# Interpreting Event-Studies from Recent Difference-in-Differences Methods

Jonathan Roth*

January 22, 2024

**Abstract**

This note discusses the interpretation of event-study plots produced by recent difference-in-differences methods. I show that even when specialized to the case of non-staggered treatment timing, the default plots produced by software for three of the most popular recent methods (de Chaisemartin and D'Haultfuille, 2020; Callaway and SantAnna, 2021; Borusyak, Jaravel and Spiess, 2024) do not match those of traditional two-way fixed effects (TWFE) event-studies: the new methods may show a kink or jump at the time of treatment even when the TWFE event-study shows a straight line. This difference stems from the fact that the new methods construct the pre-treatment coefficients asymmetrically from the post-treatment coefficients. As a result, visual heuristics for analyzing TWFE event-study plots should not be immediately applied to those from these methods. I conclude with practical recommendations for constructing and interpreting event-study plots when using these methods.

## 1 Introduction

Modern difference-in-differences (DiD) analyses typically show an event-study plot that allows the researcher to evaluate differences in trends between the treated and comparison groups both before and after the treatment.[1] Event-study plots allow the researcher to perform "visual inference" to determine whether the treated group's trends appear to have deviated from the comparison group's trends right around the time of treatment. An event-study plot is generally considered more convincing if there is a discontinuity or kink in the

*Brown University. jonathanroth@brown.edu. I am grateful to Kirill Borusyak, Clément de Chaisemartin, Xavier D'Haultfuille, Peter Hull, Pedro Sant'Anna, and Jesse Shapiro for comments.

[1] Roth (2022) identifies 70 recent papers in top economics journals displaying such plots.

coefficients close to the treatment date, so that the pattern seen cannot simply be explained by a linear violations of parallel trends (Rambachan and Roth, 2023).

Until recently, event-study plots were created by running a dynamic two-way fixed effects (TWFE) regression specification including indicators for time relative to treatment. A recent literature surveyed in de Chaisemartin and D'Haultfuille (2022) and Roth, SantAnna, Bilinski and Poe (2023) has illustrated that TWFE specifications may be difficult to interpret in settings with heterogeneous treatment effects, even under a parallel trends assumption. Most pertinently, Sun and Abraham (2021) showed that the event-study coefficient for relative time $r$ from a dynamic TWFE specification may be "contaminated" by treatment effects at a different relative time $r'$. These issues with TWFE specifications have motivated the development of a variety of new DiD methods that yield more interpretable estimands in settings with staggered treatment timing and treatment effect heterogeneity. Many of the new methods allow for the construction of an event-study plot, and so it may be tempting to simply interpret the event-study produced by these new methods analogously to those from traditional dynamic TWFE specifications in the non-staggered setting.

The main point of this note is that the default event-study plots produced by software for three of the most popular recent methods should not be interpreted in the same way as traditional dynamic TWFE specifications. Indeed, when specialized to the case with a common treatment date, the event-study plots from de Chaisemartin and D'Haultfuille (2020, henceforth, dCDH),[2] Callaway and SantAnna (2021, henceforth, CS) and Borusyak et al. (2024, henceforth, BJS) do not match those from a dynamic TWFE specification. The plots from these new methods may show a kink or jump at the time of treatment even when there is no treatment effect and the violation of parallel trends is the same in all periods, so that the traditional dynamic TWFE specification shows a straight line. The typical heuristics for visual inference developed based on dynamic TWFE specifications will thus be misleading when applied to these new estimators.

The remainder of this note is structured as follows. In Section 2, I provide a simple simulation illustrating how dynamic TWFE, dCDH, CS, and BJS produce very different event-study plots on the same data without staggered treatment timing. In Section 3, I provide a mathematical explanation showing that these differences arise because these estimators construct the pre-treatment and post-treatment event-study coefficients asymmetrically. Section 4 concludes with some recommendations.

[2]More precisely, the default plot produced by the R package for dCDH does not match the standard TWFE output; however, the default plot produced by the Stata package does. The authors have indicated to me that an updated version of the R package in forthcoming, which will match the output of the Stata package.

