# THE INTERPETATION OF 2SLS WITH A CONTINUOUS INSTRUMENT: A WEIGHTED LATE REPRESENTATION

LUIS A. F. ALVAREZ[a,*] AND RODRIGO TONETO[b]

[a] *University of São Paulo, 05508-010 São Paulo, Brazil*

[b] *Queen Mary University of London, E14NS London, United Kingdom*

Abstract. This note introduces a novel weighted local average treatment effect representation for the two-stages least-squares (2SLS) estimand in the continuous instrument with binary treatment case. Under standard conditions, we obtain weights that are nonnegative, integrate to unity, and assign larger values to instrument support points that deviate from their average. Our representation does not require instruments to be discretized nor relies on limiting arguments, such as those used in the definition of the marginal treatment effect (MTE). The pattern of the weights also has a clear interpretation. We believe these features of the representation to be useful for applied researchers when communicating their results. As a direct byproduct of our approach, we also obtain a representation of the 2SLS estimand as a weighted average of treatment effects among "marginal compliance" groups, without having to resort to the threshold-crossing representation underlying the MTE construction. The latter representation has an intuitive interpretation as well.

**Keywords:** instrumental variables, local average treatment effects; underlying weights.
**JEL classification:** C21, C23, C26.

## 1. Introduction

Instrumental variable (IV) methods constitute one of the worhorses in the research toolkit of applied economists (Angrist and Krueger, 2001; Imbens, 2014; Abadie and Cattaneo, 2018). Since the seminal work of Imbens and Angrist (1994), henceforth IA, it is well known that, in a potential outcomes framework – and under exclusion, independence, relevance and monotonicity assumptions –, the instrumental variable estimator with binary treatment and binary instrument identifies a local average treatment effect (LATE) in the subpopulation whose treatment adoption is affected by the instrument (compliers). This representation has been further extended to accomodate discrete (Angrist and Imbens, 1995) and continuous (Angrist et al., 2000) treatments. Recent research on identification with instrumental variables methods in the potential outcomes

*E-mail address*: `luisfantozzialvarez@gmail.com (Alvarez)`, `r.toneto@qmul.ac.uk (Toneto)`.

*Date*: March, 2024.

* Corresponding author.

framework has explored, *inter alia*, the nature of monotonicity assumptions under choice-theoretic perspectives (Heckman and Pinto, 2018; Mogstad et al., 2021); the interpetation of 2SLS estimands under different approaches to controlling for covariates (Kolesár et al., 2013; Blandhol et al., 2022; Słoczyński, 2022), and the interpretation of distinct weighting schemes underlying different IV estimands (Mogstad et al., 2018; Coussens and Spiess, 2021; Escanciano et al., 2023).

In this note, we revisit the interpretation of the 2SLS estimand in the case of a binary treatment and continuous instrument. In this setting, applied researchers are often faced with two distinct recommendations from the literature concerning the interpretation of 2SLS estimates. One branch advises researchers to *discretize* their instruments (Angrist and Pischke, 2009, p. 139-140), in which case available representations of IV estimands as weighted LATEs may be resorted to. Another branch recommends results be interpreted as weighted averages of marginal treatment effects (MTEs) (Heckman and Vytlacil, 2005; Heckman et al., 2006). These approaches have not been unanimously adopted, though. First, the use of continuous (undiscretized) instruments is quite common in the applied literature, as evidenced by the use of weather and distance-related instrumental variables. Second, even though the threshold-crossing model that motivates the definition of the MTE is known to be equivalent to the framework of IA (Vytlacil, 2002); and that, in the former, the MTE may be seen as a limit-form of the LATE; applied researchers are much more accostumed with non-limit forms of the LATE, often going large ways in interpreting their estimates as such, even in cases where such interpretation is not warranted (Blandhol et al., 2022).

This note aims to complement the literature by providing, to the best of our knowledge, a novel (non-limit) weighted LATE interpretation of 2SLS estimands in the continuous instrument *cum* binary treatment case. We provide conditions under which a class of Wald estimands that nests 2SLS as a particular case may be interpreted as a weighted-average of complier treatment effects, where compliance groups are defined with respect to treatment status at the (conditional on controls) average instrument value. The weights are nonnegative, integrate to unity, and assign larger values to instrument support points that deviate from the (conditional on controls) average. As we argue below, such patterns lead to weighted averages with an intuitive interpretation. Our results are derived under the same set of assumptions of IA (also Angrist et al., 1996), which we couple with the requirement that, in the specification adopted, covariates are controlled for in a sufficiently flexible manner. The latter restriction is known to be sufficient, under the IA assumptions, for 2SLS estimands to produce nonnegative weights (Blandhol et al., 2022). We further show that, as an immediate byproduct of our representation result, we obtain a representation of our class of Wald IV estimands in terms of "marginal compliance" groups, without having to resort to the the threshold-crossing model that generates the MTE representation. These weights also have an intuitive interperation, assigning larger values to compliers for which there is more variation available to estimate effects.

The remainder of this note is organized as follows. Section 2 introduces the class of Wald estimands considered and presents the representation result. Section 3 concludes. The Online Appendix contains the proof of the main result, as well as an extension of our representation to an alternative 2SLS estimand, a discussion on how to estimate the underlying weights of each representation, and a comparison with the approach of discretizing instruments.

