Title

stata.com

stteffects ra — Survival-time regression adjustment

Description

`stteffects ra` estimates the average treatment effect (ATE), the average treatment effect on the treated (ATET), and the potential-outcome means (POMs) from observational survival-time data by regression adjustment (RA). RA uses averages of treatment-specific predicted mean survival times to estimate mean survival times for each potential outcome. Contrasts of these predicted mean survival times estimate the treatment effects. `stteffects ra` offers several choices for the model used to predict mean survival time. Binary and multivalued treatments are accommodated.

See [TE] **stteffects intro** for an overview of estimating treatment effects from observational survival-time data.

Quick start

Specify `time` as observed failure time and `fail` as failure indicator

```
stset time, failure(fail)
```

ATE from a Weibull model for `time` on `x1` and `x2` with binary treatment `treat2`

```
stteffects ra (x1 x2) (treat2)
```

As above, but estimate the ATET

```
stteffects ra (x1 x2) (treat2), atet
```

As above, but estimate the potential-outcome means

```
stteffects ra (x1 x2) (treat2), pomeans
```

ATE of `treat2` using a gamma model for `time`

```
stteffects ra (x1 x2, gamma) (treat2)
```

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

```
stteffects ra (x1 x2) (treat3)
```

As above, and specify that `treat3` = 3 is the control level using the value label "MyControl" for 3

```
stteffects ra (x1 x2) (treat3), control("MyControl")
```

Menu

Statistics > Treatment effects > Survival outcomes > Regression adjustment

Syntax

```
steffects ra (omvarlist [ , omoptions ]) (tvar) [ if ] [ in ] [ , stat options ]
```

omvarlist specifies the variables that predict the survival-time variable in the outcome model.

tvar must contain integer values representing the treatment levels.

omoptions	Description
Model	
weibull	Weibull; the default
exponential	exponential
gamma	two-parameter gamma
lnormal	lognormal
ancillary(*avarlist* [, noconstant])	specify variables used to model ancillary parameter
noconstant	suppress constant from outcome model

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
noshow	do not show st setting information
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
iterinit(#)	specify starting-value iterations; 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

You must stset your data before using stteffects; see [ST] stset.

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

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 may be specified using stset; see *Weights* under *Remarks and examples* in [ST] stset. However, weights may not be specified if you are using the bootstrap prefix.

coeflegend does not appear in the dialog box.

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

Options

Model

ancillary(*avarlist* [, noconstant]) specifies the variables used to model the ancillary parameter. By default, the ancillary parameter does not depend on covariates. Specifying ancillary(*avarlist*, noconstant) causes the constant to be suppressed in the model for the ancillary parameter.

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 treatment-model parameters be displayed. By default, the results for these auxiliary parameters are not displayed.

noshow prevents stteffects ra from showing the key st variables. This option is rarely used because most people type stset, show or stset, noshow to permanently set whether they want to see these variables mentioned at the top of the output of every st command; see [ST] stset.

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

`iterinit(#)` specifies the maximum number of iterations used to calculate the starting values. This option is seldom used.

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 the 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 `stteffects` but is not shown in the dialog box:

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

Remarks and examples

stata.com

If you are not familiar with the framework for treatment-effects estimation from observational survival-time data, please see [TE] **stteffects intro**.

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

1. For each treatment level, fit a model of the survival-time outcome on the same set of covariates.
2. Compute the averages of the predicted outcomes for each subject within each treatment level.

These averages estimate the potential-outcome means (POMs). Contrasts of these averages estimate the ATEs. By restricting the computations of the averages to the subset of treated subjects, we obtain estimates of the ATETs.

Here we note only a few entry points to the vast literature on RA estimators. Imbens (2004), Imbens and Wooldridge (2009), Cameron and Trivedi (2005, chap. 25), Wooldridge (2010, chap. 21), and Vittinghoff et al. (2012, chap. 9) provide excellent general introductions to estimating ATEs and to RA estimators in particular.

Like `streg` and other survival-time commands, `stteffects ra` uses the outcome variable and the failure indicator computed by, and optionally weights specified with, `stset`. `stteffects ra` is not appropriate for data with time-varying covariates, also known as multiple-record survival-time data, or for delayed-entry data.

▷ Example 1: Estimating the ATE

Suppose we wish to study the effect of smoking on the time to a second heart attack among women aged 45–55 years. In our fictional `sheart` dataset, `atime` is the observed time in years to a second heart attack or censoring, and `fail` is the 0/1 indicator that a second heart attack was observed. (When `fail` is 1, `atime` records the time to the second heart attack; when `fail` is 0, `atime` records a censored observation of the time to a second heart attack.) We previously `stset` these data; see *A quick tour of the estimators* in [TE] **stteffects intro**.

The treatment, smoking, is stored in the 0/1 indicator `smoke`. These data also contain age at the time of the first heart attack (`age`), and indices of the level of exercise (`exercise`), diet quality (`diet`), and education (`education`) prior to the first heart attack.

