Syntax

```
truncreg  depvar  [ indepvars ]  [ if ]  [ in ]  [ weight ]  [,  options ]
```

### options Description

#### Model

- **noconstant**: suppress constant term
- **ll(varname | #)**: lower limit for left-truncation
- **ul(varname | #)**: upper limit for right-truncation
- **offset(varname)**: include `varname` in model with coefficient constrained to 1
- **constraints(constraints)**: apply specified linear constraints
- **collinear**: keep collinear variables

#### SE/Robust

- **vce(vcetype)**: `vcetype` may be oim, robust, cluster `clustvar`, opg, bootstrap, or jackknife

#### Reporting

- **level(#)***: set confidence level; default is `level(95)`
- **noksir**: perform likelihood-ratio test
- **nocnsreport**: do not display constraints
- **display_options**: control 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
- **coeflegend**: display legend instead of statistics

`indepvars` may contain factor variables; see [U] 11.4.3 Factor variables.

`depvar` and `indepvars` may contain time-series operators; see [U] 11.4.4 Time-series varlists.

bootstrap, by, fp, jackknife, mi estimate, rolling, statsby, and svy are allowed; see [U] 11.1.10 Prefix commands.

`vce(bootstrap)` and `vce(jackknife)` are not allowed with the mi estimate prefix; see [MI] mi estimate.

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

aweights are not allowed with the jackknife prefix; see [R] jackknife.

`vce()`, noskis, and weights are not allowed with the svy prefix; see [SVY] svy.

aweights, 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.
truncreg — Truncated regression

Menu
Statistics > Linear models and related > Truncated regression

Description

truncreg fits a regression model of depvar on indepvars from a sample drawn from a restricted part of the population. Under the normality assumption for the whole population, the error terms in the truncated regression model have a truncated normal distribution, which is a normal distribution that has been scaled upward so that the distribution integrates to one over the restricted range.

Options

Model

noconstant; see [R] estimation options.

ll(varname | #) and ul(varname | #) indicate the lower and upper limits for truncation, respectively. You may specify one or both. Observations with depvar \( \leq ll() \) are left-truncated, observations with depvar \( \geq ul() \) are right-truncated, and the remaining observations are not truncated. See [R] tobit for a more detailed description.

offset(varname), constraints(constraints), collinear; see [R] estimation options.

SE/Robust

vce(vcetype) specifies the type of standard error reported, which includes types that are derived from asymptotic theory (oim, opg), 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.

noskip specifies that a full maximum-likelihood model with only a constant for the regression equation be fit. This model is not displayed but is used as the base model to compute a likelihood-ratio test for the model test statistic displayed in the estimation header. By default, the overall model test statistic is an asymptotically equivalent Wald test of all the parameters in the regression equation being zero (except the constant). For many models, this option can substantially increase estimation time.

nocnsreport; see [R] estimation options.

display_options: nomitted, vsquish, noemptycells, baselevels, allbaselevels, nolabel, fvwrap(#), fvwrapon(style), cformat(%fmt), pformat(%fmt), sformat(%fmt), and nolstretch; see [R] estimation options.

Maximization

maximize_options: difficult, technique(algorithm_spec), iterate(#), [no]log, trace, gradient, showstep, hessian, showtolerance, tolerance(#), ltolerance(#), nrtolerance(#), nonrtolerance, and from(init_specs); see [R] maximize. These options are seldom used, but you may use the ltol(#) option to relax the convergence criterion; the default is 1e-6 during specification searches.

Setting the optimization type to technique(bhhh) resets the default vcetype to vce(opg).
The following option is available with `truncreg` but is not shown in the dialog box: `coeflegend`; see [R] `estimation options`.

**Remarks and examples**

Truncated regression fits a model of a dependent variable on independent variables from a restricted part of a population. Truncation is essentially a characteristic of the distribution from which the sample data are drawn. If $x$ has a normal distribution with mean $\mu$ and standard deviation $\sigma$, the density of the truncated normal distribution is

$$f(x | a < x < b) = \frac{f(x)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} = \frac{\frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

where $\phi$ and $\Phi$ are the density and distribution functions of the standard normal distribution.

Compared with the mean of the untruncated variable, the mean of the truncated variable is greater if the truncation is from below, and the mean of the truncated variable is smaller if the truncation is from above. Moreover, truncation reduces the variance compared with the variance in the untruncated distribution.

**Example 1**

We will demonstrate `truncreg` with part of the Mroz dataset distributed with Berndt (1996). This dataset contains 753 observations on women’s labor supply. Our subsample is of 250 observations, with 150 market laborers and 100 nonmarket laborers.