## 2. Weighted LATE representation of a class of Wald Estimands under a continuous instrument

Suppose the researcher is interested in assessing the causal effect of a binary treatment $D \in \{0,1\}$ on an outcome $Y$. The researcher has access to a continuous scalar instrument $Z$, and a set of controls $X$. We assume that second moments of $Y$ and $Z$ exist. We consider a Wald estimand which aims to estimate the causal effect of $D$ by leveraging covariation between $Z$ and $D$, after partialling out covariation between $Z$ and $X$. This estimand is given by:

$$\beta_{\text{Wald}} = \frac{\mathbb{E}[(Z - g^*(X))Y]}{\mathbb{E}[(Z - g^*(X))D]}\,, \tag{1}$$

where $g^*$ is the $L_2(\mathbb{P})$ projection of $Z$ on a space $\mathcal{G}$ of scalar functions of $X$ with finite second moment, i.e.

$$g^* \in \operatorname{argmin}_{g \in \mathcal{G}} \mathbb{E}[(Z - g^*(X))^2]\,. \tag{2}$$

The Wald estimand (1) explores variation of $Z$ after controlling for potential confounding due to $X$. These variables may be controlled for in a possibly nonlinear manner, for example by functions $g \in \mathcal{G}$ that vary nonlinearly in $X$. As a special case, if one considers a transformation $p$ from $X$ to $\mathbb{R}^k$, and takes $\mathcal{G} = \{\gamma' p(\cdot) : \gamma \in \mathbb{R}^k\}$, then an application of the partitioned inverse formula shows that $\beta_{\text{Wald}}$ is equal to the 2SLS estimand of $\beta$ in the linear system:

$$\begin{aligned} D &= \alpha Z + \gamma' p(X) + u \\ Y &= \beta D + \omega' p(X) + v \end{aligned}\,. \tag{3}$$

We consider the interpretation of (1) in a potential outcomes framework. Following IA (also Angrist et al., 1996), we assume that:

**Assumption 1.** *We assume that:*

*(1)* ***Potential treatments:*** *observed treatment status is given by $D = D(Z)$, where $\mathcal{D} = \{D(z) : z \in \mathcal{Z}\}$ are the potential treatment statuses associated with different values of the instrument, with $\mathcal{Z} \subseteq \mathbb{R}$ denoting the instrument support.*

(2) ***Potential outcomes:*** *observed outcomes are given by* $Y = DY(1, Z) + (1 - D)Y(0, Z)$*, where* $Y(d, z)$ *are the potential outcomes associated with treatment status* $d \in \{0, 1\}$ *and instrument value* $z \in \mathcal{Z}$*.*
(3) ***Exclusion restriction:*** *for each* $d \in \{0, 1\}$, $Y(d, z) = Y(d, z') =: Y(d)\ \forall z, z' \in \mathcal{Z}$*.*
(4) ***Conditional independence:*** *conditionally on* $X$*,* $Z$ *is independent of* $\{Y(0), Y(1), \mathcal{D}\}$*.*
(5) ***Monotonicity:*** *either* $\mathbb{P}[D(z) \leq D(z')] = 1$ *for every* $z, z' \in \mathcal{Z}$ *with* $z \leq z'$*; or* $\mathbb{P}[D(z) \geq D(z')] = 1$ *for every* $z, z' \in \mathcal{Z}$ *with* $z \leq z'$*.*

The next proposition is our main result.

**Proposition 1.** *Suppose Assumption 1 holds. Let* $\psi(X)$ *be a version of* $\mathbb{E}[Z|X]$*. In addition, suppose the following conditions hold:*

(1) ***Support condition:*** *the support* $\mathcal{Z}$ *is given by* $[\underline{z}, \overline{z}]$*, where* $\underline{z}, \overline{z} \in \mathbb{R} \cup \{-\infty, \infty\}$*.*
(2) ***Moments:*** $Y(1)$*,*$Y(0)$ *and* $Z$ *have finite second moments.*
(3) ***Relevance:*** $\mathbb{E}[(Z - g^*(X))D] \neq 0$*.*
(4) ***Flexible specification:*** $\psi \in \mathcal{G}$*.*

*We then have that the estimand* (1) *is well-defined, and that:*

$$\beta_{Wald} = \mathbb{E}[w(X, Z)\Delta(X, Z)]\,,$$

*where*

$$\Delta(x, z) = \mathbb{E}[Y(1) - Y(0)|X = x, D(z) \neq D(\psi(x))]\,,$$

*and* $w(x, z) = \frac{\omega(x,z)}{\mathbb{E}[\omega(X,Z)]}$*, with:*

$$\omega(x, z) = |z - \psi(x)|\mathbb{P}[D(z) \neq D(\psi(x))|X = x]\,.$$

*Proof.* See Online Appendix A. □

Proposition 1 shows that, under the stated assumptions, the Wald estimand (1) may be written as a weighted average of complier treatment effects. In the representation of Proposition 1, compliance is defined with respect to potential treatment status at the average instrument value $\mathbb{E}[Z|X = x]$, with weights being attached to average treatment effects in subpopulations that would change their treatment status upon being offered a shift of instrument value from $\mathbb{E}[Z|X = x]$ to $z$, for different values of $z \in [\underline{z}, \overline{z}]$. The weights in the representation are nonnegative and average to unity. Moreover, due to the monotonicity assumption, the weights $w(x, z)$ are nonincreasing in $z$ for $z < \mathbb{E}[Z|X = x]$ and nondecreasing in $z$ for $z > \mathbb{E}[Z|X = x]$. Consequently, the weighted average assigns larger weights to more extreme values of $z$.

