Not all Differences-in-differences are Equally Compatible with Outcome-based Selection Models

Clément de Chaisemartin, Xavier d'Haultfoeuille

Not all Differences-in-differences are Equally Compatible with Outcome-based Selection Models

Clément de Chaisemartin Xavier D'Haultfœuille*

We are interested in the effect of a binary treatment on an outcome. For every $(g,t) \in \{1,...,G\} \times \{1,...,T\}$, let $D_{g,t}$ denote the value of the treatment in group g at period t. For any $d \in \{0,1\}$, let $Y_{g,t}(d)$ denote the potential outcome of group g at period t if $D_{g,t} = d$. The observed outcome is $Y_{g,t} = Y_{g,t}(D_{g,t})$. Implicitly, our potential outcome notation rules out dynamic and anticipatory effects: groups' period-t outcome is only affected by their current treatment, it does not depend on their past and future treatments. We assume that groups' treatments and outcomes are random. This is in line with the standard modelling framework in panel data models, and nests as a special case the modelling framework in differences-in-differences models, where groups' treatments are often implicitly conditioned upon.

Throughout the note, we maintain the following assumptions.

Assumption 1 *(Independent groups) The vectors $(Y_{g,t}(0), Y_{g,t}(1), D_{g,t})_{t\in\{1,...,T\}}$ are mutually independent.*

Assumption 2 *(Parallel trends) For all $(g,t) \in \{1,...,G\} \times \{2,...,T\}$, $E(Y_{g,t}(0) - Y_{g,t-1}(0))$ does not vary across g.*

Assumption 1 requires that potential outcomes and treatments of different groups be independent, but it allows these variables to be correlated over time within each group. This is a commonly-made assumption in difference-in-differences (DID) analysis, where standard errors are usually clustered at the group level (see Bertrand, Duflo and Mullainathan, 2004). Assumption 2 requires that the expectation of the untreated outcome follow the same evolution over time. It is a generalization of the standard common trends assumption in DID models (see, e.g., Abadie, 2005) to settings with multiple periods and groups. Note that with identically distributed groups, Assumption 2 holds mechanically, but here we allow for potentially non-identically distributed groups.

We consider two exogeneity assumptions.

*de Chaisemartin: Sciences Po (email: clement.dechaisemartin@sciencespo.fr); D'Haultfœuille: CREST-ENSAE (email: xavier.dhaultfoeuille@ensae.fr).

Assumption 3 *(Strong exogeneity) For all* $(g,t) \in \{1,...,G\} \times \{2,...,T\}$,

$$E(Y_{g,t}(0) - Y_{g,t-1}(0)|D_{g,1},...,D_{g,T}) = E(Y_{g,t}(0) - Y_{g,t-1}(0)).$$

Assumption 3 is related to the strong exogeneity condition in panel data models. It requires that the evolution of group g's untreated outcome from $t-1$ to t be mean independent of group g's treatments from period 1 to T.

Assumption 4 *(Weaker exogeneity) For all* $(g,t) \in \{1,...,G\} \times \{2,...,T\}$,

$$E(Y_{g,t}(0) - Y_{g,t-1}(0)|D_{g,1},...,D_{g,t}) = E(Y_{g,t}(0) - Y_{g,t-1}(0)).$$

Assumption 4 is weaker than Assumption 3. It requires that the evolution of group g's untreated outcome from $t-1$ to t be mean independent of group g's treatments from period 1 to t. Assumption 4 is still stronger than the standard weak exogeneity assumption in panel data models, which only requires that $Y_{g,t}(0)-Y_{g,t-1}(0)$ be mean independent of group g's treatments from period 1 to $t-1$.

Moving from Assumption 3 to Assumption 4 restricts the DIDs one can use for identification, and the set of treatment effects that can be estimated. Assume $G=4$ and $T=3$. In what follows, we condition on $D_{1,1}=0, D_{1,2}=D_{1,3}=1$ (group 1 is untreated at period 1 and treated at periods 2 and 3), $D_{2,1}=0, D_{2,2}=1, D_{2,3}=0$ (group 2 is untreated at periods 1 and 3 and treated at period 2), $D_{3,1}=0, D_{3,2}=0, D_{3,3}=0$ (group 3 is never treated), and $D_{4,1}=1, D_{4,2}=0, D_{4,3}=0$ (group 4 is treated at period 1 and untreated at periods 2 and 3). Then, under Assumptions 1-3, a DID comparing group 1's period-one-to-three outcome evolution to that of group 2 is unbiased for group 1's treatment effect at period 3, conditional on groups' 1 and 2 period-1-to-3 treatments:

