

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

Jonathan Roth\*

January 16, 2026

## 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 several of the most popular recent methods do not match those of traditional two-way fixed effects (TWFE) event-studies. The plots produced by 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 evaluating violations of parallel trends using 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 report an event-study plot that shows the differences in trends in the outcome between the treated and comparison groups both before and after the treatment date.<sup>1</sup> Event-study plots serve two purposes. First, if the DiD design is valid, they allow the reader to evaluate the dynamic patterns of treatment effects following the implementation of the policy of interest. Second, event-study plots

---

\*Brown University. [jonathanroth@brown.edu](mailto:jonathanroth@brown.edu). I am grateful to Kirill Borusyak, Clément de Chaisemartin, Xavier D'Haultfoeuille, Peter Hull, Ziyi Liu, Pedro Sant'Anna, and Jesse Shapiro for comments. I thank the authors of all of the statistical packages studied here for their valuable (and often thankless) role in developing tools for implementing recent difference-in-differences methods and for making their code publicly available.

<sup>1</sup>Roth (2022) identifies 70 recent papers in top economics journals displaying such plots.

provide evidence about the validity of the research design. They enable the reader to perform “visual inference” to determine to what extent the estimated treatment effects may be driven by violations of the parallel trends assumption. An event-study plot is generally considered more convincing evidence in favor of a non-zero treatment effect if there is a discontinuity or kink in the coefficients close to the treatment date, so that the pattern seen cannot simply be explained by a smooth violation of parallel trends (Freyaldenhoven, Hansen, Pérez and Shapiro, 2021; Rambachan and Roth, 2023).

Until recently, event-study plots were typically created by estimating a dynamic two-way fixed effects (TWFE) regression specification, with fixed effects for unit and time and indicators for time relative to treatment (with the coefficient for the period directly before treatment typically normalized to zero). A recent literature surveyed in de Chaisemartin and D’Haultfoeuille (2022) and Roth, Sant’Anna, Bilinski and Poe (2023) has illustrated that such TWFE specifications may be difficult to interpret in settings with heterogeneous treatment effects and staggered treatment timing, even if a parallel trends assumption holds. 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  $r'$ . This undermines the first purpose of event-study plots described above, since the post-treatment event-study coefficients may not correspond directly to the dynamics of the treatment effects. These issues with TWFE specifications have motivated the development of a variety of new DiD methods that allow for a more straightforward evaluation of dynamic treatment effects, assuming parallel trends holds, in settings with staggered timing and heterogeneous treatment effects. These new methods typically produce event-studies that also show pre-treatment coefficients. It would then seem natural to also use these plots for the second purpose of evaluating the extent to which the results could be explained by violations of parallel trends, using the same heuristics as with traditional dynamic TWFE specifications in the non-staggered setting.

The main point of this note is that for the purposes of assessing violations of parallel trends, the default event-study plots produced by software for several of the most popular recent methods should not be interpreted in the same way as traditional dynamic TWFE specifications. Indeed, I show that these event-study plots do not match those of traditional dynamic TWFE specifications even when specialized to the setting with *non-staggered* treatment timing. This can be seen in Figure 1 below, which shows the event-studies produced by different methods when applied to the same simulated dataset in which there is non-staggered treatment timing. In the simulation (the details of which are described below) there is no effect of the treatment but parallel trends is violated: the outcomes for the treated group are on an upwards linear trend relative to the comparison group. As expected, the event-study

plot produced using dynamic TWFE yields coefficients that lie close to a straight line. By contrast, the event-study plot produced by the R package for implementing the method of Callaway and Sant’Anna (2021, henceforth, CS) shows a clear kink at the treatment date.<sup>2</sup> Meanwhile, the event-study plot produced by the Stata package for Borusyak, Jaravel and Spiess (2024, henceforth, BJS) shows a clear discontinuity at the treatment date. Given these starkly different results when applied to the *same* data-set with non-staggered treatment timing, it is clear that the typical heuristics for visual inference developed based on dynamic TWFE specifications will 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 recent methods can 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 and further discussion.