The representation in Proposition 1 hinges crucially on the interval support assumption on the instrument, as it ensures that the potential treatment status at the average $\mathbb{E}[Z|X = x]$ is well-defined. In addition, we note that, since the representation uses a (conditional on $X$) fixed reference potential treatment status in defining compliers, it may be easier to interpret than varying-reference-level ones (Cornelissen et al., 2016). To see this point, first note that our fixed-level representation involves overlapping subpopulations. Indeed, for any $z' > z > \psi(x)$, $\Delta(x, z')$ includes in its average all compliers averaged in $\Delta(x, z)$; the same holding true for $z' < z < \psi(x)$. Since more extreme values of the instrument are precisely those assigned larger weights in the representation, it follows that the representation assigns larger weights to the most encompassing complier subpopulations, a useful feature in the interpretation of results.

In spite of the attractiveness of our representation, it should be noted that, in some settings, it may be also useful to interpret the estimand as a weighted average of treatment effects in disjoint subpopulations. As we show below, one immediate corollary of Proposition 1 is a representation of (1) in terms of nonoverlapping "marginal compliance" groups. Interestingly, our result follows without having to resort to the threshold-crossing representation that motivates the MTE, which may be a further useful feature for applied researchers in communicating their empirical results.

In what follows, define the marginal compliance group of an individual as the variable $C$ that equals $c$ if:

$$\lim_{z \uparrow c} D(z) \neq D(c) \, ,$$

i.e. $C$ is the smallest instrument value required for an individual to change her behavior. If the individual is not a complier, we set $C = \varnothing$.

**Corollary 1.** *Suppose that the Assumptions required in Proposition 1 hold. Suppose that $C|X, C \neq \varnothing$ admits a regular conditional Lebesgue density $f^*_{C|X}(\cdot|\cdot)$. We then have that:*

$$\beta_{Wald} = \int \int_{\underline{z}}^{\overline{z}} w^*(x, c)\Delta^*(x, c) \; dc \; \mathbb{P}_X(dx) \, ,$$

*where*

$$\Delta^*(x, c) = \mathbb{E}[Y(1) - Y(0)|X = x, C = c] \, ,$$

*and* $w^*(x, c) = \frac{\omega^*(x,c)}{\int \int_{\underline{z}}^{\overline{z}} \omega^*(a,b) \; db \; \mathbb{P}_X(da)}$, *with*

$$\omega^*(x, c) = \begin{cases} \mathbb{E}[|Z - \psi(x)|\mathbf{1}\{Z < c\}|X = x]f^*_{C|X}(c|x), & \text{if } c < \psi(x) \\ \mathbb{E}[|Z - \psi(x)|\mathbf{1}\{Z \;\; \geq c\}|X = x]f^*_{C|X}(c|x), & \text{if } c \geq \psi(x) \end{cases}$$

*Proof.* Note that

$$\mathbb{E}[(Y(1) - Y(0))\mathbf{1}\{D(z) \neq D(\psi(x))\}|X = x] = \int_{\underline{z}}^{\overline{z}} \Delta^*(x, c)\mathbf{1}\{c \in [z, \psi(x)] \cup [\psi(x), z]\} f^*_C(c|x) dc \, .$$

The conclusion then follows from Fubini theorem. □

The weights $\omega^*(x,c)$ in the representation of Corollary 1 depend on two quantities. First, they are proportional to $f^*_{C|X}(c|x)$, which reflects the mass of marginal compliers at value $c$, in the subpopulation defined by $X = x$. Second, they depend on:

$$\kappa(c|x) = \begin{cases} \mathbb{E}[|Z-\psi(x)|\mathbf{1}\{Z<c\}|X=x], & \text{if } c<\psi(x) \\ \mathbb{E}[|Z-\psi(x)|\mathbf{1}\{Z\geq c\}|X=x], & \text{if } c\geq\psi(x) \end{cases}.$$

Notice that, for each $x$, the function $c \mapsto \kappa(c|x)$ is nondecreasing in $c$, for values $c < \psi(x)$, and nonincreasing in $c$, for values $c > \psi(x)$. Therefore, the term $\kappa(c|x)$ assigns larger values the closer $c$ is to $\psi(x)$. This pattern reflects the fact that marginal compliers at $\psi(x)$ are precisely those for which the most "variation" is available to estimate effects. Indeed, notice that, for a marginal complier with $C = c$, instrument values $z > c$ induce the individual to take an action, whereas values $z' < c$ induce the opposite action. If the distribution of $Z|X$ is symmetric, $\psi(x)$ coincides with the median of the conditional distribution, and $\psi(x)$ maximizes the mass of available comparisons $(z, z')$ by exactly balancing the available points below and above it.[1] More generally, the average compliance point $C = \psi(x)$ is precisely the one that balances a weighted measure of available points below and above it, as captured by the fact that $\int_{\psi(x)}^{\overline{z}} |z-\psi(x)| f_{Z|X}(z|x)dz = \int_{\underline{z}}^{\psi(x)} |\psi(x)-z| f_{Z|X}(z|x)dz$. This interpretation of the underlying weights assigning higher values to subpopulations where it is "easier" to estimate effects, in the sense of there being more available variation, is similar to existing results for linear regression in selection-on-observables settings (Angrist, 1998; Angrist and Pischke, 2009; Goldsmith-Pinkham et al., 2024).

