}_0)\to 1$ in general, where the *pseudo-validity pair set* $\mathscr{Z}_0$ could be larger than $\mathscr{Z}_{\bar{M}}$. In this case, VSIV estimators may not be asymptotically unbiased. Consider an arbitrary presumed validity pair set $\mathscr{Z}_P$, which could incorporate prior information. If no prior information is available, we set $\mathscr{Z}_P=\mathscr{Z}$. Here we show even if $\mathscr{Z}_0$ is larger than $\mathscr{Z}_{\bar{M}}$, VSIV estimators based on $\widehat{\mathscr{Z}}_0'=\widehat{\mathscr{Z}}_0\cap\mathscr{Z}_P$ have weakly lower asymptotic biases than standard LATE estimators based on $\mathscr{Z}_P$. Intuitively, VSIV estimators use the information in the data about IV validity to reduce the asymptotic bias.

Since our target parameter is the vector β_1, a natural definition of the asymptotic bias is as follows.

Definition 3.1. The asymptotic bias of an arbitrary estimator $\tilde{\beta}_1$ for the true value β_1 defined in (3.5) is defined as $\mathrm{plim}_{n\to\infty}\|\tilde{\beta}_1-\beta_1\|_2$, where $\|\cdot\|_2$ is the ℓ^2-norm on Euclidean spaces.

The next assumption extends Assumption 3.2 to $\mathscr{Z}_0$.

Assumption 3.3. For every $\mathcal{Z}_{(k,k')}\in\mathscr{Z}_0$,

$$E[g(Z_i)D_i|Z_i\in\mathcal{Z}_{(k,k')}]-E[D_i|Z_i\in\mathcal{Z}_{(k,k')}]\cdot E[g(Z_i)|Z_i\in\mathcal{Z}_{(k,k')}]\neq 0. \tag{3.10}$$

The following theorem shows that the VSIV estimators based on $\widehat{\mathscr{Z}}_0'$ exhibit a smaller asymptotic bias than standard LATE estimators based on $\mathscr{Z}_P$.

Theorem 3.2. *Suppose that Assumptions 3.1 and 3.3 hold and that $\mathbb{P}(\widehat{\mathscr{Z}}_0 = \mathscr{Z}_0) \to 1$ with $\mathscr{Z}_0 \supseteq \mathscr{Z}_{\bar{M}}$. For every presumed validity pair set $\mathscr{Z}_P$, the asymptotic bias of $\widehat{\beta}_1$ is reduced by using $\widehat{\mathscr{Z}}_0'$ in the estimation (3.3) compared to the asymptotic bias from using $\mathscr{Z}_P$.*

As shown in Proposition 4.1 below, the pseudo-validity pair set $\mathscr{Z}_0$ can be estimated consistently by $\widehat{\mathscr{Z}}_0$ under mild conditions. Compared to constructing standard IV estimators based on $\mathscr{Z}_P$, Theorem 3.2 shows that the asymptotic bias, $\text{plim}_{n\to\infty}\|\widehat{\beta}_1 - \beta_1\|_2$, can be reduced by using VSIV estimators based on $\widehat{\mathscr{Z}}_0' = \widehat{\mathscr{Z}}_0 \cap \mathscr{Z}_P$.

The arguments used for establishing the asymptotic normality of the VSIV estimators in Section 3.2 do not rely on the consistent estimation of $\mathscr{Z}_{\bar{M}}$ ($\mathbb{P}(\widehat{\mathscr{Z}}_0 = \mathscr{Z}_{\bar{M}}) \to 1$). If $\mathbb{P}(\widehat{\mathscr{Z}}_0 = \mathscr{Z}_0) \to 1$ with $\mathscr{Z}_{\bar{M}} \subsetneq \mathscr{Z}_0$, the VSIV estimators are asymptotically normal, centered at β_1 defined with $\mathscr{Z}_0$ instead of $\mathscr{Z}_{\bar{M}}$. However, note that β_1 can only be interpreted as a vector of LATEs under consistent estimation ($\mathbb{P}(\widehat{\mathscr{Z}}_0 = \mathscr{Z}_{\bar{M}}) \to 1$).

Example 3.1 (*Asymptotic Bias Reduction under VSIV Estimation*)**.** Consider a simple example where $\mathcal{Z} = \{1,2,3,4\}$ as in our application and suppose that $\mathscr{Z}_{\bar{M}} = \{(1,2)\}$. In this case, by (1.1) and (3.4),

$$\beta_1 = \left(\beta^1_{(1,2)}, \dots, \beta^1_{(1,4)}, \dots, \beta^1_{(4,1)}, \dots, \beta^1_{(4,3)}\right)^T = \left(\beta^1_{(1,2)}, 0, \dots, 0\right)^T.$$

Suppose that, by mistake, we assume Z is valid according to Assumption 2.1 and use

$$\mathscr{Z}_P = \{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$$

as an estimator for $\mathscr{Z}_{\bar{M}}$. Then by (1.1) and (3.3),

$$\widehat{\beta}_1 = \left(\widehat{\beta}^1_{(1,2)}, \widehat{\beta}^1_{(1,3)}, \widehat{\beta}^1_{(1,4)}, \widehat{\beta}^1_{(2,3)}, \widehat{\beta}^1_{(2,4)}, \widehat{\beta}^1_{(3,4)}, 0,0,0,0,0,0\right)^T, \tag{3.11}$$

where $\widehat{\beta}^1_{(1,3)}$, $\widehat{\beta}^1_{(1,4)}$, $\widehat{\beta}^1_{(2,3)}$, $\widehat{\beta}^1_{(2,4)}$, and $\widehat{\beta}^1_{(3,4)}$ may not converge to 0 in probability. However, by definition $\beta^1_{(1,3)} = 0$, $\beta^1_{(1,4)} = 0$, $\beta^1_{(2,3)} = 0$, $\beta^1_{(2,4)} = 0$, and $\beta^1_{(3,4)} = 0$. Thus, $\|\widehat{\beta}_1 - \beta_1\|_2$ may not converge to 0 in probability. The approach proposed in this paper helps reduce this asymptotic bias. We exploit the information in the data about IV validity to obtain the estimator $\widehat{\mathscr{Z}}_0$. Even if $\widehat{\mathscr{Z}}_0$ converges to a set larger than $\mathscr{Z}_{\bar{M}}$ (because we use the necessary but not sufficient conditions for IV validity), VSIV always reduces the asymptotic bias. Suppose that $\mathscr{Z}_0 = \{(1,2),(3,4)\}$, which is larger than $\mathscr{Z}_{\bar{M}}$ but smaller than $\mathscr{Z}_P$. In this case, the VSIV estimator $\widehat{\beta}_1$ constructed by using $\widehat{\mathscr{Z}}_0 \cap \mathscr{Z}_P$ converges in probability to

$$\beta_1' = \left(\beta^1_{(1,2)}, 0,0,0,0, \beta^{1\prime}_{(3,4)}, 0,0,0,0,0,0\right)^T, \tag{3.12}$$

where $\beta^{1\prime}_{(3,4)}$ is the probability limit of $\widehat{\beta}^1_{(3,4)}$. Then, clearly, VSIV reduces the probability limit of $\|\widehat{\beta}_1 - \beta_1\|_2$.

3.4. Partially valid instruments and connection to existing results

Suppose we estimate the following canonical IV regression model,

$$Y_i = \alpha_0 + \alpha_1 D_i + \epsilon_i, \tag{3.13}$$

using $g(Z_i)$ as the instrument for D_i. When the instrument Z is fully valid, the traditional IV estimator of α_1 is

$$\widehat{\alpha}_1 = \frac{n\sum_{i=1}^n g\left(Z_i\right) Y_i - \sum_{i=1}^n g\left(Z_i\right)\sum_{i=1}^n Y_i}{n\sum_{i=1}^n g\left(Z_i\right) D_i - \sum_{i=1}^n g\left(Z_i\right)\sum_{i=1}^n D_i}. \tag{3.14}$$

The asymptotic properties of $\widehat{\alpha}_1$ can be found in Imbens and Angrist (1994, p. 471) and Angrist and Imbens (1995, p. 436).

To connect VSIV estimation to canonical IV regression with fully valid instruments, consider the following special case of pairwise IV validity.