# 2 A simulation comparing the event-studies in a non-staggered setting

I begin by illustrating the differences between the event-studies produced by TWFE and the newer methods with a simulation of an extremely simple non-staggered DiD setting.[3] Units are either treated beginning at period $t = 1$ or never-treated. We observe outcomes from period $t = -15$ to period $t = 10$. There are no treatment effects in any period, so that $Y_{it}(1) = Y_{it}(0) = Y_{it}$. However, parallel trends is violated: the treated group's outcome increases linearly relative to the control groups outcome in all periods. In particular, $Y_{it} = \gamma \cdot t + \epsilon_{it}$ for units in the treated group and $Y_{it} = \epsilon_{it}$ for units in the control group, where $\epsilon_{it}$ are *iid* standard normal draws. Thus, the treated group's outcome grows by $\gamma$ more than the control group's outcome on average in each period. For the illustrative simulation, I set $\gamma = 0.5$.

Figure 1 shows event-studies for a single simulation draw from this DGP when there are 100 total units and half of them are treated. Panel (a) shows the results of a traditional dynamic TWFE specification of the form

$$Y_{it} = \alpha_i + \lambda_t + \sum_{r \neq -1} \beta_r \times 1[D_i = 1] \times 1[t = r + 1] + e_{it},$$

where $D_i$ is an indicator for treatment and $\alpha_i$ and $\lambda_t$ are unit and period fixed effects. Panel (b) shows the results for the event-study generated by the CS method, using the default settings in the `did` package in R. Panel (c) shows the event-study plot generated using the dCDH method, using the default settings in the `did_multiplegt` R package.[4] Panel (d) shows the event-study plot generated by the BJS method, using the default settings in the `did_imputation` Stata package.[5]

Despite using identical data, the four methods produce very different event-study plots. The dynamic TWFE specification shows an approximately linear pre-trend that continues into the post-treatment period, with no apparent break in trend at the treatment date. By contrast, the CS and dCDH event-study plots have relatively flat pre-treatment coefficients and then exhibit a kink at the time of the treatment, sloping upwards afterwards. One can verify that in this example the coefficient estimates are numerically the same for CS and dCDH.[6] The BJS event-study appears to show an upward pre-trend, but then exhibits a

[3] Code for the results shown in this section is available on Github, https://github.com/jonathandroth/HetEventStudies/tree/master.

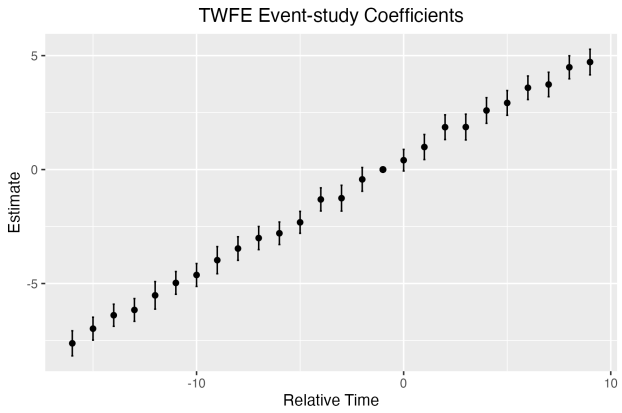
[4] See FN 2 regarding differences between the Stata and R packages.

[5] All packages were installed between January 16-18, 2024.

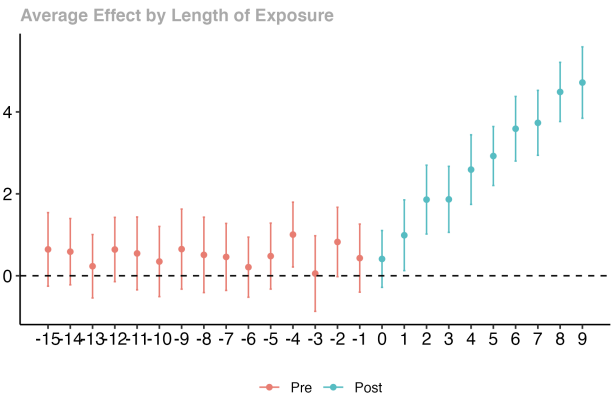
[6] The standard errors differ slightly owing to different implementations of the bootstrap.

