did2s: Two-Stage Difference-in-Differences

by Kyle Butts and John Gardner

Abstract Recent work has highlighted the difficulties of estimating difference-in-differences models when treatment timing occurs at different times for different units. This article introduces the R package **did2s** which implements the estimator introduced in Gardner (2021). The article provides an approachable review of the underlying econometric theory and introduces the syntax for the function `did2s`. Further, the package introduces a function, `event_study`, that provides a common syntax for all the modern event-study estimators and `plot_event_study` to plot the results of each estimator.

Introduction

A rapidly growing literature has identified difficulties in traditional difference-in-differences estimation when treatment turns on at different times for different groups and when the effects of treatment vary across groups and over time (Callaway and Sant'Anna, 2020; Sun and Abraham, 2020; Goodman-Bacon, 2018; Borusyak et al., 2021). Gardner (2021) proposes an estimator of the two-way fixed effects model that is quick, intuitive, and memory-efficient. The estimator relies on the standard two-way fixed effect model (see following section) and forms an intuitive estimate: the average difference in outcomes between treated and untreated units after removing fixed unit and time shocks.

This article discusses the software package **did2s** which implements the two-stage estimation approach proposed by Gardner (2021) to estimate robustly the two-way fixed effects (TWFE) model. There are two notable technical features of this package. First, `did2s` utilizes the incredibly fast package, **fixest**, which can estimate regressions with a high-number of fixed effects very quickly. Second, the package utilizes sparse matrices which allows it to estimate on larger datasets than alternative estimators. Since there a few alternative TWFE event-study estimators implemented in `R` with differing syntax and data formatting requirements, the package also has a set of functions that allow quick estimation and plotting of every alternative event study estimators using a standardized syntax. This allows for easy comparison between the results of different methods.

Difference-in-Differences Theory

Researchers commonly use the difference-in-differences (DiD) methodology to estimate the effects of treatment in the case where treatment is non-randomly assigned. Instead of random assignment giving rise to identification, the method relies on the so-called "parallel trends" assumption which says that outcomes would evolve in parallel between the treated and untreated groups *in the world where the treated were untreated*. This is formalized with the *two-way fixed effects* (TWFE) model. For example, in a static setting where average treatment effects are constant across treatment groups and over time, researchers use the static TWFE model:

$$y_{igt} = \mu_g + \eta_t + \tau D_{gt} + \varepsilon_{igt},$$

where y_{igt} is the outcome, where i denotes the individual, g denotes the initial-treatment year, and t denotes time; μ_g is a vector of time-invariant group characteristics; η_t are a vector of period shocks; and D_{gt} is an indicator for whether initial-treatment group g is receiving treatment in period t, i.e. $D_{gt} = 1(g \leq t)$. The coefficient of interest is τ which represents the (*constant*) average effect of the treatment on the treated (ATT).

Similarly, a (dynamic) event-study TWFE model could be written as:

$$y_{igt} = \mu_g + \eta_t + \sum_{k=-L}^{-2} \tau^k D_{gt}^k + \sum_{k=1}^{K} \tau^k D_{gt}^k + \varepsilon_{igt}, \tag{1}$$

where D_{gt}^k are lag/leads of treatment (k periods from initial treatment date, g). The coefficients of interests are the τ^k, which represent the average effect of being treated for k periods. For negative values of k, τ^k are known as "pretrends," and represent the average deviation in outcomes for treated units k periods away from treatment, relative to their value in the reference period. These pretrend estimates are commonly used as a test of the parallel counterfactual trends assumption.