$$\begin{aligned}
&E\left(Y_{1,3}-Y_{1,1}-(Y_{2,3}-Y_{2,1})|D_{1,1}=0,D_{1,2}=D_{1,3}=1,D_{2,1}=0,D_{2,2}=1,D_{2,3}=0\right)\\
=&E\left(Y_{1,3}(1)-Y_{1,3}(0)|D_{1,1}=0,D_{1,2}=D_{1,3}=1,D_{2,1}=0,D_{2,2}=1,D_{2,3}=0\right)\\
+&E\left(Y_{1,3}(0)-Y_{1,2}(0)+Y_{1,2}(0)-Y_{1,1}(0)|D_{1,1}=0,D_{1,2}=D_{1,3}=1\right)\\
-&E\left(Y_{2,3}(0)-Y_{2,2}(0)+Y_{2,2}(0)-Y_{2,1}(0)|D_{2,1}=0,D_{2,2}=1,D_{2,3}=0\right)\\
=&E\left(Y_{1,3}(1)-Y_{1,3}(0)|D_{1,1}=0,D_{1,2}=D_{1,3}=1,D_{2,1}=0,D_{2,2}=1,D_{2,3}=0\right)\\
+&E\left(Y_{1,3}(0)-Y_{1,2}(0)+Y_{1,2}(0)-Y_{1,1}(0)\right)\\
-&E\left(Y_{2,3}(0)-Y_{2,2}(0)+Y_{2,2}(0)-Y_{2,1}(0)\right)\\
=&E\left(Y_{1,3}(1)-Y_{1,3}(0)|D_{1,1}=0,D_{1,2}=D_{1,3}=1,D_{2,1}=0,D_{2,2}=1,D_{2,3}=0\right).
\end{aligned} \tag{1}$$

The first equality follows from Assumption 1, and from adding and subtracting $Y_{1,3}(0)$, $Y_{1,2}(0)$, and $Y_{2,2}(0)$. The second equality follows from Assumption 3. The third equality follows from Assumption 2.

Equation (1) no longer holds if one replaces Assumption 3 by Assumption 4: Assumption 4 does not ensure that $Y_{1,2}(0)-Y_{1,1}(0)$ (resp. $Y_{2,2}(0)-Y_{2,1}(0)$) is mean independent of $(D_{1,1},D_{1,2},D_{1,3})$ (resp. of $(D_{2,1},D_{2,2},D_{2,3})$). On the other hand, under Assumptions 1-2 and 4, a DID comparing group 1's period-1-to-2 outcome evolution to that of a group 3 is unbiased for group 1's treatment effect at period 2, conditional on groups' 1 and 3 period-1-to-2 treatments:

$$\begin{aligned}
&E\left(Y_{1,2}-Y_{1,1}-(Y_{3,2}-Y_{3,1})|D_{1,1}=0,D_{1,2}=1,D_{3,1}=0,D_{3,2}=0\right)\\
=&E\left(Y_{1,2}(1)-Y_{1,2}(0)|D_{1,1}=0,D_{1,2}=1,D_{3,1}=0,D_{3,2}=0\right)\\
+&E\left(Y_{1,2}(0)-Y_{1,1}(0)|D_{1,1}=0,D_{1,2}=1\right)\\
-&E\left(Y_{3,2}(0)-Y_{3,1}(0)|D_{3,1}=0,D_{3,2}=0\right)\\
=&E\left(Y_{1,2}(1)-Y_{1,2}(0)|D_{1,1}=0,D_{1,2}=1,D_{3,1}=0,D_{3,2}=0\right)\\
+&E\left(Y_{1,2}(0)-Y_{1,1}(0)\right)\\
-&E\left(Y_{3,2}(0)-Y_{3,1}(0)\right)\\
=&E\left(Y_{1,2}(1)-Y_{1,2}(0)|D_{1,1}=0,D_{1,2}=1,D_{3,1}=0,D_{3,2}=0\right). \qquad (2)
\end{aligned}$$

The first equality follows from Assumption 1, and from adding and subtracting $Y_{1,2}(0)$. The second equality follows from Assumption 4. The third equality follows from Assumption 2.

