teffects ra — Regression adjustment

Description

`teffects ra` estimates the average treatment effect (ATE), the average treatment effect on the treated (ATET), and the potential-outcome means (POMs) from observational data by regression adjustment (RA). RA estimators use contrasts of averages of treatment-specific predicted outcomes to estimate treatment effects. `teffects ra` accepts a continuous, binary, count, fractional, or nonnegative outcome and allows a multivalued treatment.

See [CAUSAL] **teffects intro** or [CAUSAL] **teffects intro advanced** for more information about estimating treatment effects from observational data.

Quick start

ATE from a linear model of `y1` on `x1` and `x2` with binary treatment `treat2`

```
teffects ra (y1 x1 x2) (treat2)
```

Same as above, but estimate the ATET

```
teffects ra (y1 x1 x2) (treat2), atet
```

Same as above, but estimate the potential-outcome means

```
teffects ra (y1 x1 x2) (treat2), pomeans
```

ATE of `treat2` using a heteroskedastic probit model for binary outcome `y2`

```
teffects ra (y2 x1 x2, hetprobit(x1 x2)) (treat2)
```

ATE of `treat2` using a Poisson model for count outcome `y3`

```
teffects ra (y3 x1 x2, poisson) (treat2)
```

ATE for each level of three-valued treatment `treat3`

```
teffects ra (y1 x1 x2) (treat3)
```

Same as above, and specify that `treat3` = 3 is the control level

```
teffects ra (y1 x1 x2) (treat3), control(3)
```

Same as above, specified using the label "MyControl" corresponding to `treat3` = 3

```
teffects ra (y1 x1 x2) (treat3), control("MyControl")
```

Menu

Statistics > Causal inference/treatment effects > Continuous outcomes > Regression adjustment

Statistics > Causal inference/treatment effects > Binary outcomes > Regression adjustment

Statistics > Causal inference/treatment effects > Count outcomes > Regression adjustment

Statistics > Causal inference/treatment effects > Fractional outcomes > Regression adjustment

Statistics > Causal inference/treatment effects > Nonnegative outcomes > Regression adjustment

Syntax

```
teffects ra (ovar omvarlist [ , omodel noconstant ]) (tvar) [ if ] [ in ] [ weight ]
    [ , stat options ]
```

ovar is a binary, count, continuous, fractional, or nonnegative outcome of interest.

omvarlist specifies the covariates in the outcome model.

tvar must contain integer values representing the treatment levels.

omodel	Description
Model	
`linear`	linear outcome model; the default
`logit`	logistic outcome model
`probit`	probit outcome model
`hetprobit(`*varlist*`)`	heteroskedastic probit outcome model
`poisson`	exponential outcome model
`flogit`	fractional logistic outcome model
`fprobit`	fractional probit outcome model
`fhetprobit(`*varlist*`)`	fractional heteroskedastic probit outcome model

omodel specifies the model for the outcome variable.

stat	Description
Stat	
`ate`	estimate average treatment effect in population; the default
`atet`	estimate average treatment effect on the treated
`pomeans`	estimate potential-outcome means

options	Description
SE/Robust	
`vce(`*vcetype*`)`	*vcetype* may be `robust`, `cluster` *clustvar*, `bootstrap`, or `jackknife`
Reporting	
`level(`#`)`	set confidence level; default is `level(95)`
`aequations`	display auxiliary-equation results
display_options	control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
Maximization	
maximize_options	control the maximization process; seldom used
Advanced	
`control(`#|*label*`)`	specify the level of *tvar* that is the control
`tlevel(`#|*label*`)`	specify the level of *tvar* that is the treatment
`coeflegend`	display legend instead of statistics

omvarlist may contain factor variables; see **[U] 11.4.3 Factor variables**.

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

Weights are not allowed with the `bootstrap` prefix; see [R] **bootstrap**.

`fweights`, `iweights`, and `pweights` are allowed; see **[U] 11.1.6 weight**.

