Title

> **xthdidregress** — Heterogeneous difference in differences for panel data

Description

xthdidregress estimates average treatment effects on the treated (ATETs) that may vary over time and over treatment cohorts. Treatment cohorts are groups subject to treatment at different points in time. xthdidregress provides four estimators: extended two-way fixed effects (TWFE), regression adjustment (RA), inverse-probability weighting (IPW), and augmented inverse-probability weighting (AIPW). See [CAUSAL] **teffects intro** for a discussion of RA, IPW, and AIPW estimators.

xthdidregress is for panel data. For repeated cross-sectional data, see [CAUSAL] **hdidregress**.

Quick start

Estimate ATETs of treatment treat on outcome y with group grpvar; use the RA estimator and model y with covariate x on xtset data

    xthdidregress ra (y x) (treat), group(grpvar)

Same as above, but use the TWFE estimator

    xthdidregress twfe (y x) (treat), group(grpvar)

Use the IPW estimator and model treat using covariate z

    xthdidregress ipw (y) (treat z), group(grpvar)

Use the AIPW estimator, model y using covariate x, and model treat using covariate z

    xthdidregress aipw (y x) (treat z), group(grpvar)

Same as above, but use the not-yet-treated group as the control group

    xthdidregress aipw (y x) (treat z), group(grpvar) ///
            controlgroup(notyet)

Same as above, but cluster at the county level

    xthdidregress aipw (y x) (treat z), group(grpvar) ///
            controlgroup(notyet) vce(cluster county)

Menu

Statistics > Causal inference/treatment effects > Continuous outcomes > Heterogeneous DID > Panel-data heterogeneous DID (TWFE)

Statistics > Causal inference/treatment effects > Continuous outcomes > Heterogeneous DID > Panel-data heterogeneous DID (RA)

Statistics > Causal inference/treatment effects > Continuous outcomes > Heterogeneous DID > Panel-data heterogeneous DID (IPW)

Statistics > Causal inference/treatment effects > Continuous outcomes > Heterogeneous DID > Panel-data heterogeneous DID (AIPW)

Syntax

Two-way fixed effects

   xthdidregress twfe (*ovar* [*omvarlist*]) (*tvar*) [*if*] [*in*] [*weight*],

     group(*groupvar*) [*options*]

Regression adjustment

   xthdidregress ra (*ovar* [*omvarlist*]) (*tvar*) [*if*] [*in*] [*weight*],

     group(*groupvar*) [*options*]

Inverse-probability weighting

   xthdidregress ipw (*ovar*) (*tvar* [*tmvarlist*]) [*if*] [*in*] [*weight*],

     group(*groupvar*) [*options*]

Augmented inverse-probability weighting

   xthdidregress aipw (*ovar* [*omvarlist*]) (*tvar* [*tmvarlist*]) [*if*] [*in*] [*weight*],

     group(*groupvar*) [*options*]

ovar is a continuous outcome of interest.

omvarlist specifies the covariates in the outcome model and may contain factor variables; see
   [U] **11.4.3 Factor variables**.

tvar must be a binary variable indicating observations subject to treatment.

tmvarlist specifies the covariates in the treatment model and may contain factor variables; see
   [U] **11.4.3 Factor variables**.

groupvar is a categorical variable that indicates the group level at which the treatment occurs.

options	Description
Model	
* **group**(*groupvar*)	specify group variable
controlgroup(*cgtype*)	specify the type of control group; default is never
cohortvar(*cvar* [, **replace**])	specify the variable name for the generated cohort
‡ **hettype**(*hetspec*)	type of heterogeneity; default is timecohort
SE/Robust	
vce(*vcetype*)	*vcetype* may be <u>cluster</u> *clustvar*, <u>robust</u>, <u>boot</u>strap, or <u>jack</u>knife
Reporting	
<u>l</u>evel(*#*)	set confidence level; default is level(95)
[**no**]**log**	suppress iteration log
nodots	suppress replication dots
display_options	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
coeflegend	display legend instead of statistics

* **group**(*groupvar*) is required.

‡ This option may be specified only when **twfe** is specified.

cgtype	Description
never	use the never-treated group as the control group; the default
notyet	use the not-yet-treated group as the control group

hetspec	Description
timecohort	heterogeneous treatment effects over time and cohort; the default
time	heterogeneous treatment effects over time
cohort	heterogeneous treatment effects over cohort

A panel variable and a time variable must be specified using **xtset**; see [XT] **xtset**.

by, **collect**, and **statsby** are allowed; see [U] **11.1.10 Prefix commands**.

fweights, **aweight**s, and **pweight**s are allowed; see [U] **11.1.6 weight**. Weights must be constant within panel.

coeflegend does not appear in the dialog box.

See [U] **20 Estimation and postestimation commands** for more capabilities of estimation commands.

Options

> Model

group(*groupvar*) specifies a group variable that indicates the group level at which the treatment
occurs. *groupvar* may be, for example, states, counties, or hospitals. group() also defines the level
clustering for the default cluster–robust standard errors. group() is required. You may specify
only one group variable.

controlgroup(*cgtype*) specifies the type of control group. A control group can be either a never-
treated group or a not-yet-treated group. A never-treated group refers to the units that are untreated
from the first to the last period. A not-yet-treated group refers to the units that are untreated up
to a specific period. *cgtype* can be one of never, referring to the never-treated group, or notyet,
referring to the not-yet-treated group. By default, *cgtype* is never.

cohortvar(*cvar* [, replace]) specifies the variable name *cvar* for the generated cohort variable.
The cohort variable is a categorical variable indicating the period when the unit is first treated.
By default, _did_cohort is used as the name of the cohort variable. If _did_cohort already
exists in the dataset, it is replaced if option cohortvar() is not specified.

