

tobit — Tobit regression

Syntax	Menu	Description	Options
Remarks and examples	Stored results	Methods and formulas	References
Also see			

Syntax

`tobit depvar [indepvars] [if] [in] [weight], ll[(#)] ul[(#)] [options]`

<i>options</i>	Description
<hr/>	
Model	
<code>noconstant</code>	suppress constant term
<code>* ll[(#)]</code>	left-censoring limit
<code>* ul[(#)]</code>	right-censoring limit
<code>offset(varname)</code>	include <i>varname</i> in model with coefficient constrained to 1
SE/Robust	
<code>vce(vctype)</code>	<i>vctype</i> may be <code>oim</code> , <code>robust</code> , <code>cluster clustvar</code> , <code>bootstrap</code> , or <code>jackknife</code>
Reporting	
<code>level(#)</code>	set confidence level; default is <code>level(95)</code>
<code>display_options</code>	control column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling
Maximization	
<code>maximize_options</code>	control the maximization process; seldom used
<code>coeflegend</code>	display legend instead of statistics

*You must specify at least one of `ll[(#)]` or `ul[(#)]`.

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`, `nestreg`, `rolling`, `statsby`, `stepwise`, and `svy` are allowed; see [\[U\] 11.1.10 Prefix commands](#).

Weights are not allowed with the `bootstrap` prefix; see [\[R\] bootstrap](#).

`aweight`s are not allowed with the `jackknife` prefix; see [\[R\] jackknife](#).

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

`aweight`s, `fweight`s, `weight`s, and `pweight`s 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.

Menu

Statistics > Linear models and related > Censored regression > Tobit regression

Description

`tobit` fits a model of *depvar* on *indepvars* where the censoring values are fixed.

Options

Model

`noconstant`; see [R] [estimation options](#).

`ll`[(#)] and `ul`[(#)] indicate the lower and upper limits for censoring, respectively. You may specify one or both. Observations with $depvar \leq ll()$ are left-censored; observations with $depvar \geq ul()$ are right-censored; and remaining observations are not censored. You do not have to specify the censoring values at all. It is enough to type `ll`, `ul`, or both. When you do not specify a censoring value, `tobit` assumes that the lower limit is the minimum observed in the data (if `ll` is specified) and the upper limit is the maximum (if `ul` is specified).

`offset(varname)`; see [R] [estimation options](#).

SE/Robust

`vce(vctype)` specifies the type of standard error reported, which includes types that are derived from asymptotic theory (`oim`), 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](#).

`display_options`: `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `nofvlabel`, `fvwrap(#)`, `fvwrapon(style)`, `cformat(%fmt)`, `pformat(%fmt)`, `sformat(%fmt)`, and `no1stretch`; see [R] [estimation options](#).

Maximization

`maximize_options`: `iterate(#)`, `[no]log`, `trace`, `tolerance(#)`, `ltolerance(#)`, `nrtolerance(#)`, and `nonrntolerance`; see [R] [maximize](#). These options are seldom used.

Unlike most maximum likelihood commands, `tobit` defaults to `nolog`—it suppresses the iteration log.

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

`coeflegend`; see [R] [estimation options](#).

Remarks and examples

[stata.com](http://www.stata.com)

Tobit estimation was originally developed by Tobin (1958). A consumer durable was purchased if a consumer's desire was high enough, where desire was measured by the dollar amount spent by the purchaser. If no purchase was made, the measure of desire was censored at zero.

► Example 1: Censored from below

We will demonstrate `tobit` with an artificial example, which in the process will allow us to emphasize the assumptions underlying the estimation. We have a dataset containing the mileage ratings and weights of 74 cars. There are no censored variables in this dataset, but we are going to create one. Before that, however, the relationship between mileage and weight in our complete data is

```
. use http://www.stata-press.com/data/r13/auto
(1978 Automobile Data)

. generate wgt = weight/1000

. regress mpg wgt
```

Source	SS	df	MS	Number of obs = 74		
Model	1591.99024	1	1591.99024	F(1, 72)	=	134.62
Residual	851.469221	72	11.8259614	Prob > F	=	0.0000
				R-squared	=	0.6515
				Adj R-squared	=	0.6467
				Root MSE	=	3.4389

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wgt	-6.008687	.5178782	-11.60	0.000	-7.041058	-4.976316
_cons	39.44028	1.614003	24.44	0.000	36.22283	42.65774

(We divided `weight` by 1,000 simply to make discussing the resulting coefficients easier. We find that each additional 1,000 pounds of weight reduces mileage by 6 mpg.)