## 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 some newer methods with a simulation of an extremely simple non-staggered DiD setting.<sup>3</sup> 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 group’s outcome in all periods. In particular,  $Y_{it}(0) = \gamma \cdot t + \epsilon_{it}$  for units in the treated group and  $Y_{it}(0) = \epsilon_{it}$  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  $\gamma = 0.5$ .

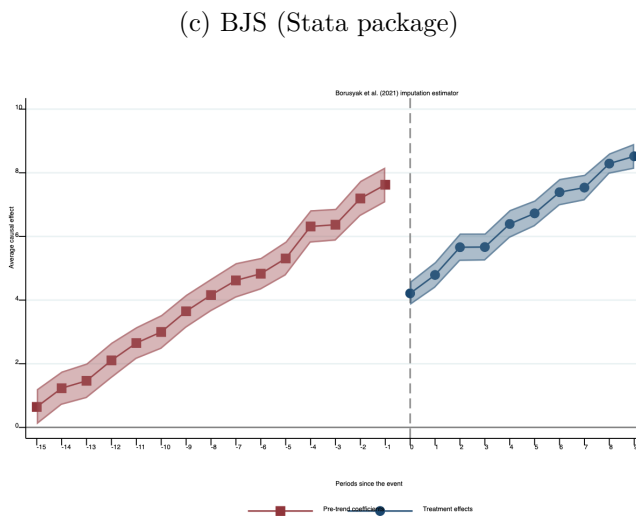
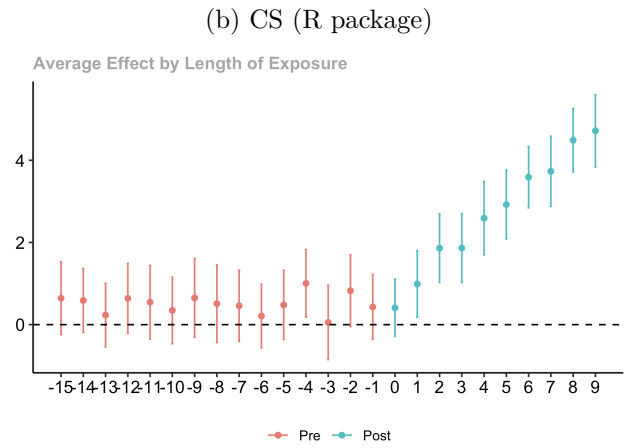
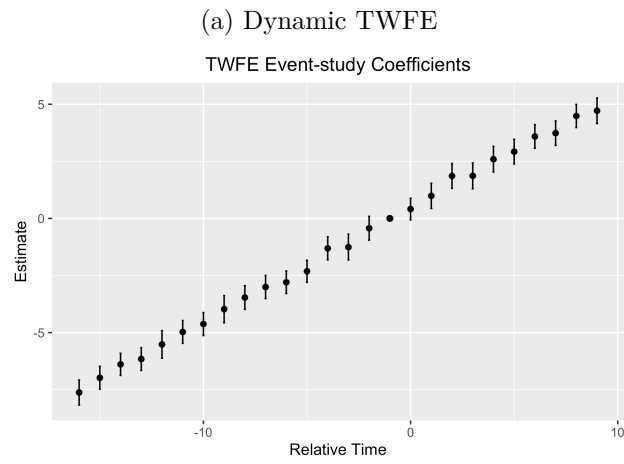
Figure 1 shows event-studies for a single simulation draw from this DGP when there are 100 total units and half of them are treated. Panel (a) shows the results of a traditional dynamic TWFE ordinary least-squares specification of the form

---

<sup>2</sup>As I describe below, the identical pattern also arises using a previous version of the R package for implementing the method of de Chaisemartin and D’Haultfœuille (2020, henceforth, dCDH). Subsequent to the release of the initial draft of this paper in January 2024, the R package for dCDH was updated to produce results matching the TWFE event-study.

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

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



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

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 by the BJS method, using the default settings in the `did_imputation` Stata package.<sup>4</sup>

As described in the introduction, despite using identical data, the three 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 event-study plot has relatively flat pre-treatment coefficients and then exhibits a kink at the time of the treatment, sloping upwards afterwards. The BJS event-study appears to show an upward pre-trend, but then exhibits a sharp downward jump at the treatment date.