> If suboption replace is specified, *cvar* is replaced.

hettype(*hetspec*) specifies time or cohort heterogeneity for the twfe estimator. By default, treatment
is interacted with time and cohort. You may choose to keep one of time or cohort interactions
using *hetspec*.

> *hetspec* may be one of timecohort for heterogeneous treatment effects over both time and cohort,
> time for heterogeneous treatment effects over time only, or cohort for heterogeneous treatment
> effects over cohort only. By default, *hetspec* is timecohort.

> SE/Robust

vce(*vcetype*) specifies the type of standard error reported, which includes types that allow for
intragroup correlation (cluster *clustvar*), that are robust to intragroup correlation among group
variable (robust), and that use bootstrap or jackknife sampling done at the panel level (bootstrap,
jackknife); see [R] *vce_option*.

> vce(cluster *clustvar*), the default, uses the variable specified in group(*groupvar*).

> Specifying vce(robust) is equivalent to specifying vce(cluster *clustvar*), where *clustvar* is
> the variable specified in the group(*groupvar*) option.

> Reporting

level(*#*); see [R] **Estimation options**.

log and nolog specify whether to display the iteration log. The iteration log is displayed by default
unless you used set iterlog off to suppress it; see set iterlog in [R] *set iter*.

nodots suppresses display of the replication dots.

display_options: noci, nopvalues, cformat(%*fmt*), pformat(%*fmt*), sformat(%*fmt*), and nol-
stretch; see [R] **Estimation options**.

The following option is available with xthdidregress but is not shown in the dialog box:

coeflegend; see [R] **Estimation options**.

Remarks and examples

It is common to study the effects of a treatment, for example, a policy or intervention, on a group. `xthdidregress` is for data where the treated groups are subject to the treatment at different points in time and they remain exposed to the treatment. For example, a health policy such as an increase in the age to purchase cigarettes is implemented in a given region, and over time, other regions decide to imitate the initiative. Another example is change in work policies across industries. Perhaps airlines implement a minimum number of hours between shifts for safety reasons. The policy is subsequently adopted by other similar industries. Some similar industries may never adopt the policy, remaining untreated, or it might be that all similar industries eventually adopt the policy.

`xthdidregress` estimates ATET parameters that change over time and treatment cohorts (groups treated at different points in time). Each one of these ATETs has the same interpretation that the parameters of a two-time two-group difference-in-differences (DID) parameter would have. Because there are multiple DID parameters, we refer to them as heterogeneous treatment effects or as heterogeneous DID. This is in contrast to estimating only one ATET, which assumes there is no variation across time or cohort. If you assume no variation across time or cohort, you may use `xtdidregress`; see [CAUSAL] **didregress**.

`xthdidregress` provides four estimators: TWFE, outlined in Wooldridge (2021); RA, IPW, and AIPW, outlined in Callaway and Sant'Anna (2021). Each one of these estimators fits a model for the outcome of interest, a model for the treatment, or a model for both. For example, RA and TWFE model the outcome; IPW models the treatment; and AIPW models both. If the model for the outcome is correctly specified, RA and TWFE are best, with TWFE being more efficient. If the treatment model is correctly specified, IPW should be best. AIPW models both treatment and outcome. If at least one of the models is correctly specified, it provides consistent estimates. Thus, it allows us to misspecify one of the models and still get consistent estimates, a property called double robustness. See [CAUSAL] **teffects intro** for a discussion of RA, IPW, and AIPW estimators.

`xthdidregress` is for panel data. For repeated cross-sectional data, see [CAUSAL] **hdidregress**. Below, we illustrate how to use `xthdidregress`. For more information about the methods used below, see [CAUSAL] **DID intro**. For general discussions about the methods, see Roth et al. (2022) and de Chaisemartin and D'Haultfœuille (forthcoming) and the references therein.

▷ Example 1: Fitting a heterogeneous DID model

We are interested in how the number of registrations of a dog breed with the American Kennel Club (AKC), `registered`, is affected by dogs being the protagonists in a movie, `movie`. We conjecture that the number of registrations increases if the dog breed appears as the protagonist in a movie. We also conjecture that registrations increase if the dog has won the Best in Show award from the Westminster Kennel Club, `best`, in the 10 years before 2034. We use simulated data, but there is some evidence of the effect of movies on dog breed registrations. See, for example, Ghirlanda, Acerbi, and Herzog (2014).

There are 141 dog breeds in our sample, which extends between the years 2031 and 2040. At the beginning of the sample, none of the breeds are featured in a movie. This changes in 2034, when four breeds are featured in movies. The next year in which we see an increase of breeds featured in movies is 2036, when 7 more breeds are featured. In 2037, there is a substantial increase, with 22 more breeds featured. There is no increase in breeds in movies thereafter. The table below illustrates this.

```
. use https://www.stata-press.com/data/r18/akc
(Fictional dog breed and AKC registration data)

. tabulate year movie
```

	Was a movie protagonist		
Year	0	1	Total
2031	141	0	141
2032	141	0	141
2033	141	0	141
2034	137	4	141
2035	137	4	141
2036	134	7	141
2037	119	22	141
2038	119	22	141
2039	119	22	141
2040	119	22	141
Total	1,307	103	1,410

We have a panel dataset. To obtain the effect of `movie` on `registered` controlling for `best`, we first `xtset` our data:

```
. xtset breed year
Panel variable: breed (strongly balanced)
 Time variable: year, 2031 to 2040
         Delta: 1 unit
```

