

**DID intro —** Introduction to difference-in-differences estimation[Description](#)[Remarks and examples](#)[References](#)[Also see](#)

## Description

Difference in differences (DID) is a method to estimate the average effect of a treatment on those who receive the treatment. The method can be applied to two types of observational data: repeated cross-sections, in which different individuals are observed at different time points, and panel data. As such, Stata provides two commands, `didregress` and `xtdidregress`. These commands estimate the average treatment effect on the treated (ATET) of a binary or continuous treatment on a continuous outcome by fitting a linear model with time fixed effects and group or panel fixed effects.

Difference-in-difference-in-differences (DDD) estimation is also available for situations in which you want to augment the DID framework to include additional control groups to obtain the ATET.

## Remarks and examples

stata.com

Remarks are presented under the following headings:

- [Introduction](#)
- [Intuition for estimating effects](#)
- [Standard error considerations](#)
- [Different types of data and specification](#)
  - [Specifying groups and time as binary indicators](#)
  - [Excluding group and time effects](#)
- [Conclusion](#)

This entry presents the intuition and some of the technical details for the estimators in `didregress` and `xtdidregress` and the diagnostics available after estimation. See [TE] `didregress` and [TE] `didregress postestimation` for details on the syntax and worked examples. For a more complete discussion and references on the subject, see [Angrist and Pischke \(2009\)](#), [Blundell and Dias \(2009\)](#), [Imbens and Wooldridge \(2009\)](#), [Lechner \(2011\)](#), [Angrist and Pischke \(2015\)](#), [Abadie and Cattaneo \(2018\)](#), and [Wing, Simon, and Bello-Gomez \(2018\)](#).

## Introduction

DID is one of the most venerable causal inference methods used by researchers. DID estimates the effect of a treatment on the treated group. When thinking about such effects, one usually thinks of comparing a group before and after the treatment, perhaps by looking at a graph such as the one below:

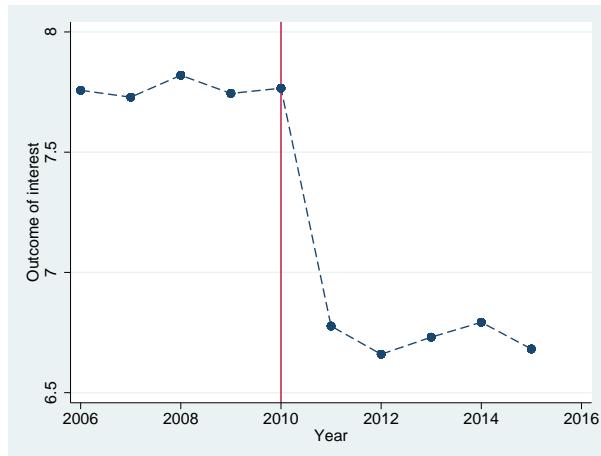


Figure 1

A treatment occurred in the year 2010. This might be a government policy, a change in medicine dosage, or any other intervention of interest. We would like to know if the treatment had an effect. It is clear from the graph that the outcome of interest changed after 2010. Is this due to the treatment or is something else occurring? Perhaps there are unobserved time effects that affect the treatment group after the treatment. For instance, there could have been a change in weather conditions or an economic downfall that affected the treatment group but was not captured in the model or the data. If this is the case, it does not suffice to look at the treatment group before and after the policy. DID addresses this by finding a control group, that is, a group that would have been subject to similar unobserved time effects and was not exposed to the treatment. Comparing the treatment group with the control group before and after the treatment might give us a better understanding of whether the treatment made a difference. A graph looking at a treatment and control group might look like

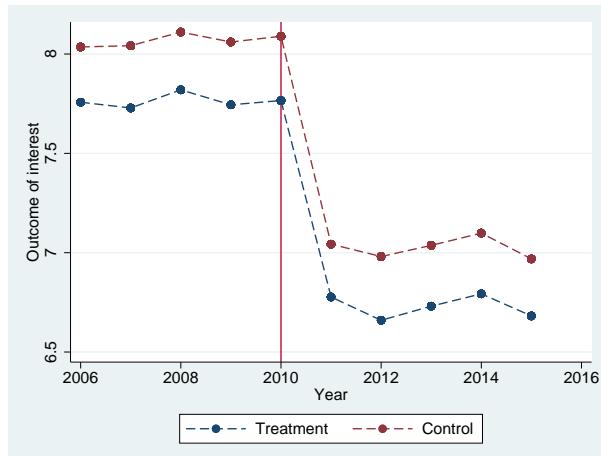


Figure 2

For both the treatment and the control group, we see that there was a decrease in the mean of the outcome after 2010. Therefore, the decrease we saw in the treatment group cannot be attributable entirely to the treatment. (In fact, these are simulated data, and we know the treatment has no effect.)

