didregress postestimation — Postestimation tools for didregress and xtdidregress	

Postestimation commands

The following postestimation commands are of special interest after didregress and xtdidregress:

Command	Description
estat trendplots	graphical diagnostics for parallel trends
estat ptrends	parallel-trends test
estat granger	Granger causality test
estat grangerplot	time-specific treatment effects
estat bdecomp	treatment-effect decomposition

The following standard postestimation commands are also available:

Command	Description
estat summarize	summary statistics for the estimation sample
estat vce	variance–covariance matrix of the estimators (VCE)
estimates	cataloging estimation results
etable	table of estimation results
forecast	dynamic forecasts and simulations
lincom	point estimates, standard errors, testing, and inference for linear combinations of parameters
nlcom	point estimates, standard errors, testing, and inference for nonlinear combinations of parameters
predict	linear predictions and residuals
predictnl	point estimates, standard errors, testing, and inference for generalized predictions
test	Wald tests of simple and composite linear hypotheses
testnl	Wald tests of nonlinear hypotheses

predict

Description for predict

`predict` creates a new variable containing predictions such as linear predictors and residuals.

Menu for predict

Statistics > Postestimation

Syntax for predict

predict [*type*] *newvar* [*if*] [*in*] [, *statistic*]

statistic	Description
Main	
xb	linear predictor; the default
residuals	residuals

Options for predict

    Main

xb, the default, calculates the linear predictor. It excludes the effect of the first group or of the panel identifier. All other effects, including the time fixed effects, are included in the linear predictor.

residuals calculates the overall residuals. It is the difference of the outcome and the linear predictor, including all group, panel, and time effects. In other words, it is not just the difference of the outcome and the linear predictor.

estat

Description for estat

estat trendplots produces two diagnostic plots for assessing the parallel-trends assumption that is required for consistent estimation of the ATET using didregress or xtdidregress. The first plot consists of two lines showing the mean of the outcome over time for the treatment and the control groups. The second plot augments the DID model to include interactions of time with an indicator of treatment and plots the predicted values of this augmented model for the treatment and control groups. Both plots include a vertical line one period before treatment.

estat ptrends performs a test of whether the linear trends in the outcome variable are parallel between control and treatment groups during the pretreatment period.

estat granger performs a test of whether treatment effects can be observed in anticipation of the treatment.

estat grangerplot produces a graph of time-specific treatment effects by plotting coefficients from leads and lags of the treatment indicator variable.

estat bdecomp performs a treatment-effect (Bacon) decomposition into all 2-by-2 treatment-effect components when there are multiple cohorts, each with a different treatment timing. Optionally, the components can be plotted against their weights. estat bdecomp requires that the data be strongly balanced. That is, units defined by the group variable must have the same number of observations, and observation times must be the same across all units.

Menu for estat

Statistics > Postestimation

Syntax for estat

Graphical diagnostics for parallel trends

estat trendplots [, *trend_options plot_options*]

Parallel-trends test

estat ptrends

Granger causality test

estat granger

Time-specific treatment effects

estat grangerplot [, *grangerplot_options*]

Treatment-effect decomposition

estat bdecomp [, *bdecomp_options*]

trend_options	Description
omeans	draw graph showing observed means
omeans(*plot_options*)	draw observed-means graph and affect its rendition
ltrends	draw graph showing linear trends
ltrends(*plot_options*)	draw linear-trends graph and affect its rendition
notitle	suppress overall title
noxline	suppress treatment-time reference line
nocommonlegend	display two individual legends
legendfrom(#)	specify which legend to use

plot_options	Description
Plot	
cline_options	affect rendition of the plotted trend lines; see [G-3] *cline_options*
line1opts(*cline_options*)	affect rendition of the line for controls
line2opts(*cline_options*)	affect rendition of the line for treated
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than by() documented in [G-3] *twoway_options*

grangerplot_options	Description
nleads(#)	number of leads
nlags(#)	number of lags
baseline(#)	baseline period
lagview	show lags instead of time units
verbose	display results of the underlying regression
post	post the results of the underlying regression in e()
noci	do not plot confidence intervals
noyline	suppress y-axis reference line
CI plot	
recastci(*plottype*)	plot confidence intervals using *plottype*
ciopts(*rcap_options*)	affect rendition of confidence intervals
level(#)	set confidence level; default is level(95)
Add plots	
addplot(*plot*)	add other plots to the graph
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than by() documented in [G-3] *twoway_options*

bdecomp_options	Description
<u>summary</u>only	show decomposition summary only
<u>full</u>decomponly	show full decomposition only
noheader	suppress the table header in the output
notable	suppress the table in the output
<u>grap</u>h	plot individual components against their weights
Plot	
copts(*marker_options*)	change look of markers for cohorts
tnopts(*marker_options*)	change look of markers for treated versus never treated
laopts(*marker_options*)	change look of markers for later versus always treated
leopts(*marker_options*)	change look of markers for later versus earlier treated
elopts(*marker_options*)	change look of markers for earlier versus later treated
atetlineopts(*cline_options*)	change look of the ATET reference line
noatetline	suppress ATET reference line
Y axis, X axis, Titles, Legend, Overall	
twoway_options	any options other than by() documented in [G-3] ***twoway_options***