In the case where treatment turns on at the same time for all treated units and under a parallel trends restriction on the error term ε_{it}, average treatment effects are identified by the TWFE and event-

study models. As noted above, when treatment turns on at different times, estimates of the above equations do not necessarily identify intuitive estimands (Callaway and Sant'Anna, 2020; Sun and Abraham, 2020; Goodman-Bacon, 2018; Borusyak et al., 2021). One way of thinking about this problem is through the Frisch–Waugh–Lovell (FWL) theorem (Frisch and Waugh, 1933). By the FWL theorem, estimating the above TWFE specification by least squares is equivalent to regressing outcomes on a "residualized" version of the treatment indicator D_{gt} or D_{gt}^k (i.e., the residual from a regression of D_{gt} or each D_{gt}^k on group and time fixed effects). To simplify the literature, this residualized treatment indicator is what creates the problem of interpreting τ or τ^k as average treatment effects, especially when treatment effects are heterogeneous (see, for example, Goodman-Bacon (2018) and Sun and Abraham (2020)).

Two-stage Difference-in-Differences Estimator

Gardner (2021) proposes an estimator to resolve the problem with the two-way fixed effects approach. Rather than attempting to estimate the group and time effects at the same time as the ATT, this approach proceeds from the observation that, under parallel trends, the group and time effects are identified from the subsample of untreated/not-yet-treated observations ($D_{gt} = 0$). This suggests a simple two-stage difference-in-differences estimator:

1. Estimate the model
$$y_{igt} = \mu_g + \eta_t + \varepsilon_{igt}$$
using the subsample of untreated/not-yet-treated observations (i.e., all observations for which $D_{gt} = 0$), retaining the estimated group and time effects in order to form the adjusted outcomes $\tilde{y}_{igt} \equiv y_{igt} - \hat{\mu}_g - \hat{\eta}_t$.
2. Regress adjusted outcomes $\tilde{y}_{igt}$ on treatment status D_{gt} or D_{gt}^k to estimate treatment effects.

To see why this procedure works, note that parallel trends implies that outcomes can be expressed as

$$\begin{aligned} y_{igt} &= \mu_g + \eta_t + \tau_{gt} D_{gt} + \varepsilon_{gt} \\ &= \mu_g + \eta_t + \bar{\tau} D_{gt} + (\tau_{gt} - \bar{\tau}) D_{gt} + \varepsilon_{gt}, \end{aligned}$$

where $\tau_{gt} = E(Y_{igt}^1 - Y_{igt}^0 \mid g, t)$ is the average treatment effect for group g in period t[1] and $\bar{\tau} = E(\tau_{gt} | D_{gt} = 1)$ is the overall average treatment effect[2]. Rearranging, this gives

$$y_{igt} - \mu_g - \eta_t = \bar{\tau} D_{gt} + (\tau_{gt} - \bar{\tau}) D_{gt} + \varepsilon_{gt}.$$

Suppose you knew the time and group fixed-effects and were able to directly observe the left-hand side (later we will estimate the left-hand side). Regressing the adjusted y variable, on D_{gt} will be a consistent estimator for $\bar{\tau}$. To see this, note that $E[(\tau_{gt} - \bar{\tau}) D_{gt} \mid D_{gt}] = 0$. Hence, the treatment dummy is uncorrelated with the omitted variables and the average treatment effect is identified in the second-stage. Since we are not able to directly observe μ_g and η_t, we estimate them using the untreated/not-yet-treated observations in the first-stage. However, since we generate the regressor in the first-stage, standard errors need adjustment to account for the added uncertainty.

This approach can be extended to dynamic models by replacing the second stage of the procedure with a regression of residualized outcomes onto the leads and lags of treatment status, D_{gt}^k, $k \in \{-L, \ldots, K\}$. Under parallel trends, the second-stage coefficients on the lags identify the overall average effect of being treated for k periods (where the average is taken over all units treated for at least that many periods). The second-stage coefficients on the leads identify the average deviation from predicted counterfactual trends among units that are k periods away from treatment, which under parallel trends should be zero for any pre-treatment value of k. Hence, the coefficients on the leads represent a test of the validity of the parallel trends assumption.

Inference

The standard variance-covariance matrix from the second-stage regression will be incorrect since it fails to account for the fact that the dependent variable is generated from the first-stage. However,

[1] i.e., the average difference between treated and untreated potential outcomes Y_{igt}^1 and Y_{igt}^0, conditional on the observed treatment-adoption times

[2] i.e., the population-weighted average of the group-time specific ATTs, τ_{gt}

this estimator takes the form of a joint generalized method of moments (GMM) estimator whose asymptotic variance is well understood (Newey and McFadden, 1986).