**Other estimators.** Although I focus on CS and BJS, two of the most popular recent methods, I now comment briefly on the event-study plots produced by other recent DiD methods. At the time of the release of the initial draft of this paper in January 2024, the `DIDmultiplert` R package for implementing the method of [de Chaisemartin and D’Haultfoeuille \(2020\)](#), dCDH) produced identical point estimates to the CS event-study described above; see Figure 2. However, the most recent version of the package produces results that match the original TWFE event-study plot in my simple example.<sup>5</sup> Two other popular approaches are those of [Sun and Abraham \(2021\)](#), henceforth SA) and [Wooldridge \(2025\)](#), henceforth W). Both of these papers propose richer TWFE specifications to capture treatment effect heterogeneity in staggered settings. However, these TWFE specifications reduce to the usual dynamic TWFE event-study in (1) when specialized to the case with non-staggered treatment timing, and thus match panel (a) of Figure 1 by construction. This can be seen in panels (c) and (d) of Figure 2, which shows implementations of SA and W from the `etwfe` and `fixest` packages in R.<sup>6</sup> Two other popular approaches are those of [Liu, Wang and Xu \(2024\)](#) and [Gardner \(2021\)](#). Like BJS, these papers rely on an “imputation” approach for inferring counterfactual outcomes, although they construct the pre-treatment coefficients in the event-study differ-