Options for estat trendplots

omeans, omeans(*plot_options*), ltrends, and ltrends(*plot_options*) specify which graphs are to be included and how they should be individually rendered. The default is omeans ltrends, meaning that both graphs are included without any modifications.

  omeans specifies that the observed-means graph be included. Specifying omeans suppresses the linear-trends model graph unless ltrends or ltrends(*plot_options*) is also specified.

  omeans(*plot_options*) specifies that the observed-means graph be included and affects its rendition. Specifying omeans(*plot_options*) suppresses the linear-trends model graph unless ltrends or ltrends(*plot_options*) is also specified.

  ltrends specifies that the linear-trends model graph be included. Specifying ltrends suppresses the observed-means graph unless omeans or omeans(*plot_options*) is also specified.

  ltrends(*plot_options*) specifies that the linear-trends model graph be included and affects its rendition. Specifying ltrends(*plot_options*) suppresses the observed-means graph unless omeans or omeans(*plot_options*) is also specified.

notitle suppresses the overall title of the rendered graph.

noxline suppresses rendering of the treatment-time reference line.

nocommonlegend suppresses the display of one common legend and renders two individual legends.

legendfrom(#) specifies which legend to use; the default is legendfrom(1), which refers to the legend of the first plot (observed means). legendfrom(#) is not allowed with the nocommonlegend option.

> Plot

cline_options affect the rendition of the plotted trend lines, including their style, size, and color; see [G-3] ***cline_options***.

line1opts(*cline_options*) affect the rendition of the plotted trend lines for the group of controls, including their style, size, and color; see [G-3] ***cline_options***.

line2opts(*cline_options*) affect the rendition of the plotted trend lines for the group of treated, including their style, size, and color; see [G-3] ***cline_options***.

> Y axis, X axis, Titles, Legend, Overall

twoway_options are any of the options documented in [G-3] ***twoway_options***, excluding by(). These include options for titling the graph (see [G-3] ***title_options***) and for saving the graph to disk (see [G-3] ***saving_option***).

Options for estat grangerplot

nleads(*#*) specifies the number of leads to be included in the model and plotted. By default, all available leads are included. The number of leads must be greater than 0.

nlags(*#*) specifies the number of lags to be included in the model and plotted. By default, all available lags are included. The number of lags must be greater than or equal to 0.

baseline(*#*) specifies the baseline period for which the corresponding lead or lag is omitted. By default, the first lead is omitted, which corresponds to the time period prior to intervention.

lagview specifies to show the values of the x axis in terms of lags. If this option is not specified, time values are shown.

verbose specifies to display the output of the underlying regression model.

post posts the results of the underlying regression as estimation results in e().

noci removes the pointwise confidence intervals. The default is to plot the confidence intervals.

noyline suppresses rendering of the reference line.

> CI plot

recastci(*plottype*) specifies that confidence intervals be plotted using *plottype*. *plottype* may be rarea, rbar, rspike, rcap, rcapsym, rline, rconnected, or rscatter; see [G-2] **graph twoway**. When recastci() is specified, the plot-rendition options appropriate to the specified *plottype* may be used in lieu of *rcap_options*. For details on those options, follow the appropriate link from [G-2] **graph twoway**.

ciopts(*rcap_options*) affects the rendition of confidence intervals; see [G-3] ***rcap_options***.

level(*#*) specifies the confidence level, as a percentage, for confidence intervals. The default is level(95) or as set by set level; see [U] **20.8 Specifying the width of confidence intervals**.

> Add plots

addplot(*plot*) provides a way to add other plots to the generated graph; see [G-3] ***addplot_option***.

⌐ Y axis, X axis, Titles, Legend, Overall ⌐

twoway_options are any of the options documented in [G-3] *twoway_options*, excluding by(). These include options for titling the graph (see [G-3] *title_options*) and for saving the graph to disk (see [G-3] *saving_option*).

Options for estat bdecomp

summaryonly displays the decomposition summary only.

fulldecomponly displays the results of the full decomposition only.

noheader suppresses the table header in the output.

notable suppresses the results table in the output.

graph specifies to plot the treatment-effect components against their weights.

