ziologit — Zero-inflated ordered logit regression

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

Description

`ziologit` fits a model for an ordinal outcome with excess zeros, a higher fraction of zeros than would be expected from a standard ordered logit model, also known as zero inflation. This model is known as a zero-inflated ordered logit (ZIOL) model. In the context of ZIOL models, zero is the lowest outcome category. The ZIOL model accounts for the zero inflation by assuming that the zero-valued outcomes come from both a logit model and an ordered logit model, allowing potentially different sets of covariates for each model.

Quick start

Zero-inflated ordered logit model of `y` on `x1` and categorical variable `a` with excess zeros modeled using `x2`

```
ziologit y x1 i.a, inflate(x2)
```

Add offset `x3` to the ordered logit model

```
ziologit y x1 i.a, inflate(x2) offset(x3)
```

Model excess zeros using only a constant

```
ziologit y x1 i.a, inflate(_cons)
```

Model excess zeros with `x2`, and offset `x5` while suppressing the constant term

```
ziologit y x1 i.a, inflate(x2, offset(x5) noconstant)
```

Account for complex sampling design using `svyset` data

```
svy: ziologit y x1 i.a, inflate(x2)
```

Menu

Statistics > Ordinal outcomes > Zero-inflated ordered logit regression

Syntax

```
ziologit depvar [indepvars] [if] [in] [weight] ,
    inflate(varlist[, noconstant offset(varname)] | _cons) [options]
```

options	Description
Model	
* `inflate()`	inflation equation that determines excess zero values
`offset(`*varname*`)`	include *varname* in model with coefficient constrained to 1
`constraints(`*constraints*`)`	apply specified linear constraints
SE/Robust	
`vce(`*vcetype*`)`	*vcetype* may be `oim`, `robust`, `cluster` *clustvar*, `opg`, `bootstrap`, or `jackknife`
Reporting	
`level(`#`)`	set confidence level; default is `level(95)`
`nocnsreport`	do not display constraints
`or`	report odds ratios
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
`collinear`	keep collinear variables
`coeflegend`	display legend instead of statistics

*`inflate(`*varlist*`[, noconstant offset(`*varname*`)] | _cons)` is required.

indepvars and *varlist* may contain factor variables; see **[U] 11.4.3 Factor variables**.

`bayes`, `bayesboot`, `bootstrap`, `by`, `collect`, `fp`, `jackknife`, `rolling`, `statsby`, and `svy` are allowed; see **[U] 11.1.10 Prefix commands**. For more details, see [BAYES] **bayes: ziologit**.

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

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

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

`collinear` and `coeflegend` do not appear in the dialog box.

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

Options

Model

`inflate(`*varlist*`[ , noconstant offset(`*varname*`)] | _cons)` specifies the inflation equation for the logit model that determines the excess zero values; this option is required. Conceptually, omitting `inflate()` would be equivalent to fitting the model with `ologit`; see [R] **ologit**.

`inflate(`*varlist*`[ , noconstant offset(`*varname*`)])` specifies the independent variables in the inflation equation. To suppress the constant in this equation, specify the `noconstant` suboption. You may optionally include an offset for this *varlist*; see `offset(`*varname*`)` in [R] **Estimation options**.

`inflate(_cons)` specifies that the inflation equation contains only an intercept. To run a zero-inflated model of *depvar* with only an intercept in both equations, type `ziologit` *depvar*`, inflate(_cons)`.

`offset(`*varname*`)`, `constraints(`*constraints*`)`; 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(#)`, `nocnsreport`; see [R] **Estimation options**.

`or` reports the estimated coefficients transformed to odds ratios, that is, e^b rather than b. Standard errors and confidence intervals are similarly transformed. This option affects how results are displayed, not how they are estimated. `or` may be specified at estimation or when replaying previously estimated results.

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: `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.

Setting the optimization type to `technique(bhhh)` resets the default *vcetype* to `vce(opg)`.

The following options are available with `ziologit` but are not shown in the dialog box:

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

Remarks and examples

ZIOL models (Kelley and Anderson 2008) are used when the outcome of interest is an ordinal response with a higher fraction of observations in the lowest level than would be expected from a standard ordered logit model. The observations in the lowest category are often referred to as zeros because they typically correspond to the absence of a behavior or trait. Examples include measurements of symptoms, diet, exercise, drug use, and other ordinal outcomes where the lowest level is "none" or "never".

