Sally Hudson, Peter Hull, and Jack Liebersohn September 4, 2017

Interpreting Instrumented Difference-in-Differences

In an instrumented difference-in-differences design (what we refer to as a DDIV), a researcher scales a difference-in-differences (DD) effect on an outcome by a DD effect on a mediating treatment variable. A canonical example is Duflo (2001), who measures the impact of Indonesian school construction on both adult labor market outcomes and educational attainment. Despite DDIV's widespread use, and in contrast to the vast literatures on both DD and instrumental variable (IV) techniques, however, DDIV identification has received little formal attention.[1] In this note, we show that DDIV estimates a convex combination of average causal effects, as in Angrist and Imbens (1995), given a set of exclusion, parallel trends, and monotonicity assumptions familiar from both DD and IV designs. By clarifying these assumptions, we hope to provide researchers with more precise justification for DDIV estimation and to highlight some potential pitfalls of causal DDIV inference.

Suppose we observe a time-varying outcome Y_{it}, a discretely- and positively-valued treatment S_{it}, and a binary instrument Z_{it} for a set of individuals i in each time period $t \in \{0, 1\}$.[2] Individuals are not exposed to the instrument until period 1, so $Z_{i0} = 0$ for all i, and we can write $Z_i = Z_{i1}$. We also let $T_t = \mathbf{1}\{t = 1\}$ denote a period indicator. The DDIV coefficient β comes from the following IV system,

$$Y_{it} = \alpha_i + \tau T_t + \beta S_{it} + \epsilon_{it} \tag{1}$$

$$\text{and} \quad S_{it} = \gamma_i + \delta T_t + \pi Z_i T_t + \eta_{it}. \tag{2}$$

As with all just-identified IV, β here can be expressed as the ratio of reduced form and first stage coefficients. The first stage parameter, π, comes from equation (2) above, and the reduced form regression is

$$Y_{it} = \mu_i + \theta T_t + \rho Z_i T_t + \nu_{it}. \tag{3}$$

The reduced form and first stage coefficients can, in turn, be obtained by first-differenced regressions of $Y_{i1} - Y_{i0}$ and $S_{i1} - S_{i0}$ on Z_i. The DDIV estimand can therefore be written:

$$\beta = \frac{E[Y_{i1} - Y_{i0} \mid Z_i = 1] - E[Y_{i1} - Y_{i0} \mid Z_i = 0]}{E[S_{i1} - S_{i0} \mid Z_i = 1] - E[S_{i1} - S_{i0} \mid Z_i = 0]}. \tag{4}$$

Let Y_{it}^s denote the potential outcome of individual i in time t if she were exposed to treatment level s. Likewise, let S_{it}^z denote an individual's potential treatment level if she were exposed to instrument value z in time t. This notation embodies two implicit *exclusion restrictions*. First, by omitting instrument superscripts from Y_{it}^s, we assume that the only way the instrument affects outcomes is through the treatment – a standard restriction in IV estimation. Second, by writing the S_{it}^z only in terms of the contemporaneous value of Z_i, we restrict attention to DDIV experiments in which the period-1 instrument doesn't affect period-0 outcomes or treatment. This restriction may rule out experiments in which the instrument is simply the interaction of time and a fixed group trait, rather than an actual

[1] A recent exception is de Chaisemartin and D'Haultfoeuille (forthcoming), which we discuss below.

[2] As with DD, it is straightforward to extend the panel data case to settings with repeated cross sections and multiple periods (see, e.g., Abadie (2005)). It is also straightforward to extend the discrete treatment case to settings with continuous treatments, as with IV (Angrist, Graddy, and Imbens, 2000). The assumption that S_{it} is positive is without loss for any lower- or upper-bounded treatment.

period-specific shock as in Duflo (2001). In the former setting, manipulating Z_i would require changing group membership, which would likely affect outcomes in both periods.[3]

In addition to the two exclusion restrictions, we make the following two assumptions:

A1 (*Parallel trends*): $S_{i1}^z - S_{i0}^0$ and $Y_{i1}^{S_{i1}^z} - Y_{i0}^{S_{i0}^0}$ are mean-independent of Z_i, for each $z = 0, 1$

A2 (*Monotonicity*): $P(S_{i1}^1 \geq S_{i1}^0) = 1$

Parallel trends is a familiar condition from the DD literature. It requires that the potential growth paths of both treatment and outcomes are independent from actual instrument assignment. The second assumption further constrains the effect of the instrument on period-1 treatment to be monotone, as in the literature on IV identification of local average treatment effects (Imbens and Angrist, 1994).