Figure 1: Comparison of event-study plots in a non-staggered setting

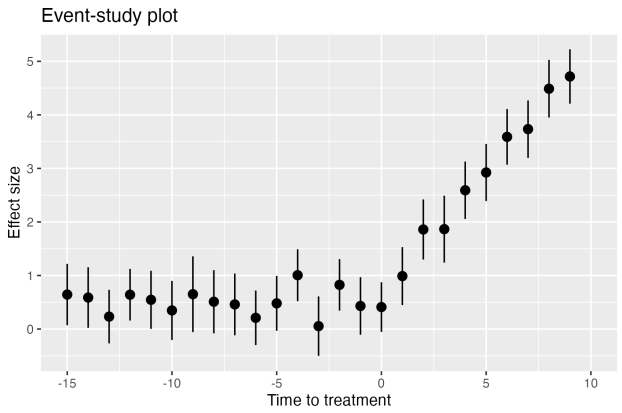
(a) Dynamic TWFE



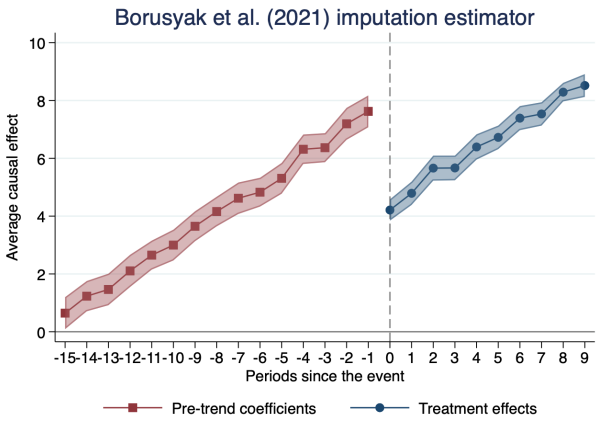

(b) CS




(c) dCDH




(d) BJS




sharp downward jump at the treatment date.

# 3 Mathematical Examination

In this section, I show that the differences between the event-plots created by dynamic TWFE and the newer estimators stems from the asymmetric way that the newer estimators construct the pre-treatment and post-treatment coefficients.

I consider a slightly generalized version of the set-up in the simulation section, in which we observe panel data for periods $t = -\underline{T}, ..., \bar{T}$. Treated units denoted by $D_i = 1$ begin treatment at period 1, whereas comparison units ($D_i = 0$) are untreated in all periods. The observed outcome is $Y_{it} = D_i Y_{it}(1) + (1 - D_i) Y_{it}(0)$. For some of our results, we will consider the setting in the simulation where $Y_{it}(1) = Y_{it}(0) = \gamma \cdot t + \epsilon_{it}$ where $E[\epsilon_{it} \mid D_i] = 0$.

## 3.1 Dynamic TWFE

Consider the dynamic TWFE specification

$$Y_{it} = \alpha_i + \lambda_t + \sum_{r \neq -1} \beta_r \times 1[D_i = 1] \times 1[t = r + 1] + e_{it}.$$

It is straightforward to show that the event-study coefficients are given by

$$\hat{\beta}_r = \left(\hat{E}[Y_{i,r+1} \mid D_i = 1] - \hat{E}[Y_{i,r+1} \mid D_i = 0]\right) - \left(\hat{E}[Y_{i0} \mid D_i = 1] - \hat{E}[Y_{i0} \mid D_i = 0]\right),$$

where $\hat{E}$ denotes the sample average. That is, $\hat{\beta}_r$ shows a 2-group, 2-period DiD estimate comparing the $D_i = 1$ and $D_i = 0$ groups between periods $r + 1$ and 0. Note that the formula for $\hat{\beta}_r$ is symmetric between pre-treatment and post-treatment periods—i.e. the same regardless of whether $r$ is greater than or less than zero.