In a DID setup, if the treated group had not received the treatment, we would expect the treated and control groups to experience the same trends. A treatment effect would imply a systematic deviation from a common trend that could be observed graphically. This is what we observe:

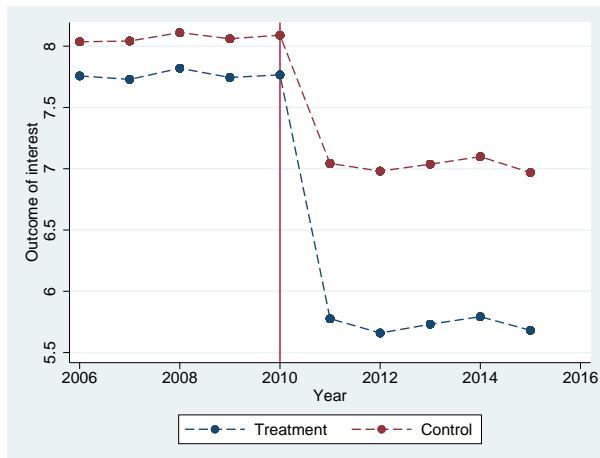


Figure 3

Here both groups experienced a decrease after 2010, but the treatment group decrease was more substantial. The difference in the decreases across groups may indicate the effect of the treatment.

Researchers may motivate their analysis with such graphs. However, graphical evidence is not enough. We need statistical validation, so we fit a model.

The DID strategy relies on two differences. The first is a difference across time periods. Separately for the treatment group and the control group, we compute the difference of the outcome mean before and after the treatment. This across-time difference eliminates time-invariant unobserved group characteristics that confound the effect of the treatment on the treated group. But eliminating group-invariant unobserved characteristics is not enough to identify an effect. There may be time-varying unobserved confounders with an effect on the outcome mean, even after we control for time-invariant unobserved group characteristics. Therefore, we incorporate a second difference—a difference between the treatment group and the control group. DID eliminates time-varying confounders by comparing the treatment group with a control group that is subject to the same time-varying confounders as the treatment group.

The ATET is then consistently estimated by differencing the mean outcome for the treatment and control groups over time to eliminate time-invariant unobserved characteristics and also differencing the mean outcome of these groups to eliminate time-varying unobserved effects common to both groups. These two differences give the method its name and highlight its intuitive appeal. More appealing is the fact that you can get the effect of interest, the ATET, from one parameter in a linear regression.

When talking about DID, people cite [Snow \(1849\)](#) and [Snow \(1855\)](#) as the first known applications. Snow claimed that cholera was not transmitted by contaminated air or contaminated blood as was thought by some academics of his time. Snow hypothesized the disease was communicated via water that had been polluted with sewage. Below, he describes how he came up with an idea for a natural experiment to validate his hypothesis:

“In Thomas Street, Horsleydown, there are two courts close together, consisting of a number of small houses or cottages, inhabited by poor people. The houses occupy one side of each court or alley—the south side of Trusscott’s Court, and the north

side of the other, which is called Surrey Buildings, being placed back to back, with an intervening space, divided into small back areas, in which are situated the privies of both the courts, communicating with the same drain, and there is an open sewer which passes the further end of both courts. Now, in Surrey's buildings the cholera has committed fearful devastation, whilst in the adjoining court there has been but one fatal case, and another case that ended in recovery. In the former court the slops of dirty water poured down by the inhabitants into a channel in front of the houses got into the well from which they obtained their water, this being the only difference . . .”