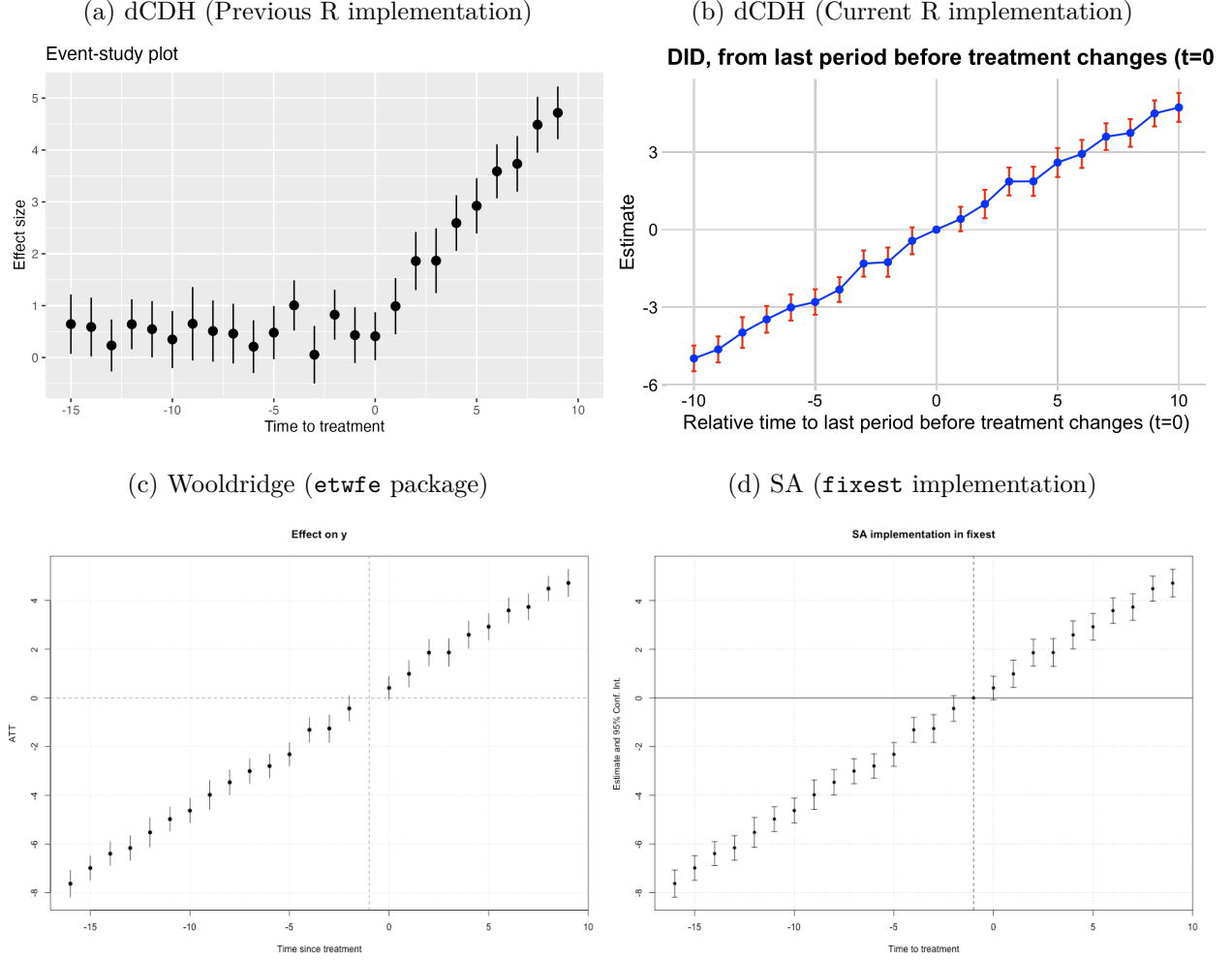
---

<sup>4</sup>All packages were re-installed using the latest versions as of October 31, 2025. I obtained similar results as of the initial draft of this paper in January 2024, with one exception owing to a software update described below (see Figure 2 and surrounding discussion).

<sup>5</sup>The dCDH package only reports as many pre-trends coefficients as there are post-treatment periods. Thus, it matches the TWFE event-study for the 10 latest pre-treatment periods but does not show the earlier pre-trends coefficients.

<sup>6</sup>For `etwfe`, we must specify the option `cgroup = "never"` in order to get pre-trends estimates.

Figure 2: Comparison of event-studies from other methods in the non-staggered setting



ently. In Appendix Section A, I discuss the results produced by the packages for these papers (`fect` and `did2s`). Both of these methods also yield an event-study that shows a jump at the date of treatment.

### 3 Mathematical Examination

In this section, I show that the differences between the event-plots created by dynamic TWFE and CS and BJS 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  $t = -\underline{T}, \dots, \bar{T}$ . 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  $Y_{it}(1) = Y_{it}(0) = \gamma \cdot t \cdot D_i + \epsilon_{it}$  and  $E[\epsilon_{it} | D_i] = 0$ .

### 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 estimated by OLS are given by

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

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 in probability to population means, and thus  $\hat{\beta}_r$  is consistent for

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

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

I now turn to the construction of the CS event-study estimates. CS constructs 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 are “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.

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

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

Note that for pre-treatment coefficients ( $r < 0$ ), we compare outcomes in consecutive periods (periods  $r + 1$  vs  $r$ ), whereas for post-treatment coefficients ( $r \geq 0$ ), we compare outcomes over longer horizons (periods  $r + 1$  vs 0). Under mild regularity conditions, it follows that  $\hat{\beta}_r^{CS}$  is consistent for

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

In our simulation DGP, we thus have that

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

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

### 3.3 BJS

Like the CS 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 regression 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 regression. They then average the  $\hat{\tau}_{it}$  for treated units at a given lag from treatment. For the pre-treatment coefficients, they simply run a dynamic TWFE regression 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).<sup>7</sup>

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,  $-T$ . Specifically, let  $\bar{Y}_i^{pre} = \frac{1}{T+1} \sum_{t=-T}^0 Y_{it}$  be

---

<sup>7</sup>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  $T$  pre-treatment coefficients.



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

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

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

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

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

Under our simulation DGP, this corresponds to<sup>8</sup>

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

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 the methods considered in Figure 1 should not be interpreted in the same way as traditional dynamic TWFE event-study plots for the purposes of evaluating parallel trends violations. 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. Sensitivity analyses that compare the pre-treatment and post-treatment coefficients from these event-studies (e.g. [Manski and Pepper, 2018](#); [Rambachan and Roth, 2023](#)) may likewise yield misleading conclusions given the asymmetric construction of the pre- and post-treatment coefficients.