The estimator takes the form of a two-stage GMM estimator with the following two moment conditions:

$$m(\theta) = (Y - X_{10}'\gamma)X_{10} \tag{2}$$
$$g(\gamma, \theta) = (Y - X_1'\gamma - X_2'\theta)X_2, \tag{3}$$

where X_1 is the matrix of unit and time fixed effects, X_{10} corresponds to the matrix of, with rows corresponding to observations for which $D_{gt} = 1$ replaced with zeros (as only observations with $D_{gt} = 0$ are used in the first stage) and X_2 is the matrix of treatment variable(s). The first equation corresponds with the first stage and the second equation corresponds with the second stage. From Theorem 6.1 of Newey and McFadden (1986), the asymptotic variance of the two-stage estimator is

$$V = G_\theta^{-1} E\left[(g + G_\gamma \psi)(g + G_\gamma \psi)'\right] G_\theta^{-1'}, \tag{4}$$

where from our moment conditions, we have:

$$G_\theta = -E\left(X_2 X_2'\right),$$
$$G_\gamma = -E\left(X_2 X_1'\right),$$
$$\psi = E(X_{10}X_{10}')^{-1}\varepsilon_{10}X_{10}.$$

This can be estimated using

$$\left(X_2'X_2\right)^{-1}\left(\sum_{g=1}^{G} W_g'W_g\right)\left(X_2'X_2\right)^{-1}, \tag{5}$$

where

$$W_g = X_{2g}'\hat{\varepsilon}_{2g} - \hat{\varepsilon}_{10g}'X_{1g}\left(X_{1g}'X_{1g}\right)^{-1}\left(X_{1g}'X_{2g}\right)$$

and matrices indexed by g correspond to the gth cluster.

The did2s Package

The did2s Command

The command `did2s` implements the two-stage difference-in-differences estimator following (Gardner, 2021). The general syntax is

```
did2s(data, yname, first_stage, second_stage, treatment, cluster_var,
  weights = NULL, bootstrap = FALSE, n_bootstraps = 250, verbose = TRUE)
```

The option `data` specifies the data set that contains the variables for the analysis. The option `yname` is the variable name of the outcome variable. The option `weights` specifies the variable name of estimation weights, but is not required. The options `first_stage` and `second_stage` are user-provided one-sided formula for the first- and the second-stage estimators respectively. As discussed above, the first stage should consist of the unit/group and time fixed effects as well as time-invariant covariates. The second stage should consist of, in the static case, the treatment indicator or, in the dynamic case, the relative time indicators. The formula are used in the `fixest::feols` function from **fixest** and therefore there are two features non-standard formula options worth mentioning (Bergé, 2018). First, fixed effects can be inserted after the covariates, e.g. `~ x1 | fe_1 + fe_2`, which will make estimation much faster. Second, the function `fixest::i` can be used for treatment indicators instead of `factor`. The advantage of this is that you can specify the reference values, e.g. for event-study indicators where researchers typically want to drop time $t = -1$, `~ i(rel_year, ref = c(-1))` would be the correct second-stage formula.

The option `treatment` is the variable name of a $0/1$ variable that denotes when treatment is active for a given unit, D_{gt} in the above notation. Observations with $D_{gt} = 0$ will be used to estimate the first stage, which removes the problem of treatment effects contaminating estimation of the unit and time fixed-effects. However, as an important note, if you suspect anticipation effects before treatment begins, the `treatment` variable should be shifted forward by x periods for observations in order to prevent the aforementioned contamination. For example, if you suspect that units could adjust 1 period ahead of treatment, then the `treatment` should begin one period ahead. These anticipation effects can be estimated, after adjusting the `treatment` variable, by using a reference year of say, $t = -2$ and looking at the estimate for relative year -1.