`mpg` in our data ranges from 12 to 41. Let us now pretend that our data were censored in the sense that we could not observe a mileage rating below 17 mpg. If the true `mpg` is 17 or less, all we know is that the `mpg` is less than or equal to 17:

```
. replace mpg=17 if mpg<=17
(14 real changes made)

. tobit mpg wgt, ll

Tobit regression
```

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wgt	-6.87305	.7002559	-9.82	0.000	-8.268658	-5.477442
_cons	41.49856	2.05838	20.16	0.000	37.39621	45.6009
/sigma	3.845701	.3663309			3.115605	4.575797

```
Log likelihood = -164.25438
```

	Number of obs	=	74
	LR chi2(1)	=	72.85
	Prob > chi2	=	0.0000
	Pseudo R2	=	0.1815

```
Obs. summary:      18 left-censored observations at mpg<=17
                   56 uncensored observations
                   0 right-censored observations
```

The `replace` before estimation was not really necessary—we remapped all the mileage ratings below 17 to 17 merely to reassure you that `tobit` was not somehow using uncensored data. We typed `ll` after `tobit` to inform `tobit` that the data were left-censored. `tobit` found the minimum of `mpg` in our data and assumed that was the censoring point. We could also have dispensed with `replace` and typed `ll(17)`, informing `tobit` that all values of the dependent variable 17 and below are really censored at 17. In either case, at the bottom of the table, we are informed that there are, as a result, 18 left-censored observations.

On these data, our estimate is now a reduction of 6.9 mpg per 1,000 extra pounds of weight as opposed to 6.0. The parameter reported as `/sigma` is the estimated standard error of the regression; the resulting 3.8 is comparable with the estimated root mean squared error reported by `regress` of 3.4.



□ Technical note

You would never want to throw away information by purposefully censoring variables. The `regress` estimates are in every way preferable to those of `tobit`. Our example is designed solely to illustrate the relationship between `tobit` and `regress`. If you have uncensored data, use `regress`. If your data are censored, you have no choice but to use `tobit`.



▷ Example 2: Censored from above

`tobit` can also fit models that are censored from above. This time, let's assume that we do not observe the actual mileage rating of cars yielding 24 mpg or better—we know only that it is at least 24. (Also assume that we have undone the change to mpg we made in the [previous example](#).)

```
. use http://www.stata-press.com/data/r13/auto, clear
(1978 Automobile Data)
. generate wgt = weight/1000
. regress mpg wgt
(output omitted)
. tobit mpg wgt, ul(24)
```

```
Tobit regression                                Number of obs =          74
                                                LR chi2(1)           =         90.72
                                                Prob > chi2          =         0.0000
                                                Pseudo R2            =         0.2589

Log likelihood = -129.8279
```

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wgt	-5.080645	.43493	-11.68	0.000	-5.947459	-4.213831
_cons	36.08037	1.432056	25.19	0.000	33.22628	38.93445
/sigma	2.385357	.2444604			1.898148	2.872566

```
Obs. summary:          0 left-censored observations
                      51 uncensored observations
                      23 right-censored observations at mpg>=24
```



▷ Example 3: Two-limit tobit model

tobit can also fit models that are censored from both sides (the so-called two-limit tobit):

```
. tobit mpg wgt, ll(17) ul(24)
```

```
Tobit regression                               Number of obs   =          74
                                                LR chi2(1)      =          77.60
                                                Prob > chi2     =          0.0000
Log likelihood = -104.25976                    Pseudo R2       =          0.2712
```

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wgt	-5.764448	.7245417	-7.96	0.000	-7.208457	-4.320438
_cons	38.07469	2.255917	16.88	0.000	33.57865	42.57072
/sigma	2.886337	.3952143			2.098676	3.673998

```
Obs. summary:          18 left-censored observations at mpg<=17
                      33 uncensored observations
                      23 right-censored observations at mpg>=24
```

◀

Stored results

tobit stores the following in `e()`:

Scalars

```
e(N)                number of observations
e(N_unc)            number of uncensored observations
e(N_llc)            number of left-censored observations
e(N_rlc)            number of right-censored observations
e(llopt)            contents of ll(), if specified
e(ulopt)            contents of ul(), if specified
e(k_aux)            number of auxiliary parameters
e(df_m)             model degrees of freedom
e(df_r)             residual degrees of freedom
e(r2_p)             pseudo-R-squared
e(chi2)              $\chi^2$ 
e(ll)               log likelihood
e(ll_0)             log likelihood, constant-only model
e(N_clust)          number of clusters
e(F)                F statistic
e(p)                significance
e(rank)             rank of e(V)
e(converged)        1 if converged, 0 otherwise
```