Under A1, the DDIV first stage coefficient identifies

$$\begin{aligned} &E[S_{i1} - S_{i0} \mid Z_i = 1] - E[S_{i1} - S_{i0} \mid Z_i = 0] \\ &= E[S_{i1}^1 - S_{i0}^0 \mid Z_i = 1] - E[S_{i1}^0 - S_{i0}^0 \mid Z_i = 0] \\ &= E[S_{i1}^1 - S_{i1}^0] \end{aligned} \tag{5}$$

noting again that $Z_{i0} = 0$ for all i. Similarly, the reduced form identifies

$$\begin{aligned} &E[Y_{i1} - Y_{i0} \mid Z_i = 1] - E[Y_{i1} - Y_{i0} \mid Z_i = 0] \\ &= E[Y_{i1}^{S_{i1}^1} - Y_{i0}^{S_{i0}^0} \mid Z_i = 1] - E[Y_{i1}^{S_{i1}^0} - Y_{i0}^{S_{i0}^0} \mid Z_i = 0] \\ &= E[Y_{i1}^{S_{i1}^1} - Y_{i1}^{S_{i1}^0}], \end{aligned} \tag{6}$$

so that

$$\beta = \frac{E[Y_{i1}^{S_{i1}^1} - Y_{i1}^{S_{i1}^0}]}{E[S_{i1}^1 - S_{i1}^0]}, \tag{7}$$

assuming the denominator is nonzero (a further *first stage* condition). By A2 and Theorem 1 in Angrist and Imbens (1995), DDIV thus identifies

$$\beta = \sum_{s>0} \omega_s E[Y_{i1}^s - Y_{i1}^{s-1} \mid S_{i1}^1 \geq s > S_{i1}^0] \tag{8}$$

where

$$\omega_s = \frac{P(S_{i1}^1 \geq s > S_{i1}^0)}{\sum_{r>0} P(S_{i1}^1 \geq r > S_{i1}^0)} \tag{9}$$

are a set of weights with $0 \leq \omega_s \leq 1$ and $\sum_{s>0} \omega_s = 1$. Thus, under the assumptions, the DDIV estimator captures a proper weighted average of period-1 causal responses to a unit change in treatment, for those whose treatment status is affected by the instrument. Angrist and Imbens (1995) refer to this object as an *average causal response.*

The parallel trends assumption, assumed to hold for both the outcome and the treatment, is critical for this result. Parallel trends will clearly be satisfied under the usual independence assumption for IV, but it also holds under certain violations. For example, researchers may

[3] Abdulkadiroğlu et al. (2016) consider causal inference in DDIV-like experiments when the first exclusion restriction is violated, while the approach in de Chaisemartin and D'Haultfoeuille (forthcoming) may be more appropriate for "time and group" designs in which the second exclusion restriction fails.

employ DDIV techniques if they believe potential outcomes and treatments have an additive structure where the instrument is correlated with time-invariant individual characteristics but independent from time-varying shocks, e.g.

$$Y_{it}^s = \alpha_i + \tau_t + \beta_{it}^s \tag{10}$$

with $\beta_{i1}^s - \beta_{i0}^r \perp\!\!\!\perp Z_i$ for all s and r but α_i not necessarily independent from the instrument. In practice, researchers employing DD methods often provide evidence to support parallel trends, through pre-trend or other placebo tests. A1 may be similarly supported in DDIV applications by applying these tests to both the outcome and endogenous variable.

Although not shown here, it is straightforward to verify that the basic DDIV logic carries through when the parallel trends assumption is weakened to hold conditionally on covariates. As in Angrist (1998), DDIV regressions that control for a saturated model of individual controls X_i and time effects identify a variance-weighted average of conditional-on-X_i average causal response functions. Two-step weighting procedures, in the spirit of Abadie (2003, 2005) may moreover be developed to non-parametrically recover the unconditional average causal response function under a conditional version of A1. We leave the formal characterization of these estimators to future work.

References

Abadie, A. "Semiparametric Instrumental Variables Estimation of Treatment Response Functions," *Journal of Econometrics*, 113 (2003), 231-263.

Abadie, A. "Semiparametric Difference-in-Difference Estimators," *Review of Economic Studies*, 72 (2005), 1-19.

Abdulkadiroğlu, A., J. D. Angrist, P. D. Hull, and P. A. Pathak. "Charters Without Lotteries: Testing Takeovers in New Orleans and Boston," *American Economic Review*, 106 (2016), 1878-1920.

Angrist, J. D. "Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants," *Econometrica*, 66 (1998), 249-288.

Angrist, J. D., K. Graddy., and G. W. Imbens. "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *Review of Economic Studies*, 67 (2000), 499-527.

Angrist, J. and G. W. Imbens. "Two-stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90 (1995), 431-442.

de Chaisemartin, C. and X. D'Haultfoeuille. "Fuzzy Differences-in-Differences," *Review of Economic Studies*, forthcoming.

Duflo, E. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 91 (2001), 795-813.

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