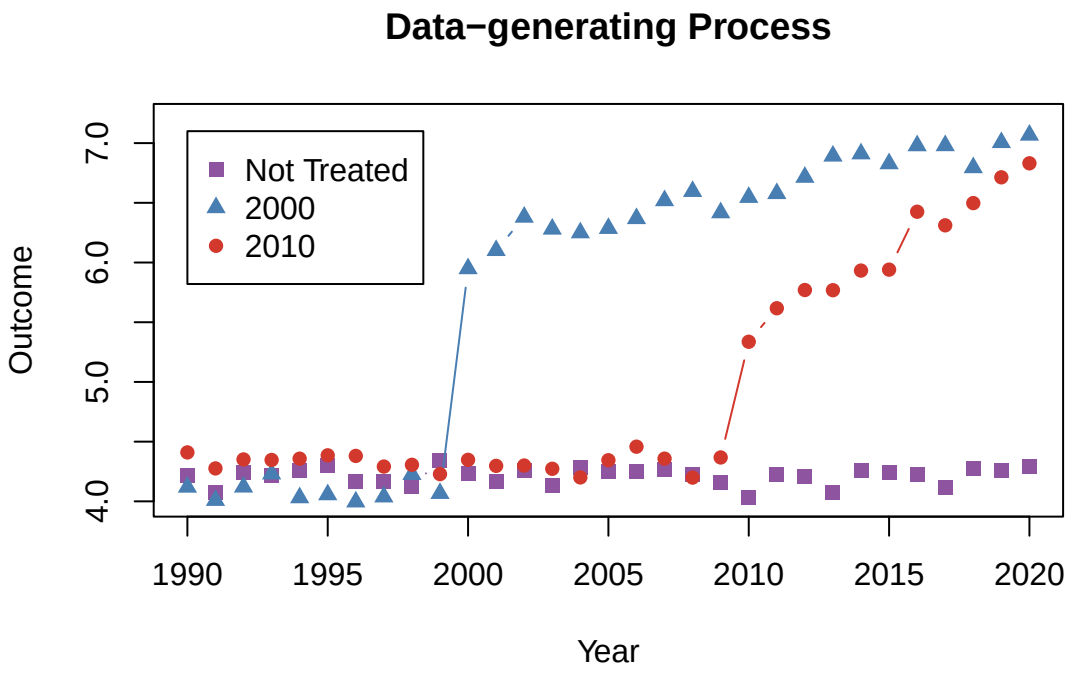

Figure 1: Example data with heterogeneous and dynamic treatment effects. Each line represents the average outcome in a given year for each group. In the absence of treatment, all three groups would exhibit parallel trends.

The option `cluster_var` is the variable name of the cluster variable. The error term for observations that are in the same cluster will be allowed to have an arbitrary correlation, but terms in different clusters will be assumed to have 0 interdependence in their error term structure. Since panel data at the very least likely has interdependence within a *unit*, `cluster_var` is a required option. If instead of asymptotic standard errors, bootstrapped standard errors are preffered, the option `bootstrap = TRUE` will perform a block boostrap `n_bootstraps` times with the block at the level of `cluster_var`. The last option is `verbose` which denotes whether or not the function should display info on the function call to the user.

Example usage

For basic usage, I will use the simulated dataset, `df_het`, that comes with the **did2s** package with the command

```
data(df_het, package = "did2s")
```

The data-generating process is displayed in Figure 1 Each line represents the mean outcome for each treatment group and the never-treated group. In the absence of treatment, each group were simulated to be on parallel trends. There is heterogeneity in treatment effects both within a treatment group over time and across treatment groups.