Definition 3.2. Suppose that the instrument Z is pairwise valid for the treatment D with the largest validity pair set $\mathscr{Z}_{\bar{M}}$. If there is a validity pair set

$$\mathscr{Z}_M = \{(z_{k_1}, z_{k_2}), (z_{k_2}, z_{k_3}), \dots, (z_{k_{M-1}}, z_{k_M})\}$$

for some $M > 0$, then the instrument Z is called a **partially valid instrument** for the treatment D. The set $\mathcal{Z}_M = \{z_{k_1}, \dots, z_{k_M}\}$ is called a **validity value set** of Z.

Assumption 3.4. The validity value set $\mathcal{Z}_M$ satisfies that

$$E[g(Z_i)D_i|Z_i \in \mathcal{Z}_M] - E[D_i|Z_i \in \mathcal{Z}_M] \cdot E[g(Z_i)|Z_i \in \mathcal{Z}_M] \neq 0. \tag{3.15}$$

Suppose that Z is partially valid for the treatment D with a validity value set $\mathcal{Z}_M$ and that there is a consistent estimator $\widehat{\mathcal{Z}}_0$ of $\mathcal{Z}_M$. We then construct a VSIV estimator for α_1 in (3.13) by estimating the model

$$Y_i 1\left\{Z_i \in \widehat{\mathcal{Z}}_0\right\} = \gamma_0 1\left\{Z_i \in \widehat{\mathcal{Z}}_0\right\} + \gamma_1 D_i 1\left\{Z_i \in \widehat{\mathcal{Z}}_0\right\} + \epsilon_i 1\left\{Z_i \in \widehat{\mathcal{Z}}_0\right\}, \tag{3.16}$$

using $g(Z_i)1\{Z_i \in \widehat{\mathcal{Z}_0}\}$ as the instrument for $D_i 1\{Z_i \in \widehat{\mathcal{Z}_0}\}$. We obtain the VSIV estimator for α_1 in (3.13) by

$$\hat{\theta}_1 = \frac{n_z \sum_{i=1}^n g\left(Z_i\right) Y_i 1\left\{Z_i \in \widehat{\mathcal{Z}_0}\right\} - \sum_{i=1}^n g\left(Z_i\right) 1\left\{Z_i \in \widehat{\mathcal{Z}_0}\right\} \sum_{i=1}^n Y_i 1\left\{Z_i \in \widehat{\mathcal{Z}_0}\right\}}{n_z \sum_{i=1}^n g\left(Z_i\right) D_i 1\left\{Z_i \in \widehat{\mathcal{Z}_0}\right\} - \sum_{i=1}^n g\left(Z_i\right) 1\left\{Z_i \in \widehat{\mathcal{Z}_0}\right\} \sum_{i=1}^n D_i 1\left\{Z_i \in \widehat{\mathcal{Z}_0}\right\}}, \tag{3.17}$$

where $n_z = \sum_{i=1}^n 1\{Z_i \in \widehat{\mathcal{Z}_0}\}$. We can see that $\hat{\theta}_1$ is a generalized version of $\hat{\alpha}_1$ in (3.14), because when the instrument is fully valid, we can just let $\widehat{\mathcal{Z}_0} = \mathcal{Z}$ and then $\hat{\theta}_1 = \hat{\alpha}_1$.

Theorem 3.3. *Suppose that the instrument Z is partially valid for the treatment D according to Definition 3.2 with a validity value set $\mathcal{Z}_M = \{z_{k_1}, \ldots, z_{k_M}\}$, and that the estimator $\widehat{\mathcal{Z}_0}$ for $\mathcal{Z}_M$ satisfies $\mathbb{P}(\widehat{\mathcal{Z}_0} = \mathcal{Z}_M) \to 1$. Under Assumptions 3.1 and 3.4, it follows that $\hat{\theta}_1 \xrightarrow{p} \theta_1$, where*

$$\theta_1 = \frac{E\left[g\left(Z_i\right) Y_i | Z_i \in \mathcal{Z}_M\right] - E\left[Y_i | Z_i \in \mathcal{Z}_M\right] E\left[g\left(Z_i\right) | Z_i \in \mathcal{Z}_M\right]}{E\left[g\left(Z_i\right) D_i | Z_i \in \mathcal{Z}_M\right] - E\left[D_i | Z_i \in \mathcal{Z}_M\right] E\left[g\left(Z_i\right) | Z_i \in \mathcal{Z}_M\right]}.$$

Also, $\sqrt{n}(\hat{\theta}_1 - \theta_1) \xrightarrow{d} N(0, \Sigma_1)$, where Σ_1 is provided in (D.25) in the Appendix. In addition, the quantity θ_1 can be interpreted as the weighted average of $\{\beta_{k_2,k_1}, \ldots, \beta_{k_M,k_{M-1}}\}$ defined as in (2.2). Specifically, $\theta_1 = \sum_{m=1}^{M-1} \mu_m \beta_{k_{m+1},k_m}$ with

$$\mu_m = \frac{\left[p\left(z_{k_{m+1}}\right) - p\left(z_{k_m}\right)\right] \sum_{l=m}^{M-1} \mathbb{P}\left(Z_i = z_{k_{l+1}} | Z_i \in \mathcal{Z}_M\right) \left\{g\left(z_{k_{l+1}}\right) - E\left[g\left(Z_i\right) | Z_i \in \mathcal{Z}_M\right]\right\}}{\sum_{l=1}^{M} \mathbb{P}\left(Z_i = z_{k_l} | Z_i \in \mathcal{Z}_M\right) p\left(z_{k_l}\right) \left\{g\left(z_{k_l}\right) - E\left[g\left(Z_i\right) | Z_i \in \mathcal{Z}_M\right]\right\}},$$

$p(z_k) = E[D_i | Z_i = z_k]$, and $\sum_{m=1}^{M-1} \mu_m = 1$.

Theorem 3.3 is an extension of Theorem 2 of Imbens and Angrist (1994) to the case where the instrument is partially but not fully valid. To establish a connection to existing results, Theorem 3.3 assumes consistent estimation of the validity value set such that $\mathbb{P}(\widehat{\mathcal{Z}_0} = \mathcal{Z}_M) \to 1$. If $\widehat{\mathcal{Z}_0}$ converges to a larger set than $\mathcal{Z}_M$, the properties of VSIV estimation follow from the results in Section 3.3 because partially valid instruments are a special case of pairwise valid instruments.

4. Definition and estimation of $\mathscr{Z}_0$

Here we discuss the definition and the estimation of $\mathscr{Z}_0$ based on the testable implications in Kitagawa (2015), Mourifié and Wan (2017), Kédagni and Mourifié (2020), and Sun (2023) for pairwise IV validity. We show that under weak assumptions, the proposed estimator $\widehat{\mathscr{Z}_0}$ is consistent for the pseudo-validity pair set $\mathscr{Z}_0$ in the sense that $\mathbb{P}(\widehat{\mathscr{Z}_0} = \mathscr{Z}_0) \to 1$. As a consequence, when $\mathscr{Z}_0 = \mathscr{Z}_{\bar{M}}$, the largest validity pair set can be estimated consistently.

Specifically, we first construct two sets, $\mathscr{Z}_1$ and $\mathscr{Z}_2$, of pairs of instrument values such that $\mathscr{Z}_1$ satisfies the testable implications in Kitagawa (2015), Mourifié and Wan (2017), and Sun (2023), and $\mathscr{Z}_2$ satisfies the testable implications in Kédagni and Mourifié (2020). We then construct $\mathscr{Z}_0$ as the intersection of these two sets, $\mathscr{Z}_0 = \mathscr{Z}_1 \cap \mathscr{Z}_2$ (see Appendices D.4 and E.2). Lemma 4.1 shows that $\mathscr{Z}_1 \subseteq \mathscr{Z}_2$ when D is binary. For multivalued D, we are not aware of such a result.