In the first edition (1849) of the text, Snow reports the deaths from cholera from September 23, 1848 to August 25, 1849 for five London districts. The number of deaths is higher in the South and East districts relative to the other three districts, arising from the source of their water supply. Snow obtains a clear motivation for his theory. In the second edition (1855), he collects data before and after a pump with contaminated water in Broadstreet, London, is closed. It is then that he can compare a treated with a control group before and after a treatment to establish a treatment effect.

## Intuition for estimating effects

We can build our intuition about the causal inference implied by the DID setup by using the potential outcomes framework described in [TE] teffects intro and [TE] teffects intro advanced. We consider individual-level data for which we sample different individuals at different points in time, a repeated cross-section. The treatment occurs at the group level. For example, the treatment may occur at the state, county, or hospital level. All individuals in a given state, county, or hospital either are treated or are controls at a given point in time. We index individuals by  $i$ , groups by  $g$ , and time by  $t$ . We are interested in the effect of a treatment,  $D_{igt} \in \{0, 1\}$ , on an outcome,  $Y_{igt}$ . Suppose the potential outcome mean of an individual in group  $g$  at time  $t$  that does not receive the treatment is given by the following:

$$E \{Y_{igt} (0) | g, t\} = \gamma_g + \gamma_t \quad (1)$$

Above,  $\gamma_g$  denotes the group effects and  $\gamma_t$  denotes the time effects. Also suppose the potential outcome mean for someone who receives the treatment is given by the following:

$$E \{Y_{igt} (1) | g, t\} = \gamma_g + \gamma_t + \delta \quad (2)$$

The potential outcome framework described above allows us to think of the regression model:

$$Y_{igt} = \gamma_g + \gamma_t + D_{gt}\delta + \varepsilon_{igt}$$

A regression estimate of  $\delta$ , the coefficient on the indicator of treatment, consistently estimates the ATET in this simplified framework, if we can make standard assumptions needed for linear regression and for treatment-effects estimators, discussed in [TE] teffects intro advanced, plus one additional assumption.

To introduce this additional assumption, it is sometimes more intuitive to look at a two-period, two-group example. In this case,  $g \in \{0, 1\}$ , where 0 is the control group and 1 is the treatment group, and  $t \in \{0, 1\}$ , where 0 is the period before the treatment and 1 occurs after the treatment. To guarantee a consistent estimate of the ATET, we need to make the parallel-trends assumption:

$$E(Y_{i01} | D_{gt} = 1) - E(Y_{i00} | D_{gt} = 1) = E(Y_{i01} | D_{gt} = 0) - E(Y_{i00} | i D_{gt} = 0)$$

The parallel-trends assumption is stated in terms of the potential outcomes of not being treated,  $Y_{i0t}$ , conditional on treatment,  $D_{gt}$ . It implies that if the treated had not received the treatment, the groups defined by  $D_{gt} = 1$  and  $D_{gt} = 0$  would have the same trends. For this to be true, we need group effects to be time invariant and time effects to be group invariant. The simple framework described in (1) and (2) satisfies the parallel-trends assumption.

The parallel-trends assumption has a graphical representation. Let's think again about the case with multiple time periods. The parallel-trends assumption implies the paths of the outcome variable over time are parallel between the control and treatment groups prior to the date of the treatment. We can visually check this assumption by plotting the means of the outcome over time for both groups or by visualizing the results of the linear trends model. For instance, we might use a graph like the one in figure 2, where we plotted the means over time. After fitting a model using `didregress` and `xtdidregress`, you can get both the mean outcome plot and the trends plot by typing

```
. estat trendplots
```

Another way to think about the parallel-trends assumption in the pretreatment period is that treatment and control groups do not change their behavior in anticipation of the treatment. We can think of the parallel-trends assumption as implying that there should be no treatment effect in anticipation of the treatment. To test this assumption, we could fit a Granger-type causality model where we augment our model with dummies for each pretreatment–treatment period for the treated observations. A joint test of the coefficients on these dummies against 0 can be used as a test of the null hypothesis that no anticipatory effects have taken place. We can perform this test by typing

```
. estat granger
```

## Standard error considerations

While a standard linear regression model can be used to estimate the ATET, the best estimate of the standard error requires some consideration. Many standard error estimates have been proposed, and each one performs differently depending on the type of DID model being fit and the structure of the data. Here we provide a discussion of some of the issues centered on the available standard error estimates for `didregress` and `xtdidregress`. For a more complete discussion, see [Cameron and Miller \(2015\)](#) and [MacKinnon \(2019\)](#) and the references therein.