---

<sup>8</sup>In the calculation below, we use the fact that

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

A natural follow-up question is: to what extent can the event-study plots for these alternative methods be modified to be more comparable to conventional event-study plots? I discuss CS and BJS in turn, followed by some discussion of additional considerations.

**CS.** For CS, there is a straightforward answer, which is to use “long-differences” for the pre-treatment coefficients as well as the post-treatment coefficients (i.e. always use the period before treatment as the baseline). This can be implemented, for example, in the `did` R and `csdid` Stata packages using the options `base_period = "universal"` and `long2`, respectively.<sup>9</sup> Using these settings, the event-study estimates are numerically equivalent to the dynamic TWFE specification in the non-staggered setting considered here.

**BJS.** For BJS, the answers are somewhat less obvious. The challenge arises from the fact that, by construction, BJS uses *all* of the pre-treatment data in predicting post-treatment counterfactuals for treated units (as they show, this is efficient under spherical errors).<sup>10</sup> Any analysis that does a similar imputation in the pre-treatment period will either (a) be looking at *in-sample* prediction fit—i.e. looking at residuals  $Y_{it} - \hat{Y}_{it}$ , where  $Y_{it}$  was included in the training data for creating  $\hat{Y}_{it}$ , or (b) using predictions trained only on a sub-set of the pre-treatment data, in contrast to the predictions for post-treatment outcomes, which use all of the pre-treatment data. Nevertheless, it may be possible to construct event-study plots using imputation estimators that are more similar to the standard TWFE event-study plot than the BJS implementation. Below I discuss a few possible options, along with their pros and cons; this strikes me as an area where additional innovation in future work may also be possible.

In my ongoing non-staggered example, one natural approach to make the BJS event-study coefficients more symmetric would be to use the *average* of the pre-treatment periods as the reference for both the pre-treatment and post-treatment event-study coefficients. That is, we could construct

$$\hat{\beta}_r^{BJS,new} = \left( \hat{E}[Y_{i,r+1} \mid D_i = 1] - \hat{E}[Y_{i,r+1} \mid D_i = 0] \right) - \left( \hat{E}[\bar{Y}_i^{pre} \mid D_i = 1] - \hat{E}[\bar{Y}_i^{pre} \mid D_i = 0] \right) \text{ for all } r.$$

---

<sup>9</sup>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.

<sup>10</sup>The same is true of other imputation approaches such as [Gardner \(2021\)](#) and [Liu et al. \(2024\)](#). In our non-staggered setting, the post-treatment effects estimated by these approaches are equivalent to those for BJS; the methods only differ in how they create the pre-treatment coefficients. Thus, the recommendations for modifying the pre-treatment coefficients for BJS also apply to these other methods.

This would match the usual TWFE event-study up to a *vertical shift* of all the coefficients. That is,  $\hat{\beta}_r^{BJS,new} = \hat{\beta}_r^{TWFE} + \hat{c}$  for a value  $\hat{c}$  that doesn't depend on  $r$ . However, it is worth noting that this construction guarantees that the modified pre-trends coefficients *average* to zero (i.e.  $\sum_{r<0} \hat{\beta}_r^{BJS,new} = 0$ ). Visual inference based on comparing the *average* of the event-study coefficients before and after treatment may thus be misleading, since the average in the pre-treatment period is zero by construction. Nonetheless, using this more symmetric imputation approach seems preferable to me than the default BJS approach if one wants output resembling a typical event-study, provided that the researcher keeps in mind that they should only be looking at the relative differences between the coefficients and not their averages before and after treatment. How can this approach be extended to the staggered setting? Liu (2025) proposes a notion which he refers to as “block bias” which measures the pre-treatment violations of parallel trends for each adoption cohort (relative to a fixed comparison group) in a staggered setting.<sup>11</sup> Applied to our non-staggered example, the pre-trends estimated by the block bias approach align with the modified estimator  $\hat{\beta}_r^{BJS,new}$ . Perhaps the most natural approach for imputation estimators in staggered settings is then to create an event-study where the pre-trends estimates are formed by taking a (weighted) average of the block biases across adoption cohorts.