Lemma 4.1. *If $D \in \{0,1\}$, then $\mathscr{Z}_1 \subseteq \mathscr{Z}_2$, and the testable restrictions defining $\mathscr{Z}_1$ are sharp.*[20]

Lemma 4.1 shows that when $D \in \{0,1\}$, the testable implications of Kédagni and Mourifié (2020) are implied by those of Kitagawa (2015), Mourifié and Wan (2017), and Sun (2023). This result may be of independent interest. Moreover, Lemma 4.1 establishes the sharpness of the testable implications used to define $\mathscr{Z}_1$. This follows from the arguments in Proposition 1.1(i) of Kitagawa (2015).

Lemma 4.1 implies that $\mathscr{Z}_0 = \mathscr{Z}_1 \cap \mathscr{Z}_2 = \mathscr{Z}_1$ when D is binary. Therefore, we define $\mathscr{Z}_0$ based on the testable implications proposed in Kitagawa (2015), Mourifié and Wan (2017), and Sun (2023). These testable implications were originally proposed for full IV validity. In the following, we extend them to Definition 2.1. To describe the testable restrictions, we use the notation of Sun (2023). Define conditional probabilities

$$P_z(B, C) = \mathbb{P}(Y \in B, D \in C | Z = z)$$

for all Borel sets $B, C \in \mathcal{B}_{\mathbb{R}}$ and all $z \in \mathcal{Z}$. With $\mathscr{Z}_{\bar{M}} = \{(z_{k_1}, z_{k'_1}), \ldots, (z_{k_{\bar{M}}}, z_{k'_{\bar{M}}})\}$, for every $m \in \{1, \ldots, \bar{M}\}$, it follows that

$$P_{z_{k_m}}(B, \{1\}) \leq P_{z_{k'_m}}(B, \{1\}) \text{ and } P_{z_{k_m}}(B, \{0\}) \geq P_{z_{k'_m}}(B, \{0\}) \tag{4.1}$$

for all $B \in \mathcal{B}_{\mathbb{R}}$. By definition, for all $B, C \in \mathcal{B}_{\mathbb{R}}$,

$$\mathbb{P}(Y \in B, D \in C | Z = z) = \frac{\mathbb{P}(Y \in B, D \in C, Z = z)}{\mathbb{P}(Z = z)}.$$

[20] The statement in *Lemma* 4.1 holds for ordered or unordered $D \in \mathcal{D} = \{d_1, d_2\}$.

Define the function spaces

$$\mathcal{G}_P = \left\{ \left(1_{\mathbb{R}\times\mathbb{R}\times\{z_k\}}, 1_{\mathbb{R}\times\mathbb{R}\times\{z_{k'}\}} \right) : k, k' \in \{1, \dots, K\}, k \neq k' \right\},$$
$$\mathcal{H} = \left\{ (-1)^d \cdot 1_{B\times\{d\}\times\mathbb{R}} : B \text{ is a closed interval in } \mathbb{R}, d \in \{0,1\} \right\}, \text{ and}$$
$$\bar{\mathcal{H}} = \left\{ (-1)^d \cdot 1_{B\times\{d\}\times\mathbb{R}} : B \text{ is a closed, open, or half-closed interval in } \mathbb{R}, d \in \{0,1\} \right\}. \tag{4.2}$$

Similarly to Sun (2023), by Lemma B.7 in Kitagawa (2015), we use all closed intervals $B \subseteq \mathbb{R}$ to construct $\mathcal{H}$ instead of all Borel sets.

Suppose we have access to an i.i.d. sample $\{(Y_i, D_i, Z_i)\}_{i=1}^n$ distributed according to some probability distribution P in $\mathcal{P}$, that is, $P(G) = \mathbb{P}((Y_i, D_i, Z_i) \in G)$ for all measurable G. The closure of $\mathcal{H}$ in $L^2(P)$ is equal to $\bar{\mathcal{H}}$ by Lemma C.1 of Sun (2023). For every $(h, g) \in \bar{\mathcal{H}} \times \mathcal{G}_P$ with $g = (g_1, g_2)$, we define

$$\phi(h, g) = \frac{P(h \cdot g_2)}{P(g_2)} - \frac{P(h \cdot g_1)}{P(g_1)}$$

and

$$\sigma^2(h, g) = \Lambda(P) \cdot \left\{ \frac{P(h^2 \cdot g_2)}{P^2(g_2)} - \frac{P^2(h \cdot g_2)}{P^3(g_2)} + \frac{P(h^2 \cdot g_1)}{P^2(g_1)} - \frac{P^2(h \cdot g_1)}{P^3(g_1)} \right\}, \tag{4.3}$$

where $\Lambda(P) = \prod_{k=1}^K P(1_{\mathbb{R}\times\mathbb{R}\times\{z_k\}})$ and $P^m(g_j) = [P(g_j)]^m$ for $m \in \mathbb{N}$ and $j \in \{1, 2\}$. We denote the sample analog of ϕ as

$$\widehat{\phi}(h, g) = \frac{\widehat{P}(h \cdot g_2)}{\widehat{P}(g_2)} - \frac{\widehat{P}(h \cdot g_1)}{\widehat{P}(g_1)},$$

where $\widehat{P}$ is the empirical probability measure corresponding to P so that for every measurable function v (by abuse of notation),

$$\widehat{P}(v) = \frac{1}{n} \sum_{i=1}^n v(Y_i, D_i, Z_i). \tag{4.4}$$

For every $(h, g) \in \bar{\mathcal{H}} \times \mathcal{G}_P$ with $g = (g_1, g_2)$, define the sample analog of $\sigma^2(h, g)$ as

$$\widehat{\sigma}^2(h, g) = \frac{T_n}{n} \cdot \left\{ \frac{\widehat{P}(h^2 \cdot g_2)}{\widehat{P}^2(g_2)} - \frac{\widehat{P}^2(h \cdot g_2)}{\widehat{P}^3(g_2)} + \frac{\widehat{P}(h^2 \cdot g_1)}{\widehat{P}^2(g_1)} - \frac{\widehat{P}^2(h \cdot g_1)}{\widehat{P}^3(g_1)} \right\},$$

where $T_n = n \cdot \prod_{k=1}^K \widehat{P}(1_{\mathbb{R}\times\mathbb{R}\times\{z_k\}})$. By (1.1), $\widehat{\sigma}^2$ is well defined. By similar arguments as in the proof of Lemma 3.1 in Sun (2023), σ^2 and $\widehat{\sigma}^2$ are uniformly bounded in (h, g). The following lemma reformulates the testable restrictions in (4.1) in terms of ϕ. Below, we use this reformulation to define $\mathscr{Z}_0$ and the corresponding estimator $\widehat{\mathscr{Z}_0}$.

Lemma 4.2. *Suppose that the instrument Z is pairwise valid for the treatment D with the largest validity pair set $\mathscr{Z}_{\bar{M}} = \{(z_{k_1}, z_{k'_1}), \dots, (z_{k_{\bar{M}}}, z_{k'_{\bar{M}}})\}$. For every $m \in \{1, \dots, \bar{M}\}$, $\sup_{h\in\mathcal{H}} \phi(h, g) = 0$ with $g = (1_{\mathbb{R}\times\mathbb{R}\times\{z_{k_m}\}}, 1_{\mathbb{R}\times\mathbb{R}\times\{z_{k'_m}\}})$.*

Lemma 4.2 reformulates the necessary conditions based on Kitagawa (2015), Mourifié and Wan (2017), and Sun (2023) for the validity pair set $\mathscr{Z}_{\bar{M}}$. Define

$$\mathcal{G}_0 = \left\{ g \in \mathcal{G}_P : \sup_{h\in\mathcal{H}} \phi(h, g) = 0 \right\} \text{ and } \widehat{\mathcal{G}_0} = \left\{ g \in \mathcal{G}_P : \sqrt{T_n} \left| \sup_{h\in\mathcal{H}} \frac{\widehat{\phi}(h, g)}{\xi_0 \vee \widehat{\sigma}(h, g)} \right| \leq \tau_n^g \right\}, \tag{4.5}$$

where $1/\min_{g\in\mathcal{G}_P} \tau_n^g \to 0$ in probability and $\max_{g\in\mathcal{G}_P} \tau_n^g/\sqrt{n} \to 0$ in probability as $n \to \infty$, and ξ_0 is a small positive number.[21] Here, we allow τ_n^g to be different for every g. This added flexibility helps improve the finite sample performance of VSIV estimation. We discuss our explicit choice of τ_n^g in more detail in Section 5.