⌐ Plot ⌐

copts(*marker_options*) affect the rendition of the markers for the plotted cohort coefficients, including their style, size, and color; see [G-3] *marker_options*.

tnopts(*marker_options*) affect the rendition of the markers for the plotted coefficients for treated versus never treated, including their style, size, and color; see [G-3] *marker_options*.

laopts(*marker_options*) affect the rendition of the markers for the plotted coefficients for later versus always treated, including their style, size, and color; see [G-3] *marker_options*.

leopts(*marker_options*) affect the rendition of the markers for the plotted coefficients for later versus earlier treated, including their style, size, and color; see [G-3] *marker_options*.

elopts(*marker_options*) affect the rendition of the markers for the plotted coefficients for earlier versus later treated, including their style, size, and color; see [G-3] *marker_options*.

atetlineopts(*cline_options*) affect the rendition of the plotted ATET reference line, including its style, size, and color; see [G-3] *cline_options*.

noatetline suppresses rendering of the ATET reference line.

⌐ Y axis, X axis, Titles, Legend, Overall ⌐

twoway_options are any of the options documented in [G-3] *twoway_options*, excluding by(). These include options for titling the graph (see [G-3] *title_options*) and for saving the graph to disk (see [G-3] *saving_option*).

Remarks and examples

In the following, we discuss the treatment-effect decomposition approach implemented in `estat bdecomp`. To see examples of `estat trendplots`, `estat ptrends`, and `estat granger`, see examples 1, 4, 5, and 6 in [CAUSAL] **didregress**. An example of `estat grangerplot` is shown in example 7 in [CAUSAL] **didregress**.

A simple 2-by-2 DID design is characterized by two time periods (a period before and a period after a treatment or intervention occurs) and two groups (a treatment and a control group). Under this design, all treated units receive the treatment at the same time. A staggered DID design, on the other hand, involves multiple time points at which the treatment is administered. Goodman-Bacon (2021) has shown that, in this case, the ATET is a weighted average of all possible 2-by-2 DID estimators that compare treatment cohorts with each other. Decomposing an overall ATET into these 2-by-2 components provides a diagnostic tool to assess the validity of the two-way fixed-effects DID estimator implemented in `didregress` and `xtdidregress` when treatment timing varies.

Consider the case of a DID design with two treatment times so that the data consist of three groups: an earlier-treated group, a later-treated group, and a never-treated group. Under this scenario, there would be four 2-by-2 comparisons. Two of these come from comparing the group that was treated first with the never-treated group and comparing the group that was treated second with the never-treated group. The third compares the earlier-treated group with the later-treated group (before the later-treated group was treated). The fourth compares the later-treated group with the earlier-treated group (beginning at the point where the earlier-treated group received treatment). This fourth comparison is of particular concern if the 2-by-2 DID estimate substantially differs from the overall ATET estimate and if the corresponding weight is considerably large. In that case, this 2-by-2 estimate induces bias in the overall ATET estimate. The Bacon decomposition implemented in `estat bdecomp` offers insight into which 2-by-2 comparisons are driving the overall ATET estimate. With this decomposition, we can identify sources of variation that potentially bias the overall effect estimate. In cases of suspected bias, users may consider using `hdidregress` or `xthdidregress` to account for treatment-effect heterogeneity.

The decomposition requires that the data be strongly balanced with respect to the observed times per unit. Specifically, units defined by the group variable that is specified in the `group()` option of `didregress` or `xtdidregress` must have an equal number of observations, and observation times given by the variable specified in the `time()` option of `didregress` or `xtdidregress` must be the same across all units.

For further theoretical and empirical details, see Goodman-Bacon (2021). A summary of the Bacon treatment-effect decomposition is provided in Baker, Larcker, and Wang (2022). Applications of the Bacon decomposition can be found in Rabideau et al. 2021 and Strasseri, Oggenfuss, and Wolter (2022).

▷ Example 1: DID treatment-effect decomposition for designs with variation in treatment timing

We use a (fictional) dataset where the units are dog breeds for which registration numbers at the American Kennel Club are observed over the course of 10 years. We wish to find out whether the registration numbers change in response to a breed being featured in a widely shown movie. Here is an excerpt of the data:

```
. use https://www.stata-press.com/data/r19/akc
(Fictional dog breed and AKC registration data)
. list in 981/1000, sepby(breed) abbreviate(10) noobs
```

year	breed	movie	best	registered
2031	Old English Sheepdog	0	0	1324
2032	Old English Sheepdog	0	0	878
2033	Old English Sheepdog	0	0	699
2034	Old English Sheepdog	0	0	645
2035	Old English Sheepdog	0	0	1762
2036	Old English Sheepdog	0	0	1182
2037	Old English Sheepdog	1	0	1816
2038	Old English Sheepdog	1	0	4171
2039	Old English Sheepdog	1	0	3755
2040	Old English Sheepdog	1	0	4054
2031	Otterhound	0	0	692
2032	Otterhound	0	0	598
2033	Otterhound	0	0	670
2034	Otterhound	0	0	1642
2035	Otterhound	0	0	1311
2036	Otterhound	0	0	572
2037	Otterhound	0	0	1167
2038	Otterhound	0	0	620
2039	Otterhound	0	0	511
2040	Otterhound	0	0	1065