Under mild regularity conditions, the sample means converge to population means, and thus $\hat{\beta}_r$ is consistent for

$$\beta_r = (E[Y_{i,r+1} \mid D_i = 1] - E[Y_{i,r+1} \mid D_i = 0]) - (E[Y_{i0} \mid D_i = 1] - E[Y_{i0} \mid D_i = 0]).$$

Under the DGP in our simulations, we thus have

$$\beta_r = \gamma \cdot (r + 1),$$

and so the population event-study coefficients lie on a straight line.

## 3.2 CS and dCDH

I now turn to the construction of the CS and dCDH event-study estimates, which recall are numerically equivalent in this simple non-staggered setting.[7] CS/dCDH construct event-studies by aggregating DiD comparisons of treated and not-yet-treated units. However, by default, the construction is asymmetric across the pre-treatment and post-treatment periods. In particular, the pre-treatment coefficients average "short-differences", i.e. comparisons of consecutive periods, whereas the post-treatment coefficients are "long-differences", i.e. comparisons relative to the period before treatment.

[7] Specifically, I am referring to the dCDH estimates in R; see FN 2.

Formally, in our setting with common treatment timing, we have that

$$\hat{\beta}_r^{CS/dCDH} = \begin{cases} \left(\hat{E}[Y_{i,r+1} \mid D_i = 1] - \hat{E}[Y_{i,r+1} \mid D_i = 0]\right) - \left(\hat{E}[Y_{i,r} \mid D_i = 1] - \hat{E}[Y_{i,r} \mid D_i = 0]\right) & \text{if } r < 0 \\ \left(\hat{E}[Y_{i,r+1} \mid D_i = 1] - \hat{E}[Y_{i,r+1} \mid D_i = 0]\right) - \left(\hat{E}[Y_{i0} \mid D_i = 1] - \hat{E}[Y_{i0} \mid D_i = 0]\right) & \text{if } r \geqslant 0 \end{cases}$$

Note that for $r < 0$, the comparison period is period $r-1$, whereas for $r \geqslant 0$ the comparison period is always period 0. Under mild regularity conditions, it follows that $\hat{\beta}_r^{CS/dCDH}$ is consistent for

$$\beta_r^{CS/dCDH} = \begin{cases} (E[Y_{i,r+1} \mid D_i = 1] - E[Y_{i,r+1} \mid D_i = 0]) - (E[Y_{i,r} \mid D_i = 1] - E[Y_{i,r} \mid D_i = 0]) & \text{if } r < 0 \\ (E[Y_{i,r+1} \mid D_i = 1] - E[Y_{i,r+1} \mid D_i = 0]) - (E[Y_{i0} \mid D_i = 1] - E[Y_{i0} \mid D_i = 0]) & \text{if } r \geqslant 0 \end{cases}$$

In our simulation DGP, we thus have that

$$\beta_r^{CS/dCDH} = \begin{cases} \gamma & \text{if } r < 0 \\ \gamma \cdot (r+1) & \text{if } r \geqslant 0 \end{cases}$$

so the population version of the CS/dCDH event-study coefficients show a kink at zero if $\gamma \neq 0$.

## 3.3 BJS

Like the CS/dCDH event-study approach, the BJS event-study is also constructed in a way that is asymmetric between the pre-treatment and post-treatment periods. For the post-treatment effects, they use what they call an imputation approach: they fit a TWFE model using only untreated $(i,t)$ pairs, then form individual treatment effect estimates of the form $\hat{\tau}_{it} = Y_{it} - \hat{Y}_{it}$, where $\hat{Y}_{it}$ is the prediction from the TWFE model. They then average the $\hat{\tau}_{it}$ for units at a given lag from treatment. For the pre-treatment model, they simply run a dynamic TWFE specification using only $(i,t)$ pairs that are untreated, with dynamic indicators for the number of periods until treatment (with the earliest pre-treatment period normalized to zero).[8]