The set $\mathcal{G}_0$ is different from the contact sets defined in Beare and Shi (2019), Sun and Beare (2021), and Sun (2023) in different contexts because of the presence of the map sup. We refer to Linton et al. (2010) and Lee et al. (2018) for further discussions of contact set estimation. Define $\mathscr{Z}_0$ as the collection of all (z, z') associated with some $g \in \mathcal{G}_0$:

$$\mathscr{Z}_0 = \left\{ (z_k, z_{k'}) \in \mathscr{Z} : g = (1_{\mathbb{R}\times\mathbb{R}\times\{z_k\}}, 1_{\mathbb{R}\times\mathbb{R}\times\{z_{k'}\}}) \in \mathcal{G}_0 \right\}. \tag{4.6}$$

Note that $\mathscr{Z}_0$ is the set of all pairs that satisfy the testable implications in (4.1). For example, if $K = 4$ and $\mathcal{G}_0 = \{(1_{\mathbb{R}\times\mathbb{R}\times\{z_1\}}, 1_{\mathbb{R}\times\mathbb{R}\times\{z_2\}}), (1_{\mathbb{R}\times\mathbb{R}\times\{z_3\}}, 1_{\mathbb{R}\times\mathbb{R}\times\{z_4\}})\}$, then $\mathscr{Z}_0 = \{(z_1, z_2), (z_3, z_4)\}$.

We use $\widehat{\mathcal{G}_0}$ to construct the estimator of $\mathscr{Z}_0$, denoted by $\widehat{\mathscr{Z}_0}$, which is defined as the set of all (z, z') associated with some $g \in \widehat{\mathcal{G}_0}$:

$$\widehat{\mathscr{Z}_0} = \left\{ (z_k, z_{k'}) \in \mathscr{Z} : g = (1_{\mathbb{R}\times\mathbb{R}\times\{z_k\}}, 1_{\mathbb{R}\times\mathbb{R}\times\{z_{k'}\}}) \in \widehat{\mathcal{G}_0} \right\}. \tag{4.7}$$

Note that (4.7) is the sample analog of (4.6). The following proposition establishes consistency of $\widehat{\mathscr{Z}_0}$.

[21] In practice, we use $\xi_0 = 10^{-100}$.

Proposition 4.1. *Under Assumption 3.1, $\mathbb{P}(\widehat{\mathcal{G}_0} = \mathcal{G}_0) \to 1$, and thus $\mathbb{P}(\widehat{\mathscr{Z}_0} = \mathscr{Z}_0) \to 1$.*

Proposition 4.1 is related to the contact set estimation in Sun (2023). Since, by definition, $\mathcal{G}_0 \subseteq \mathcal{G}_P$ and $\mathcal{G}_P$ is a finite set, we can use techniques similar to those in Sun (2023) to obtain the stronger result in Proposition 4.1, that is, $\mathbb{P}(\hat{\mathcal{G}}_0 = \mathcal{G}_0) \to 1$.

Remark 4.1 (*Uniqueness of $\mathscr{Z}_{\bar{M}}$ and $\mathscr{Z}_0$*)**.** By definition, $\mathscr{Z}_{\bar{M}}$ is the collection of all valid pairs of values of Z, while $\mathscr{Z}_0$ is the collection of all pairs that satisfy the testable restrictions in Lemma 4.2. Thus, both $\mathscr{Z}_{\bar{M}}$ and $\mathscr{Z}_0$ are unique, and clearly $\mathscr{Z}_{\bar{M}} \subseteq \mathscr{Z}_0$.

Remark 4.2 (*Local Violations of the Testable Restrictions*)**.** The theory for VSIV estimation relies on selection consistency, $\mathbb{P}(\widehat{\mathscr{Z}_0} = \mathscr{Z}_0) \to 1$. A potential concern with this approach is that selection consistency is theoretically only possible if the violations of the testable restrictions are well-separated from zero. However, VSIV estimation remains useful even in the presence of local (to-zero) violations. It will always yield asymptotic bias reductions from removing pairs of instrument values for which the violations are well-separated from zero. An important feature of VSIV estimation is that it proceeds pairwise, so that the presence of pairs corresponding to local violations does not impact the selection performance for pairs corresponding to well-separated violations. We explore the performance of our method when we change the magnitude of the violations in Appendix F.

5. Simulations and application

In this section, we evaluate the finite sample performance of VSIV estimation in Monte Carlo simulations designed based on an empirical application. First, we discuss the empirical context. Second, we present the results from the simulation study and determine the choice of the tuning parameter τ_n^g. Finally, we present the results from the empirical application.

5.1. Empirical context

We revisit the analysis of the causal effect of college education on earnings. We use the dataset analyzed by Heckman et al. (2001) and Li et al. (2024), consisting of data on 1,230 white males from the National Longitudinal Survey of Youth of 1979 (NLSY). The outcome of interest (Y) is the log wage and the treatment (D) is a dummy variable for college enrollment. Li et al. (2024) use the maximum parental education (E) as an instrument for college enrollment. Since the overall sample size is relatively small, we consider a coarsened version of this instrument: $Z = 1\{E < 12\} + 2 \cdot 1\{E = 12\} + 3 \cdot 1\{12 < E < 16\} + 4 \cdot 1\{E \geq 16\}$.

The parental education instrument likely violates the exclusion restriction due to its potential positive effect on earnings, especially at lower levels of parental education. Therefore, Li et al. (2024) consider a relaxation of IV validity, building on Manski and Pepper (2000, 2009). As discussed in Section 2, their relaxation allows the exclusion restriction to be violated for IV values below a cutoff, which they estimate to be $E = 11$ (Table 2, Column (3)). The results in Li et al. (2024) suggest that the parental education instrument is partially invalid, especially for pairs of instrument values corresponding to lower levels of parental education, thus providing an ideal setting for illustrating the usefulness of VSIV estimation.

We emphasize that there are two major differences between our analysis here and the one in Li et al. (2024). First, we consider a different relaxation of IV validity. Unlike Li et al. (2024), we do not impose any assumptions on how IV validity fails for invalid pairs. Second, Li et al. (2024) focus on partial identification of the average treatment effect. By contrast, we consider the estimation of LATEs for all pairs of instrument values that are not screened out based on the testable implications.

5.2. Simulation evidence

5.2.1. Data generating processes

Here we describe the data generating processes (DGPs) that we use in the simulations. All DGPs are calibrated to the joint empirical distribution of (D, Z) in the application. Denote by $\hat{\mathbb{P}}$ the empirical probability measure of $\mathbb{P}$. In the empirical application, we have $\hat{\mathbb{P}}(Z = 1) = 0.1317$, $\hat{\mathbb{P}}(Z = 2) = 0.4716$, $\hat{\mathbb{P}}(Z = 3) = 0.1495$, $\hat{\mathbb{P}}(Z = 4) = 0.2472$, $\hat{\mathbb{P}}(D = 1|Z = 1) = 0.1420$, $\hat{\mathbb{P}}(D = 1|Z = 2) = 0.3086$, $\hat{\mathbb{P}}(D = 1|Z = 3) = 0.5054$, and $\hat{\mathbb{P}}(D = 1|Z = 4) = 0.7796$. Given the prior information that Z may be positively correlated with D, we assume that $\mathscr{Z}_P = \{(1,2), (1,3), (1,4), (2,3), (2,4), (3,4)\}$.[22] We consider five data generating processes (DGPs (0)–(4)), where Definition 2.1 holds for each pair in $\mathscr{Z}_P$ under DGP (0) and is violated for every pair under DGPs (1)–(4). These DGPs are constructed following those in Kitagawa (2015) and Sun (2023).