We list the data for two breeds, the Old English Sheepdog and the Otterhound. We can see that the Old English Sheepdog is featured in a `movie` that is released in the year 2037 and, thus, is a treated unit. The Otterhound breed has no such luck and belongs to the group that is never treated. Our outcome variable `registered` measures the number of registrations at the American Kennel Club. We use `xtdidregress` to fit a two-way fixed-effects DID model:

```
. xtset breed
Panel variable: breed (balanced)
. xtdidregress (registered) (movie), group(breed) time(year)
Treatment and time information

Time variable: year
Control:       movie = 0
Treatment:     movie = 1
```

	Control	Treatment
Group		
breed	119	22
Time		
Minimum	2031	2034
Maximum	2031	2037

```
Difference-in-differences regression                  Number of obs = 1,410
Data type: Longitudinal
                               (Std. err. adjusted for 141 clusters in breed)
```

	Coefficient	Robust std. err.	t	P>|t|	[95% conf. interval]
ATET					
movie (1 vs 0)	2129.655	78.16241	27.25	0.000	1975.124 2284.186

```
Note: ATET estimate adjusted for panel effects and time effects.
Note: Treatment occurs at different times.
```

We can see that our overall ATET estimate is around 2,130, which is to say that we expect the number of dog registrations to increase by 2,130 if a member of their breed is featured as a movie protagonist. We can now use `estat bdecomp` to decompose this number into its constituent 2-by-2 parts:

```
. estat bdecomp

DID treatment-effect decomposition

ATET = 2129.655                              Number of obs    =   1,410
                                             Number of groups =     141
                                             Number of cohorts =      4

ATET decomposition summary          ATET component             Weight

Treated vs never treated                 2166.021            0.970810
Treated earlier vs later                936.70729            0.013157
Treated later vs earlier                906.60588            0.016033


Full ATET decomposition             2x2 coefficient            Weight

Treated vs never treated
  2034 vs never treated                 1673.5688            0.157125
  2036 vs never treated                 1999.2515            0.140290
  2037 vs never treated                 2315.6702            0.673394
Treated earlier vs later
  2034 vs 2036                          580.59722            0.001132
  2034 vs 2037                          836.17778            0.008488
  2036 vs 2037                          1291.9333            0.003537
Treated later vs earlier
  2036 vs 2034                            542.825            0.001886
  2037 vs 2034                          831.70833            0.011318
  2037 vs 2036                          1448.7167            0.002829

Note: Number of cohorts includes never treated.
Note: The ATET reported by xtdidregress is a weighted average of the ATET
      components. If any component is substantially different from the ATET
      reported by xtdidregress and the weight is large, consider accounting
      for treatment-effect heterogeneity by using xthdidregress.
```
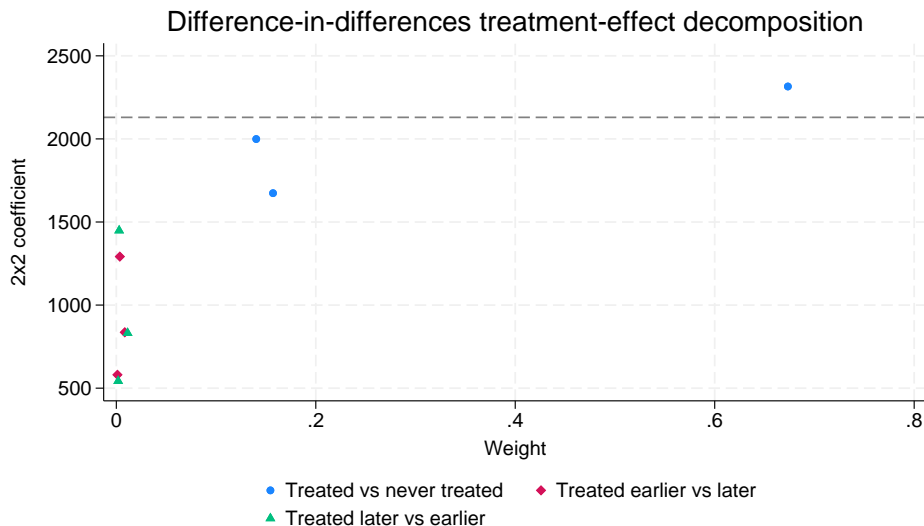
In the table header, we see that we have 1,410 observations from 141 groups (breeds) and that we have four cohorts. The four cohorts consist of groups of breeds that are treated at three different times (2034, 2036, and 2037) and a group that is never treated. The first table shows a decomposition summary where the estimated ATET is decomposed into components due to comparisons between treated and never treated, between earlier treated and later treated, and between later treated and earlier treated. The results in the second table decompose the summary results further into the individual 2-by-2 comparisons.