First, we will calculate a static difference-in-differences estimate using the `did2s` function.

```
static = did2s(data = df_het, yname = "dep_var", treatment = "treat",
      first_stage = ~ 0 | unit + year, second_stage = ~ i(treat, ref = FALSE),
      cluster_var = "unit", verbose = FALSE)

summary(static)

#> OLS estimation, Dep. Var.: dep_var
#> Observations: 31,000
#> Standard-errors: Custom
#>             Estimate Std. Error t value Pr(>|t|))
#> treat::TRUE    2.263   0.033879  66.795 < 2.2e-16 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#> RMSE: 1.4417   Adj. R2: 0.35288
```

Since the returning object is a `fixest` object, all the accompanying output commands from **fixest** are available to use. For example, we can create regression tables:

```
fixest::etable(static, fitstat = c("n"), tex = TRUE,
               title = "Estimate of Static TWFE Model",
               notes = "Standard errors clustered at unit level.
Estimated using Two-Stage Difference-in-Differences.
proposed by Gardner (2021).")
```

Table 1: Estimate of Static TWFE Model

Dependent Variable:	dep_var
Model:	(1)
Variables	
treat = TRUE	2.263***
	(0.0339)
Fit statistics	
Observations	31,000

Custom standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Notes: Standard errors clustered at unit level. Estimated using Two-Stage Difference-in-Differences. proposed by Gardner (2021).

However, since there are dynamic treatment effects in this example, it is much better to estimate the dynamic effects themselves using an event-study specification. We will then plot the results using `fixest::iplot` which plots coefficients corresponding to an `i()` variable. Note that `rel_year` is coded as `Inf` for never-treated units, so this has to be noted in the reference part of the formula.

```
es = did2s(data = df_het, yname = "dep_var", treatment = "treat",
      first_stage = ~ 0 | unit + year,
      second_stage = ~ i(rel_year, ref = c(-1, Inf)),
      cluster_var = "unit", verbose = FALSE)

fixest::iplot(es, main = "Event study: Staggered treatment",
              xlab = "Relative time to treatment", col = "steelblue",
              ref.line = -0.5)

# Add the (mean) true effects
true_effects = tapply((df_het$te + df_het$te_dynamic), df_het$rel_year, mean)
true_effects = head(true_effects, -1)
points(-20:20, true_effects, pch = 20, col = "grey60", trans.val = 0.4)

#> Warning in plot.xy(xy.coords(x, y), type = type, ...): "trans.val" is not a
#> graphical parameter

# Legend
legend(x=-20, y=3, col = c("steelblue", "grey60"),
       pch = c(20, 20),
       legend = c("Two-stage estimate", "True effect"))
```

The event study estimates are found in Figure 2 and match closely to the true average treatment effects. For comparison to traditional OLS estimation of the event-study specification, Figure 3 plots point estimates from both methods. As pointed out by Sun and Abraham (2020), treatment effect heterogeneity between groups causes the pre-trend estimates to be estimated incorrectly. In the below figure, the OLS estimates appear to show violations of pretrends.

```
twfe = feols(dep_var ~ i(rel_year, ref=c(-1, Inf)) | unit + year, data = df_het)

fixest::iplot(list(es, twfe), sep = 0.2, ref.line = -0.5,
      col = c("steelblue", "#82b446"), pt.pch = c(20, 18),
```