**Remark 1** (Connection with Theorem 2 of IA)**.** Consider a setting where we only include an intercept among the controls (in which case we take $\mathcal{G} = \mathbb{R}$), and that $Z$ is discretely supported, with $K$ support points $c_1 < c_2 < \ldots < c_K$. In this case, Theorem 2 of IA applies, and we obtain the following representation for the Wald estimand:

$$\beta_{\text{Wald}} = \frac{\sum_{k=1}^{K-1} \omega_k \mathbb{E}[Y(1)-Y(0)|D(c_{k+1}) \neq D(c_k)]}{\sum_{k=1}^{K-1} \omega_k}, \tag{4}$$

where $\omega_k = |\mathbb{P}[D=1|Z=c_{k+1}] - \mathbb{P}[D=1|Z=c_k]|\mathbb{E}[(Z-\mathbb{E}[Z])\mathbf{1}\{Z\geq c_k\}]$. By noting that, under the monotonicity assumption, $|\mathbb{P}[D=1|Z=c_{k+1}] - \mathbb{P}[D=1|Z=c_k]| = \mathbb{P}[D(c_{k+1}) \neq D(c_k)]$, and that, for $c_k < \mathbb{E}[Z]$, $\mathbb{E}[(Z-\mathbb{E}[Z])\mathbf{1}\{Z\geq c_k\}] = \mathbb{E}[|Z-\mathbb{E}[Z]|\mathbf{1}\{Z<c_k\}]$, we conclude that the IA weights may be rewritten as:

[1] To see this, note that the mass of available comparisons $\int_{\psi}^{\overline{z}}\int_{\underline{z}}^{\psi} f_{Z|X}(z'|x)f_{Z|X}(z|x)dz'dz$ is maximized at the choice $\psi^*$ such that $\int_{\underline{z}}^{\psi^*} f_{Z|X}(z'|x)dz' = \int_{\psi^*}^{\overline{z}} f_{Z|X}(z|x)dz$, i.e. the median.

$$\omega_k = \begin{cases} \mathbb{P}[D(c_{k+1}) \neq D(c_k)]\mathbb{E}[|Z - \mathbb{E}[Z]|\mathbf{1}\{Z < c_k\}], & \text{if } c_k < \mathbb{E}[Z] \\ \mathbb{P}[D(c_{k+1}) \neq D(c_k)]\mathbb{E}[|Z - \mathbb{E}[Z]|\mathbf{1}\{Z \geq c_k\}], & \text{if } c_k \geq \mathbb{E}[Z] \end{cases}.$$

This shows that, in the case without covariates, the representation in Corollary 1 may be seen as a continuous counterpart of representation (4). Indeed, in a "heuristic limit" where the number of support points $K$ becomes dense in an interval $[\underline{z}, \overline{z}]$ of the real line, one would expect (4) to converge to our non-overlapping representation. Corollary 1 makes this argument precise, by properly defining what a complier is in the continuous case and offering a formal proof of the non-overlapping representation without relying on "heuristic limits", while also extending the representation to a class of Wald estimands that allows for the inclusion of covariates.

**Remark 2** (On the version of the monotonicity assumption)**.** We derive Proposition 1 under the requirement that the sign of the monotonicity be the same across the support of $X$. This contrasts with a weak monotonicity assumption (Słoczyński, 2022), whereby the direction of the monotonicity may vary across the different subpopulations defined by the values of $X$. In the binary (discrete scalar) instrument setting, Słoczyński (2022) (Blandhol et al., 2022, Proposition 10) show that, for 2SLS estimands of the system (3) to identify weighted averages of LATEs with nonnegative weights, the weak monotonicity assumption is not sufficient, whereas our stronger version is. This insufficiency is essentially due to the "inflexibility" of the first-stage of (3), which does not allow the sign of the relation between $Z$ and $D$ to vary with $X$. In the discrete $X$ setting, a "saturate-and-weight" strategy (Angrist and Pischke, 2009), which consists of running 2SLS with a full set of dummies for every support point of $X$ while also overidentifying the first-stage by including as instruments the interacton between each of these dummies with $Z$, is known to produce a weighted average of LATEs with nonnegative weights under weak monotonicity and a binary instrument. Online Appendix B shows the same is true in the continuous scalar instrument setting: we are able to obtain both overlapping and non-overlapping representations for the saturate-and-weight estimand under the weaker monotonicity requirement.