In our special case of non-staggered treatment timing, the post-treatment event-study coefficients for BJS correspond to simple DiDs between period $r+1$ and the *average* outcome in the pre-treatment period. By contrast, the pre-treatment coefficients correspond to DiDs between period $r+1$ and the earliest period, period $-\underline{T}$. Specifically, let $\bar{Y}_i^{pre} = \frac{1}{\underline{T}+1} \sum_{t=-\underline{T}}^{0} Y_{it}$

[8]The Stata package requires the researcher to specify how many pre-treatment coefficients they would like to calculate, and all earlier periods are used as the omitted category. The omitted category is thus the earliest period if one requests $\underline{T}$ pre-treatment coefficients.

be $i$'s average outcome in periods $t \leqslant 0$. Then we have

$$\hat{\beta}_r^{BJS} = \begin{cases} \Big(\hat{E}[Y_{i,r+1} \mid D_i = 1] - \hat{E}[Y_{i,r+1} \mid D_i = 0]\Big) - \Big(\hat{E}[Y_{i,-\underline{T}} \mid D_i = 1] - \hat{E}[Y_{i,-\underline{T}} \mid D_i = 0]\Big) & \text{if } r < 0 \\ \Big(\hat{E}[Y_{i,r+1} \mid D_i = 1] - \hat{E}[Y_{i,r+1} \mid D_i = 0]\Big) - \Big(\hat{E}[\bar{Y}_i^{pre} \mid D_i = 1] - \hat{E}[\bar{Y}_i^{pre} \mid D_i = 0]\Big) & \text{if } r \geqslant 0 \end{cases}$$

This makes clear that the construction is asymmetric, with $\hat{\beta}_r^{BJS}$ using a comparison to period $-\underline{T}$ for the pre-treatment coefficients and to the average of the pre-period for the post-treatment coeffiients.

Under mild regularity conditions, we thus have that $\hat{\beta}_r^{BJS}$ is consistent for

$$\beta_r^{BJS} = \begin{cases} (E[Y_{i,r+1} \mid D_i = 1] - E[Y_{i,r+1} \mid D_i = 0]) - \big(E[Y_{i,-\underline{T}} \mid D_i = 1] - E[Y_{i,-\underline{T}} \mid D_i = 0]\big) & \text{if } r < 0 \\ (E[Y_{i,r+1} \mid D_i = 1] - E[Y_{i,r+1} \mid D_i = 0]) - \big(E[\bar{Y}_i^{pre} \mid D_i = 1] - E[\bar{Y}_i^{pre} \mid D_i = 0]\big) & \text{if } r \geqslant 0 \end{cases}$$

Under our simulation DGP, this corresponds to[9]

$$\beta_r^{BJS} = \begin{cases} \gamma \cdot (r + 1 - (-\underline{T})) & \text{if } r < 0 \\ \gamma \cdot (r + 1 - \frac{1}{2}(-\underline{T})) & \text{if } r \geqslant 0 \end{cases}$$

so the population version of BJS will exhibit a jump at zero if $\gamma \neq 0$.

# 4 Practical takeaways and recommendations

The first practical takeaway from these results is that the default event-studies from CS, dCDH, and BJS should not be interpreted in the same way as traditional dynamic TWFE event-study plots. Owing to the asymmetric construction of the pre-treatment and post-treatment coefficients, a kink or jump in the plot may arise even if there is no treatment effect and parallel trends is equally violated in all periods. One should therefore not apply the typical heuristics for visual inference—e.g. looking for a discontinuity or kink at the treatment date—that may be familiar from TWFE event-studies in the non-staggered case.