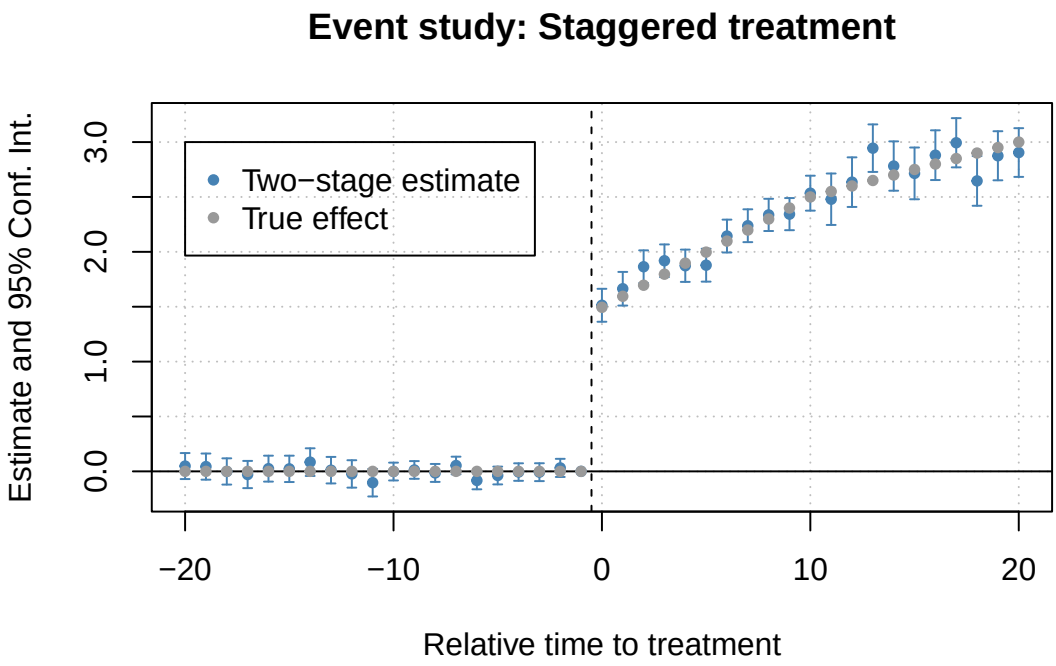

Figure 2: Event-study Estimate of TWFE Model. Standard Errors clustered at unit level. Estimated using the Two-Stage Difference-in-Differences proposed by Gardner (2021).

```
      xlab = "Relative time to treatment",
      main = "Event study: Staggered treatment (comparison)")

# True Effects
points(-20:20, true_effects, pch = 20, col = "grey60", trans.val = 0.4)

#> Warning in plot.xy(xy.coords(x, y), type = type, ...): "trans.val" is not a
#> graphical parameter

# Legend
legend(x=-20, y=3, col = c("steelblue", "#82b446", "grey60"), pch = c(20, 18, 20),
       legend = c("Two-stage estimate", "TWFE", "True Effect"))
```

The `event_study` and `plot_event_study` command

The command `event_study` presents a common syntax that estimates the event-study TWFE model for robust estimators recommended by the literature and returns all the estimates in a data.frame for easy plotting by the command `plot_event_study`. The general syntax is

```
event_study(data, yname, idname, tname, gname,
            xformla = NULL, horizon = NULL, weights = NULL)
```

The option `data` specifies the data set that contains the variables for the analysis. The four other required options are all names of variables: `yname` corresponds with the outcome of interest; `idname` is the variable corresponding to the unit, i; `tname` is the variable corresponding to the time, t; and `gname` is a variable indicating the period when treatment first starts. There are three additional optional parameters. First `xformla` is a (base R) formula corresponding to additional covariates. The option `horizon` is an integer vector of length two whose first element is the earliest pre-period effect and second element is the latest post-effect to include in the estimates. Last, `weights` is the variable name for estimation weights.

The available estimators are given in Table 2.

The result of `event_study` is a tibble in a `tidy` format (Robinson et al., 2021) that contains point estimates and standard errors for each relative time indicator for each individual estimator. To see the results of `event_study`, we return to the `df_het` dataset. The results of `event_study` is a dataframe with event-study term, the estimate, standard error, and a column containing a character for which estimator is used. This output dataframe will in turn be passed to `plot_event_study` for easy comparison.