**Remark 3** (On the role of the flexible specification assumption)**.** Proposition 1 is derived under the requirement that the function class $\mathcal{G}$ is flexible enough so as to contain the conditional expectation function $\mathbb{E}[Z|X = x]$. It has been shown that this assumption is sufficient, in the IA setup, for 2SLS estimands of the system (3) to be represented as weighted averages of complier treatment effects with nonnegative weights (Blandhol et al., 2022). In the case that $X$ has a finite number of support points, the assumption is satisfied by relying on a saturated $\mathcal{G}$, i.e. by considering $p(X)$ as a vector of indicator functions of all possible values $x$ in the support of $X$. In settings where $X$ is more complex, e.g. it contains continuous entries or a finite but very large number of support points, it may be preferable to rely on machine-learning methods that estimate representation (1) while flexibly controlling for $X$ (Belloni et al., 2012; Chernozhukov et al., 2018).

**Remark 4** (Estimation of the weights)**.** In Online Appendix C, we discuss how the underlying weights in both representations may be estimated, which may be a useful tool for applied researchers in understanding which subpopulations receive larger weights in their target estimand. We also show how an estimator of the conditional probability function $\mathbb{P}[D = 1|Z, X]$, which is required in the computation of the weights, may be used for testing an implication of the monotonicity hypothesis in Assumption 1.

**Remark 5** (Comparison with discretization of continuous instruments)**.** Online Appendix D compares the estimand discussed in the main text with those obtained from discretizing a continuous $Z$ and using as instruments the indicators for the discretized support points. While, in the case without covariates, the latter approach enables the researcher to be agnostic about the encoding of $Z$ that ensures the validity of the monotonicity hypothesis in Assumption 1, these benefits may be outweighed in settings with covariates, where the researcher would have to saturate and interact $X$ with the dummies for the discretized $Z$ (and use these interactions as instruments) to ensure a representation with nonnegative weights. When this approach results in many instruments, the researcher may incur in estimation and inference problems. In contrast, we note that it is always possible to ensure the "correct" (monotonicity-ensuring) recoding of $Z$ by using a flexible estimator of $v(Z) = \mathbb{P}[D = 1|Z]$ as the instrument.

## 3. Conclusion

This note introduces a weighted LATE representation for the 2SLS estimand in the binary treatment with continuous instrument case. We have shown our non-limit representation has an intuitive interpretation. As a direct byproduct of our representation, we have also obtained a weighted LATE representation in terms of non-overlapping, marginal compliance groups. This alternative representation also has an intuitive interpretation. In the Online Appendix, we further show how the weights in both representations may be estimated, which may be a useful tool for applied researchers seeking to communicate which compliers are assigned larger weights by their estimand.

## Data availability

No data was used for the research described in the article.

## Acknowledgements

The authors thank an anonymous referee for their helpful comments and suggestions. This work was supported by Fapesp grant 2022/16691-1, and by Queen Mary University of London HSS Principal's Research Studentship.

## References

Abadie, A. and M. D. Cattaneo (2018). Econometric methods for program evaluation. *Annual Review of Economics 10*(1), 465–503.

Angrist, J. D. (1998). Estimating the labor market impact of voluntary military service using social security data on military applicants. *Econometrica 66*(2), 249–288.

Angrist, J. D., K. Graddy, and G. W. Imbens (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *The Review of Economic Studies 67*(3), 499–527.

Angrist, J. D. and G. W. Imbens (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association 90*(430), 431–442.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*(434), 444–455.

Angrist, J. D. and A. B. Krueger (2001, December). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives 15*(4), 69–85.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica 80*(6), 2369–2429.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2022). When is tsls actually late? Technical report, National Bureau of Economic Research.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018, 01). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1–C68.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2016). From late to mte: Alternative methods for the evaluation of policy interventions. *Labour Economics 41*, 47–60. SOLE/EALE conference issue 2015.

Coussens, S. and J. Spiess (2021). Improving inference from simple instruments through compliance estimation. *arXiv preprint arXiv:2108.03726*.

Escanciano, J. C., C. Caetano, and G. Caetano (2023). Robust identification in regression discontinuity designs with covariates.

Freyberger, J. and Y. Rai (2018). Uniform confidence bands: Characterization and optimality. *Journal of Econometrics 204*(1), 119–130.

Goldsmith-Pinkham, P., P. Hull, and M. Kolesár (2024). Contamination bias in linear regressions.

Heckman, J. J. and R. Pinto (2018). Unordered monotonicity. *Econometrica 86*(1), 1–35.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *The review of economics and statistics 88*(3), 389–432.

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation 1. *Econometrica 73*(3), 669–738.

Imbens, G. W. (2014). Instrumental Variables: An Econometrician's Perspective. *Statistical Science 29*(3), 323 – 358.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Kolesár, M. et al. (2013). Estimation in an instrumental variables model with treatment effect heterogeneity. Technical report.

Mikusheva, A. and L. Sun (2021, 12). Inference with Many Weak Instruments. *The Review of Economic Studies 89*(5), 2663–2686.

Mogstad, M., A. Santos, and A. Torgovitsky (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica 86*(5), 1589–1619.

Mogstad, M., A. Torgovitsky, and C. R. Walters (2021, November). The causal interpretation of two-stage least squares with multiple instrumental variables. *American Economic Review 111*(11), 3663–98.

Słoczyński, T. (2022). When should we (not) interpret linear iv estimands as late?

Vytlacil, E. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica 70*(1), 331–341.