From the summary results, we see that the component due to comparisons between treated and never treated is close to the overall effect estimate and that its weight is large. Specifically, the weight is around 0.97, which indicates that 97% of the overall ATET estimate is due to comparisons between treated and never treated. The other two components are not close to the overall effect. This would be a potential concern, especially for the later- versus earlier-treated comparisons, if the corresponding weight was considerably large. However, the weights are small here. The later- versus earlier-treated and earlier-versus later-treated components together account for only 3% of the overall ATET.

In the full decomposition results, we see the three components from the summary broken down further. For example, if we look at the treated versus never-treated component, we see that around two thirds of that component is driven by the group that received treatment last. We also see that the 2-by-2 coefficients are increasing, which provides some indication that the treatment effect may be changing over time.

In this example, the number of treatment times is relatively small. However, if that number is large, it can be instructive to plot the 2-by-2 coefficients against their weights. We do this by using the `graph` option of `estat bdecomp`. Because we do not need to see the numerical results again, we also specify the `notable` and `noheader` options:

```
. estat bdecomp, graph notable noheader
```


Difference-in-differences treatment-effect decomposition

Legend:
- Treated vs never treated
- Treated earlier vs later
- Treated later vs earlier

The blue circles are the 2-by-2 coefficients from the treated versus never-treated component; the red diamonds are the 2-by-2 coefficients from the earlier- versus later-treated component; and the green triangles are the 2-by-2 coefficients from the later- versus earlier-treated component. The triangles and diamonds, while being substantially different from the overall ATET (indicated by the dashed reference line), are all very close to zero weight. This indicates that the contribution of these coefficients to the overall ATET is negligible.

◁

Stored results

estat ptrends stores the following results for the test of linear trends in r():

Scalars
r(N)	number of observations
r(F)	test statistic
r(df_r)	number of degrees of freedom of the residuals for the F distribution under H_0
r(p)	p-value
r(df_m)	number of degrees of freedom of the test for the F distribution under H_0

estat granger stores the following results for the test of treatment anticipation in r():

Scalars
r(N)	number of observations
r(F)	test statistic
r(df_r)	number of degrees of freedom of the residuals for the F distribution under H_0
r(p)	p-value
r(df_m)	number of degrees of freedom of the test for the F distribution under H_0

estat grangerplot, when used with option post, stores results from the underlying regression model in e() and r().

estat bdecomp stores the following results in r():

Scalars
r(N_groups)	number of panels
r(N_obs)	number of observations
r(N_cohorts)	number of cohorts
r(atet)	overall ATET
r(atet_between_b)	between component
r(atet_within_b)	within component
r(atet_between_w)	weight of between component
r(atet_within_w)	weight of within component

Macros
r(ttimes)	times of treatment

Matrices
r(atet_comp_b)	ATET components
r(atet_comp_w)	weights of ATET components
r(coefs)	individual 2-by-2 coefficients
r(wgts)	weights of individual 2-by-2 coefficients

Methods and formulas

The tests performed with estat ptrends and estat granger are based on augmented difference-in-differences (DID) models. With estat ptrends, we augment the DID model with terms that capture the differences in slopes between treated and controls. With estat granger, we augment the model by interacting the dummy variable that marks treated observations with dummy variables for time periods prior to the treatment to capture any potential anticipatory treatment effects.

Let's consider the case of panel data for individuals over time, in which individuals belong to a group s. Groups could be states, occupational categories, districts, etc. Let y_{its} be the outcome of individual i, who belongs to group s, at time t, where $i = 1, \ldots, N$, $t = 1, \ldots, T$, and $s = 1, \ldots, S$.

We can write the DID model for such setups as follows:

$$y_{ist} = \gamma_i + \gamma_t + \mathbf{x}_{ist}\boldsymbol{\beta} + D_{st}\delta + \epsilon_{ist}$$

Here γ_i are individual fixed effects, γ_t are time fixed effects, $\mathbf{x}_{ist}$ are covariates, D_{st} is a variable that is 1 if an individual belongs to a group s that is treated at time t and is 0 otherwise, and ϵ_{ist} is an error term. The coefficient δ represents the average treatment effect on the treated (ATET).

To simplify the exposition below, we rewrite the model as follows:

$$y_{ist} = \gamma_i + \gamma_t + \mathbf{x}_{ist}\boldsymbol{\beta} + D_{st}\delta + \epsilon_{ist}$$
$$y_{ist} = \mathrm{DID}_{ist} + \epsilon_{ist} \tag{1}$$

The linear-trends model that is used for the parallel-trends test with `estat ptrends` augments the above model with two more terms. Let $d_{t,0} = 1(d_t = 0)$ be a variable indicating pretreatment time periods, and let $d_{t,1} = 1(d_t = 1)$ be a variable indicating posttreatment time periods. Also, let w_i be a variable that is 1 if the individual belongs to a treated group and is 0 otherwise. The augmentation terms then consist of two 3-way interactions between $d_{t,0}$, w_i, and t, and $d_{t,1}$, w_i, and t:

$$y_{ist} = \mathrm{DID}_{ist} + w_i d_{t,0} t \zeta_1 + w_i d_{t,1} t \zeta_2 + \epsilon_{ist} \tag{2}$$