[Bertrand, Duflo, and Mullainathan \(2004\)](#) show that the standard errors for DID estimates are inconsistent if they do not account for the serial correlation of the outcome of interest. Because the outcomes studied usually vary at the group and time levels, it makes sense to correct for serial correlation. The authors show that using cluster–robust standard errors at the group level where treatment occurs provides correct coverage in the presence of serial correlation when the number of groups is not too small. [Bester, Conley, and Hansen \(2011\)](#) further show that using cluster–robust standard errors and using critical values of a  $t$  distribution with  $G - 1$  degrees of freedom, where  $G$  is the number of groups, is asymptotically valid for a fixed number of groups and a growing sample size. In other words, you do not need the number of groups to be arbitrarily large, that is, to grow asymptotically. Cluster–robust standard errors with  $G - 1$  degrees of freedom are the default standard errors of `didregress` and `xtdidregress`.

The results of [Bertrand, Duflo, and Mullainathan \(2004\)](#) and [Bester, Conley, and Hansen \(2011\)](#) demonstrate that we could still obtain reliable standard errors even when the number of groups was not large. But what about data with a very small number of groups? Several simulation and theoretical results suggest that cluster–robust standard errors may still have poor coverage when the number of groups is very small or when the number of treated groups is small relative to the number of control groups. These scenarios with small group sizes are common in DID studies, and another method of standard error estimation should be considered in these situations.

For cases where the number of groups is small, we provide three alternatives. The first alternative is to use the wild cluster bootstrap that imposes the null hypothesis that the ATET is 0. Cameron, Gelbach, and Miller (2008) and MacKinnon and Webb (2018) show that the wild cluster bootstrap provides better inference than using cluster-robust standard errors with  $t(G - 1)$  critical values. The second alternative comes from Imbens and Kolesár (2016), who show that with a small number of groups, you may use bias-corrected standard errors with the degrees of freedom adjustment proposed by Bell and McCaffrey (2002). For the third alternative, you may use aggregation type methods like those proposed by Donald and Lang (2007); they show that their method works well when the number of groups is small but the number of individuals in each group is large.

When the disparity between treatment and control groups is large, for example, because there is only one treated group or because the group sizes vary greatly, cluster-robust standard errors and the other methods mentioned above underperform. Yet the bias-corrected and cluster-bootstrap methods provide an improvement over the cluster-robust standard errors.

## Different types of data and specification

We allow for DID estimation of ATETs with two types of data: repeated cross-sections and panel data. The type of estimator to use depends on the assumptions you would like to make about the model and what type of data you have. Below, we walk you through the different types of models you may want to fit and how to fit them with `didregress` and `xtdidregress`.

Suppose you have a repeated cross-section of individuals  $i$  over a period of time  $t$ . The treatment is implemented at the state level,  $s$ , and it is denoted by  $D_{st}$ , where  $D_{st} = 1$  for all observations that are subject to the treatment in state  $s$  at time  $t$ . The indexing of the treatment makes clear that all individuals in the state at a given time are either treated or untreated. The model to be fit is given by

$$y_{ist} = \gamma_s + \gamma_t + \mathbf{z}_{ist}\beta + D_{st}\delta + \varepsilon_{ist} \quad (3)$$

In the expression above,  $y_{ist}$  is the outcome,  $y$  in our data;  $\mathbf{z}_{ist}$  are individual-level characteristics, given by  $\mathbf{z1}$  and  $\mathbf{z2}$ ; and  $D_{st}$  is given by the variable  $d$ . To obtain the ATET, we type

```
. didregress (y z1 z2) (d), group(state) time(year)
```

Within the first set of parentheses, we list the outcome and covariates  $z_{ist}$  from (3). In the second set of parentheses, we specify the treatment variable. Group and time effects are included in the model through the `group()` and `time()` options, respectively. Also, by default, cluster-robust standard errors are computed at the `group()` level. The command above is equivalent to typing

```
. regress y z1 z2 i.year i.state d, vce(cluster state)
```