We chose regression adjustment, `ra`, to fit our model (see [CAUSAL] **teffects ra** for details on this model). In the first set of parentheses, we define the outcome, `registered`, and any covariates that affect the outcome directly. In the second set of parentheses, we define the observation-level treatment variable, `movie`. Other models allow us to define covariates that affect the treatment, but this is not the case of the `ra` model. After the comma, we need to define the group variable in `group()`; this is a required option. The group variable defines at which level the treatment occurs and also identifies the clustering variable, which in this case is `breed`.

```
. xthdidregress ra (registered best) (movie), group(breed)
note: variable _did_cohort, containing cohort indicators formed by treatment
      variable movie and group variable breed, was added to the dataset.

Computing ATET for each cohort and time:
Cohort 2034 (9): ......... done
Cohort 2036 (9): ......... done
Cohort 2037 (9): ......... done

Treatment and time information

Time variable: year
Time interval: 2031 to 2040
Control:       _did_cohort = 0
Treatment:     _did_cohort > 0
```

	_did_cohort
Number of cohorts	4
Number of obs	
Never treated	1190
2034	40
2036	30
2037	150

```
Heterogeneous-treatment-effects regression          Number of obs    = 1,410
                                                    Number of panels =   141
Estimator:        Regression adjustment
Panel variable:   breed
Treatment level:  breed
Control group:    Never treated
```

(Std. err. adjusted for 141 clusters in breed)

Cohort	ATET	Robust std. err.	z	P>|z|	[95% conf. interval]	
2034						
year						
2032	-254.8927	266.1024	-0.96	0.338	-776.4439	266.6584
2033	-257.5329	217.9389	-1.18	0.237	-684.6852	169.6194
2034	701.1318	127.0935	5.52	0.000	452.0331	950.2304
2035	1099.044	282.0704	3.90	0.000	546.196	1651.892
2036	1367.632	225.8702	6.05	0.000	924.9343	1810.329
2037	2008.294	237.2396	8.47	0.000	1543.313	2473.275
2038	2472.624	278.2949	8.88	0.000	1927.176	3018.072
2039	2689.615	504.3324	5.33	0.000	1701.142	3678.088
2040	3110.97	568.916	5.47	0.000	1995.915	4226.025
2036						
year						
2032	216.0259	122.9107	1.76	0.079	-24.87472	456.9265
2033	-172.5154	372.0776	-0.46	0.643	-901.7741	556.7433
2034	-218.0495	504.5267	-0.43	0.666	-1206.904	770.8045
2035	621.033	156.1306	3.98	0.000	315.0227	927.0434
2036	999.0781	180.1055	5.55	0.000	646.0779	1352.078
2037	1003.333	250.5916	4.00	0.000	512.1829	1494.484
2038	1556.669	451.6914	3.45	0.001	671.3697	2441.967
2039	2590.674	662.6979	3.91	0.000	1291.81	3889.538
2040	2225.712	486.9917	4.57	0.000	1271.225	3180.198
2037						
year						
2032	-114.582	160.0972	-0.72	0.474	-428.3668	199.2028
2033	-127.9856	183.3941	-0.70	0.485	-487.4315	231.4603
2034	33.40901	168.0312	0.20	0.842	-295.9262	362.7442
2035	130.3495	166.2261	0.78	0.433	-195.4477	456.1468
2036	-10.48288	167.5059	-0.06	0.950	-338.7884	317.8226
2037	1717.016	268.5592	6.39	0.000	1190.65	2243.383
2038	2086.798	278.0215	7.51	0.000	1541.886	2631.71
2039	2473.611	268.186	9.22	0.000	1947.976	2999.246
2040	2835.117	378.6699	7.49	0.000	2092.938	3577.296

Note: ATET computed using covariates.

Notice the note below the command. A variable with the name _did_cohort has been generated. Using the group variable and the observation-level treatment, xthdidregress generated treatment-time cohorts. The new variable creates treatment groups based on the time when a group was first treated. For instance, if a Boxer and a Rottweiler are featured in movies in 2034, they are grouped in the 2034 cohort. The variable also contains a category for a control group. In this case, the control group is formed by the breeds that are not featured in a movie. Cohorts are an important input for estimation and for postestimation commands. You do not need to adhere to the default name, _did_cohort, and may provide your own name using the cohortvar() option.

Next appears a table that gives you a sense of the treatment groups and time. You see the time variable, year, and its range, 2031 to 2040. Then we see what defines a treated or a control group.

The table after provides group-level information about the cohort-time groups. The first row tells you the number of cohorts. Following the number of cohorts is a tabulation showing how many observations are in each cohort. For instance, 1,190 observations are never treated in our data. The table gives you a sense of the amount of information available in each cohort and might hint at the variability of cohort-level estimates.

The next table presents the ATET estimates. The first panel shows the ATETs for the 2034 cohort. We first have the 2032 ATET of -255, with a confidence interval that includes 0. This is as expected; before treatment, the effect should be 0. We should interpret the ATET to mean that among the breeds about which a movie was made in 2034, the expected AKC registrations in 2032 are 255 lower than if the breed had never been in a movie. At treatment onset, in 2034, we observe a treatment effect for the number of registrations of 701, and on the last year of our data, for the 2034 cohort, we see a treatment effect of 3,111 registrations. We interpret the results for the remaining cohorts similarly.