Under this specification, the coefficient ζ_1 captures the differences in slopes between treatment group and control group in pretreatment periods, while ζ_2 captures the differences in slopes in posttreatment periods. If ζ_1 is 0, the linear trends in the outcome are parallel during pretreatment periods. The same is true for ζ_2 with respect to the posttreatment period; however, posttreatment differences in trends are not relevant for assessing the parallel-trends assumption. `estat ptrends` uses a Wald test of ζ_1 against 0 to assess whether the linear trends are parallel prior to treatment. Thus, the null hypothesis of this test is that the linear trends are parallel.

`estat granger` performs a Granger-type causality test to assess whether treatment effects are observed prior to the treatment. To illustrate this, suppose the treatment took place at time $t = j$. We could express D_{st} as $D_{st} = 1(t \geq j)w_i$. The Granger-type test augments the model with counterfactual treatment-time indicators. For example, if the treatment occurred at time $j - 1$, then we could construct a new treatment as $1(t_{it} \geq j - 1)w_i$, and if we have sufficient time points, we could construct another counterfactual treatment as $1(t_{it} \geq j - 2)w_i$, and so on. These terms are referred to as leads in the DID literature. The model used by `estat granger` uses the model in (1) and augments it with all leads leaving out one for identification purposes. Let J index the time at which the treatment occurs.

$$y_{ist} = \mathrm{DID}_{ist} + \sum_{j=2}^{J-1} 1(t_{it} \geq j)w_i \lambda_j + \nu_{ist} \tag{3}$$

The test result is then obtained by performing a joint Wald test on the coefficients λ_j. Thus, the null hypothesis for this test is that the coefficients in λ_j are jointly 0, which is to say there are no anticipatory effects.

estat grangerplot fits a generalization of the DID model in (1) and plots the estimated coefficients from this model, including their 95% confidence intervals. The model is similar to (3), but this model parameterizes the leads differently and includes lags in addition to leads. Let I_s be the time of treatment, $m < 0$ be the number of time periods prior to I_s, $q \geq 0$ be the number of periods after I_s, and b be the baseline period. The model is

$$y_{ist} = \gamma_i + \gamma_t + \mathbf{x}_{ist}\boldsymbol{\beta} + \sum_{k=m,k\neq b}^{q} B_{st}^k w_i \lambda_k + \epsilon_{ist}$$

where

$$B_{st}^k = \begin{cases} 1(t_{it} \leq I_s + k), \text{if } k = m \\ 1(t_{it} = I_s + k), \text{if } m < k < q \\ 1(t_{it} \geq I_s + k), \text{if } k = q \end{cases}$$

This yields a model with $|m|$ leads and q lags. By default, estat grangerplot uses all available leads and lags. If, without loss of generality, we set the base to the period prior to treatment, $b = -1$ (the default), then with $t = 1, \ldots, T$ and $I_s = J$, a maximum of $n_{\text{leads}} = J - 2$ leads and $n_{\text{lags}} = T - J + 1$ lags is available. Notice that, if all available leads and lags are used, B_{st}^k reduces to $B_{st}^k = 1(t_{it} = I_s + k)$ because there are no periods before or after $I_s + k$. With fewer than available leads and lags, that is, $|m| < n_{\text{leads}}$ or $q < n_{\text{lags}}$, notice that the corresponding indicator variables capture the periods beyond the endpoints that correspond to m and q.

At a minimum, the model has to include a single lead. In that case, we have that $m = -1$ and $q = 0$. After omitting the base, we have that $k = 0$ and B_{st}^k reduces to $B_{st}^k = 1(t_{it} \geq I_s)$. Notice that $B_{st}^k w_i$ now yields our original treatment indicator D_{st}. In other words, the model with a single lead and no lags yields the special case of the DID model in (1). Notice also that the model in (3) is a special case, too. It is equivalent to the model fit by estat grangerplot with all available leads and no lags. However, (3) uses a different parameterization because the indicator variables are constructed differently. estat grangerplot plots the coefficients λ_k against the corresponding time periods.

The estat trendplots command produces two plots. The first plot is simply plotting the observed means for each treatment group at each point in time. The second plot is based on the model in (2), which is the model used for the parallel-trends test, but this model centers the continuous time variable around its minimum value:

$$y_{ist} = \text{DID}_{ist} + w_i d_{t,0}\{t - \min(t)\}\zeta_1 + w_i d_{t,1}\{t - \min(t)\}\zeta_2 + \mu_{ist}$$

Centering around the minimum time value provides a common reference point at the first observed time point such that deviations from parallelism are easily detectable. The graph then shows the predicted values from this model, evaluated at all observed time points for each of the treatment groups and at the means of the covariates.