## Online Appendix of "The interpetation of 2SLS with a continuous instrument: a weighted LATE representation"

## Appendix A. Proof of Proposition 1

Without loss of generality, we consider the monotonicity assumption in the direction $\mathbb{P}[D(z) \leq D(z')] = 1$ for every $z \leq z'$. Otherwise, we can redefine the instrument as $Z^* = -Z$ and the following proof would also apply.

First, we note that, by the moment and relevance assumptions, the numerator and denominator of $\beta_{\text{Wald}}$ are well-defined. Next, we observe that the numerator is given by:

$$\begin{aligned}\mathbb{E}[(Z - g^*(X))Y] =& \\ \mathbb{E}[(Z - g^*(X))Y(0)] + \mathbb{E}[(Z - g^*(X))D(Y(1) - Y(0))] =& \\ \mathbb{E}[(Z - g^*(X))\mathbb{E}[Y(0)|X]] + \mathbb{E}[(Z - g^*(X))D(Y(1) - Y(0))]&,\end{aligned}$$

where the last equality uses iterated expectations, followed by the independence assumption. Now, it follows from the flexible specification assumption that $g^* = \psi$, from which we have that $\mathbb{E}[(Z - g^*(X))\mathbb{E}[Y(0)|X]] = 0$.

As for the second term, we note that, for support points $(x, z)$:

$$\begin{aligned}\mathbb{E}[(Z - g^*(X))D(Y(1) - Y(0))|Z = z, X = x] =& \\ (z - \psi(x))\mathbb{E}[D(z)(Y(1) - Y(0))|X = x] =& \\ (z - \psi(x))\mathbb{E}[(D(z) - D(\psi(x)))(Y(1) - Y(0))|X = x]+& \\ (z - \psi(x))\mathbb{E}[D(\psi(x))(Y(1) - Y(0))|X = x]&,\end{aligned} \tag{5}$$

where the last equality added and subtracted $D(\psi(x))$ inside the expectation. Observe that, for $\mathbb{P}_X$-almost every $x$, $D(\psi(x))$ is well-defined, since $\mathbb{P}[\psi(X) \in [\underline{z}, \overline{z}]] = 1$. Now, suppose $z > \psi(x)$. In this case, by the monotonicity assumption, $D(z) - D(\psi(x))$ takes either value 0 (always- or never-taker in the comparison between instrument values $\psi(x)$ and $z$), or 1 (complier). Consequently, in this case:

$$\begin{aligned}(z - \psi(x))\mathbb{E}[(D(z) - D(\psi(x)))(Y(1) - Y(0))|X = x] =& \\ \underbrace{(z - \psi(x))}_{=|z-\psi(x)|}\mathbb{P}[D(z) \neq D(\psi(x))|X = x]\mathbb{E}[Y(1) - Y(0)|X = x, D(z) \neq D(\psi(x))]&.\end{aligned}$$

Symmetrically, if $z < \psi(x)$, $D(z) - D(\psi(x))$ takes either value 0 or $-1$, with the latter corresponding to compliance at the comparison between instrument values $\psi(x)$ and $z$. Consequently, we have that, in this case:

$$(z-\psi(x))\mathbb{E}[(D(z)-D(\psi(x)))(Y(1)-Y(0))|X=x] =$$
$$\underbrace{-(z-\psi(x))}_{=|z-\psi(x)|}\mathbb{P}[D(z)\neq D(\psi(x))|X=x]\mathbb{E}[Y(1)-Y(0)|X=x, D(z)\neq D(\psi(x))]\,.$$

Combining the above results, and using that, for $\xi(x) := \mathbb{E}[D(\psi(x))(Y(1)-Y(0))|X=x]$, $\mathbb{E}[(Z-\psi(X))\xi(X)] = 0$, we have that:

$$\mathbb{E}[(Z-g^*(X))Y] = \mathbb{E}[\omega(X,Z)\Delta(X,Z)]\,.$$

A similar argument then shows that the denominator of $\beta_{\text{Wald}}$ is equal to $\mathbb{E}[\omega(X,Z)]$, which proves the desired result.

## Appendix B. Representation for the saturate-and-weight strategy

We consider the case where the support of $X$ is discrete, and $p(X)$ consists of a set of dummies for every support point of $X$. In this case, the "saturate-and-weight" (SW) estimand is given by the 2SLS estimand $\beta$ of the system:

$$\begin{aligned} D &= \psi'(Zp(X)) + \gamma' p(X) + u \\ Y &= \beta D + \omega' p(X) + v \end{aligned},$$

Note that, due to the saturated and interacted nature of the first stage, the $L_2(\mathbb{P})$-projection of $D$ onto $Zp(X)$ and $p(X)$, which we denote by $L_D(Z,X)$, is such that, for every support point $x$, $L_D(Z,x)$ is equal to the $L_2$-projection of $D$ onto an intercept and $Z$, in the subpopulation defined by $X=x$, i.e. $L_D(Z,X) = a(X) + b(X)Z$, where $a(x)$ and $b(x)$ are the intercept and slope of the $L_2(\mathbb{P})$-projection onto an intercept and $Z$ in the subpopulation with $X=x$.