Under Assumptions 1-2 and 4, we also have that a "backward" DID comparing group 3's period-2-to-1 outcome evolution to that of a group 4 is unbiased for group 4's treatment effect at period 1, conditional on groups' 1 and 4 period-1-to-2 treatments:

$$\begin{aligned}
&E\left(Y_{3,1}-Y_{3,2}-(Y_{4,2}-Y_{4,1})|D_{3,1}=0,D_{3,2}=0,D_{4,1}=1,D_{4,2}=0\right)\\
=&E\left(Y_{4,1}(1)-Y_{4,1}(0)|D_{3,1}=0,D_{3,2}=0,D_{4,1}=1,D_{4,2}=0\right)\\
+&E\left(Y_{3,1}(0)-Y_{3,2}(0)|D_{3,1}=0,D_{3,2}=0\right)\\
-&E\left(Y_{4,2}(0)-Y_{4,1}(0)|D_{4,1}=1,D_{4,2}=0\right)\\
=&E\left(Y_{4,1}(1)-Y_{4,1}(0)|D_{3,1}=0,D_{3,2}=0,D_{4,1}=1,D_{4,2}=0\right)\\
+&E\left(Y_{3,1}(0)-Y_{3,2}(0)\right)\\
-&E\left(Y_{4,2}(0)-Y_{4,1}(0)\right)\\
=&E\left(Y_{4,1}(1)-Y_{4,1}(0)|D_{3,1}=0,D_{3,2}=0,D_{4,1}=1,D_{4,2}=0\right). \qquad (3)
\end{aligned}$$

The first equality follows from Assumption 1, and from adding and subtracting $Y_{4,1}(0)$. The second equality follows from Assumption 4. The third equality follows from Assumption 2.

More generally, under Assumptions 1-2 and 4, we can estimate the treatment effects of only two sets of (g,t) cells. The first set are (g,t) cells such that $D_{g,t}=1, D_{g,t-1}=0$, whose treatment effect can be estimated using a "forward first-difference DID" comparing g's $t-1$-to-t outcome evolution to that of groups untreated at $t-1$ and t. The second set are (g,t) cells such that

$D_{g,t} = 1, D_{g,t+1} = 0$, whose treatment effect can be estimated using a "backward first-difference DID" comparing g's $t+1$-to-t outcome evolution to that of groups untreated at $t+1$ and t. Under Assumptions 1-3, we can estimate the treatment effects of a potentially much larger set of cells. Specifically, we can estimate the treatment effects of all (g, t) cells such that $D_{g,t} = 1, D_{g,t'} = 0$ for some $t' \neq t$, using a "forward or backward long-difference DID" comparing g's t'-to-t outcome evolution to that of groups untreated at t' and t.

Assumption 4 has less identifying power than Assumption 3, but it is more compatible than Assumption 3 with models where groups get treated because their expected benefit from treatment is larger than the cost, the so-called Roy selection model (see Roy, 1951).

Assumption 5 *(Roy selection with independent shocks)*

1. *For* $d \in \{0, 1\}$, $Y_{g,1}(d) = \varepsilon_{g,1}(d)$ *and* $Y_{g,t}(d) = Y_{g,t-1}(d) + \varepsilon_{g,t}(d)$ *for* $t \geq 2$, *with* $E(\varepsilon_{g,t}(d)) = \lambda_t(d)$ *for all* g, *and* $(\varepsilon_{g,t}(0), \varepsilon_{g,t}(1))$ *mutually independent across* t.

2. $D_{g,1} = 0$ *and* $D_{g,t} = 1\{E(Y_{g,t}(1) - Y_{g,t}(0)|D_{g,1}, ..., D_{g,t-1}, Y_{g,1}, ..., Y_{g,t-1}) \geq c_{g,t}\}$ *for* $t \geq 2$.

Point 2 of Assumption 5 is a Roy selection equation, where groups predict their period-t treatment effect using all their past treatments and outcomes. If Assumption 5 holds, Assumption 4 holds while Assumption 3 fails. Accordingly, first-difference DIDs are unbiased while long-difference DIDs are biased. Of course, Assumption 4 does not hold in any model with Roy selection. For instance, assuming that the shocks are independent across t is critical: Assumptions 4 and 3 both fail with correlated shocks.[1] Still, this shows that Assumption 4 is compatible with some types of Roy selection, something which has not be shown yet for Assumption 3.

Assumption 4 is also more compatible than Assumption 3 with models where groups get treated because they experience a negative shock, the so-called Ashenfelter's dip.

Assumption 6 *(Ashenfelter's dip with independent shocks)*

1. $Y_{g,1}(0) = \varepsilon_{g,1}(0)$ *and* $Y_{g,t}(0) = Y_{g,t-1}(0) + \varepsilon_{g,t}(0)$ *for* $t \geq 2$, *with* $E(\varepsilon_{g,t}(0)) = \lambda_t(0)$ *for all* g, *and* $\varepsilon_{g,t}(0)$ *mutually independent across* t.