We can use `stteffects ra` to estimate the ATE by RA. We model the mean survival time using the default Weibull model, controlling for `age`, `exercise`, `diet`, and `education`, and we specify that `smoke` is the treatment variable.

```
. use https://www.stata-press.com/data/r17/sheart
(Time to second heart attack (fictional))
. stteffects ra (age exercise diet education) (smoke)
        Failure _d: fail
  Analysis time _t: atime
Iteration 0:   EE criterion =  1.006e-14
Iteration 1:   EE criterion =  2.305e-25
Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : regression adjustment
Outcome model  : Weibull
Treatment model: none
Censoring model: none
```

_t	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
smoke (Smoker vs Nonsmoker)	-1.956657	.3331787	-5.87	0.000	-2.609676	-1.303639
POmean						
smoke Nonsmoker	4.243974	.2620538	16.20	0.000	3.730358	4.75759

When every woman smoked in the population of women aged 45–55 years who have had a heart attack, the average time to a second heart attack is estimated to be 1.96 years less than when no women in the population of interest smoked. The estimated average time to a second heart attack when no women in the population of interest smoked is 4.24 years. In other words, if every woman in the population of interest smoked, then the average time to a second heart attack would fall by an estimated 46% relative to the case when no women smoked.

◁

▷ Example 2: Changing the outcome model

Instead of a Weibull model for the outcome model, we could have used an exponential, a gamma, or a lognormal model. By way of comparison, we use a gamma model and the same covariates to estimate the ATE.

```
. stteffects ra (age exercise diet education, gamma) (smoke)
        Failure _d: fail
  Analysis time _t: atime
Iteration 0:   EE criterion =  6.212e-25
Iteration 1:   EE criterion =  7.858e-31
Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : regression adjustment
Outcome model  : gamma
Treatment model: none
Censoring model: none
```

_t	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATE						
smoke (Smoker vs Nonsmoker)	-1.801787	.2924388	-6.16	0.000	-2.374956	-1.228617
POmean						
smoke Nonsmoker	3.994327	.2258257	17.69	0.000	3.551717	4.436937

The estimated ATE of -1.80 and control-level POM of 3.99 are similar to those of -1.96 and 4.24 obtained from the Weibull model in example 1. The ratio of the estimated ATE to the control-level POM indicates a 45% reduction instead of the 46% reduction obtained from the Weibull model.

◁

▷ Example 3: Estimating the ratio of the ATE to the control-level POM

The ratio of the ATE to the control-level POM measures the importance of the effect. In example 1, we computed the point estimate of this ratio from the output, but we were left without a confidence interval. In this example, we use `nlcom` to compute a point estimate and a confidence interval.

Below, we refit the model from example 1, specifying the `coeflegend` option to learn the parameter names. We use the parameter names in `nlcom` to estimate the ratio of the ATE to the control-level POM.

```
. steffects ra (age exercise diet education) (smoke), coeflegend
        Failure _d: fail
  Analysis time _t: atime
Iteration 0:   EE criterion =  1.006e-14
Iteration 1:   EE criterion =  2.305e-25
Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : regression adjustment
Outcome model  : Weibull
Treatment model: none
Censoring model: none
```

_t	Coefficient	Legend
ATE		
smoke (Smoker vs Nonsmoker)	-1.956657	_b[ATE:r1vs0.smoke]
POmean		
smoke Nonsmoker	4.243974	_b[POmean:0.smoke]

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

_t	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
_nl_1	-.4610437	.0598709	-7.70	0.000	-.5783885	-.3436988

The output shows that when every woman smoked, the average time to a second heart attack falls by an estimated 46% relative to the case when no women smoked, as we computed earlier. We also obtain a 95% confidence interval of 34% to 58% for this estimate.

◁

▷ Example 4: Estimating the ATET

Intuitively, the ATET measures the effect of the treatment on an at-risk subpopulation. Sometimes the subpopulation that gets the treatment defines such an at-risk subpopulation. The ATET has the added benefit that it can be estimated under weaker conditions than the ATE; see *Assumptions and trade-offs* in [TE] **steffects intro**.

```
. stteffects ra (age exercise diet education) (smoke), atet
        Failure _d: fail
  Analysis time _t: atime
Iteration 0:   EE criterion =  1.006e-14
Iteration 1:   EE criterion =  2.970e-26
Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : regression adjustment
Outcome model  : Weibull
Treatment model: none
Censoring model: none
```

_t	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATET						
smoke (Smoker vs Nonsmoker)	-1.527476	.2489203	-6.14	0.000	-2.015351	-1.039602
POmean						
smoke Nonsmoker	3.436937	.2217808	15.50	0.000	3.002255	3.87162

When every woman in the subpopulation smoked, the average time to a second heart attack is estimated to be 1.53 years less than when no women in the subpopulation smoked. The estimated average time to a second heart attack when no women in the subpopulation smoked is 3.44 years.