Macros

e(cmd)	tobit
e(cmdline)	command as typed
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(offset)	linear offset variable
e(chi2type)	LR; type of model χ^2 test
e(vce)	<i>vcetype</i> specified in <code>vce()</code>
e(vcetype)	title used to label Std. Err.
e(properties)	b V
e(predict)	program used to implement <code>predict</code>
e(footnote)	program and arguments to display footnote
e(asbalanced)	factor variables <code>fvset</code> as <code>asbalanced</code>
e(asobserved)	factor variables <code>fvset</code> as <code>asobserved</code>

Matrices

e(b)	coefficient vector
e(V)	variance–covariance matrix of the estimators
e(V_modelbased)	model-based variance

Functions

e(sample)	marks estimation sample
-----------	-------------------------

James Tobin (1918–2002) was an American economist who after education and research at Harvard moved to Yale, where he was on the faculty from 1950 to 1988. He made many outstanding contributions to economics and was awarded the Nobel Prize in 1981 “for his analysis of financial markets and their relations to expenditure decisions, employment, production and prices”. He trained in the U.S. Navy with the writer in Herman Wouk, who later fashioned a character after Tobin in the novel *The Caine Mutiny* (1951): “A mandarin-like midshipman named Tobit, with a domed forehead, measured quiet speech, and a mind like a sponge, was ahead of the field by a spacious percentage.”

Methods and formulas

See [Methods and formulas](#) in [R] [intreg](#).

See [Tobin \(1958\)](#) for the original derivation of the tobit model. An introductory description of the tobit model can be found in, for instance, [Wooldridge \(2013, sec. 17.2\)](#), [Davidson and MacKinnon \(2004, 484–486\)](#), [Long \(1997, 196–210\)](#), and [Maddala and Lahiri \(2006, 333–336\)](#). [Cameron and Trivedi \(2010, chap. 16\)](#) discuss the tobit model using Stata examples.

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

`tobit` also supports estimation with survey data. For details on VCEs with survey data, see [SVY] [variance estimation](#).

References

- Amemiya, T. 1973. Regression analysis when the dependent variable is truncated normal. *Econometrica* 41: 997–1016.
- . 1984. Tobit models: A survey. *Journal of Econometrics* 24: 3–61.
- Burke, W. J. 2009. Fitting and interpreting Cragg’s tobit alternative using Stata. *Stata Journal* 9: 584–592.
- Cameron, A. C., and P. K. Trivedi. 2010. *Microeconometrics Using Stata*. Rev. ed. College Station, TX: Stata Press.
- Cong, R. 2000. `sg144`: Marginal effects of the tobit model. *Stata Technical Bulletin* 56: 27–34. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 189–197. College Station, TX: Stata Press.
- Davidson, R., and J. G. MacKinnon. 2004. *Econometric Theory and Methods*. New York: Oxford University Press.
- Drukker, D. M. 2002. Bootstrapping a conditional moments test for normality after tobit estimation. *Stata Journal* 2: 125–139.
- Goldberger, A. S. 1983. Abnormal selection bias. In *Studies in Econometrics, Time Series, and Multivariate Statistics*, ed. S. Karlin, T. Amemiya, and L. A. Goodman, 67–84. New York: Academic Press.
- Hurd, M. 1979. Estimation in truncated samples when there is heteroscedasticity. *Journal of Econometrics* 11: 247–258.
- Long, J. S. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage.
- Maddala, G. S., and K. Lahiri. 2006. *Introduction to Econometrics*. 4th ed. New York: Wiley.
- McDonald, J. F., and R. A. Moffitt. 1980. The use of tobit analysis. *Review of Economics and Statistics* 62: 318–321.
- Shiller, R. J. 1999. The ET interview: Professor James Tobin. *Econometric Theory* 15: 867–900.
- Stewart, M. B. 1983. On least squares estimation when the dependent variable is grouped. *Review of Economic Studies* 50: 737–753.
- Tobin, J. 1958. Estimation of relationships for limited dependent variables. *Econometrica* 26: 24–36.
- Wooldridge, J. M. 2013. *Introductory Econometrics: A Modern Approach*. 5th ed. Mason, OH: South-Western.

Also see

- [R] **tobit postestimation** — Postestimation tools for tobit
- [R] **heckman** — Heckman selection model
- [R] **intreg** — Interval regression
- [R] **ivtobit** — Tobit model with continuous endogenous regressors
- [R] **regress** — Linear regression
- [R] **truncreg** — Truncated regression
- [SVY] **svy estimation** — Estimation commands for survey data
- [XT] **xtintreg** — Random-effects interval-data regression models
- [XT] **xttobit** — Random-effects tobit models
- [U] **20 Estimation and postestimation commands**