Next, we note that, by the mechanics of 2SLS and the Frisch-Waugh-Lovell theorem that:

$$\beta_{SW} = \frac{\mathbb{E}[Y(L_D(Z,X)-\tilde{L}_D(Z,X))]}{\mathbb{E}[(L_D(Z,X)-\tilde{L}_D(Z,X))^2]}\,,$$

where $\tilde{L}_D(Z,X)$ is the $L_2(\mathbb{P})$-projection of $L_D(Z,X)$ onto $p(X)$. Since $p(X)$ is saturated, this coincides with the conditional expectation of $L_D(Z,X)$ given $X$, i.e. $\tilde{L}_D(Z,X) = \mathbb{E}[L_D(Z,X)|X] = a(X) + b(X)\mathbb{E}[Z|X]$. Consequently, we have that:

$$\begin{aligned} \beta_{SW} &= \frac{\mathbb{E}[Yb(X)(Z-\mathbb{E}[Z|X])]]}{\mathbb{E}[b(X)^2(Z-\mathbb{E}[Z|X])^2]} \\ &= \frac{\mathbb{E}[\beta_{\text{IV}}(X)b(X)^2\mathbb{V}[Z|X]]}{\mathbb{E}[b(X)^2\mathbb{V}[Z|X]]}\,, \end{aligned} \tag{6}$$

where $\beta_{\text{IV}}(X) = \frac{\mathbb{E}[Y(Z-\mathbb{E}[Z|X])|X]}{b(X)\mathbb{V}[Z|X]} = \frac{\text{cov}(Y,Z|X)}{\text{cov}(D,Z|X)}$. Notice that, for each $x$, $\beta_{\text{IV}}(x)$ corresponds to the 2SLS estimand, in the subpopulation defined by $X = x$, of the system.

$$\begin{aligned} D &= a_x + b_x Z + \nu \\ Y &= \alpha_x + \beta_x D + \epsilon \end{aligned}, \tag{7}$$

Consequently, if the Assumptions required by Proposition 1 hold **in the subpopulation defined by $\boldsymbol{X = x}$** (notice from (7) that, once we condition on this subpopulation, we are effectively considering a setting where the controls only include an intercept, i.e. $\mathcal{G} = \mathbb{R}$), then the representations in Proposition 1 and Corollary 1 hold for $\beta_{\text{IV}}(x)$. If the conditions hold in every subpopulation defined by $X = x$, for every $x$ in the support of $X$, we may plug the resulting representations of $\beta_{\text{IV}}(x)$ onto (6) to obtain a weighted LATE representation of the saturate-and-weight estimand. Importantly, since we require a representation to be valid separately for each subpopulation IV estimand $\beta_{\text{IV}}(x)$, for every support point $x$ of $X$, the direction of the monotonicity may vary across subpopulations. The outer weights that aggregate across different values of $X$ are proportional both to the square of the strength of the subpopulation first-stage, as captured by the subpopulation slope $b(X)$, as well as proportional to the conditional variance $\mathbb{V}[Z|X]$. Both features indicate that the saturate-and-weight estimand attaches larger weights to subpopulations defined by the support points of $X$ where the subpopulation IV is easier to estimate.

## Appendix C. Estimation of the weights

Suppose the researcher has access to a sample $\{Y_i, X_i, Z_i, D_i\}_{i=1}^n$ from the model described in Section 2 of the main text. She may be interested in estimating the weights $\omega(x, z)$ and $\omega^*(x, c)$, so as to have a better understanding of which complier subpopulations are assigned larger weights in the representation. To estimate $\omega(x, z)$, the researcher must have an estimator $\hat{\psi}$ of $\psi$, the $L_2(\mathbb{P})$-projection of $Z$ on $X$. For the 2SLS estimands of system (3) in the main text, this may be implemented by setting $\hat{\psi}(X) = \hat{\theta}' p(X)$, where $\hat{\theta}$ is the estimated coefficient of a linear regression of $Y_i$ on $p(X_i)$. More generally, $\hat{\psi}$ will be the estimator the researcher uses when estimating the parameter (1) in the main text, e.g. by considering the sample analog of the minimisation problem given by equation 2 in the main text, with the possible inclusion of a regularizing penalty if the complexity of $\mathcal{G}$ is large. The researcher also requires an estimator of $\mathbb{P}[D(z) \neq D(\psi(x))|X = x]$. Observe that, by the monotonicity hypothesis in Assumption (1) in the main text, this quantity is equal to $|\mathbb{P}[D = 1|Z = z, X = x] - \mathbb{P}[D = 1|Z = \psi(x), X = x]|$. Therefore, given a flexible estimator $\hat{\tau}(x, z)$ of the conditional probability function $\mathbb{P}[D = 1|Z = z, X = x]$, the weights $\omega(x, z)$ may be estimated, in the sample points $(X_i, Z_i)$, as:

$$\hat{\omega}(X_i, Z_i) = \frac{|\hat{\tau}(Z_i, X_i) - \hat{\tau}(\hat{\psi}(X_i), X_i)||Z_i - \hat{\psi}(X_i)|}{\frac{1}{n}\sum_{j=1}^n |\hat{\tau}(Z_j, X_j) - \hat{\tau}(\hat{\psi}(X_j), X_j)||Z_j - \hat{\psi}(X_j)|}, \quad i = 1, \ldots, n.$$