For all DGPs, we specify $U \sim \text{Unif}(0,1)$, $V \sim \text{Unif}(0,1)$, $W \sim \text{Unif}(0,1)$, $Z = 1\{U \leq 0.1317\} + 2 \times \{0.1317 < U \leq 0.6033\} + 3 \times \{0.6033 < U \leq 0.7528\} + 4 \times 1\{U > 0.7528\}$, $D_1 = 1\{V \leq 0.1420\}$, $D_2 = 1\{V \leq 0.3086\}$, $D_3 = 1\{V \leq 0.5054\}$, $D_4 = 1\{V \leq 0.7796\}$, and $D = \sum_{z=1}^4 1\{Z = z\} \times D_z$. We let $N_Z \sim N(0,1)$, and let $N_{10a}(\sigma) \sim N(-1, \sigma^2)$, $N_{10b}(\sigma) \sim N(-0.5, \sigma^2)$, $N_{10c}(\sigma) \sim N(0, \sigma^2)$, $N_{10d}(\sigma) \sim N(0.5, \sigma^2)$, $N_{10e}(\sigma) \sim N(1, \sigma^2)$, and $N_C(\sigma) = 1\{W \leq 0.15\} \times N_{10a}(\sigma) + 1\{0.15 < W \leq 0.35\} \times N_{10b}(\sigma) + 1\{0.35 < W \leq 0.65\} \times N_{10c}(\sigma) + 1\{0.65 < W \leq 0.85\} \times N_{10d}(\sigma) + 1\{W > 0.85\} \times N_{10e}(\sigma)$ for $\sigma > 0$. The DGPs (0)–(4) are specified as follows.

(0): $N_{dz} = N_Z$ for each $d \in \{0,1\}$ and all $z \in \{1,2,3,4\}$, $Y = \sum_{z=1}^4 1\{Z = z\} \times (\sum_{d=0}^1 1\{D = d\} \times N_{dz})$

[22] This is consistent with the fact that the empirical proportions $\hat{\mathbb{P}}(D = 1|Z = z)$ are increasing in z.

Table 5.1
Validity pair set estimation for DGP (0).

n	c	(1, 2)	(1, 3)	(1, 4)	(2, 3)	(2, 4)	(3, 4)
1230	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.002	0.000	0.000	0.001
	0.5	0.000	0.053	0.198	0.003	0.140	0.217
	0.6	0.003	0.585	0.726	0.264	0.733	0.821
	0.7	0.138	0.901	0.953	0.736	0.967	0.985
	0.8	0.542	0.983	0.996	0.956	0.998	1.000
	0.9	0.865	0.997	1.000	0.996	1.000	1.000
	1	0.968	1.000	1.000	1.000	1.000	1.000
2460	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.006
	0.5	0.000	0.182	0.370	0.018	0.290	0.568
	0.6	0.018	0.802	0.898	0.482	0.908	0.956
	0.7	0.344	0.977	0.995	0.901	0.998	0.997
	0.8	0.821	0.998	1.000	0.989	1.000	1.000
	0.9	0.968	1.000	1.000	1.000	1.000	1.000
	1	0.997	1.000	1.000	1.000	1.000	1.000

(1): For $(z_1,z_2)\in\mathscr{Z}_P$, $N_{1z_1}\sim N(\mu_{(z_1,z_2)},1)$, $N_{1z_2}\sim N(0,1)$, $N_{0z_1}\sim N(0,1)$, $N_{0z_2}\sim N(\mu_{(z_1,z_2)},1)$ with $\mu_{(1,2)}=-0.9$, $\mu_{(1,3)}=-1.1$, $\mu_{(1,4)}=-1.3$, $\mu_{(2,3)}=-0.9$, $\mu_{(2,4)}=-1.1$, $\mu_{(3,4)}=-0.9$, $N_{dz}\sim N(0,1)$ for $d\in\{0,1\}$ and $z\in\{1,2,3,4\}\setminus\{z_1,z_2\}$, $Y=\sum_{z=1}^4 1\{Z=z\}\times(\sum_{d=0}^1 1\{D=d\}\times N_{dz})$

(2): For $(z_1,z_2)\in\mathscr{Z}_P$, $N_{1z_1}\sim N(0,1)$, $N_{1z_2}\sim N(0,\sigma^2_{(z_1,z_2)})$, $N_{0z_1}\sim N(0,\sigma^2_{(z_1,z_2)})$, $N_{0z_2}\sim N(0,1)$ with $\sigma_{(1,2)}=3$, $\sigma_{(1,3)}=5$, $\sigma_{(1,4)}=7$, $\sigma_{(2,3)}=3$, $\sigma_{(2,4)}=5$, $\sigma_{(3,4)}=3$, $N_{dz}\sim N(0,1)$ for $d\in\{0,1\}$ and $z\in\{1,2,3,4\}\setminus\{z_1,z_2\}$, $Y=\sum_{z=1}^4 1\{Z=z\}\times(\sum_{d=0}^1 1\{D=d\}\times N_{dz})$

(3): For $(z_1,z_2)\in\mathscr{Z}_P$, $N_{1z_1}\sim N(0,1)$, $N_{1z_2}\sim N(0,\sigma^2_{(z_1,z_2)})$, $N_{0z_1}\sim N(0,\sigma^2_{(z_1,z_2)})$, $N_{0z_2}\sim N(0,1)$ with $\sigma_{(1,2)}=0.5$, $\sigma_{(1,3)}=0.45$, $\sigma_{(1,4)}=0.4$, $\sigma_{(2,3)}=0.5$, $\sigma_{(2,4)}=0.45$, $\sigma_{(3,4)}=0.5$, $N_{dz}\sim N(0,1)$ for $d\in\{0,1\}$ and $z\in\{1,2,3,4\}\setminus\{z_1,z_2\}$, $Y=\sum_{z=1}^4 1\{Z=z\}\times(\sum_{d=0}^1 1\{D=d\}\times N_{dz})$

(4): For $(z_1,z_2)\in\mathscr{Z}_P$, $N_{1z_1}\sim N_C(\sigma_{(z_1,z_2)})$, $N_{1z_2}\sim N(0,1)$, $N_{0z_1}\sim N(0,1)$, $N_{0z_2}\sim N_C(\sigma_{(z_1,z_2)})$ with $\sigma_{(1,2)}=0.08$, $\sigma_{(1,3)}=0.04$, $\sigma_{(1,4)}=0.02$, $\sigma_{(2,3)}=0.08$, $\sigma_{(2,4)}=0.04$, $\sigma_{(3,4)}=0.08$, $N_{dz}\sim N(0,1)$ for $d\in\{0,1\}$ and $z\in\{1,2,3,4\}\setminus\{z_1,z_2\}$, $Y=\sum_{z=1}^4 1\{Z=z\}\times(\sum_{d=0}^1 1\{D=d\}\times N_{dz})$

The random variables U, V, W, and those generated from $N(\mu,\sigma^2)$ for some μ and σ are mutually independent.

5.2.2. Choice of τ_n^g

As shown in (4.5), the tuning parameter τ_n^g should satisfy that $1/\min_{g\in\mathcal{G}_P}\tau_n^g\to 0$ in probability and $\max_{g\in\mathcal{G}_P}\tau_n^g/\sqrt{n}\to 0$ in probability as $n\to\infty$. In our simulations, for every $g=(1_{\mathbb{R}\times\mathbb{R}\times\{z_k\}},1_{\mathbb{R}\times\mathbb{R}\times\{z_{k'}\}})$, we set τ_n^g as

$$\tau_n^g=c\cdot\frac{(n\widehat{\mathbb{P}}(Z\in\{z_k,z_{k'}\}))^{1/5}}{|\widehat{\mathbb{P}}(D=1|Z=z_k)-\widehat{\mathbb{P}}(D=1|Z=z_{k'})|^{1/5}} \tag{5.1}$$

for some constant $c>0$.[23] The numerator of (5.1) captures the relevant sample size, and the denominator captures the idea that if the difference is large, violations are harder to detect. This (heuristic) adjustment in the denominator does not affect the asymptotic properties but works well in simulations.

5.2.3. Simulation results

We consider two different sample sizes: $n=1230$ as in the empirical application and $n=2460$ to explore how the properties of our methods improve as the sample size increases. We report results for τ_n^g in (5.1) with $c\in\{0.1,0.2,\dots,1\}$. For each simulation, we use 1,000 Monte Carlo iterations. To calculate the supremum in $\sqrt{T_n}|\sup_{h\in\mathcal{H}}\hat{\phi}(h,g)/(\xi_0\vee\hat{\sigma}(h,g))|$ for every g, we use the approach employed by Kitagawa (2015) and Sun (2023). Specifically, we compute the supremum based on the closed intervals $[a,b]$ with the realizations of $\{Y_i\}$ as endpoints, that is, intervals $[a,b]$ where $a,b\in\{Y_i\}$ and $a\le b$.