A related suggestion by Li and Strezhnev (2025) is to adopt a “leave-one-out” approach to pre-trends estimation using imputation estimators, where for relative time  $r < 0$ , one computes averages of the form  $Y_{it} - \hat{Y}_{it}^{loo}$ , where the prediction  $\hat{Y}_{it}^{loo}$  does not use information from relative time  $r$ . This approach does not align with the standard TWFE event-study when specialized to the common timing case—the pre-trend coefficients are based on predictions using all-but-one pre-treatment period but the post-treatment coefficients are based on predictions using all the pre-treatment periods—but the differences are small when the number of pre-treatment periods is large.<sup>12</sup>

An alternative approach recommended by BJS is to completely separate the testing and estimation steps.<sup>13</sup> The pre-treatment coefficients in the BJS default 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  $r < 0$ , and so can be used for testing the null of parallel pre-trends.<sup>14</sup> However, if one adopts this approach, then I would

<sup>11</sup>Liu (2025) also discusses how these block biases can be used in conjunction with the sensitivity analysis approach of Rambachan and Roth (2023).

<sup>12</sup>Another approach, which is implemented in the `fect` and `did2s` packages, reports averages of the form  $Y_{it} - \hat{Y}_{it}$  for *all* relative time periods  $r$ , where  $\hat{Y}_{it}$  is the same as for BJS. As discussed in more detail in Appendix Section A, this produces an event-study that shrinks the coefficients  $\hat{\beta}_r^{BJS,new}$  towards zero in the pre-treatment period, and thus may produce event-study plots with discontinuities at the treatment date.

<sup>13</sup>The documentation for BJS' Stata package states, “A pre-trend test (for the assumptions of parallel trends and no anticipation) is a separate exercise.”

<sup>14</sup>See Roth (2022) and Bilinski and Hatfield (2018) regarding pitfalls of simply looking at the significance of

humbly suggest putting the BJS pre-treatment estimates on a different plot from the post-treatment estimates to avoid making misleading visual inferences. If one prefers to have a typical-looking event-study plot, then a straightforward way of separating the testing and estimation steps would be for the researcher to report an event-study for a method that treats the periods completely symmetrically (e.g. using the `did` package with the options described above) to visually evaluate the plausibility of the parallel trends assumption, and to then subsequently report the potentially more-precise post-treatment estimates from BJS that assume parallel trends in all periods. One can also conduct sensitivity analysis for these post-treatment estimates directly using the approach in [Rambachan and Roth \(2023\)](#); see [Liu \(2025\)](#) for suggestions for implementing this approach in staggered settings in conjunction with imputation estimators like BJS.

**Additional discussion.** It is worth emphasizing that the discussion here does not threaten the validity of the post-treatment event-study estimates for recent methods *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 (or typical sensitivity analyses).

Finally, it is worth noting that even if a method produces an event-study plot that matches the usual TWFE event-study when specialized to the non-staggered setting, additional complications may arise when interpreting these results in staggered settings. One particularly important issue for interpretation in the staggered setting, which is discussed in detail in [Liu \(2025\)](#), is that the composition of units that contribute to an event-study coefficient will often change across different relative time periods. Indeed, if a panel is balanced in calendar time, then with staggered treatment adoption it is necessarily *imbalanced* in relative time. One therefore has to be careful in interpreting aggregated event-studies since the composition of units that contribute to the coefficient for relative time  $r$  will be different from that for relative time  $r'$ . To assess the importance of changing composition, it will often be useful to supplement aggregated event-studies with plots of the time series of the mean outcome by adoption cohort. It may also be useful to report aggregated event-studies that enforce some balancing criteria—e.g. reporting an event-study that is restricted

---

the pre-treatment coefficients, however. As argued in those papers, researchers should ideally also look at the *magnitudes* of the pre-trends estimates and their confidence intervals. However, evaluating the magnitudes of potential bias from pre-trends is challenging if the pre-treatment and post-treatment coefficients are not constructed symmetrically.

to units who are observed at least 3 periods before and after treatment.