◁

▷ Example 2: Visualizing estimation results

In the example above, in the output table, we had three cohorts and nine time periods. There is a lot of information to process, and it can get even more daunting if we had more cohorts and time periods. To better visualize the results, we can use estat atetplot:
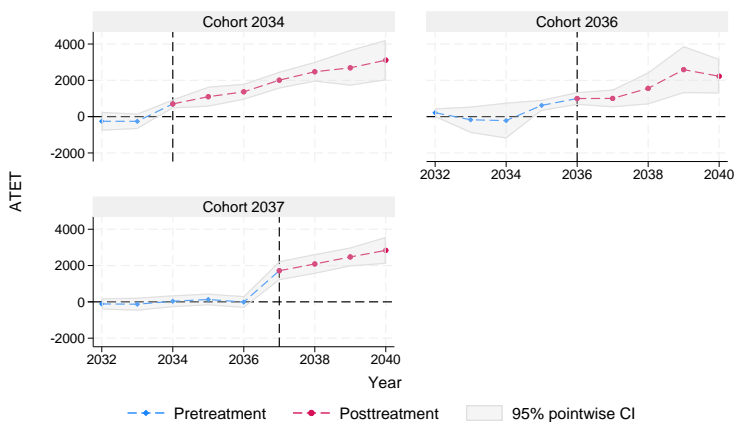
```
. estat atetplot
```


Figure 1. ATETs by cohort over time

The graph shows the pretreatment and the posttreatment ATETs for each cohort and their pointwise confidence intervals. ATETs increase over time after treatment and are flat before treatment.

◁

▷ Example 3: Less heterogeneity; aggregating and summarizing treatment effects

So far, we have allowed treatment effects to change over cohort and over time. But we might want to obtain only one treatment effect for each cohort, abstracting away from time variation within cohorts. You would get this using the postestimation command `estat aggregation`.

```
. estat aggregation, cohort
```

ATET over cohort   Number of obs = 1,410

(Std. err. adjusted for 141 clusters in breed)

Cohort	ATET	Robust std. err.	z	P>|z|	[95% conf. interval]	
2034	1921.33	187.2787	10.26	0.000	1554.271	2288.389
2036	1675.093	130.4929	12.84	0.000	1419.332	1930.855
2037	2278.136	166.5283	13.68	0.000	1951.746	2604.525

We now have 3 treatment effects to analyze instead of 27. For the 2034 cohort, we have a treatment effect of 1,921 registrations for the periods after treatment. For the 2036 cohort, the effect is 1,675 registrations, and for the 2037 cohort, it is 2,278. We cannot see how the treatment evolves over time for each cohort, but we have a sense of the average effect over time for each of them.

We could instead want to see the treatment effect at each point in time, abstracting from cohort-level variation.

```
. estat aggregation, time
```

ATET over time   Number of obs = 1,410

(Std. err. adjusted for 141 clusters in breed)

Time	ATET	Robust std. err.	z	P>|z|	[95% conf. interval]	
2034	701.1318	127.0935	5.52	0.000	452.0331	950.2304
2035	1099.044	282.0704	3.90	0.000	546.196	1651.892
2036	1209.68	170.2043	7.11	0.000	876.0858	1543.275
2037	1672.655	202.1854	8.27	0.000	1276.379	2068.932
2038	2084.658	214.5072	9.72	0.000	1664.232	2505.084
2039	2528.847	225.8763	11.20	0.000	2086.138	2971.557
2040	2802.171	291.8412	9.60	0.000	2230.173	3374.17

We see the treatment effects for each one of the posttreatment periods. As before, we have the option to look at the effects graphically. We just need to use the `graph` option.

```
. estat aggregation, time graph
```
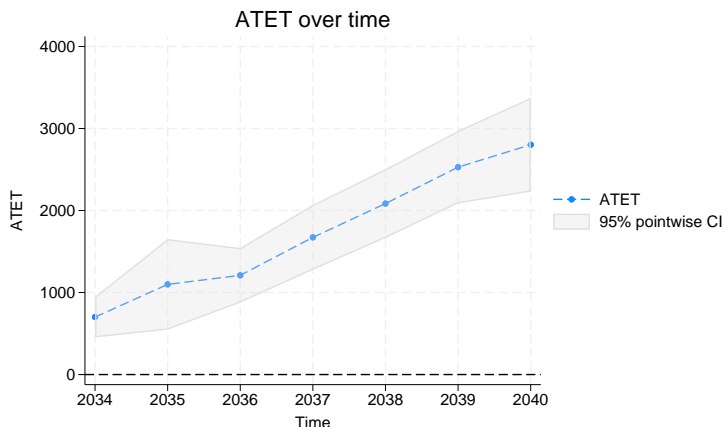

Figure 2. ATETs over time

◁

▷ Example 4: Dynamic treatment effects

We could also ask what the evolution of the treatment effect is after treatment. For instance, we might want to know what happens one period after the onset of treatment, two periods after treatment, and so forth. It might be the case that treatment effects vanish over time or even change patterns. We might also want to see whether, before treatment, we observe a treatment effect or a pattern that might suggest that there is anticipation of treatment. `estat aggregation` allows us to answer these questions by using the `dynamic` option.

```
. estat aggregation, dynamic graph
```

Duration of exposure ATET   Number of obs = 1,410

                                  (Std. err. adjusted for 141 clusters in breed)

Exposure	ATET	Robust std. err.	z	P>|z|	[95% conf. interval]	
-5	-114.582	160.0972	-0.72	0.474	-428.3668	199.2028
-4	-70.65034	156.3185	-0.45	0.651	-377.029	235.7283
-3	-.9117242	153.0999	-0.01	0.995	-300.982	299.1585
-2	12.79653	144.8216	0.09	0.930	-271.0486	296.6417
-1	30.71473	132.8508	0.23	0.817	-229.668	291.0975
0	1434.409	206.3277	6.95	0.000	1030.014	1838.804
1	1759.461	224.0229	7.85	0.000	1320.385	2198.538
2	2147.486	221.903	9.68	0.000	1712.564	2582.408
3	2651.452	284.8928	9.31	0.000	2093.073	3209.832
4	2366.805	267.4253	8.85	0.000	1842.661	2890.949
5	2689.615	504.3324	5.33	0.000	1701.142	3678.088
6	3110.97	568.916	5.47	0.000	1995.915	4226.025