`coeflegend` does not appear in the dialog box.

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

Options

Model

`noconstant`; see [R] **Estimation options**.

Stat

stat is one of three statistics: `ate`, `atet`, or `pomeans`. `ate` is the default.

`ate` specifies that the average treatment effect be estimated.

`atet` specifies that the average treatment effect on the treated be estimated.

`pomeans` specifies that the potential-outcome means for each treatment level be estimated.

SE/Robust

`vce(`*vcetype*`)` specifies the type of standard error reported, which includes types that are robust to some kinds of misspecification (`robust`), that allow for intragroup correlation (`cluster` *clustvar*), and that use bootstrap or jackknife methods (`bootstrap`, `jackknife`); see [R] ***vce_option***.

Reporting

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

`aequations` specifies that the results for the outcome-model or the treatment-model parameters be displayed. By default, the results for these auxiliary parameters are not displayed.

display_options: noci, nopvalues, noomitted, vsquish, noemptycells, baselevels, allbaselevels, nofvlabel, fvwrap(#), fvwrapon(*style*), cformat(%*fmt*), pformat(%*fmt*), sformat(%*fmt*), and nolstretch; see [R] **Estimation options**.

Maximization

maximize_options: iterate(#), [no]log, and from(*init_specs*); see [R] **Maximize**. These options are seldom used.

init_specs is one of

matname [, skip copy]

#[, #...], copy

Advanced

control(#|*label*) specifies the level of *tvar* that is the control. The default is the first treatment level. You may specify the numeric level # (a nonnegative integer) or the label associated with the numeric level. control() may not be specified with statistic pomeans. control() and tlevel() may not specify the same treatment level.

tlevel(#|*label*) specifies the level of *tvar* that is the treatment for the statistic atet. The default is the second treatment level. You may specify the numeric level # (a nonnegative integer) or the label associated with the numeric level. tlevel() may only be specified with statistic atet. tlevel() and control() may not specify the same treatment level.

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

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

Remarks and examples

Remarks are presented under the following headings:

Overview
Video example

Overview

Regression adjustment (RA) estimators use the contrasts of the averages of treatment-specific predicted outcomes to estimate treatment effects. RA estimators use a two-step approach to estimating treatment effects:

1. They fit separate regression models of the outcome on a set of covariates for each treatment level.
2. They compute the averages of the predicted outcomes for each subject and treatment level. These averages reflect the POMs. The contrasts of these averages provide estimates of the ATEs. By restricting the computations of the means to the subset of treated subjects, we obtain the ATETs.

RA estimators are consistent as long as the treatment is independent of the potential outcomes after conditioning on the covariates. In fact, teffects ra uses an estimation technique that implements both steps at once so that we do not need to correct the standard errors in the second step to reflect the uncertainty surrounding the predicted outcomes.

We will illustrate the use of teffects ra by using data from a study of the effect of a mother's smoking status during pregnancy (mbsmoke) on infant birthweight (bweight) as reported by Cattaneo (2010). This dataset also contains information about each mother's age (mage), education level (medu), marital status (mmarried), whether the first prenatal exam occurred in the first trimester (prenatal1), and whether this baby was the mother's first birth (fbaby).

Example 1: Estimating the ATE

We begin by using teffects ra to estimate the average treatment effect of smoking, controlling for first-trimester exam status, marital status, mother's age, and first-birth status. In Stata, we type

```
. use https://www.stata-press.com/data/r19/cattaneo2
(Excerpt from Cattaneo (2010) Journal of Econometrics 155: 138-154)
. teffects ra (bweight prenatal1 mmarried mage fbaby) (mbsmoke)
Iteration 0:  EE criterion = 7.734e-24
Iteration 1:  EE criterion = 1.196e-25
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	-239.6392	23.82402	-10.06	0.000	-286.3334	-192.945
POmean						
mbsmoke Nonsmoker	3403.242	9.525207	357.29	0.000	3384.573	3421.911

The average birthweight if all mothers were to smoke would be 240 grams less than the average of 3,403 grams that would occur if none of the mothers had smoked.

The previous results showed us the average amount by which infants' weights are affected by their mothers' decision to smoke. We may instead be interested in knowing the average amount by which the weight of babies born to smoking mothers was decreased as a result of smoking. The ATET provides us with the answer.

Example 2: Estimating the ATET

To obtain the ATET rather than the ATE, we use the `atet` option:

```
. teffects ra (bweight prenatal1 mmarried mage fbaby) (mbsmoke), atet

Iteration 0:  EE criterion = 7.629e-24
Iteration 1:  EE criterion = 2.697e-26

Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATET						
mbsmoke (Smoker vs Nonsmoker)	-223.3017	22.7422	-9.82	0.000	-267.8755	-178.7278
POmean						
mbsmoke Nonsmoker	3360.961	12.75749	263.45	0.000	3335.957	3385.966

The average birthweight is 223 grams less when all the mothers who smoke do so than the average of 3,361 grams that would have occurred if none of these mothers had smoked.

The ATET differs from the ATE because the distribution of the covariates among mothers who smoke differs from the distribution for nonsmoking mothers. For example, in [CAUSAL] **teffects intro**, we remarked that in our sample, mothers who smoked tended to be older than those who did not. The differing distributions of covariates also affect the estimated POMs.

◁

By default, `teffects ra` reports the ATE, which is the difference between the two POMs in the case of a binary treatment variable. Sometimes, we want to know the estimated POMs themselves. We might also want to see the actual regression equations used to estimate the POMs. Obtaining this information is easy, as the next example illustrates.

Example 3: Estimating the POMs

Here we use the `pomeans` option to display the POMs and the `aequations` option to display the estimated regression coefficients for the treated and untreated subjects.

```
. teffects ra (bweight prenatal1 mmarried mage fbaby) (mbsmoke),
> pomeans aequations

Iteration 0:  EE criterion = 7.734e-24
Iteration 1:  EE criterion = 2.850e-26

Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
POmeans						
mbsmoke						
Nonsmoker	3403.242	9.525207	357.29	0.000	3384.573	3421.911
Smoker	3163.603	21.86351	144.70	0.000	3120.751	3206.455
OME0						
prenatal1	64.40859	27.52699	2.34	0.019	10.45669	118.3605
mmarried	160.9513	26.6162	6.05	0.000	108.7845	213.1181
mage	2.546828	2.084324	1.22	0.222	-1.538373	6.632028
fbaby	-71.3286	19.64701	-3.63	0.000	-109.836	-32.82117
_cons	3202.746	54.01082	59.30	0.000	3096.886	3308.605
OME1						
prenatal1	25.11133	40.37541	0.62	0.534	-54.02302	104.2457
mmarried	133.6617	40.86443	3.27	0.001	53.5689	213.7545
mage	-7.370881	4.21817	-1.75	0.081	-15.63834	.8965804
fbaby	41.43991	39.70712	1.04	0.297	-36.38461	119.2644
_cons	3227.169	104.4059	30.91	0.000	3022.537	3431.801

The nonsmoker POM for infant birthweight is 3,403 grams; that means that if none of the women in our sample smoked during pregnancy, the expected average birthweight would be 3,403 grams. The POM if all mothers did smoke during pregnancy is 3,164 grams, a difference of 240 grams, as we established in example 1. The coefficients for the equation labeled `OME0` represent the linear equation used to estimate the nontreated POM, and the coefficients for the equation labeled `OME1` represent the linear equation used to estimate the treated POM. The coefficients are identical to those we would obtain using `regress`, but the standard errors differ slightly because `teffects ra` does not make the small-sample adjustment that `regress` does.

◁

We often express statistics as percentages to alleviate scaling issues and aid interpretation. In the present context, we may wish to express an ATE as a percentage of the untreated POM to gain a more intuitive measure of efficacy.

Example 4: Reporting the ATE as a percentage

Sometimes, we are interested in reporting the estimated treatment effect as a percentage of the untreated POM. We continue to use the same model as in the previous examples, but we specify the `coeflegend` option so that `teffects ra` reports the names of the parameters. Knowing the correct names to use, we can then use `nlcom` to obtain the percentage change along with its delta-method-based standard error. We type