`didregress` also allows us to specify a DDD model for situations in which we would like to augment the DID framework to include another control group. For instance, starting from (3), let's assume the treatment occurs at the state level but also varies for people older than 65 versus people who are younger, defined by a binary variable  $g$ . The DDD model is now given by

$$y_{isgt} = \gamma_s + \gamma_g + \gamma_t + \gamma_s\gamma_t + \gamma_g\gamma_t + \gamma_s\gamma_g + \mathbf{z}_{isgt}\beta + D_{sgt}\delta + \varepsilon_{isgt}$$

To fit this model, we would type

```
. didregress (y z1 z2) (d), group(state g) time(year)
```

Although the model has a large set of interactions and looks much more complex than the DID model, the only difference in what we type is the addition of `g` to the set of group variables. The cluster-robust standard errors are computed at the highest level of clustering, in this case, `state`.

In some cases, data do not include a time component but rather multiple grouping variables across which differences may be taken. You could fit DDD models for three groups, say, `state`, `g1`, and `g2`, by typing

```
. didregress (y z1 z2) (d), group(state g1 g2)
```

Or, you could fit DID models without time by typing

```
. didregress (y z1 z2) (d), group(g1 g2)
```

With these last two specifications, diagnostics that rely on time variables are not available.

With panel data, the basic model is given by

$$y_{ist} = \gamma_i + \gamma_t + \mathbf{z}_{ist}\beta + D_{st}\delta + \varepsilon_{ist} \quad (4)$$

Again,  $i$  denotes the individual,  $t$  is the time period, and  $s$  is a group-level index. Individuals are assumed to be nested within the group, and treatment occurs at the group and time levels.

The specification is analogous to the one in (3) except that instead of  $\gamma_s$  we have  $\gamma_i$ . We assume the model in (4) is fit so that data comprising a panel of individuals is observed across the time for which individuals are nested within treatment group. Suppose that our data now include a panel identifier variable, `id`, which corresponds to  $i$  in (4). To fit the model above, we would type

```
. xtset id
. xtdidregress (y z1 z2) (d), group(state) time(year)
```

In this model, the group variable is different from the `xtset` identifying variable. The group variable defines the level of clustering, whereas the `xtset` variable defines the panel identifier. In other words, we are fitting a fixed-effects model with individual fixed-effects, not one with state fixed effects. Of course, the group and `xtset` variables could be the same, but they do not need to be.

## Specifying groups and time as binary indicators

The model discussed above is referred to as a generalized DID model by [Wing, Simon, and Bello-Gomez \(2018\)](#). It is a generalization of the two-period, two-group specification that is usually discussed in the literature. This is also the way [Angrist and Pischke \(2009\)](#) write the model.

It is not uncommon, however, to write the model in terms of binary indicators. Instead of having group effects and time effects in the model for all groups and times in the data, you could instead include only a posttreatment effect and a treatment-group effect. To do this, you would create and include an indicator in your model that takes the value of 1 if the observation belongs to a treated group and 0 otherwise as well as an indicator that is 1 if the period considered occurs after the treatment. You can specify this model by using the `nogteffects` option and including your own indicators.

Say, for example, you create a time-constant indicator, `gtreated`, that is 1 if a group is treated and 0 otherwise, as well as an indicator, `post`, that is 1 after the policy was implemented and 0 before. You would type

```
. didregress (y z1 z2 i.post i.gtreated) (d), group(state) time(year) nogteffects
```

The syntax is equivalent for `xtdidregress`.

The model you specify in this case is given by

$$y_{ist} = 1\{s = \text{treated}\} + 1\{t = \text{post}\} + \mathbf{z}_{ist}\beta + D_{st}\delta + \varepsilon_{ist}$$

where  $1\{s = \text{treated}\}$  is an indicator function that is 1 if group  $s$  receives the treatment and 0 otherwise. Similarly,  $1\{t = \text{post}\}$  indicates that we are in the period for which the treatment is active. In the expression above, we still specify the `group()` and `time()` options to obtain the correct standard errors, to validate the assumptions of the DID specification, and to obtain group- and time-level statistics.