Tables 5.1–5.5 show the empirical probabilities with which each element of $\mathscr{Z}_P$ is selected to be in $\widehat{\mathscr{Z}}_0$ in the simulations. The results show that choosing c is subject to a trade-off between the ability of our method to screen out invalid pairs and its ability to include valid pairs. Given the nature of the method, screening out invalid pairs is particularly important since LATE estimators

[23] If $\mathbb{P}(D=1|Z=z_k)-\mathbb{P}(D=1|Z=z_{k'})=0$ for some pair $(z_k,z_{k'})$, we still have that $\tau_n^g/\sqrt{n}\to 0$ in probability as $n\to\infty$ for every $g\in\mathcal{G}_P$, and τ_n^g satisfies the two conditions above in this case.

Table 5.2
Validity pair set estimation for DGP (1).

n	c	(1, 2)	(1, 3)	(1, 4)	(2, 3)	(2, 4)	(3, 4)
1230	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.002	0.011	0.000	0.001	0.017
	0.7	0.000	0.021	0.104	0.020	0.028	0.108
	0.8	0.000	0.086	0.298	0.126	0.148	0.311
	0.9	0.010	0.199	0.538	0.390	0.357	0.583
	1	0.051	0.375	0.743	0.689	0.592	0.787
2460	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.004	0.000	0.000	0.001
	0.7	0.000	0.003	0.058	0.003	0.010	0.019
	0.8	0.001	0.010	0.219	0.037	0.046	0.093
	0.9	0.001	0.037	0.405	0.182	0.164	0.276
	1	0.005	0.118	0.648	0.509	0.352	0.533

Table 5.3
Validity pair set estimation for DGP (2).

n	c	(1, 2)	(1, 3)	(1, 4)	(2, 3)	(2, 4)	(3, 4)
1230	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.001	0.000	0.000	0.004
	0.7	0.000	0.000	0.039	0.000	0.010	0.027
	0.8	0.000	0.003	0.188	0.024	0.063	0.134
	0.9	0.000	0.012	0.416	0.192	0.189	0.342
	1	0.008	0.037	0.630	0.479	0.418	0.564
2460	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.001	0.000	0.000	0.000
	0.7	0.000	0.000	0.030	0.000	0.002	0.004
	0.8	0.000	0.000	0.185	0.007	0.021	0.050
	0.9	0.000	0.000	0.436	0.041	0.093	0.135
	1	0.000	0.000	0.652	0.185	0.241	0.282

based on these pairs are inconsistent. Our method with $c = 0.6$ detects invalid pairs almost perfectly while selecting most of the valid pairs with high probability. With $c = 0.6$, as n increases from 1230 to 2460, the selection rates for valid pairs are increasing and those for invalid pairs are decreasing to 0. Overall, the simulation results show that the proposed method performs well in identifying the validity pair set in finite samples.

Table 5.6 shows the coverage rates of the confidence intervals constructed based on the asymptotic distribution in Theorem 3.1 for DGP (0) under which all pairs are valid. We construct the confidence interval as follows: If a pair is in the estimated validity pair set, then the confidence interval is constructed based on the asymptotic distribution; if the pair is not in the estimated validity pair set, then the confidence interval is $\mathbb{R}$, since in this case the LATE is not identified. The results in Table 5.6 show that the coverage rates are close to $1 - \alpha$, where $\alpha = 0.05$, for most pairs when $c = 0.6$. Since the sample size for each pair is relatively small in our simulations for both choices of n, the coverage rates can be somewhat higher than 0.95 because of the way we construct the confidence intervals if the pair is not in the estimated validity pair set.

In Appendix F.1, we provide additional simulation results for a variant of DGP (0) with a more balanced design. The results in Tables F.1 and F.2 show that in this case for $c = 0.6$, the selection rates for all pairs are high and converging to one, and the coverage rates are converging to 95%.

In Tables 5.7–5.10, we explore the asymptotic bias reduction property of VSIV estimation by presenting the Root Mean Square Errors (RMSEs) of $\sqrt{n}(\widehat{\beta}^1_{(k,k')} - \beta^1_{(k,k')})$ for every pair in $\mathscr{Z}_P$ under DGPs (1)–(4) under which all instrument pairs are invalid. As shown in Theorem 3.1, $\sqrt{n}(\widehat{\beta}^1_{(k,k')} - \beta^1_{(k,k')}) \to 0$ in probability if $(z_k, z_{k'})$ is invalid. Our simulation results show that overall, the RMSEs are

Table 5.4
Validity pair set estimation for DGP (3).

n	c	(1, 2)	(1, 3)	(1, 4)	(2, 3)	(2, 4)	(3, 4)
1230	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.000	0.000	0.000	0.002
	0.7	0.000	0.005	0.022	0.002	0.000	0.050
	0.8	0.000	0.033	0.141	0.064	0.004	0.282
	0.9	0.000	0.164	0.337	0.270	0.035	0.668
	1	0.003	0.436	0.630	0.556	0.182	0.922
2460	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.000	0.000	0.000	0.000
	0.7	0.000	0.000	0.001	0.000	0.000	0.001
	0.8	0.000	0.000	0.006	0.002	0.000	0.013
	0.9	0.000	0.003	0.050	0.055	0.000	0.193
	1	0.000	0.058	0.214	0.237	0.006	0.593

Table 5.5
Validity pair set estimation for DGP (4).

n	c	(1, 2)	(1, 3)	(1, 4)	(2, 3)	(2, 4)	(3, 4)
1230	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.000	0.000	0.000	0.030
	0.7	0.000	0.001	0.001	0.036	0.002	0.302
	0.8	0.002	0.023	0.051	0.269	0.054	0.700
	0.9	0.043	0.173	0.239	0.606	0.241	0.914
	1	0.180	0.433	0.530	0.843	0.508	0.983
2460	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.000	0.000	0.000	0.016
	0.7	0.000	0.000	0.000	0.031	0.000	0.220
	0.8	0.000	0.002	0.015	0.261	0.021	0.613
	0.9	0.010	0.030	0.136	0.649	0.150	0.865
	1	0.114	0.168	0.381	0.883	0.395	0.962

getting closer to 0 as n increases for $c = 0.6$. For larger c, the RMSEs for some invalid pairs may not be small enough in finite samples. Also note that for larger c, the RMSEs may not be decreasing in the sample size due to the rescaling by $\sqrt{n}$.

According to these simulation results, we suggest using $c = 0.6$ in applications. The results for other values of c may also be presented for consideration.

5.3. Empirical results

In this section, we apply VSIV to estimate the returns of college education. We choose the tuning parameter τ_n^g using formula (5.1) in Section 5.2.2. The tuning parameter τ_n^g depends on the user-specified constant c. Table 5.11 presents the estimated validity pair set for a grid of values for c. The simulation results in the previous section suggest choosing $c = 0.6$. For this choice of c, only the pairs $(2,4)$ and $(3,4)$ are selected, while the other pairs are screened out, so that the estimated validity pair set is $\widehat{\mathscr{Z}_0} = \{(2,4),(3,4)\}$.[24] This result is consistent with the results in Li et al. (2024) in that for small values of Z, the exclusion condition may fail.

The resulting VSIV estimates reported in the last row of Table 5.11 are $\widehat{\beta}^1_{(2,4)} = 0.542$ and $\widehat{\beta}^1_{(3,4)} = 0.652$. The corresponding 95% confidence intervals based on heteroskedasticity-robust standard errors are $[0.38493, 0.69954]$ and $[0.28062, 1.0233]$, respectively.

[24] In Appendix F.3, we present the results of a standard IV validity test.