```
. use https://www.stata-press.com/data/r19/cattaneo2
(Excerpt from Cattaneo (2010) Journal of Econometrics 155: 138-154)
. teffects ra (bweight prenatal1 mmarried mage fbaby) (mbsmoke), coeflegend
Iteration 0:  EE criterion = 7.734e-24
Iteration 1:  EE criterion = 1.196e-25
Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : regression adjustment
Outcome model  : linear
Treatment model: none
```

bweight	Coefficient	Legend
ATE		
mbsmoke (Smoker vs Nonsmoker)	-239.6392	_b[ATE:r1vs0.mbsmoke]
POmean		
mbsmoke Nonsmoker	3403.242	_b[POmean:0.mbsmoke]

```
. nlcom _b[ATE:r1vs0.mbsmoke] / _b[POmean:0.mbsmoke]
       _nl_1: _b[ATE:r1vs0.mbsmoke] / _b[POmean:0.mbsmoke]
```

bweight	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
_nl_1	-.070415	.0069245	-10.17	0.000	-.0839867	-.0568433

The average birthweight falls by an estimated 7.0% when every mother smokes relative to the case when no mothers smoke. We also obtain a 95% confidence interval of a 5.7% to 8.4% reduction.

Birthweights cannot be negative, though it is possible for a linear regression model to make negative predictions. A common way to enforce nonnegative predictions is to use an exponential conditional-mean model, which is commonly fitted using the Poisson quasimaximum likelihood estimator, as discussed in Cameron and Trivedi (2005, sec. 5.7), Wooldridge (2010, sec. 18.2), and Pawitan (2001, chap. 14). `teffects ra` provides an option to use this model rather than linear regression for the outcomes.

Example 5: Modeling nonnegative outcomes

Now we refit our model of smoking behavior on birthweight, but we specify the `poisson` option in the outcome-model equation so that `teffects ra` uses the Poisson exponential model rather than linear regression:

```
. teffects ra (bweight prenatal1 mmarried mage fbaby, poisson) (mbsmoke)

Iteration 0:  EE criterion = 3.950e-17
Iteration 1:  EE criterion = 1.244e-23

Treatment-effects estimation                    Number of obs     =      4,642
Estimator      : regression adjustment
Outcome model  : Poisson
Treatment model: none
```

bweight	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
mbsmoke (Smoker vs Nonsmoker)	-239.6669	23.83757	-10.05	0.000	-286.3877	-192.9462
POmean						
mbsmoke Nonsmoker	3403.178	9.526006	357.25	0.000	3384.508	3421.849

In this case, using a model that forces outcomes to be nonnegative did not make any substantive difference. In this dataset, nearly 90% of babies weigh at least 2,700 grams, and even the smallest baby weighs 340 grams. When the dependent variable is so large, the predictions from Poisson and linear regression models are remarkably similar.

◁

We now consider models for fractional outcomes. Fractional responses concern outcomes between 0 and 1. These responses may be averaged 0/1 outcomes such as participation rates, or they may be variables that are naturally on a 0 to 1 scale such as pollution levels, patient oxygen saturation, and Gini coefficients (income inequality measures).

Example 6: Modeling fractional outcomes

We will illustrate the use of `teffects ra` with the outcome-model option `fprobit` by using simulated data. The observations are 543 cities at least 200 miles apart. The data contain information about each city's level of industrialization (`industrial`), average annual rainfall in millimeters (`rainfall`), whether or not the city has a metro or train (`train`), and traffic congestion measured by an index (`traffic`).

Our outcome is the level of pollution (`pollution`) measured on a 0 to 1 scale. Values of `pollution` between 0 and 0.3 have no public health implications, but values greater than 0.7 imply that people with breathing or health problems should remain indoors. We study the effect of a tax on gas-guzzler cars (`guzzler`) on air pollution. A tax that is effective in reducing pollution improves public health.

We estimate the ATE of a gas-guzzler tax on pollution, controlling for average yearly rainfall, traffic congestion, the level of industrialization, and whether the city has a train or a metro by using a fractional probit model.