## References

- Bergé, Laurent**, *fixest: Fast Fixed-Effects Estimations* 2025. R package version 0.13.2 (CRAN).
- Bilinski, Alyssa and Laura A. Hatfield**, “Seeking evidence of absence: Reconsidering tests of model assumptions,” *arXiv:1805.03273 [stat]*, May 2018.
- Borusyak, Kirill and Xavier Jaravel**, *DID\_IMPUTATION: Stata module to perform treatment effect estimation and pre-trend testing in event studies* 2023. Statistical Software Components S458957, revised 22 Nov 2023. Available via `ssc install did_imputation`.
- , —, and **Jann Spiess**, “Revisiting Event-Study Designs: Robust and Efficient Estimation,” *The Review of Economic Studies*, November 2024, *91* (6), 3253–3285.
- Butts, Kyle**, *did2s: Two-Stage Difference-in-Differences Following Gardner (2021)* 2025. R package version 1.2.0 (CRAN).
- 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.
- and —, *did: Treatment Effects with Multiple Periods and Groups* 2022. R package version 2.1.2 (CRAN).
- Ciccia, Diego, Felix Knau, Méline Malezieux, Doulo Sow, Shuo Zhang, and Clément de Chaisemartin**, *DIDmultiplt: Estimators DID with Multiple Groups and Periods* 2024. R package version 2.0.0.
- de Chaisemartin, Clément and Xavier D’Haultfœuille**, “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.

- Freyaldenhoven, Simon, Christian Hansen, Jorge Pérez Pérez, and Jesse M. Shapiro**, “Visualization, identification, and estimation in the linear panel event-study design.,” *Advances in Economics and Econometrics: Twelfth World Congress*, 2021, *Forthcoming*.
- Gardner, John**, “Two-stage differences in differences,” *Working Paper*, 2021.
- Li, Zikai and Anton Strezhnev**, “Benchmarking parallel trends violations in regression imputation difference-in-differences,” July 2025.
- Liu, Licheng, Ye Wang, and Yiqing Xu**, “A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data,” *American Journal of Political Science*, 2024, 68 (1), 160–176. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/ajps.12723>.
- , **Ziyi Liu, Ye Wang, Yiqing Xu, Tianzhu Qin, Shiyun Hu, and Rivka Lipkowitz**, *fect: Fixed Effects Counterfactual Estimators* 2025. R package version 2.0.5.
- Liu, Ziyi**, “Cohort-Anchored Robust Inference for Event-Study with Staggered Adoption,” September 2025. arXiv:2509.01829 [econ].
- Manski, Charles F. and John V. Pepper**, “How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity Using Bounded-Variation Assumptions,” *The Review of Economics and Statistics*, 2018, 100 (2), 232–244.
- McDermott, Grant**, *etwfe: Extended TwoWay Fixed Effects* 2025. R package version 0.6.0.
- 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.
- Wooldridge, Jeffrey M.**, “Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators,” *Empirical Economics*, August 2025.

## A Event-study plots based on in-sample imputation

Another approach to constructing event-studies for imputation estimators, which is implemented in the `fect` and `did2s` packages, reports averages of the form  $Y_{it} - \hat{Y}_{it}$  for *all* relative time periods  $r$ , where  $\hat{Y}_{it}$  is the same as for BJS described above. [Li and Strezhnev \(2025\)](#) study this approach in detail, and argue that it systematically under-estimates the magnitude of the pre-trends. The issue stems from the fact that the pre-treatment outcomes for treated units are used to form  $\hat{Y}_{it}$ , and thus  $Y_{it} - \hat{Y}_{it}$  is an *in-sample* prediction error for pre-treatment observations. When applied to the non-staggered setting, the results in [Li and Strezhnev \(2025\)](#) imply that this in-sample approach yields pre-treatment coefficients equal to  $\frac{N_0}{N} \hat{\beta}_r^{BJS, new}$  for  $r < 0$ , where  $N_0$  is the number of untreated units. Thus, the pre-treatment coefficients  $\hat{\beta}_r^{BJS, new}$  are multiplied by the factor  $\frac{N_0}{N} < 1$ , which is very small when most of the units are treated. Appendix Figure 1 illustrates this attenuation by plotting the event-studies produced by `fect` and `did2s` in our simulated data: there is a clear jump in the coefficients around the treatment date.

Appendix Figure 1: Event-studies from methods using in-sample imputation (`fect` and `did2s`)