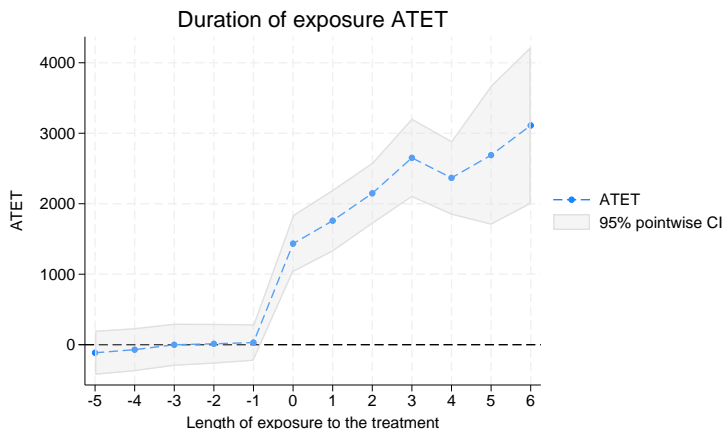Note: Exposure is the number of periods since the first treatment time.

Figure 3. ATET dynamics

From the output table, we see that 5 periods before treatment onset, there is no significant treatment effect or, for that matter, no significant effect at any time of exposure before treatment. This suggests there is no anticipation of treatment. At treatment onset, 0, there is a positive effect, which seems to increase as exposure time to treatment augments.

◁

▷ Example 5: TWFE estimation

The literature on heterogeneous DID started by pointing out the problems that arise when one assumes erroneously that the treatment effects are homogeneous. It suggested that TWFE estimation was inadequate. Wooldridge (2021) suggests that fixed-effects estimation can be used if we extend it to include interactions between treatment-time cohorts and time.

Another important insight of Wooldridge (2021) is that you can use pooled ordinary least squares and add panel-level averages of covariates and obtain the same point estimates as one would get with fixed-effects estimation in the context of DID estimation. This is an extension of the intuition by Mundlak (1978). xthdidregress and hdidregress fit pooled ordinary least-squares models using these ideas. Below, we present the results we obtain using the twfe estimator.

```
. xthdidregress twfe (registered best) (movie), group(breed)
note: variable _did_cohort, containing cohort indicators formed by treatment
      variable movie and group variable breed, was added to the dataset.
Computing ATETs using margins ...

Treatment and time information

Time variable: year
Time interval: 2031 to 2040
Control:       _did_cohort = 0
Treatment:     _did_cohort > 0
```

	_did_cohort
Number of cohorts	4

Number of obs	
Never treated	1190
2034	40
2036	30
2037	150

```
Heterogeneous-treatment-effects regression        Number of obs    =  1,410
                                                   Number of panels =    141

Estimator:        Two-way fixed effects
Panel variable:   breed
Treatment level:  breed
Control group:    Never treated
Heterogeneity:    Cohort and time
                         (Std. err. adjusted for 141 clusters in breed)
```

Cohort	ATET	Robust std. err.	t	P>|t|	[95% conf. interval]	
2034						
year						
2034	469.2023	148.8998	3.15	0.002	174.8195	763.5852
2035	823.8532	211.7491	3.89	0.000	405.2138	1242.493
2036	1108.669	179.8404	6.16	0.000	753.1144	1464.223
2037	1752.287	283.487	6.18	0.000	1191.818	2312.756
2038	2216.617	173.4446	12.78	0.000	1873.708	2559.526
2039	2433.608	521.4074	4.67	0.000	1402.757	3464.458
2040	2854.963	494.1892	5.78	0.000	1877.924	3832.001
2036						
year						
2036	1336.121	96.75296	13.81	0.000	1144.835	1527.406
2037	1343.004	297.7487	4.51	0.000	754.3383	1931.669
2038	1896.339	399.2574	4.75	0.000	1106.985	2685.692
2039	2930.344	591.0712	4.96	0.000	1761.765	4098.924
2040	2565.382	578.9303	4.43	0.000	1420.806	3709.958
2037						
year						
2037	1750.126	216.4288	8.09	0.000	1322.234	2178.017
2038	2119.908	217.1685	9.76	0.000	1690.554	2549.262
2039	2506.72	303.9392	8.25	0.000	1905.816	3107.624
2040	2868.227	313.1679	9.16	0.000	2249.077	3487.377

```
Note: ATET computed using covariates.
```

The output is almost the same as the one for the `ra` estimator in example 1. There are a couple of noteworthy differences. First, the estimator fits an extended TWFE regression and then uses `margins`

to compute the treatment effects. The command tells us `Computing ATETs using margins` when it begins this more computationally intensive calculation.

Second, the ATET parameters are shown for each cohort only at the time of treatment exposure and for the periods thereafter but not for the pretreatment periods. As discussed in Wooldridge (2021), these are the parameters identified using the parallel-trends assumption he derives.

As we did before, we could use `estat aggregation` to explore different ways of looking at our treatment effects and `estat atetplot` to visualize the estimated ATETs.

◁

▷ Example 6: Reducing model complexity

When we fit the `ra` model, we had to estimate ATET parameters for each cohort over time. The complexity of the model grows with the number of cohorts and the number of time periods. As is described in *Methods and formulas*, the `ra` estimator uses a different subset of the data to obtain each parameter. To get a reliable estimator of each parameter, you need sufficient data for each subsample. Sometimes, there are few observations for a given cohort in a given set of time periods.