With ordinal outcomes, larger values of the response variable represent higher outcome levels, but the precise numeric value is irrelevant. For consistency, we will refer to the lowest category as zero regardless of whether that outcome takes the number 0 in the dataset.

The presence of a high fraction of zeros, also known as zero inflation, is driven by the fact that the lowest-level observations represent two latent (unobservable) groups: subjects with excess zeros and subjects with conditional zeros. Excess zeros are subjects who are not susceptible to the outcome, whereas conditional zeros are susceptible to higher outcomes even though they currently demonstrate zero response. For example, if the outcome is alcohol consumption, a teetotaler who completely abstains from alcohol would be an excess zero, whereas a drinker who chose not to drink on the day of the survey would be a conditional zero.

To determine susceptibility (that is, to identify excess zeros), ZIOL regression uses a logit model for the inflation equation. For the main regression that determines the ordinal outcome of susceptible subjects, ZIOL uses an ordered logit model. The equation for the main regression is sometimes referred to as the intensity equation. These two equations can have different sets of covariates and different offsets. Parameters can be reported as coefficients or odds ratios with the `or` option. See *Methods and formulas* for more details.

ZIOL models are similar in spirit to zero-inflated ordered probit (ZIOP) models (see [R] **zioprobit**), which use a probit model for the inflation equation and an ordered probit model for the intensity equation. Unlike ZIOP models, the results from ZIOL models can be presented as odds ratios. As such, ZIOL models often appeal to epidemiologists and health researchers, whereas ZIOP models are traditionally used by economists and social scientists. In general, the two types of models tend to produce similar predictions and yield similar inference.

In the context of ZIOL models, the excess-zero or "always-zero" group is often called the nonsusceptible group, and, likewise, the other group is called the susceptible group. In the context of ZIOP models, the terms "nonparticipation" and "participation" are used to refer to the groups. We will use both terminologies interchangeably depending on the context.

The ordered logit model is not nested within the ZIOL model, which means that a likelihood-ratio test cannot be used to check whether a standard ordered logit model is adequate. Instead, one can use information criteria; see example 2 in [R] **zip**.

Example 1: Zero-inflated ordered logit model

We use the fictional data on cigarette consumption from example 1 in [R] **zioprobit**. In that example, we fit a ZIOP model to these data; here we fit a ZIOL model. Both models are appropriate for analyzing this zero-inflated ordinal outcome, but the ZIOP model is more likely to be used by an economist studying consumer purchasing patterns, while the ZIOL model is more likely to be used by an epidemiologist studying smoking behavior.

The outcome of interest, `tobacco`, represents daily cigarette consumption as an ordinal response with four levels: 0 indicates "no cigarettes", and responses 1 through 3 indicate increasing cigarette consumption. More than half of the respondents reported no cigarette consumption, and we suspect that these respondents belong to one of two latent groups. Individuals in the first group are excess zeros, which is to say that they are genuine nonsmokers and thus are not susceptible to cigarette consumption. Individuals in the second group are would-be smokers with no current smoking activity but who might smoke, say, if the price of cigarettes falls or their income increases. These individuals are called conditional zeros because they demonstrate zero cigarette consumption, conditional on being susceptible to smoking. We suspect that these two types of zeros are driven by different patterns of behavior; hence, we choose the ZIOL model over the traditional ordered logit model.

Here the inflation (logit) equation corresponds to the decision to smoke. The intensity (ordered logit) equation corresponds to the level of cigarette consumption by a smoker or would-be smoker. The intensity equation includes covariates for education in years (`education`), annual income in tens of thousands of dollars (`income`), age in decades (`age`), and gender (`female`). The inflation equation includes all the covariates from the intensity equation, plus indicators for whether either of the respondent's parents smoked (`parent`) and whether the respondent's religion discourages smoking (`religion`). Covariates for the intensity equation are specified in the `ziologit` command statement directly after the dependent variable `tobacco`. Covariates for the inflation equation are specified in the required `inflate()` option.

```
. use https://www.stata-press.com/data/r19/tobacco
(Fictional tobacco consumption data)

. ziologit tobacco education income age i.female,
> inflate(education income age i.female i.parent i.religion)

Iteration 0:  Log likelihood = -15877.562  (not concave)
Iteration 1:  Log likelihood = -13082.531  (not concave)
Iteration 2:  Log likelihood = -12316.574  (not concave)
Iteration 3:  Log likelihood = -11686.438  (not concave)
Iteration 4:  Log likelihood = -11214.794
Iteration 5:  Log likelihood = -9857.4914
Iteration 6:  Log likelihood = -8355.1808
Iteration 7:  Log likelihood = -7805.3902
Iteration 8:  Log likelihood = -7658.2562
Iteration 9:  Log likelihood = -7652.5121
Iteration 10: Log likelihood = -7652.4891
Iteration 11: Log likelihood = -7652.4891

Zero-inflated ordered logit regression          Number of obs =  15,000
                                                Wald chi2(4)  = 2143.69
Log likelihood = -7652.4891                     Prob > chi2   =  0.0000
```