## Excluding group and time effects

The DID commands also allow you to forgo the group and time effects that are generated by default. In this case, you would have to add your own group and time effects or omit them entirely by using the `nogteffects` option, which excludes group and time effects from your specification. For DDD models, you may also use the `nointeract` option, which will exclude group and time interactions from your model. Combining both options excludes group and time effects altogether. You may type

```
. didregress (y z) (d), group(s g) time(t) nogteffects nointeract
```

You would then fit

$$y_{isgt} = \beta_0 + \mathbf{z}_{igst}\beta + D_{sgt}\delta + \varepsilon_{ist}$$

You could still include group and time interactions by adding them in the first set of parentheses.

## Conclusion

`didregress` and `xtdidregress` compute the ATET using DID and DDD for panel data and repeated cross-sections. Both commands offer standard error computations that address some of the issues that researchers face, such as a small number of groups. After fitting models with these commands, diagnostics and tests to validate the assumptions and internal validity of the DID and DDD results are available.

## References

- Abadie, A., and M. D. Cattaneo. 2018. Econometric methods for program evaluation. *Annual Review of Economics* 10: 465–503. <https://doi.org/10.1146/annurev-economics-080217-053402>.
- Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton, NJ: Princeton University Press.
- . 2015. *Mastering Metrics: The Path from Cause to Effect*. Princeton, NJ: Princeton University Press.
- Bell, R. M., and D. F. McCaffrey. 2002. Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology* 28: 169–181.
- Bertrand, M., E. Duflo, and S. Mullainathan. 2004. How much should we trust difference-in-differences estimates? *Quarterly Journal of Economics* 119: 249–275. <https://doi.org/10.1162/003355304772839588>.
- Bester, C. A., T. G. Conley, and C. B. Hansen. 2011. Inference with dependent data using cluster covariance estimators. *Journal of Econometrics* 165: 137–151. <https://doi.org/10.1016/j.jeconom.2011.01.007>.
- Blundell, R. W., and M. C. Dias. 2009. Alternative approaches to evaluation in empirical microeconomics. *Journal of Human Resources* 44: 565–640. <https://doi.org/10.1007/s10258-002-0010-3>.

- Cameron, A. C., J. B. Gelbach, and D. L. Miller. 2008. Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics* 90: 414–427. <https://doi.org/10.1162/rest.90.3.414>.
- Cameron, A. C., and D. L. Miller. 2015. A practitioner's guide to cluster-robust inference. *Journal of Human Resources* 50: 317–372. <https://doi.org/10.3368/jhr.50.2.317>.
- Donald, S. G., and K. Lang. 2007. Inference with difference-in-differences and other panel data. *Review of Economics and Statistics* 89: 221–233. <https://doi.org/10.1162/rest.89.2.221>.
- Imbens, G. W., and M. Kolesár. 2016. Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics* 98: 701–712. [https://doi.org/10.1162/REST\\_a\\_00552](https://doi.org/10.1162/REST_a_00552).
- Imbens, G. W., and J. M. Wooldridge. 2009. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47: 5–86. <https://doi.org/10.1257/jel.47.1.5>.
- Lechner, M. 2011. The estimation of causal effects by difference-in-difference methods. *Foundations and Trends in Econometrics* 4: 165–224. <http://doi.org/10.1561/0800000014>.
- MacKinnon, J. G. 2019. How cluster-robust inference is changing applied econometrics. *Canadian Journal of Economics* 52: 851–881. <https://doi.org/10.1111/caje.12388>.
- MacKinnon, J. G., and M. D. Webb. 2018. The wild bootstrap for few (treated) clusters. *Econometrics Journal* 21: 114–135. <https://doi.org/10.1111/ectj.12107>.
- Roodman, D., J. G. MacKinnon, M. O. Nielsen, and M. D. Webb. 2019. *Fast and wild: Bootstrap inference in Stata using boottest*. *Stata Journal* 19: 4–60.
- Snow, J. 1849. *On the Mode of Communication of Cholera*. London: Churchill.
- . 1855. *On the Mode of Communication of Cholera*. 2nd ed. London: Churchill.
- Wing, C., K. Simon, and R. A. Bello-Gomez. 2018. Designing difference in difference studies: Best practices for public health policy research. *Annual Review of Public Health* 39: 453–469. <https://doi.org/10.1146/annurev-publhealth-040617-013507>.

## Also see

[TE] **teffects** — Treatment-effects estimation for observational data

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

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