2. $D_{g,1} = 0$ *and* $D_{g,t} = 1\{E(Y_{g,t}(0)|D_{g,1}, ..., D_{g,t-1}, (1-D_{g,1})Y_{g,1}, ..., (1-D_{g,t-1})Y_{g,t-1}) \leq m_{g,t}\}$ *for* $t \geq 2$.

[1] The model $Y_{g,t}(d) = Y_{g,t-1}(d) + \varepsilon_{g,t}(d)$ is also critical, but without it may be hard to rationalize any parallel trends assumption. For instance, if $Y_{g,t}(0) = \rho Y_{g,t-1}(0) + \varepsilon_{g,t}(0)$ with $\rho < 1$, parallel trends on $Y_{g,t}(0)$ requires that groups have the same expectation of $(\rho - 1)Y_{g,t-1}(0) + \varepsilon_{g,t}(0)$, which is hard to rationalize without imposing that they have the same expectation of $Y_{g,t-1}(0)$ and $\varepsilon_{g,t}(0)$, a strong assumption at odds with the DID logic. If $Y_{g,t}(0) = \rho Y_{g,t-1}(0) + \varepsilon_{g,t}(0)$, it may be preferable to impose a sequential ignorability assumption, see e.g. Robins (1986) and Bojinov, Rambachan and Shephard (2021)

In Point 2 of Assumption 6, groups get treated at period t if their predicted $Y_{g,t}(0)$ given their past untreated outcomes is below some threshold. Under Assumption 6, if group g is untreated at period $t-1$,

$$D_{g,t} = 1\{Y_{g,t-1}(0) \leq m_{g,t}\} = 1\{\varepsilon_{g,t-1} \leq m_{g,t} - Y_{g,t-2}(0)\}.$$

Accordingly g gets treated at period-t if its period-$t-1$ shock is low (lower than $m_{g,t} - Y_{g,t-2}(0)$) and remains untreated otherwise, the so-called Ashenfelter's dip. If Assumption 6 holds, Assumption 4 holds while Assumption 3 fails. Again, assuming independent shocks is critical for this result to hold, but this at least shows that Assumption 4 is compatible with some types of Ashenfelter's dip.

Note that when $T = 2$, the selection equation in Point 2 of Assumption 6 is similar to a selection equation previously considered by Ashenfelter and Card (1985) and Ghanem, Sant'Anna and Wüthrich (2022). Under that selection equation and a model similar to that in Point 1 of Assumption 6, Ghanem, Sant'Anna and Wüthrich (2022) show that the standard parallel trends assumption with two periods holds.

To conclude, under a strong-exogeneity assumption one can use first- and long-difference DIDs, to estimate the treatment effects of a large set of (g,t) cells. On the other hand, under a weaker exogeneity assumption, one can only use first-difference DIDs to estimate the treatment effects of a more restricted set of (g,t) cells. However, the weaker-exogeneity assumption may be more plausible, as under some assumptions on the outcome equation, it is compatible with two prominent selection models, namely Roy selection and Ashenfelter's dip. Those results have consequences for estimators recently proposed in the heterogeneity-robust DID literature. The heterogeneity-robust DID estimator proposed by de Chaisemartin and D'Haultfœuille (2020) for the joiners only leverages first-difference DIDs. The authors show that their estimator is unbiased for joiners' average treatment effect under Assumption 3. Under Assumption 4, one could show that their estimator is consistent for that parameter, though it is no longer unbiased. Accordingly, the consistency of their estimator relies on an exogeneity assumption compatible with some forms of Roy selection and Ashenfelter's dip.

References

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

Ashenfelter, Orley, and David Card. 1985. "USING THE LONGITUDINAL STRUCTURE OF EARNINGS TO ESTIMATE THE EFFECT OF TRAINING PROGRAMS." *The Review of Economics and Statistics*, 67(4): 648–660.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How much should we trust differences-in-differences estimates?" *The Quarterly Journal of Economics*, 119(1): 249–275.

Bojinov, Iavor, Ashesh Rambachan, and Neil Shephard. 2021. "Panel experiments and dynamic causal effects: A finite population perspective." *Quantitative Economics*, 12(4): 1171–1196.

de Chaisemartin, Clement, and Xavier D'Haultfœuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review*, 110(9): 2964–96.

Ghanem, Dalia, Pedro HC Sant'Anna, and Kaspar Wüthrich. 2022. "Selection and parallel trends." *arXiv preprint arXiv:2203.09001.*

Robins, James. 1986. "A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect." *Mathematical modelling*, 7(9-12): 1393–1512.

Roy, Andrew Donald. 1951. "Some thoughts on the distribution of earnings." *Oxford economic papers*, 3(2): 135–146.