tobacco	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
tobacco						
education	.9239328	.0202141	45.71	0.000	.8843139	.9635516
income	1.289223	.0285828	45.10	0.000	1.233202	1.345244
age	-1.393183	.0352799	-39.49	0.000	-1.462331	-1.324036
female						
Female	-.7190566	.0755389	-9.52	0.000	-.8671101	-.5710032
inflate						
education	-.1602718	.0046406	-34.54	0.000	-.1693671	-.1511765
income	-.1922987	.007553	-25.46	0.000	-.2071024	-.177495
age	.313469	.0152296	20.58	0.000	.2836196	.3433184
female						
Female	-.4400725	.0516756	-8.52	0.000	-.5413548	-.3387902
parent						
Smoking	1.282534	.0527674	24.31	0.000	1.179112	1.385956
religion						
Discourag..	-.5387562	.0832958	-6.47	0.000	-.7020129	-.3754995
_cons	2.088609	.0984716	21.21	0.000	1.895609	2.28161
/cut1	5.327041	.1431499			5.046472	5.607609
/cut2	14.6561	.3250778			14.01896	15.29324
/cut3	20.28054	.4446689			19.409	21.15207

The first section of the coefficient table, labeled `tobacco`, corresponds to the tobacco intensity equation and contains coefficients from the ordered logit model fit to susceptible subjects. The second section, labeled `inflate`, corresponds to the inflation equation and contains coefficients from the logit model for susceptibility. ZIOL models allow different sets of covariates in the inflation and intensity equations, but when a variable is included in both equations, it is possible for the coefficients to have opposite signs. The third section of the coefficient table contains the cutpoints from the ordered logit model. See *Remarks and examples* in [R] **ologit** for an explanation of cutpoints.

Compared with the ZIOP model from example 1 in [R] **zioprobit**, the coefficients here are larger in absolute value, but the z scores and p-values are similar. This is consistent with the fact that the logistic distribution used by ZIOL has heavier tails than the normal distribution used by ZIOP. Perhaps the most important difference between these two models is that the estimated coefficients from the ZIOP model are difficult to interpret, but the coefficients from the ZIOL model can be exponentiated and reported as odds ratios with the `or` option.

We replay our results specifying `or`:

```
. ziologit, or
Zero-inflated ordered logit regression                Number of obs =  15,000
                                                      Wald chi2(4)  = 2143.69
Log likelihood = -7652.4891                           Prob > chi2   =  0.0000
```

tobacco	Odds ratio	Std. err.	z	P>|z|	[95% conf.	interval]
tobacco						
education	2.519178	.0509228	45.71	0.000	2.421323	2.620989
income	3.629966	.1037545	45.10	0.000	3.432202	3.839125
age	.2482837	.0087594	-39.49	0.000	.2316957	.2660593
female						
Female	.4872117	.0368034	-9.52	0.000	.420164	.5649584
inflate						
education	.8519122	.0039533	-34.54	0.000	.8441989	.8596959
income	.8250604	.0062317	-25.46	0.000	.8129364	.8373652
age	1.368163	.0208365	20.58	0.000	1.327928	1.409618
female						
Female	.6439897	.0332785	-8.52	0.000	.5819593	.712632
parent						
Smoking	3.605764	.1902668	24.31	0.000	3.251484	3.998647
religion						
Discourag..	.5834735	.0486009	-6.47	0.000	.4955867	.6869461
_cons	8.07368	.7950279	21.21	0.000	6.656599	9.792435
/cut1	5.327041	.1431499			5.046472	5.607609
/cut2	14.6561	.3250778			14.01896	15.29324
/cut3	20.28054	.4446689			19.409	21.15207

```
Note: Estimates are transformed only in the first 2 equations to odds ratios.
Note: _cons estimates baseline odds.
```

Examining the odds ratios from the tobacco intensity equation, we see that a one-unit increase in income, which corresponds to an increase of $10,000 in annual income, raises a smoker's odds of increased cigarette consumption by a factor of 3.63. Looking at the inflation equation, we see that a one-unit increase in income lowers the odds of being susceptible to smoking by a factor of 0.825. This suggests that wealthier individuals are less likely to smoke (consistent with income acting as a proxy for health consciousness), but if they do decide to smoke, they tend to smoke more cigarettes (perhaps because they can afford them).