```
. use https://www.stata-press.com/data/r19/pollution
(Simulated Urban Pollution Data)
. teffects ra (pollution rainfall i.traffic industrial i.train, fprobit) (guzzler)
Iteration 0:  EE criterion = 3.023e-16
Iteration 1:  EE criterion = 9.917e-32
Treatment-effects estimation                    Number of obs     =        534
Estimator      : regression adjustment
Outcome model  : fractional probit
Treatment model: none
```

pollution	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
guzzler (tax vs no tax)	-.0960214	.0113896	-8.43	0.000	-.1183447	-.0736981
POmean						
guzzler no tax	.3879346	.0101733	38.13	0.000	.3679952	.407874

The POM if no city were to implement a gas-guzzler tax is an air pollution index of 0.39. If all cities implement a gas-guzzler tax, the air pollution index would decrease by 0.096 relative to a scenario where no city implements the tax.

Video example

Treatment effects: Regression adjustment

Stored results

teffects ra stores the following in e():

Scalars

e(N)	number of observations
e(n*j*)	number of observations for treatment level *j*
e(N_clust)	number of clusters
e(k_eq)	number of equations in e(b)
e(k_levels)	number of levels in treatment variable
e(treated)	level of treatment variable defined as treated
e(control)	level of treatment variable defined as control
e(converged)	1 if converged, 0 otherwise

Macros

e(cmd)	teffects
e(cmdline)	command as typed
e(depvar)	name of outcome variable
e(tvar)	name of treatment variable
e(subcmd)	ra
e(omodel)	linear, logit, probit, hetprobit, poisson, flogit, fprobit, or fhetprobit
e(stat)	statistic estimated, ate, atet, or pomeans
e(wtype)	weight type
e(wexp)	weight expression
e(title)	title in estimation output
e(clustvar)	name of cluster variable
e(tlevels)	levels of treatment variable
e(vce)	*vcetype* specified in vce()
e(vcetype)	title used to label Std. err.
e(properties)	b V
e(estat_cmd)	program used to implement estat
e(predict)	program used to implement predict
e(marginsnotok)	predictions disallowed by margins
e(asbalanced)	factor variables fvset as asbalanced
e(asobserved)	factor variables fvset as asobserved

Matrices

e(b)	coefficient vector
e(V)	variance–covariance matrix of the estimators

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

teffects ra implements a smooth treatment-effects estimator. All smooth treatment-effects estimators are documented in [CAUSAL] **teffects aipw**.

References

Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Cattaneo, M. D. 2010. Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics* 155: 138–154. https://doi.org/10.1016/j.jeconom.2009.09.023.

Drukker, D. M. 2016. Exact matching on discrete covariates is the same as regression adjustment. *The Stata Blog: Not Elsewhere Classified.* https://blog.stata.com/2016/08/16/exact-matching-on-discrete-covariates-is-the-same-as-regression-adjustment/.

Huber, C. 2015. Introduction to treatment effects in Stata: Part 1. *The Stata Blog: Not Elsewhere Classified.* https://blog.stata.com/2015/07/07/introduction-to-treatment-effects-in-stata-part-1/.

Pawitan, Y. 2001. *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford: Oxford University Press.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Also see

[CAUSAL] **teffects postestimation** — Postestimation tools for teffects

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

[ERM] **eregress** — Extended linear regression

[U] 20 Estimation and postestimation commands