Online Appendix of "The interpretation 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 β_{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 β_{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 β 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 $\boldsymbol{X} = \boldsymbol{x}$** (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 ψ, 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 Λ 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 ϕ_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).

References

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.

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

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

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

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

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