◁

Stored results

ziologit stores the following in e():

Scalars

e(N)	number of observations
e(N_zero)	number of zeros or lowest-category observations
e(k_cat)	number of categories
e(k)	number of parameters
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(k_dv)	number of dependent variables
e(df_m)	model degrees of freedom
e(ll)	log likelihood
e(N_clust)	number of clusters
e(chi2)	χ^2
e(p)	p-value for model test
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)	ziologit
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(offset1)	offset
e(offset2)	offset for inflate()
e(chi2type)	Wald or LR; type of model χ^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(marginsok)	predictions allowed by margins
e(marginsnotok)	predictions disallowed by margins
e(marginsdefault)	default predict() specification for margins
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(cat)	category values
e(V)	variance–covariance matrix of the estimators
e(V_modelbased)	model-based variance

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 modern class of zero-inflated models was originally developed by Lambert (1992) to address the problem of excess zeros in count data (see [R] **zip**). This model framework has been extended to ordinal outcomes with the ZIOL model by Kelley and Anderson (2008).

The precise numeric value of an ordinal outcome is irrelevant, so without loss of generality we consider an ordinal response variable Y with levels coded as 0, 1, 2, …, H. The first step in the ZIOL model is to determine susceptibility. Let $s_j = 1$ if the jth individual is susceptible to exhibiting a nonzero response, and let $s_j = 0$ if the jth individual is an excess zero. (Similarly, using alternative terminology, let individuals with $s_j = 1$ belong to the participation group and those with $s_j = 0$ belong to the nonparticipation group.) The ZIOL inflation equation uses a logit model to determine the probability of susceptibility (or participation) as

$$\Pr\left(s_j = 1|\mathbf{z}_j\right) = F\left(\mathbf{z}_j\boldsymbol{\gamma}\right) \tag{1}$$

$\mathbf{z}_j$ is a vector of covariates that determines susceptibility, $\boldsymbol{\gamma}$ is a vector of coefficients that have to be estimated, and $F(\cdot)$ is the logistic distribution function: $F(x) = e^x/(1+e^x)$.

Next, conditioning on $s_j = 1$, outcome intensity levels $\tilde{y}_j$ are modeled using an ordered logit model whose levels may also include 0. The corresponding probabilities are given by

$$\Pr\left(\tilde{y}_j = h|s_j = 1,\, \mathbf{x}_j\right) = F\left(\kappa_h - \mathbf{x}_j\boldsymbol{\beta}\right) - F\left(\kappa_{h-1} - \mathbf{x}_j\boldsymbol{\beta}\right) \qquad h = 0, 1, \ldots, H \tag{2}$$

$\mathbf{x}_j$ is a vector of covariates that determine intensity (which can be different from $\mathbf{z}_j$), $\boldsymbol{\beta}$ is a vector of coefficients to be estimated, and cutpoints κ_h are boundary parameters to be estimated (subject to $\kappa_{-1} = -\infty$, $\kappa_H = +\infty$). These cutpoints take the place of an intercept.

The observed response variable is $y_j = s_j\tilde{y}_j$. Thus, a zero outcome occurs when $s_j = 0$ (the individual is an excess zero) or when $s_j = 1$ and $\tilde{y}_j = 0$ (the individual is a conditional zero). To observe a positive y_j, there is a joint requirement that $s_j = 1$ and $\tilde{y}_j > 0$.