. use http://www.stata-press.com/data/r13/laborsub
. describe

Contains data from http://www.stata-press.com/data/r13/laborsub.dta

obs: 250
vars: 6
size: 1,750

<table>
<thead>
<tr>
<th>variable name</th>
<th>storage</th>
<th>type</th>
<th>format</th>
<th>display</th>
<th>value label</th>
<th>variable label</th>
</tr>
</thead>
<tbody>
<tr>
<td>lfp</td>
<td>byte</td>
<td>%9.0g</td>
<td>label</td>
<td>%9.0g</td>
<td>1 if woman worked in 1975</td>
<td></td>
</tr>
<tr>
<td>whrs</td>
<td>int</td>
<td>%9.0g</td>
<td>label</td>
<td>%9.0g</td>
<td>Wife’s hours of work</td>
<td></td>
</tr>
<tr>
<td>k16</td>
<td>byte</td>
<td>%9.0g</td>
<td>label</td>
<td>%9.0g</td>
<td># of children younger than 6</td>
<td></td>
</tr>
<tr>
<td>k618</td>
<td>byte</td>
<td>%9.0g</td>
<td>label</td>
<td>%9.0g</td>
<td># of children between 6 and 18</td>
<td></td>
</tr>
<tr>
<td>wa</td>
<td>byte</td>
<td>%9.0g</td>
<td>label</td>
<td>%9.0g</td>
<td>Wife’s age</td>
<td></td>
</tr>
<tr>
<td>we</td>
<td>byte</td>
<td>%9.0g</td>
<td>label</td>
<td>%9.0g</td>
<td>Wife’s educational attainment</td>
<td></td>
</tr>
</tbody>
</table>

Sorted by:
We first perform ordinary least-squares estimation on the market laborers.

```
.regress whrs kl6 k618 wa we if whrs > 0
```

```
Source | SS       | df | MS       | Number of obs = 150 
-------+----------+----+----------+--------------------- 
Model   | 7326995.15 | 4  | 1831748.79 | F( 4, 145) = 2.80 
        |          |    |          | Prob > F = 0.0281  
Residual| 94793104.2 | 145| 653745.546 | R-squared = 0.0717 
        |          |    |          | Adj R-squared = 0.0461 
Total   | 102120099 | 149| 685369.794 | Root MSE = 808.55 
```

```
whrs   Coef. Std. Err. t  P>|t|  [95% Conf. Interval] 
-------+-----------------------+-------+-------+-------------------+------------------- 
kl6    -421.4822  167.9734 -2.51 0.013  -753.4748  -89.48953 
k618   -104.4571  54.18616 -1.93 0.056  -211.5538   2.639668 
wa     -4.784917  9.690502 -0.49 0.622  -23.9378  14.36797 
we     9.353195  31.23793  0.30 0.765   71.0937  71.0937 
_cons  1629.817  615.1301  2.65 0.009  414.0371  2845.597 
```

Now we use `truncreg` to perform truncated regression with truncation from below zero.

```
.truncreg whrs kl6 k618 wa we, ll(0)
```

```
Fitting full model:
Iteration 0:  log likelihood = -1205.6992
Iteration 1:  log likelihood = -1200.9873
Iteration 2:  log likelihood = -1200.9159
Iteration 3:  log likelihood = -1200.9157
Iteration 4:  log likelihood = -1200.9157
Truncated regression
Limit:  lower = 0 upper = +inf
Wald chi2(4) = 10.05
Log likelihood = -1200.9157
Prob > chi2 = 0.0395
```

```
whrs   Coef. Std. Err. z  P>|z|  [95% Conf. Interval] 
-------+-----------------------+-------+-------+-------------------+------------------- 
k16    -803.0042  321.3614 -2.50 0.012  -1432.861  -173.1474 
k618  -172.875  88.72898 -1.93 0.051  -346.7806   1.030579 
wa    -4.784917  9.690502 -0.49 0.622  -23.9378  14.36797 
we     9.353195  31.23793  0.30 0.765   71.0937  71.0937 
_cons  1586.26  912.355  1.74 0.082  -201.9233  3374.442 
```

```
/sigma  983.7262  94.44303  10.42 0.000   798.6213  1168.831 
```
If we assume that our data were censored, the tobit model is

```
.tobit whrs k16 k618 wa we, ll(0)
```

Tobit regression

|                | Coef.     | Std. Err.  | t     | P>|t|   | [95% Conf. Interval] |
|----------------|-----------|------------|-------|-------|----------------------|
| whrs           |           |            |       |       |                      |
| k16            | -827.7657 | 214.7407   | -3.85 | 0.000 | -1250.731 to -404.8008 |
| k618           | -140.0192 | 74.22303   | -1.89 | 0.060 | -286.2129 to 6.174547 |
| wa             | -24.97919 | 13.25639   | -1.88 | 0.061 | -51.08969 to 1.131317 |
| we             | 103.6896  | 41.82393   | 2.48  | 0.014 | 21.31093 to 186.0683  |
| _cons          | 589.0001  | 841.5467   | 0.70  | 0.485 | -1068.556 to 2246.556 |
| /sigma         | 1309.909  | 82.7335    |       |       | 1146.953 to 1472.865  |

Obs. summary: 100 left-censored observations at whrs<=0
150 uncensored observations
0 right-censored observations

Technical note

Whether truncated regression is more appropriate than the ordinary least-squares estimation depends on the purpose of that estimation. If we are interested in the mean of wife’s working hours conditional on the subsample of market laborers, least-squares estimation is appropriate. However if we are interested in the mean of wife’s working hours regardless of market or nonmarket labor status, least-squares estimates could be seriously misleading.