We can ameliorate this problem by reducing the amount of heterogeneity we assume. For the `twfe` estimator, the complexity of the model comes from the interactions between the observation-level treatment with cohort and time and the interactions between the observation-level treatment, cohort, time, and covariates. This allows us to decide which interactions to include in our model. We could, for instance, allow for heterogeneity at the cohort level instead of at the cohort and time level. We use the `hettype()` option with the argument `cohort()` to do this:

```
. xthdidregress twfe (registered best) (movie), group(breed) hettype(cohort)
note: variable _did_cohort, containing cohort indicators formed by treatment
      variable movie and group variable breed, was added to the dataset.
```

Computing ATETs using **margins** ...

Treatment and time information

```
Time variable: year
Time interval: 2031 to 2040
Control:       _did_cohort = 0
Treatment:     _did_cohort > 0
```

	_did_cohort
Number of cohorts	4
Number of obs	
Never treated	1190
2034	40
2036	30
2037	150

```
Heterogeneous-treatment-effects regression          Number of obs    = 1,410
                                                     Number of panels =   141

Estimator:        Two-way fixed effects
Panel variable:   breed
Treatment level:  breed
Control group:    Never treated
Heterogeneity:    Cohort
```

(Std. err. adjusted for 141 clusters in breed)

Cohort	ATET	Robust std. err.	t	P>|t|	[95% conf. interval]	
2034	1662.492	108.002	15.39	0.000	1448.966	1876.017
2036	1978.645	54.21043	36.50	0.000	1871.468	2085.822
2037	2276.223	70.63244	32.23	0.000	2136.579	2415.867

Note: ATET computed using covariates.

You fit a regression model with fewer terms and obtain treatment effects only at the cohort level. You could also have the treatment effect change over time but not over cohort by typing cohort(time).

For the estimators proposed by Callaway and Sant'Anna (2021), heterogeneity is built in, so we need to estimate all the ATET parameters.

◁

Stored results

xthdidregress stores the following in e():

Scalars
e(N)	number of observations
e(N_clust)	number of clusters
e(N_panels)	number of panels
e(tmin)	first time period
e(tmax)	last time period
e(rank)	rank of e(V)

Macros
e(cmd)	xthdidregress
e(cmdline)	command as typed
e(clustvar)	name of cluster variable
e(control_group)	control group
e(het_type)	heterogeneity type for twfe estimator
e(cohortvar)	name of cohort variable
e(ovar)	name of outcome variable
e(wtype)	weight type
e(wexp)	weight expression
e(marginsnotok)	predictions disallowed by margins
e(ivar)	panel identifier variable
e(timevar)	time variable
e(treatname)	name of treatment variable
e(estat_cmd)	program used to implement estat
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. err.
e(method)	estimator method
e(properties)	b V

Matrices
e(b)	coefficient vector
e(V)	variance–covariance matrix of the estimators
e(cohort_count)	matrix with cohort count information

Functions
e(sample)	marks estimation sample

In addition to the above, the following is stored in r():

Matrices
r(table)	matrix containing the coefficients with their standard errors, test statistics, *p*-values, and confidence intervals

Note that results stored in r() are updated when the command is replayed and will be replaced when any r-class command is run after the estimation command.

Methods and formulas

Methods and formulas are presented under the following headings:

The model
The RA, IPW, and AIPW estimators
　　Panel data
The TWFE estimator

The model

xthdidregress estimates the treatment-effects parameters using the DID approach 1) with multiple periods, 2) with different timing when a unit is first treated, and 3) when the treatment effects can be heterogeneous across time and cohort. In particular, xthdidregress provides the RA, IPW, and AIPW estimators, outlined in Callaway and Sant'Anna (2021), and the TWFE estimator, outlined in Wooldridge (2021).

We observe data $\{y_{it}, \mathbf{x}_{it}, d_{it}, \mathbf{z}_{it}\}$ for unit i at time t with unit i, $i = 1, \ldots, N$, and time t, with $t = 1, \ldots, T$, where

1. y_{it} is the observed outcome;

2. $\mathbf{x}_{it}$ are covariates for the outcome model;

3. d_{it} is an indicator that equals one if an observation is treated or zero otherwise; and

4. $\mathbf{z}_{it}$ are the covariates for the treatment-assignment model.

Denote $y_{it}(g)$ as the potential outcome for unit i at time t if it was first treated at time g. Denote $y_{it}(0)$ as the potential outcome for unit i and t if it is never treated. All the units can be grouped into cohorts, and each cohort starts the treatment simultaneously. For example, cohort g means all the units start treatment at time g. Let G_{ig} be an indicator that equals one if unit i is first treated at time g. Then the units in cohort g can be denoted by $G_{ig} = 1$. When a unit i is never treated, we denote $G_{i0} = 1$. Thus, cohort 0 indicates all the units that are never treated. We assume that once a unit is treated, it will remain treated.

To reveal how the heterogeneous treatment effects evolve across cohorts and time, we are interested in estimating the ATET for each combination of cohort and time. Let $\theta(g, t)$ be the ATET for cohort g at time t, which is defined as

$$\theta(g,t) = \mathbf{E}\left\{y_t(g) - y_t(0) | G_g = 1\right\} \tag{ATET}$$

where G_g equals 1 if a unit belongs to cohort g. All the four estimators provided in xthdidregress and hdidregress estimate $\theta(g, t)$ in equation (ATET). We cannot directly estimate $\theta(g, t)$ using equation (ATET) because the potential outcomes $y_t(g)$ and $y_t(0)$ are not observable.