The distribution of Y is given by

$$\begin{aligned}\Pr(Y) &= \begin{cases}\Pr(y_j = 0|\mathbf{z}_j, \mathbf{x}_j) \\ \Pr(y_j = h|\mathbf{z}_j, \mathbf{x}_j) \quad h = 1, 2, \ldots, H\end{cases} \\ &= \begin{cases}\Pr(s_j = 0|\mathbf{z}_j) + \Pr(s_j = 1|\mathbf{z}_j)\Pr(\tilde{y}_j = 0|s_j = 1, \mathbf{x}_j) \\ \Pr(s_j = 1|\mathbf{z}_j)\Pr(\tilde{y}_j = h|s_j = 1, \mathbf{x}_j) \qquad\qquad h = 1, 2, \ldots, H\end{cases}\end{aligned} \tag{3}$$

The probability of zero outcome has been inflated because it is the sum of the probability of zero intensity from the ordered logit model and the probability of nonsusceptibility from the logit model.

Substituting (1) and (2) in (3), we get

$$\begin{aligned}\Pr(Y) &= \begin{cases}\Pr(y_j = 0|\mathbf{z}_j, \mathbf{x}_j) \\ \Pr(y_j = h|\mathbf{z}_j, \mathbf{x}_j) \quad h = 1, 2, \ldots, H-1 \\ \Pr(y_j = H|\mathbf{z}_j, \mathbf{x}_j)\end{cases} \\ &= \begin{cases}\left\{1 - F\left(\mathbf{z}_j\boldsymbol{\gamma}\right)\right\} + F(\mathbf{z}_j\boldsymbol{\gamma})\,F(\kappa_0 - \mathbf{x}_j\boldsymbol{\beta}) \\ F(\mathbf{z}_j\boldsymbol{\gamma})\left\{F(\kappa_h - \mathbf{x}_j\boldsymbol{\beta}) - F(\kappa_{h-1} - \mathbf{x}_j\boldsymbol{\beta})\right\} \quad h = 1, 2, \ldots, H-1 \\ F(\mathbf{z}_j\boldsymbol{\gamma})\left\{1 - F(\kappa_{H-1} - \mathbf{x}_j\boldsymbol{\beta})\right\}\end{cases}\end{aligned} \tag{4}$$

If the `offset()` option is specified, $\mathbf{x}_j\boldsymbol{\beta}$ in the intensity equation is replaced with $\mathbf{x}_j\boldsymbol{\beta} + \text{offset}_j^{\beta}$. If the `offset()` suboption is specified in the `inflate()` option, $\mathbf{z}_j\boldsymbol{\gamma}$ in the inflation equation is replaced with $\mathbf{z}_j\boldsymbol{\gamma} + \text{offset}_j^{\gamma}$.

The log-likelihood function is

$$\ln L = \sum_{j=1}^{N} w_j \sum_{h=0}^{H} I\left(y_j = h\right) \ln\left\{\Pr\left(y_j = h|\mathbf{z}_j,\, \mathbf{x}_j\right)\right\}$$

where w_j is an optional weight for the jth observation and

$$I\left(y_j = h\right) = \begin{cases}1 & \text{if } y_j = h \\ 0 & \text{otherwise}\end{cases}$$

The choice between the ZIOL model and the ordered logit model cannot be made using a likelihood-ratio test because the two hypotheses are not nested in the usual sense of parameter restrictions. The inflation effect is removed, and all subjects are deemed susceptible when $\mathbf{z}_j\boldsymbol{\gamma} \to \infty$, a condition that cannot be imposed. To compare the fits of nonnested models, you can use `estat ic` to display information criteria (see [R] **estat ic**).

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

`ziologit` also supports estimation with survey data. For details on VCEs with survey data, see [SVY] **Variance estimation**.

References

Kelley, M. E., and S. J. Anderson. 2008. Zero inflation in ordinal data: Incorporating susceptibility to response through the use of a mixture model. *Statistics in Medicine* 27: 3674–3688. https://doi.org/10.1002/sim.3267.

Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34: 1–14. https://doi.org/10.2307/1269547.

Also see

[R] **ziologit postestimation** — Postestimation tools for ziologit

[R] **logit** — Logistic regression, reporting coefficients

[R] **ologit** — Ordered logistic regression

[R] **zioprobit** — Zero-inflated ordered probit regression

[BAYES] **bayes: ziologit** — Bayesian zero-inflated ordered logit regression

[SVY] **svy estimation** — Estimation commands for survey data

[U] 20 Estimation and postestimation commands

Stata, Stata Press, Mata, NetCourse, and NetCourseNow are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. StataNow is a trademark of StataCorp LLC. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2025 StataCorp LLC, College Station, TX, USA. All rights reserved.

For suggested citations, see the FAQ on citing Stata documentation.