While the formulas above are shown for the case of panel data, these methods work the same way for data that consist of repeated cross-sections.

The `estat bdecomp` command may be used in cases where units get treatment at different times. The command performs a decomposition of the estimated DID treatment effect into components that come from all possible 2-by-2 cohort comparisons. Cohorts are defined by the times of treatment administration and also include a group of never treated and potentially a group of always treated. Cohorts are also referred to as timing groups. `estat bdecomp` results are based on the derivation in Goodman-Bacon (2021).

For a DID model without covariates, we have the two-way fixed-effects model

$$y_{it} = \gamma_i + \gamma_t + D_{it}\delta^{\text{DD}} + \epsilon_{it}$$

where δ^{DD} is the overall treatment effect from the DID model. Let t_i denote the time when unit i receives treatment. If unit i was never treated, we set $t_i = U > T$ and say that the unit belongs to cohort U (where T is the latest time period for which we have data). If unit i was treated in period $k \in \{1, 2, \ldots T\}$, that is, $t_i = k$, we say that the unit belongs to cohort k. Let K denote the set of all cohorts (including U).

Goodman-Bacon (2021) show that we can decompose the estimated $\hat{\delta}^{\text{DD}}$ into the weighted sum of the DID coefficients that comes from all possible 2-by-2 cohort pairs,

$$\hat{\delta}^{\text{DD}} = \sum_{k \neq U} w_{kU}\hat{\delta}^{2\times2}_{kU} + \sum_{k \neq U}\sum_{l>k} \left[w^k_{kl}\hat{\delta}^{2\times2,k}_{kl} + w^l_{kl}\hat{\delta}^{2\times2,l}_{kl}\right]$$

where $\hat{\delta}^{2\times2}_{kU}$ is the coefficient from the 2-by-2 DID model that uses cohorts k and U as treatment and control, respectively, and $\hat{\delta}^{2\times2,k}_{kl}$ is the coefficient from the model that uses cohorts k and l as treatment and control, respectively. The weights w are given by

$$w_{kU} = \frac{(s_k + s_U)^2 s_{kU}(1 - s_{kU})\overline{D}_k(1 - \overline{D}_k)}{\hat{V}^D}$$

$$w^k_{kl} = \frac{\left[(s_k + s_l)(1 - \overline{D}_l)\right]^2 s_{kl}(1 - s_{kl})\frac{\overline{D}_k - \overline{D}_l}{1 - \overline{D}_l}\frac{1 - \overline{D}_k}{1 - \overline{D}_l}}{\hat{V}^D}$$

$$w^l_{kl} = \frac{\left[(s_k + s_l)\overline{D}_k\right]^2 s_{kl}(1 - s_{kl})\frac{\overline{D}_l}{\overline{D}_k}\frac{\overline{D}_k - \overline{D}_l}{\overline{D}_k}}{\hat{V}^D}$$

where $s_j \equiv \sum_i 1(t_i = j)/N$ is the share of cohort $j \in K$, $s_{ab} \equiv s_a/(s_a + s_b)$ is the relative size of cohort a with respect to cohort b, $\overline{D}_j \equiv \sum_t 1(t \geq j)/T$ is the proportion of time spent in treatment by cohort $j \in K$, and $\hat{V}^D$ is the sample variance of the treatment indicator demeaned with respect to both group and time fixed effects. The weights w sum to unity: $\sum_{k \neq U} w_{kU} + \sum_{k \neq U}\sum_{l>k}\left[w^k_{kl} + w^l_{kl}\right] = 1$.

For a model that includes time-varying covariates, $\mathbf{x}_{it}$, a similar decomposition result holds. Consider the model

$$y_{it} = \gamma_i + \gamma_t + \mathbf{x}_{it}\boldsymbol{\beta} + D_{it}\delta^{\text{DD}|\mathbf{x}} + \epsilon_{it}$$

where $\delta^{\text{DD}|\mathbf{x}}$ denotes the covariate-adjusted ATET. In this case, we can decompose $\hat{\delta}^{\text{DD}|\mathbf{x}}$ into a within and a between component,

$$\hat{\delta}^{\text{DD}|\mathbf{x}} = \Omega\hat{\delta}^p_w + (1 - \Omega)\hat{\delta}^d_b$$

To estimate the within and between components, we first fit a linear model that partials the covariates out of D_{it}; that is,

$$\widetilde{D}_{it} = \tilde{\mathbf{x}}_{it}\hat{\Gamma} + \tilde{d}_{it} = \tilde{p}_{it} + \tilde{d}_{it} \tag{4}$$

where $\widetilde{D}_{it}$ and $\tilde{\mathbf{x}}_{it}$ are respectively the treatment indicator and the covariates demeaned with respect to both group and time fixed effects. The resulting residuals can then be separated into a within component, $\tilde{d}_{i(k)t}$, and a between component, $\tilde{d}_{kt}$, so that $\tilde{d}_{it} = \tilde{d}_{i(k)t} + \tilde{d}_{kt}$.