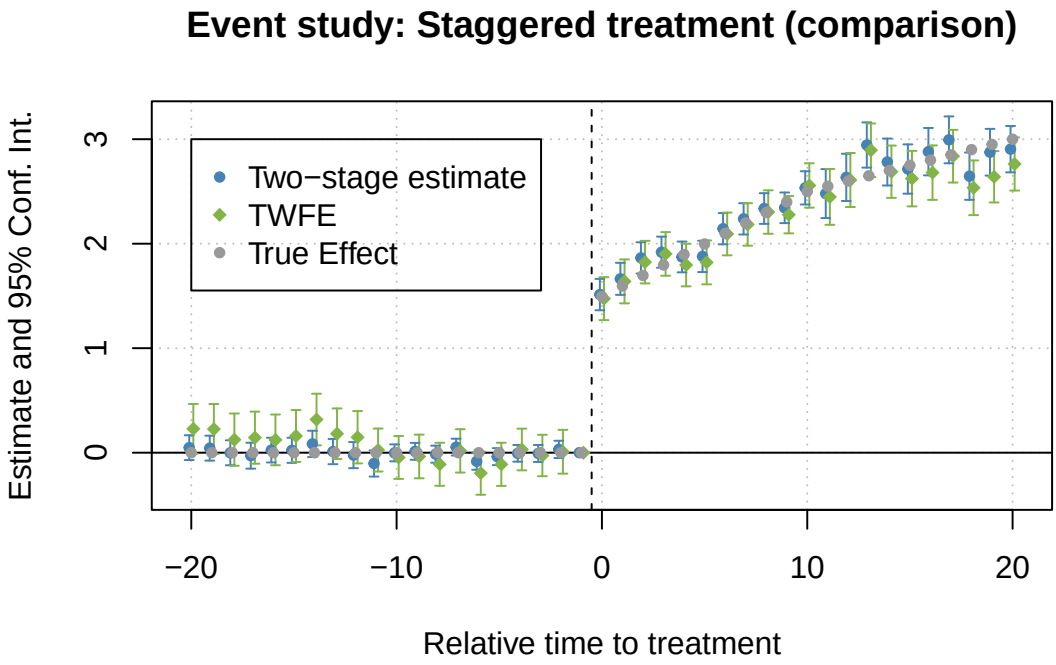

Figure 3: Event-study Estimate of TWFE Model. Standard Errors clustered at unit level. Estimated using the Two-Stage Difference-in-Differences proposed by Gardner (2021) and a Traditional TWFE model.

Table 2: Estimators used in codeevent_study

Estimator	Corresponding R Function
Gardner (2021)	did2s::did2s()
Borusyak, Jaravel, and Spiess (2021)	didimputation::did_imputation()
Callaway and Sant'Anna (2021)	did::att_gt()
Roth and Sant'Anna (2021)	staggered::staggered()
Sun and Abraham (2020)	fixest::sunab()