Next, we will describe the RA, IPW, and AIPW estimators.

The RA, IPW, and AIPW estimators

To define treatment effect, we need a control group. There are two ways to define the control group. One way is to use the units that are never treated as the control group. Let C^{NEV} be an indicator that equals one if a unit belongs to the never-treated group. In particular, $C^{\text{NEV}} = G_0$. Another way is to use the units not in cohort g and not yet treated at time t as the control group. Let $C_{g,t}^{\text{NY}}$ be an indicator that equals one if a unit belongs to the not-yet-treated group by time t. In particular, $C_{g,t}^{\text{NY}} = (1 - G_g)(1 - d_t)$. To simplify, we indicate control, in both cases, as $C_{g,t}^*$.

The definitions of the RA, IPW, and AIPW estimators depend on the definition of $C_{g,t}^*$, which can either be C^{NEV} or $C_{g,t}^{\text{NY}}$. However, regardless of the control group's choice, the estimators' definitions can always be written using the general notation $C_{g,t}^*$.

Panel data

The estimands for RA, IPW, and AIPW are defined as follows.

The RA estimand is

$$\theta_{\mathrm{RA}}(g,t) = \mathbf{E}\left[\frac{G_g}{\mathbf{E}(G_g)}\left\{y_t - y_{g-1} - m_{g,t}(\mathbf{x})\right\}\right] \tag{RA}$$

where $m_{g,t}(\mathbf{x}) = \mathbf{E}(y_t - y_{g-1}|\mathbf{x}, C_{g,t}^* = 1)$.

The IPW estimand is

$$\theta_{\mathrm{IPW}}(g,t) = \mathbf{E}\left(\left[\frac{G_g}{\mathbf{E}(G_g)} - \frac{\frac{p_{g,t}(\mathbf{z})C_{g,t}^*}{1-p_{g,t}(\mathbf{z})}}{\mathbf{E}\left\{\frac{p_{g,t}(\mathbf{z})C_{g,t}^*}{1-p_{g,t}(\mathbf{z})}\right\}}\right](y_t - y_{g-1})\right) \tag{IPW}$$

where

$$p_{g,t}(\mathbf{z}) = Pr(G_g = 1|\mathbf{z}, G_g + C_{g,t}^* = 1) \tag{Pz}$$

The AIPW estimand is

$$\theta_{\mathrm{AIPW}}(g,t) = \mathbf{E}\left(\left[\frac{G_g}{\mathbf{E}(G_g)} - \frac{\frac{p_{g,t}(\mathbf{z})C_{g,t}^*}{1-p_{g,t}(\mathbf{z})}}{\mathbf{E}\left\{\frac{p_{g,t}(\mathbf{z})C_{g,t}^*}{1-p_{g,t}(\mathbf{z})}\right\}}\right]\left\{y_t - y_{g-1} - m_{g,t}(\mathbf{x})\right\}\right) \tag{AIPW}$$

Callaway and Sant'Anna (2021) show that the estimand for RA, IPW, and AIPW is the same as $\theta(g,t)$ in equation (ATET). In other words,

$$\theta(g,t) = \theta_{\mathrm{RA}}(g,t) = \theta_{\mathrm{IPW}}(g,t) = \theta_{\mathrm{AIPW}}(g,t)$$

However, unlike the nonestimable $\theta(g,t)$ in equation (ATET), the estimands in equation (RA), (IPW), and (AIPW) are estimable because only the observed random variables such as $\mathbf{x}$, y, $\mathbf{z}$, G_g, and $C_{g,t}^*$ are used.

The identification of the estimators sheds light on how to estimate $\theta(g,t)$. The estimator can be generally divided into three steps:

1. Restrict the sample to time t and t_0, and keep only the units in cohort g or in control group $C_{g,t}^*$, where $t_0 = g - 1$ if $t \geq g$ or $t_0 = t - 1$ if $t < g$. Only the paired observations are used. In other words, unit i will be used if it is observed at both times t and t_0.

2. Use a parametric model to estimate the nuisance function for $p_{g,t}(\mathbf{z})$ and $m_{g,t}(\mathbf{x})$. In practice, we use a linear regression model to fit $m_{g,t}(\mathbf{x})$ and a logit regression model to fit $p_{g,t}(\mathbf{z})$. In particular, do the following:

    a. Run a linear regression of $y_t - y_{t_0}$ on $\mathbf{x}$ when $C_{g,t}^* = 1$.

    b. Denote the linear predictor as $\widehat{m}_{g,t}(\mathbf{x})$.

    c. Run a logit regression of G_g on $\mathbf{z}$.

    d. Denote the prediction of the probability of a positive outcome as $\widehat{p}_{g,t}(\mathbf{z})$.

3. Plug in the nuisance function estimates $\widehat{m}_{g,t}(\mathbf{x})$, $\widehat{p}_{g,t}(\mathbf{z})$, or both into the estimating equation in equation (RA), (IPW), or (AIPW). Notice that the expectation operator $\mathbf{E}(\cdot)$ is replaced by the sample average.

The variance–covariance matrix for $\widehat{\theta}(g,t)$ for each cohort g and time t is computed using the influence function approach proposed in theorem 2 in Callaway and Sant'Anna (2021). The influence function approach is numerically equivalent to the generalized method-of-moments approach. However, it is much faster because it avoids computing the covariance matrix for the parameters in the nuisance functions $m_{g,t}(\mathbf{x})$ and $p_{g,t}(\mathbf{z})$. For more discussions on influence functions, see Hampel et al. (1986), Newey and McFadden (1994), and Jann (2020).