Truncation and censoring are different concepts. A sample has been censored if no observations have been systematically excluded but some of the information contained in them has been suppressed. In a truncated distribution, only the part of the distribution above (or below, or between) the truncation points is relevant to our computations. We need to scale it up by the probability that an observation falls in the range that interests us to make the distribution integrate to one. The censored distribution used by tobit, however, is a mixture of discrete and continuous distributions. Instead of rescaling over the observable range, we simply assign the full probability from the censored regions to the censoring points. The truncated regression model is sometimes less well behaved than the tobit model. Davidson and MacKinnon (1993) provide an example where truncation results in more inconsistency than censoring.
**Stored results**

`truncreg` stores the following in `e()`:

Scalars

- `e(N)`: number of observations
- `e(N_bf)`: number of obs. before truncation
- `e(chi2)`: model $\chi^2$
- `e(k_eq)`: number of equations in `e(b)`
- `e(k_eq_model)`: number of equations in overall model test
- `e(k_aux)`: number of auxiliary parameters
- `e(df_m)`: model degrees of freedom
- `e(ll)`: log likelihood
- `e(ll_0)`: log likelihood, constant-only model
- `e(N_clust)`: number of clusters
- `e(sigma)`: estimate of sigma
- `e(p)`: significance
- `e(rank)`: rank of `e(V)`
- `e(ic)`: number of iterations
- `e(rc)`: return code
- `e(converged)`: 1 if converged, 0 otherwise

Macros

- `e(cmd)`: `truncreg`
- `e(cmdline)`: command as typed
- `e(llopt)`: contents of `ll()`, if specified
- `e(ulopt)`: contents of `ul()`, if specified
- `e(depvar)`: name of dependent variable
- `e(wtype)`: weight type
- `e(wexp)`: weight expression
- `e(title)`: title in estimation output
- `e(clustvar)`: name of cluster variable
- `e(offset1)`: offset
- `e(chi2type)`: Wald or LR; type of model $\chi^2$ test
- `e(vce)`: `vcetype` specified in `vce()`
- `e(vcetype)`: title used to label Std. Err.
- `e(opt)`: type of optimization
- `e(which)`: max or min; whether optimizer is to perform maximization or minimization
- `e(ml_method)`: type of ml method
- `e(user)`: name of likelihood-evaluator program
- `e(technique)`: maximization technique
- `e(properties)`: `b  V`
- `e(predict)`: program used to implement `predict`
- `e(asbalanced)`: factor variables `fvset` as asbalanced
- `e(asobserved)`: factor variables `fvset` as asobserved

Matrices

- `e(b)`: coefficient vector
- `e(Cns)`: constraints matrix
- `e(ilog)`: iteration log (up to 20 iterations)
- `e(gradient)`: gradient vector
- `e(V)`: variance–covariance matrix of the estimators
- `e(V_modelbased)`: model-based variance
- `e(means)`: means of independent variables
- `e(dummy)`: indicator for dummy variables

Functions

- `e(sample)`: marks estimation sample

**Methods and formulas**

Greene (2012, 833–839) and Davidson and MacKinnon (1993, 534–537) provide introductions to the truncated regression model.
Let \( y = X\beta + \epsilon \) be the model. \( y \) represents continuous outcomes either observed or not observed. Our model assumes that \( \epsilon \sim N(0, \sigma^2 I) \).

Let \( a \) be the lower limit and \( b \) be the upper limit. The log likelihood is

\[
\ln L = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{j=1}^{n} (y_j - x_j\beta)^2 - \sum_{j=1}^{n} \log \left\{ \Phi\left( \frac{b - x_j\beta}{\sigma} \right) - \Phi\left( \frac{a - x_j\beta}{\sigma} \right) \right\}
\]

This command supports the Huber/White/sandwich estimator of the variance and its clustered version using \textit{vce(robust)} and \textit{vce(cluster clustvar)}, respectively. See [P] \texttt{.robust}, particularly \textit{Maximum likelihood estimators} and \textit{Methods and formulas}.

\texttt{truncreg} also supports estimation with survey data. For details on VCEs with survey data, see [SVY] \texttt{variance estimation}.

References


Also see

[R] \texttt{truncreg postestimation} — Postestimation tools for truncreg

[R] \texttt{regress} — Linear regression

[R] \texttt{tobit} — Tobit regression

[MI] \texttt{estimation} — Estimation commands for use with mi estimate

[SVY] \texttt{svy estimation} — Estimation commands for survey data

[U] 20 Estimation and postestimation commands