As for the weights $\omega^*(x,c)$, observe that $\kappa(c|x) = \mathbb{E}[(Z-\psi(x))\mathbf{1}\{Z \geq c\}|X=x]$. Let $\hat{\kappa}(c|x)$ be an estimator of $\kappa(c|x)$, e.g. the prediction at $X=x$ from a linear regression of $(Z_i - \hat{\psi}(X_i))\mathbf{1}\{Z_i \geq c\}$ on a vector of transformations of $X_i$. Next, note that, by the monotonicity assumption in Assumption 1, $|\mathbb{P}[D=1|Z=z',X=x] - \mathbb{P}[D=1|Z=z,X=x]| = \int_z^{z'} f^*_{C|X}(c|x)dc$ for any $z' > z$. Consequently, if we rely on an estimator $\hat{\tau}(z,x)$ that is differentiable in $z$, and given a grid of $K$ points $c_1 < c_2 < \ldots < c_K$, we may estimate the weights as:

$$\hat{\omega}^*(Z_i, c_k) = \frac{\hat{\kappa}(c_k|X_i)\left|\frac{\partial}{\partial z}\hat{\tau}(c_k, X_i)\right|}{\frac{1}{nK}\sum_{j=1}^n \sum_{l=1}^K \hat{\kappa}(c_l|X_j)\left|\frac{\partial}{\partial z}\hat{\tau}(c_l, X_j)\right|}, \quad i=1,\ldots,n, \quad k=1,\ldots,K.$$

**Remark 6** (Testing the monotonicity assumption)**.** The estimator of the conditional probability function $\mathbb{P}[D=1|Z,X]$ may be used to test an implication of the monotonicity hypothesis in Assumption 1. To see this, note that an implication of this version of the monotonicity hypothesis is that the random function $z \mapsto \mathbb{P}[D=1|Z=z,X]$ is either nondecreasing with probability one or nonincreasing with probability one. This implication is testable. Indeed, if one adopts a smooth parametric form $\mathbb{P}[D=1|Z,X] = \Lambda(\phi_0' T(Z,X))$, where $\Lambda$ is a differentiable function with $\Lambda'(u) > 0$ for every $u \in \mathbb{R}$, and $T(Z,X)$ is a vector of continuous transformations of $(Z,X)$ differentiable in $Z$, then the implication is equivalent to: $\phi_0' \frac{\partial}{\partial z} T(z,x) \geq 0$ for every support point $(z,x)$ of $(Z,X)$, or $\phi_0' \frac{\partial}{\partial z} T(z,x) \leq 0$ for every support point $(z,x)$ of $(Z,X)$. These restrictions can be tested by constructing a uniform confidence band for the function $(z,x) \mapsto \phi_0' \frac{\partial}{\partial z} T(z,x)$. Freyberger and Rai (2018) provide a general method for constructing such confidence band, when a consistent and asymptotically normal estimator of $\phi_0$ is available. If the band contains at least one nonpositive or nonnegative function, then we do not reject the null that the testable implication holds at one minus the confidence level of the band. If the band **only** contains functions whose sign does not change across $(z,x)$, i.e only nonnegative or nonpositive functions, then we **reject** the null that the implication does **not** hold at at one minus the confidence level of the band.

## Appendix D. Comparison of the approach in the main text with discretization of continuous instruments

Consider the case where the instrument is discretely supported, with **unordered** support points $\{z_1, \ldots, z_K\}$. Suppose that the monotonicity assumption holds for **some** ordering of the instrument, which is unknown. In the case without covariates, Angrist and Imbens (1995) show that the 2SLS estimand that uses the full set of dummies $\mathbf{1}\{Z = z_j\}$, $j=1,\ldots,K$, as instruments identifies a weighted average of LATEs, without having to take a stance on the correct (monotonicity-ensuring) ordering of the points. This suggests that, in some settings without covariates, discretizing and then saturating a continuous instrument may be advantageous, as it would allow the researcher to be agnostic about the correct order ensuring monotonicity, whereas Assumption 1 in the main text requires $Z$ to be encoded in a way such that potential treatment statuses $z \mapsto D(z)$

are either almost surely monotone nonincreasing or almost-surely monotone nondecreasing in $z$. These benefits do not directly extend to specifications that control for covariates $X$, though. Indeed, in a discrete unordered instrument setting with controls, Blandhol et al. (2022) argue that, either the researcher has to saturate-and-weight $X$, interacting the indicators for the support points of $X$ with the dummies for the support points of $Z$ (and using these interactions as instruments), or the researcher should recode $Z$ so as to reflect the monotonicity-ensuring order in $D$, and then consider estimation of the system with a scalar instrument as in (3). If the researcher does not proceed in either way, then some of the underlying weights in the LATE representation of the estimand may be negative. From the estimation viewpoint, the saturate-and-weight strategy, which only works for discrete/discretized $X$, may generate a many-weak instruments problem (Mikusheva and Sun, 2021). In contrast, we note that, if the monotonicity hypothesis in Assumption 1 holds for some unknown recoding of $Z$, then, by taking $v(Z) = \mathbb{P}[D = 1|Z]$ as the instrument, we automatically obtain the monotonicity-ensuring encoding.[2] This suggests that the undiscretized approach discussed in the main text is often appropriate, provided that we correctly encode $Z$ (which possibly involves estimating $v$ in a flexible way).

[2]This idea is originally discussed in page 470 of IA.