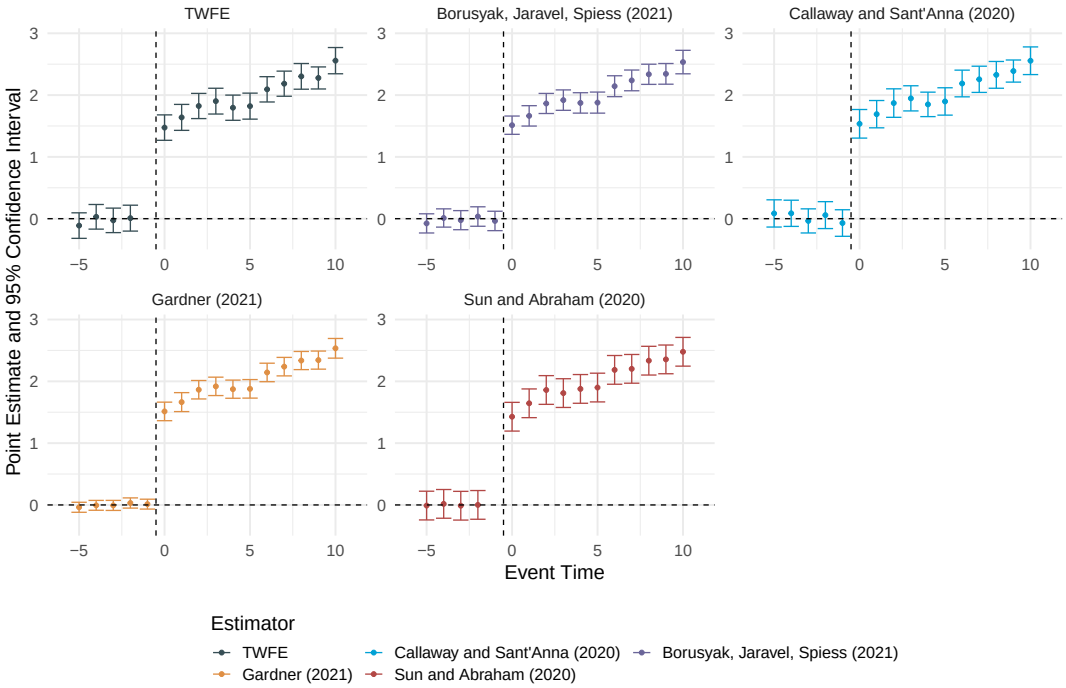

Figure 4: Event-study Estimators

```
data(df_het, package = "did2s")
out = event_study(
  data = df_het, yname = "dep_var", idname = "unit",
  tname = "year", gname = "g"
)
```

```
#> Estimating TWFE Model
```

```
#> Estimating using Gardner (2021)
```

```
#> Estimating using Callaway and Sant'Anna (2020)
```

```
#> Estimating using Sun and Abraham (2020)
```

```
#> Estimating using Borusyak, Jaravel, Spiess (2021)
```

```
#> Estimatng using Roth and Sant'Anna (2021)
```

```
head(out)
```

```
#> # A tibble: 6 x 4
#>    term estimate std.error estimator
#>   <dbl>    <dbl>     <dbl> <chr>
#> 1   -20    0.228     0.122 TWFE
#> 2   -19    0.226     0.123 TWFE
#> 3   -18    0.125     0.128 TWFE
#> 4   -17    0.144     0.127 TWFE
#> 5   -16    0.123     0.124 TWFE
#> 6   -15    0.160     0.127 TWFE
```

```
plot_event_study(out, horizon = c(-5,10))
```

Conclusion

This article introduced the package `did2s` which provides a fast, memory-efficient, and robust way to estimate two-way fixed effect models. The package also includes the `event_study` and `plot_event_study` functions to allow for a single syntax for the various estimators introduced in the literature. A companion package in Stata is also available with similar syntax for the `did2s` function.

Bibliography

L. Bergé. Efficient estimation of maximum likelihood models with multiple fixed-effects: the R package FENmlm. *CREA Discussion Papers*, (13), 2018. [p3]

K. Borusyak, X. Jaravel, and J. Spiess. Revisiting event study designs: Robust and efficient estimation. Technical report, 2021. [p1, 2]

B. Callaway and P. H. Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 2020. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2020.12.001. URL https://www.sciencedirect.com/science/article/pii/S0304407620303948. Citation Key: CALLAWAY2020. [p1, 2]

R. Frisch and F. V. Waugh. Partial time regressions as compared with individual trends. *Econometrica*, 1(4):387–401, 1933. ISSN 00129682, 14680262. URL http://www.jstor.org/stable/1907330. [p2]

J. Gardner. Two-Stage Difference-in-Differences. Technical report, 2021. [p1, 2, 3]

A. Goodman-Bacon. *Difference-in-Differences with Variation in Treatment Timing*. Number w25018. Sep 2018. doi: 10.3386/w25018. URL http://www.nber.org/papers/w25018.pdf. [p1, 2]

W. Newey and D. McFadden. Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier, 1 edition, 1986. URL https://EconPapers.repec.org/RePEc:eee:ecochp:4-36. [p3]

D. Robinson, A. Hayes, and S. Couch. *broom: Convert Statistical Objects into Tidy Tibbles*, 2021. URL https://CRAN.R-project.org/package=broom. R package version 0.7.9. [p6]

L. Sun and S. Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, Dec 2020. ISSN 0304-4076. doi: 10.1016/j.jeconom.2020.09.006. URL https://www.sciencedirect.com/science/article/pii/S030440762030378X. [p1, 2, 5]

Kyle Butts
University of Colorado Boulder

https://www.kylebutts.com/
ORCiD: 0000-0002-9048-8059
buttskyle96@gmail.com

John Gardner
University of Mississippi

https://jrgcmu.github.io/
jrgardne@olemiss.edu