Table 5.6
Coverage rates of the confidence intervals for DGP (0).

n	c	(1, 2)	(1, 3)	(1, 4)	(2, 3)	(2, 4)	(3, 4)
1230	0.1	1.000	1.000	1.000	1.000	1.000	1.000
	0.2	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	1.000	1.000	1.000	1.000	1.000
	0.4	1.000	1.000	1.000	1.000	1.000	1.000
	0.5	1.000	0.999	0.994	0.999	0.997	0.991
	0.6	1.000	0.984	0.964	0.995	0.968	0.955
	0.7	0.997	0.974	0.960	0.979	0.953	0.949
	0.8	0.985	0.969	0.959	0.971	0.952	0.948
	0.9	0.978	0.969	0.959	0.970	0.952	0.948
	1	0.972	0.969	0.959	0.970	0.952	0.948
2460	0.1	1.000	1.000	1.000	1.000	1.000	1.000
	0.2	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	1.000	1.000	1.000	1.000	1.000
	0.4	1.000	1.000	1.000	1.000	1.000	1.000
	0.5	1.000	0.995	0.979	1.000	0.983	0.979
	0.6	1.000	0.966	0.964	0.988	0.940	0.966
	0.7	0.986	0.953	0.954	0.967	0.934	0.965
	0.8	0.961	0.952	0.954	0.964	0.934	0.965
	0.9	0.954	0.952	0.954	0.963	0.934	0.965
	1	0.953	0.952	0.954	0.963	0.934	0.965

Table 5.7
RMSEs for DGP (1).

n	c	(1, 2)	(1, 3)	(1, 4)	(2, 3)	(2, 4)	(3, 4)
1230	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.940	0.332	0.000	0.204	3.067
	0.7	0.000	2.751	1.843	3.642	0.877	8.122
	0.8	0.000	7.177	3.444	11.820	2.392	15.625
	0.9	9.110	12.447	4.930	23.042	4.150	23.877
	1	21.949	19.690	6.233	32.068	6.184	29.984
2460	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.421	0.000	0.000	0.681
	0.7	0.000	1.996	1.431	1.498	0.734	4.616
	0.8	2.945	3.824	3.467	7.503	1.616	10.312
	0.9	2.945	7.404	5.354	17.387	3.253	19.648
	1	7.884	14.607	7.466	32.231	5.749	30.738

6. Conclusion

We propose an approach for estimating LATEs when the instruments are partially invalid. We focus on settings where the instruments are pairwise valid. Under pairwise validity, there are two types of instrument value pairs: Pairs for which the LATE assumptions hold and pairs for which the LATE assumptions fail due to violations of exclusion, independence, monotonicity, or combinations thereof. Pairwise validity is a natural starting point and useful in applications in which it is difficult to determine which LATE assumptions fail and how. However, in settings where researchers have information about how exactly the LATE assumptions fail, it would be interesting to consider intermediate cases that incorporate such information. Under restrictions on which LATE assumptions fail and how, it is often possible to derive nontrivial bounds on LATEs (e.g., Huber, 2014; Noack, 2021; Kédagni, 2023; Cui et al., 2024).

Throughout this paper, we focus on average treatment effects for the compliers. However, under pairwise validity, both marginal potential outcome distributions are identified for each $(z, z') \in \mathscr{Z}_{\bar{M}}$. Therefore, another interesting direction for future research is to extend VSIV to allow for the estimation of distributional treatment effects for the compliers, such as local quantile treatment effects (e.g., Abadie et al., 2002; Frölich and Melly, 2013; Melly and Wüthrich, 2017).

Table 5.8
RMSEs for DGP (2).

n	c	(1, 2)	(1, 3)	(1, 4)	(2, 3)	(2, 4)	(3, 4)
1230	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.492	0.000	0.000	2.559
	0.7	0.000	0.000	6.044	0.000	2.065	4.397
	0.8	0.000	1.514	13.078	4.721	5.465	9.519
	0.9	0.000	3.599	20.644	14.404	9.314	14.755
	1	3.964	6.682	26.301	23.569	14.271	20.772
2460	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.159	0.000	0.000	0.000
	0.7	0.000	0.000	6.373	0.000	0.989	2.103
	0.8	0.000	0.000	15.145	2.702	3.455	6.309
	0.9	0.000	0.000	22.974	5.737	7.433	10.970
	1	0.000	0.000	28.708	13.680	10.664	16.494

Table 5.9
RMSEs for DGP (3).

n	c	(1, 2)	(1, 3)	(1, 4)	(2, 3)	(2, 4)	(3, 4)
1230	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.000	0.000	0.000	0.333
	0.7	0.000	0.307	0.366	0.384	0.000	1.571
	0.8	0.000	0.928	0.921	3.482	0.149	4.226
	0.9	0.000	2.612	1.429	7.000	0.460	6.845
	1	0.340	4.145	2.119	10.064	1.239	8.501
2460	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.000	0.000	0.000	0.000
	0.7	0.000	0.000	0.045	0.000	0.000	0.022
	0.8	0.000	0.000	0.262	0.175	0.000	0.631
	0.9	0.000	0.194	0.668	2.536	0.000	3.089
	1	0.000	1.160	1.327	5.498	0.288	6.092

Acknowledgments

We are grateful to the Editor (Elie Tamer), Associate Editor, two anonymous referees, Zheng Fang, Martin Huber, Toru Kitagawa, Julian Martinez-Iriarte, Désiré Kédagni, Ismael Mourifié, Xiaoxia Shi, and all seminar and conference participants for their insightful suggestions and comments. We thank Désiré Kédagni for sharing the data for the empirical application with us. Sun acknowledges funding by the National Natural Science Foundation of China [Grant Number 72103004]. Wüthrich is also affiliated with CESifo. The usual disclaimer applies.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2025.106009.

Table 5.10
RMSEs for DGP (4).

n	c	(1, 2)	(1, 3)	(1, 4)	(2, 3)	(2, 4)	(3, 4)
1230	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.000	0.000	0.000	1.976
	0.7	0.000	0.024	0.035	2.166	0.304	5.750
	0.8	1.301	1.212	1.118	7.129	1.119	9.226
	0.9	4.496	3.836	2.472	10.658	2.444	10.591
	1	8.876	5.881	3.687	12.397	3.532	11.038
2460	0.1	0.000	0.000	0.000	0.000	0.000	0.000
	0.2	0.000	0.000	0.000	0.000	0.000	0.000
	0.3	0.000	0.000	0.000	0.000	0.000	0.000
	0.4	0.000	0.000	0.000	0.000	0.000	0.000
	0.5	0.000	0.000	0.000	0.000	0.000	0.000
	0.6	0.000	0.000	0.000	0.000	0.000	1.407
	0.7	0.000	0.000	0.000	1.994	0.000	5.094
	0.8	0.000	0.398	0.786	6.534	0.755	8.282
	0.9	1.803	1.366	1.715	10.912	1.790	9.792
	1	5.893	3.564	3.012	12.921	2.941	10.315

Table 5.11
Validity pair set estimation in application.

c	(1, 2)	(1, 3)	(1, 4)	(2, 3)	(2, 4)	(3, 4)
0.1	✗	✗	✗	✗	✗	✗
0.2	✗	✗	✗	✗	✗	✗
0.3	✗	✗	✗	✗	✗	✗
0.4	✗	✗	✗	✗	✗	✗
0.5	✗	✗	✗	✗	✗	✗
0.6	✗	✗	✗	✗	✓	✓
0.7	✗	✓	✗	✓	✓	✓
0.8	✗	✓	✓	✓	✓	✓
0.9	✗	✓	✓	✓	✓	✓
1	✓	✓	✓	✓	✓	✓
$\widehat{\beta}^{1}_{(k,k')}$	–	–	–	–	0.542	0.652