The within component of $\tilde{d}_{it}$ is defined as $\tilde{d}_{i(k)t} = (d_{it} - \bar{d}_i) - (\bar{d}_{kt} - \bar{d}_k)$, where $d_{it} = D_{it} - \hat{\Gamma}\mathbf{x}_{it}$ is the treatment status adjusted by covariates only; and $\bar{d}_i$, $\bar{d}_k$, and $\bar{d}_{kt}$ are respectively the averages of d_{it} by individual, cohort, and cohort by year. This component captures deviations of d_{it} from cohort averages and cohort-by-year averages.

The between component of $\tilde{d}_{it}$ is defined as $\tilde{d}_{kt} = (\bar{d}_{kt} - \bar{d}_k) - (\bar{d}_t - \bar{\bar{d}})$, where $\bar{d}_t$ and $\bar{\bar{d}}$ are the averages of d_{it} by year and in the whole sample, respectively.

The within and between components of the DID treatment effect, $\hat{\delta}_w^p$ and $\hat{\delta}_b^d$, can be obtained by separately regressing the outcome variable y_{it} on $\tilde{d}_{i(k)t}$ and $\tilde{d}_{kt}$, respectively. Thus, $y_{it} = a_w + \tilde{d}_{i(k)t}\hat{\delta}_w^p + \nu_{it}$ and $y_{it} = a_b + \tilde{d}_{kt}\hat{\delta}_b^d + \eta_{it}$. The weight Ω is given by the ratio of the variances $\Omega = \hat{V}_w^d/\hat{V}^d$, where $\hat{V}_w^d$ and $\hat{V}^d$ are the sample variances of $\tilde{d}_{i(k)t}$ and $\tilde{d}_{it}$, respectively.

The between component $\hat{\delta}_b^d$ can be decomposed further into 2-by-2 components

$$\hat{\delta}_b^d = \sum_k \sum_{l>k} w_{kl}^{b|\mathbf{x}} \hat{\delta}_{kl}^{2\times 2|d}$$

where

$$\hat{\delta}_{kl}^{2\times 2|d} = \frac{\hat{V}_{kl}^D \hat{\delta}_{kl}^{2\times 2} - \hat{V}_{b,kl}^p \hat{\delta}_{b,kl}^p}{\hat{V}_{b,kl}^d}$$

and

$$w_{kl}^{b|\mathbf{x}} = (s_k + s_l)^2 \frac{\hat{V}_{b,kl}^d}{\hat{V}_b^d}$$

In the formulas above, $\hat{V}_b^d$ and $\hat{V}_{b,kl}^d$ are the sample variances of $\tilde{d}_{kt}$ in the whole sample and in the subsample consisting of cohorts k and l, respectively. $\hat{V}_{kl}^D$ and $\hat{V}_{b,kl}^p$ are the sample variances of $\widetilde{D}_{it}$ and $\tilde{p}_{it}$, respectively, in the subsample consisting of cohorts k and l. Finally, $\hat{\delta}_{kl}^{2\times 2}$ and $\hat{\delta}_{b,kl}^p$ are the coefficients from a regression of y_{it} on $\widetilde{D}_{it}$ and from a regression of y_{it} on $\tilde{p}_{it}$, respectively, in the subsample consisting of cohorts k and l.

Thus, the full decomposition for a DID specification with controls is

$$\hat{\delta}^{\text{DD}|\mathbf{x}} = \Omega\hat{\delta}_w^p + (1 - \Omega) \sum_k \sum_{l>k} w_{kl}^{b|\mathbf{x}} \hat{\delta}_{kl}^{2\times 2|d}$$

References

Baker, A. C., D. F. Larcker, and C. C. Y. Wang. 2022. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics* 144: 370–395. https://doi.org/10.1016/j.jfineco.2022.01.004.

Goodman-Bacon, A. 2021. Difference-in-differences with variation in treatment timing. *Journal of Econometrics* 225: 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014.

Rabideau, B., M. D. Eisenberg, R. Reid, and N. Sood. 2021. Effects of employer-offered high-deductible plans on low-value spending in the privately insured population. *Journal of Health Economics* 76: 102424. https://doi.org/10.1016/j.jhealeco.2021.102424.

Strasseri, M., C. Oggenfuss, and S. C. Wolter. 2022. Much ado about nothing? School curriculum reforms and students' educational trajectories. CESifo Working Paper 9912, CESifo Group, Munich, Germany. https://www.cesifo.org/DocDL/cesifo1_wp9912.pdf.

Also see

[CAUSAL] **didregress** — Difference-in-differences estimation

[CAUSAL] **DID intro** — Introduction to difference-in-differences estimation

[U] **20 Estimation and postestimation commands**