The TWFE estimator

A TWFE estimator fits

$$y_{it} = \alpha_i + \gamma_t + \mathbf{x}_{it}\beta + d_{it}\tau + \varepsilon_{it}$$

Wooldridge (2021) extends this model to incorporate interactions between the treatment, d_{it}, cohort, G_{ig}, posttreatment periods, and covariates. We define indicators for posttreatment period as f_s with s going from q to T, where q is the first time period we observe treatment. For instance, f_q equals 1 if we are in time period q and 0 otherwise. To simplify this notation, we show the model without covariates. The extended fixed-effects model is given by

$$y_{it} = \eta + \sum_{g=q}^{T} G_{ig}\alpha_g + \sum_{s=q}^{T} f_s\gamma_s + \sum_{g=q}^{T}\sum_{s=g}^{T} d_{it}G_{ig}f_s\tau_{gs} + \varepsilon_{it} \tag{TWFE}$$

We can fit equation (TWFE) using pooled ordinary least squares or a within estimator. We are going to use the estimator proposed by Mundlak (1978). This gives the same point estimates as using the within estimator of xtreg ..., fe for the parameters in equation (TWFE) but has different degrees of freedom because of the additional terms added by the Mundlak approach. Unlike within estimation, the Mundlak approach works for both repeated cross-sectional data and panel data. Also, it has good properties to obtain partial effect under various data-generating processes, as pointed out in Wooldridge (2019).

Above, the τ_{gs} are the cohort-time treatment effects. When we have covariates, we interact them with all the relevant variables in the model. To get the treatment effects in this case, we need to control for the variation in the covariates. We can obtain both effects using margins.

In particular, we are going to type

```
. margins, dydx(d) at(year=q ... year=T) over(cohort) vce(unconditional)
```

where d is the treatment indicator, year indicates treatment times at which treatment will be evaluated using at(), and cohort is the treatment-time cohorts. We use vce(unconditional) to account for the variation in the covariates.

With the hettype() option, we reduce the complexity of (TWFE). In particular, if we ask for hettype(time), we have

$$y_{it} = \eta + \sum_{g=q}^{T} G_{ig}\alpha_g + \sum_{s=q}^{T} f_s\gamma_s + \sum_{s=q}^{T} d_{it}f_s\tau_s + \varepsilon_{it}$$

Now treatment varies over time but not over cohort, that is, τ_s. If we use the `hettype(cohort)` option, we have

$$
y_{it} = \eta + \sum_{g=q}^{T} G_{ig}\alpha_g + \sum_{s=q}^{T} f_s\gamma_s + \sum_{g=q}^{T} d_{it}G_{ig}\tau_g + \varepsilon_{it}
$$

Now treatment varies over cohort but not over time, that is, τ_g.

When the `controlgroup(notyet)` option is specified, the G_{ig} indicator excludes the last treated cohort. As discussed in Wooldridge (2021), when every group is eventually treated, we cannot identify the treatment effect for this cohort. It is therefore sensible to use the last treated cohort as a control group. When some of the units in our sample are never treated, we can always identify all cohorts, and the `twfe` estimator will always revert to using `controlgroup(never)`.

Acknowledgments

References

Callaway, B., and P. H. C. Sant'Anna. 2021. Difference-in-differences with multiple time periods. *Journal of Econometrics* 225: 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001.

de Chaisemartin, C., and X. D'Haultfœuille. Forthcoming. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *Econometrics Journal.* https://doi.org/10.1093/ectj/utac017.

Ghirlanda, S., A. Acerbi, and H. Herzog. 2014. Dog movie stars and dog breed popularity: A case study in media influence on choice. *PLOS ONE* 9: e106565. https://doi.org/10.1371/journal.pone.0106565.

Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. 1986. *Robust Statistics: The Approach Based on Influence Functions.* New York: Wiley.

Jann, B. 2020. Influence functions continued. A framework for estimating standard errors in reweighting, matching, and regression adjustment. Working Papers 35, University of Bern Social Sciences. https://ideas.repec.org/p/bss/wpaper/35.html.

Mundlak, Y. 1978. On the pooling of time series and cross section data. *Econometrica* 46: 69–85. https://doi.org/10.2307/1913646.

Newey, W. K., and D. L. McFadden. 1994. Large sample estimation and hypothesis testing. In Vol. 4 of *Handbook of Econometrics*, ed. R. F. Engle and D. L. McFadden, 2111–2245. Amsterdam: Elsevier. https://doi.org/10.1016/S1573-4412(05)80005-4.

Roth, J., P. H. C. Sant'Anna, A. Bilinski, and J. Poe. 2022. What's trending in difference-in-differences? A synthesis of the recent econometrics literature. ArXiv Working Paper No. arXiv:2201.01194. https://doi.org/10.48550/arXiv.2201.01194.

Wooldridge, J. M. 2019. Correlated random effects models with unbalanced panels. *Journal of Econometrics* 211: 137–150. https://doi.org/10.1016/j.jeconom.2018.12.010.

———. 2021. Two-way fixed effects, the two-way Mundlak regression, and difference-in-differences estimators. Working paper, Department of Economics, Michigan State University, East Lansing, MI. https://dx.doi.org/10.2139/ssrn.3906345.

Also see

[CAUSAL] **hdidregress postestimation** — Postestimation tools for hdidregress and xthdidregress

[CAUSAL] **hdidregress** — Heterogeneous difference in differences

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

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

[CAUSAL] **teffects intro** — Introduction to treatment effects for observational data

[CAUSAL] **teffects intro advanced** — Advanced introduction to treatment effects for observational data

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