◁

▷ Example 5: Fixed or random censoring time

The time to censoring in survival-time data can be random or deterministic, although it must be independent of treatment assignment and the potential outcomes; see Kalbfleisch and Prentice (2002, chap. 3) for the standard case and see *The correct adjustment for censoring assumption* under *Assumptions and trade-offs* in [TE] **stteffects intro** for the treatment-effects case.

The RA estimator and the likelihood-adjusted-censoring version of the inverse-probability-weighted RA estimator can accommodate a fixed time to censoring; see *The correct adjustment for censoring assumption* in [TE] **stteffects intro**. (The estimators that handle censoring by weighting cannot accommodate a fixed time to censoring because the weights are not well defined with a fixed time to censoring.)

We have fictional data on the time to rearrest among men aged 25–35 who were previously in prison for a felony conviction (`rtime`). The time to censoring is fixed in these data because individuals were followed for a maximum of five years.

Some of the young men chose to enter a vocational training program before release from prison; `train` is 1 for participants and 0 for nonparticipants. The dataset also contains `fail` (which is 1 if the observed time is a failure time and 0 if it is time to censoring), age at the time of the first arrest (`age`), an index of the parents' socioeconomic level (`parental`), and the number of years behind in school at the time of the first arrest (`edeficit`).

We estimate the ATET because we wish to allow the gains from the training program to be related to an unobservable characteristic that affects who self-selects into the program; see *Average treatment effect on the treated* in [TE] **stteffects intro**.

We model the outcome as a function of age, parental, and edeficit.

```
. use https://www.stata-press.com/data/r17/recid2, clear
(Time to rearrest (fictional))
. stteffects ra (age parental edeficit) (train), atet
        Failure _d: fail
  Analysis time _t: rtime
Iteration 0:   EE criterion =  1.875e-23
Iteration 1:   EE criterion =  4.785e-32
Survival treatment-effects estimation           Number of obs     =      2,000
Estimator      : regression adjustment
Outcome model  : Weibull
Treatment model: none
Censoring model: none
```

_t	Coefficient	Robust std. err.	z	P>|z|	[95% conf.	interval]
ATET						
train (Student vs Nonstudent)	2.440919	.4689057	5.21	0.000	1.52188	3.359957
POmean						
train Nonstudent	2.062029	.1231492	16.74	0.000	1.820661	2.303397

When everyone who selected the training got the training, the average time to rearrest is 2.44 years later than the average rearrest time if none of those who selected the training got the training. The average rearrest time if none of those who selected the training got the training is 2.06 years. In other words, the average time to rearrest increases from about 2.06 years to about 4.50 years for the subpopulation of young men who self-selected into the prerelease vocational training program.

◁

Stored results

stteffects 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)	stteffects
e(cmdline)	command as typed
e(dead)	_d
e(depvar)	_t
e(tvar)	name of treatment variable
e(subcmd)	ra
e(omodel)	outcome model: weibull, exponential, gamma, or lognormal
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

The methods and formulas for the RA estimators implemented in stteffects ra are given in *Methods and formulas* of [TE] **stteffects ipwra**.

References

Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.

Bai, X., A. A. Tsiatis, and S. M. O'Brien. 2013. Doubly robust estimators of treatment-specific survival distributions in observational studies with stratified sampling. *Biometrics* 69: 830–839. https://doi.org/10.1111/biom.12076.

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.

Cattaneo, M. D., D. M. Drukker, and A. D. Holland. 2013. Estimation of multivalued treatment effects under conditional independence. *Stata Journal* 13: 407–450.

Guo, S., and M. W. Fraser. 2015. *Propensity Score Analysis: Statistical Methods and Applications*. 2nd ed. Thousand Oaks, CA: SAGE.

Imbens, G. W. 2004. Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics* 86: 4–29. https://doi.org/10.1162/003465304323023651.

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.

Kalbfleisch, J. D., and R. L. Prentice. 2002. *The Statistical Analysis of Failure Time Data*. 2nd ed. New York: Wiley.

Rosenbaum, P. R., and D. B. Rubin. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70: 41–55. https://doi.org/10.2307/2335942.

Rubin, D. B. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66: 688–701. https://doi.org/10.1037/h0037350.

Tsiatis, A. A. 2006. *Semiparametric Theory and Missing Data*. New York: Springer.

Vittinghoff, E., D. V. Glidden, S. C. Shiboski, and C. E. McCulloch. 2012. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. 2nd ed. New York: Springer.

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

Also see

[TE] **stteffects postestimation** — Postestimation tools for stteffects

[TE] **stteffects intro** — Introduction to treatment effects for observational survival-time data

[ST] **streg** — Parametric survival models

[ST] **stset** — Declare data to be survival-time data

[U] 20 Estimation and postestimation commands