References

Abadie, A., Angrist, J., Imbens, G., 2002. Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. Econometrica 70 (1), 91–117.
Angrist, J.D., Evans, W.N., 1998. Children and their parents' labor supply: Evidence from exogenous variation in family size. Am. Econ. Rev. 88 (3), 450–477.
Angrist, J.D., Fernández-Val, I., 2013. ExtrapoLATE-ing: External validity and overidentification in the LATE framework. In: Acemoglu, D., Arellano, M., Dekel, E. (Eds.), Advances in Economics and Econometrics: Tenth World Congress. In: Econometric Society Monographs, Cambridge University Press, pp. 401–434.
Angrist, J.D., Imbens, G.W., 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. J. Amer. Statist. Assoc. 90 (430), 431–442.
Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. J. Amer. Statist. Assoc. 91 (434), 444–455.
Angrist, J.D., Krueger, A.B., 1991. Does compulsory school attendance affect schooling and earnings? Q. J. Econ. 106 (4), 979–1014.
Angrist, J.D., Pischke, J.-S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.
Angrist, J.D., Pischke, J.-S., 2014. Mastering Metrics: The Path from Cause to Effect. Princeton University Press.
Armstrong, T.B., Kolesár, M., 2021. Sensitivity analysis using approximate moment condition models. Quant. Econ. 12 (1), 77–108.
Balke, A., Pearl, J., 1997. Bounds on treatment effects from studies with imperfect compliance. J. Amer. Statist. Assoc. 92 (439), 1171–1176.
Beare, B.K., Shi, X., 2019. An improved bootstrap test of density ratio ordering. Econ. Stat. 10, 9–26.
Bound, J., Jaeger, D.A., Baker, R.M., 1995. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. J. Amer. Statist. Assoc. 90 (430), 443–450.
Brinch, C.N., Mogstad, M., Wiswall, M., 2017. Beyond LATE with a discrete instrument. J. Political Econ. 125 (4), 985–1039.
Buckles, K.S., Hungerman, D.M., 2013. Season of birth and later outcomes: Old questions, new answers. Rev. Econ. Stat. 95 (3), 711–724.
Card, D., 1993. Using geographic variation in college proximity to estimate the return to schooling. Working Paper Series, Working Paper 4483, National Bureau of Economic Research.
Carneiro, P., Heckman, J.J., Vytlacil, E.J., 2011. Estimating marginal returns to education. Am. Econ. Rev. 101 (6), 2754–2781.
Carr, T., Kitagawa, T., 2021. Testing instrument validity with covariates. arXiv:2112.08092.
Conley, T.G., Hansen, C.B., Rossi, P.E., 2012. Plausibly exogenous. Rev. Econ. Stat. 94 (1), 260–272.
Cui, Y., Kédagni, D., Wu, H., 2024. Robust identification in randomized experiments with noncompliance. arXiv:2408.03530.
Dahl, C.M., Huber, M., Mellace, G., 2023. It is never too LATE: A new look at local average treatment effects with or without defiers. Econom. J. 26 (3), 378–404.
De Chaisemartin, C., 2017. Tolerating defiance? Local average treatment effects without monotonicity. Quant. Econ. 8 (2), 367–396.
Farbmacher, H., Guber, R., Klaassen, S., 2022. Instrument validity tests with causal forests. J. Bus. Econom. Statist. 40 (2), 605–614.

Folland, G.B., 1999. Real Analysis: Modern Techniques and Their Applications. John Wiley & Sons.
Frandsen, B., Lefgren, L., Leslie, E., 2023. Judging judge fixed effects. Am. Econ. Rev. 113 (1), 253–277.
Frölich, M., 2007. Nonparametric IV estimation of local average treatment effects with covariates. J. Econometrics 139 (1), 35–75.
Frölich, M., Melly, B., 2013. Unconditional quantile treatment effects under endogeneity. J. Bus. Econom. Statist. 31 (3), 346–357.
Goff, L., 2024. A vector monotonicity assumption for multiple instruments. J. Econometrics 241 (1), 105735.
Goh, G., Yu, J., 2022. Causal inference with some invalid instrumental variables: A quasi-Bayesian approach. Oxf. Bull. Econ. Stat. 84 (6), 1432–1451.
Heckman, J., Tobias, J.L., Vytlacil, E., 2001. Four parameters of interest in the evaluation of social programs. South. Econ. J. 68 (2), 210–223.
Heckman, J., Tobias, J.L., Vytlacil, E., 2003. Simple estimators for treatment parameters in a latent-variable framework. Rev. Econ. Stat. 85 (3), 748–755.
Heckman, J.J., Vytlacil, E., 2005. Structural equations, treatment effects, and econometric policy evaluation. Econometrica 73 (3), 669–738.
Huber, M., 2014. Sensitivity checks for the local average treatment effect. Econom. Lett. 123 (2), 220–223.
Huber, M., Mellace, G., 2015. Testing instrument validity for LATE identification based on inequality moment constraints. Rev. Econ. Stat. 97 (2), 398–411.
Huber, M., Wüthrich, K., 2019. Local average and quantile treatment effects under endogeneity: A review. J. Econ. Methods 8 (1), 20170007.
Imbens, G.W., 2014. Instrumental variables: An econometrician's perspective. Statist. Sci. 29 (3), 323–358.
Imbens, G.W., Angrist, J.D., 1994. Identification and estimation of local average treatment effects. Econometrica 62 (2), 467–475.
Imbens, G.W., Rubin, D.B., 1997. Estimating outcome distributions for compliers in instrumental variables models. Rev. Econ. Stud. 64 (4), 555–574.
Imbens, G.W., Rubin, D.B., 2015. Causal Inference in Statistics, Social, and Biomedical Sciences. Cambridge University Press.
Jiang, H., Sun, Z., 2023. Testing partial instrument monotonicity. Econom. Lett. 233, 111400.
Kane, T.J., Rouse, C.E., 1993. Labor market returns to two- and four-year colleges: Is a credit a credit and do degrees matter? Working Paper Series, Working Paper 4268, National Bureau of Economic Research.
Kédagni, D., 2023. Identifying treatment effects in the presence of confounded types. J. Econometrics 234 (2), 479–511.
Kédagni, D., Mourifié, I., 2020. Generalized instrumental inequalities: Testing the instrumental variable independence assumption. Biometrika 107 (3), 661–675.
Kitagawa, T., 2015. A test for instrument validity. Econometrica 83 (5), 2043–2063.
Kowalski, A.E., 2023. How to examine external validity within an experiment. J. Econ. Manag. Strat. 32 (3), 491–509.
Lee, S., Song, K., Whang, Y.-J., 2018. Testing for a general class of functional inequalities. Econometric Theory 34 (5), 1018–1064.
Li, L., Kédagni, D., Mourifié, I., 2024. Discordant relaxations of misspecified models. Quant. Econ. 15 (2), 331–379.
Linton, O., Song, K., Whang, Y.-J., 2010. An improved bootstrap test of stochastic dominance. J. Econometrics 154 (2), 186–202.
Manski, C.F., Pepper, J.V., 2000. Monotone instrumental variables: With an application to the returns to schooling. Econometrica 68 (4), 997–1010.
Manski, C.F., Pepper, J.V., 2009. More on monotone instrumental variables. Econom. J. 12 (suppl_1), S200–S216.
Masten, M.A., Poirier, A., 2021. Salvaging falsified instrumental variable models. Econometrica 89 (3), 1449–1469.
Melly, B., Wüthrich, K., 2017. Local quantile treatment effects. In: Koenker, R., Chernozhukov, V., He, X., Peng, L. (Eds.), Handbook of Quantile Regression. Chapman and Hall/CRC, pp. 145–164.
Mogstad, M., Santos, A., Torgovitsky, A., 2018. Using instrumental variables for inference about policy relevant treatment parameters. Econometrica 86 (5), 1589–1619.
Mogstad, M., Torgovitsky, A., Walters, C.R., 2021. The causal interpretation of two-stage least squares with multiple instrumental variables. Am. Econ. Rev. 111 (11), 3663–3698.
Mourifié, I., Wan, Y., 2017. Testing local average treatment effect assumptions. Rev. Econ. Stat. 99 (2), 305–313.
Nevo, A., Rosen, A.M., 2012. Identification with imperfect instruments. Rev. Econ. Stat. 94 (3), 659–671.
Noack, C., 2021. Sensitivity of LATE estimates to violations of the monotonicity assumption. arXiv:2106.06421.
Slichter, D., 2014. Testing instrument validity and identification with invalid instruments. Working Paper.
Sun, Z., 2023. Instrument validity for heterogeneous causal effects. J. Econometrics 237 (2, Part A), 105523.
Sun, Z., Beare, B.K., 2021. Improved nonparametric bootstrap tests of Lorenz dominance. J. Bus. Econom. Statist. 39 (1), 189–199.
Thornton, R.L., 2008. The demand for, and impact of, learning HIV status. Am. Econ. Rev. 98 (5), 1829–1863.
Wüthrich, K., 2020. A comparison of two quantile models with endogeneity. J. Bus. Econom. Statist. 38 (2), 443–456.