A natural follow-up question is to what extent the event-study plots for these alternative methods can be modified to be more comparable to conventional event-study plots. For CS/dCDH, there is a straightforward solution, which is to use "long-differences" for the pre-treatment coefficients as well as the post-treatment coefficients (i.e. always use the period

[9] In the calculation below, we use the fact that

$$E[\bar{Y}_i^{pre} \mid D_i = 1] = \gamma \frac{1}{\underline{T}+1}(-\underline{T} + ... + 0) = \gamma \frac{1}{\underline{T}+1}\left(-\frac{1}{2}\underline{T}(\underline{T}+1)\right) = -\frac{1}{2}\underline{T} \cdot \gamma.$$

before treatment as the baseline). This can be implemented, for example, in the `did` R and Stata packages using the options `base_period = "universal"` and `long2`, respectively.[10] Using the default options in the `did_multiplegt` Stata package also yields comparisons using "long differences", in contrast to the results for the R package shown above. Using these settings, the event-study estimates are numerically equivalent to the dynamic TWFE specification in the non-staggered setting considered here.

For BJS, the answers are less obvious. By design, the BJS approach uses *all* pre-treatment periods to compute counterfactual post-treatment outcomes—which increases efficiency under spherical errors—but this seems to lead to an inherent asymmetry between the post-treatment estimates (which assume parallel trends for all periods) and the pre-treatment estimates (which try to test this assumption). It is worth noting that the pre-treatment coefficients in the BJS event-study are still valid tests of pre-treatment parallel trends in the sense that if parallel trends holds in all pre-treatment periods, the $\beta_r^{BJS}$ should be zero for all $r < 0$. One possible approach is then to treat the testing of pre-treatment parallel trends, and estimation of post-treatment effects *assuming* parallel trends in all periods as completely separate exercises. This motivates putting the BJS pre-treatment estimates on a different plot from the post-treatment estimates to avoid making misleading visual inferences.[11] Alternatively, the researcher might report an event-study for a method that treats the periods symmetrically (e.g. using the `did` package with the options described above) before reporting the potentially more-precise estimates from BJS that assume parallel trends in all periods.

It is also worth emphasizing that the discussion here does not threaten the validity of the post-treatment event-study estimates *if* the appropriate parallel trends assumption holds: in this case, all of the recent methods yield interpretable treatment effect estimates even under heterogeneous treatment effects. Rather, the asymmetric construction of the pre-treatment and post-treatment event-study coefficients makes it challenging to visualize whether the estimated post-treatment coefficients could be explained by a violation of parallel trends rather than a treatment effect, at least using the conventional visual heuristics.

[10] A blog post by Callaway (2021) provides further discussion of the differences between these two options. Similar to our ongoing example, he considers a DGP with a linear violation of parallel trends, although it only has one post-treatment period and thus does not exhibit the kink seen in our example. Nevertheless, he concludes that long-run violations of parallel trends are easier to visualize using the universal-base option, but argues that using short-differences may make it easier to detect violations of the no anticipation assumption.

[11] Indeed, the documentation for BJS' Stata packages states that "A pre-trend test (for the assumptions of parallel trends and no anticipation) is a separate exercise."

# References


**Borusyak, Kirill, Xavier Jaravel, and Jann Spiess**, "Revisiting Event Study Designs: Robust and Efficient Estimation," *Review of Economic Studies*, 2024, *forthcoming.*

**Callaway, Brantly**, "Universal vs. Varying Base Period in Event Studies," November 2021. https://bcallaway11.github.io/posts/event-study-universal-v-varying-base-period.

_ **and Pedro H. C. SantAnna**, "Difference-in-Differences with multiple time periods," *Journal of Econometrics*, 2021, *225* (2), 200–230.

**de Chaisemartin, Clément and Xavier D'Haultfuille**, "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 2020, *110* (9), 2964–2996.

_ **and** _ , "Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey," *Econometrics Journal*, 2022.

**Rambachan, Ashesh and Jonathan Roth**, "A More Credible Approach to Parallel Trends," *The Review of Economic Studies*, October 2023, *90* (5), 2555–2591.

**Roth, Jonathan**, "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends," *American Economic Review: Insights*, 2022, *4* (3), 305–322.

_ **, Pedro H. C. SantAnna, Alyssa Bilinski, and John Poe**, "Whats trending in difference-in-differences? A synthesis of the recent econometrics literature," *Journal of Econometrics*, August 2023, *235* (2), 2218–2244.

**Sun, Liyang and Sarah Abraham**, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 2021, *225* (2), 175–199.