predict after gsem — Generalized linear predictions, etc.

Description Menu Syntax Options
Remarks and examples Reference Also see

Description

predict is a standard postestimation command of Stata. This entry concerns use of predict after gsem. See [SEM] **predict after sem** if you fit your model with sem.

predict after gsem creates new variables containing observation-by-observation values of estimated observed response variables, linear predictions of observed response variables, latent class probabilities, or endogenous or exogenous continuous latent variables.

Menu

Statistics > SEM (structural equation modeling) > Predictions

Syntax

Syntax for predicting observed endogenous outcomes and other statistics

```
predict [type] {stub*|newvarlist} [if] [in] [, statistic options]
```

Syntax for obtaining estimated continuous latent variables and their standard errors

```
predict [type] {stub*|newvarlist} [if] [in], lstatistic [loptions]
```

Syntax for obtaining ML scores

```
predict [type] stub* [if] [in], scores
```

The default is to predict observed endogenous variables with empirical Bayes means predictions of the continuous latent variables. If the model includes a categorical latent variable, the default is class-specific predictions of the observed endogenous variables.

statistic	Description
Main	
mu	expected value of *depvar*; the default
pr	probability (synonym for mu when μ is a probability)
eta	expected value of linear prediction of *depvar*
density	density function at *depvar*
distribution	distribution function at *depvar*
survival	survivor function at *depvar*
expression(*exp*)	calculate prediction using *exp*
classpr	latent class probability
classposteriorpr	posterior latent class probability

options	Description
Main	
conditional(*ctype*)	compute *statistic* conditional on estimated continuous latent variables; default is conditional(ebmeans)
marginal	compute *statistic* marginally with respect to the latent variables
pmarginal	compute mu marginally with respect to the posterior latent class probabilities
nooffset	make calculation ignoring offset or exposure
† outcome(*depvar* [*#*])	specify observed response variable (default all)
* class(*lclspec*)	specify latent class (default all)
Integration	
int_options	integration options

† outcome(*depvar #*) is allowed only if *depvar* has family multinomial, ordinal, or bernoulli. Predicting other generalized responses requires specifying only outcome(*depvar*).
outcome(*depvar #*) may also be specified as outcome(*#.depvar*) or outcome(*depvar ##*).
outcome(*depvar* #3) means the third outcome value. outcome(*depvar* #3) would mean the same as outcome(*depvar* 4) if outcomes were 1, 3, and 4.

* class(*lclspec*) is allowed only for models with categorical latent variables. For models with one categorical latent variable, *lclspec* can be a class value, such as class(2) or its equivalent factor-variable notation class(2.C), assuming the categorical latent variable is C. For models with two or more categorical latent variables, *lclspec* may only be in factor-variable notation, such as class(2.C#1.D) for categorical latent variables C and D.

ctype	Description
ebmeans	empirical Bayes means of latent variables; the default
ebmodes	empirical Bayes modes of latent variables
fixedonly	prediction for the fixed portion of the model only

lstatistic	Description
Main	
latent	empirical Bayes prediction of all latent variables
latent(*varlist*)	empirical Bayes prediction of specified latent variables

loptions	Description
Main	
ebmeans	empirical Bayes means of latent variables; the default
ebmodes	empirical Bayes modes of latent variables
se(*stub** | *newvarlist*)	standard errors of empirical Bayes estimates
Integration	
int_options	integration options

int_options	Description
`intpoints(#)`	use # quadrature points to compute marginal predictions and empirical Bayes means
`iterate(#)`	set maximum number of iterations in computing statistics involving empirical Bayes estimators
`tolerance(#)`	set convergence tolerance for computing statistics involving empirical Bayes estimators

Options

Main

`mu`, the default, calculates the expected value of the outcomes.

`pr` calculates predicted probabilities and is a synonym for `mu`. This option is available only for multinomial, ordinal, and Bernoulli outcomes.

`eta` calculates the fitted linear prediction.

`density` calculates the density function. This prediction is computed using the current values of the observed variables, including the dependent variable.

`distribution` calculates the distribution function. This prediction is computed using the current values of the observed variables, including the dependent variable. This option is not allowed for multinomial outcomes.

`survival` calculates the survivor function. This prediction is computed using the current values of the observed variables, including the dependent variable. This option is only allowed for exponential, gamma, loglogistic, lognormal, and Weibull outcomes.

`expression(`*exp*`)` specifies the prediction as an expression. *exp* is any valid Stata expression, but the expression must contain a call to one of the two special functions unique to this option:

1. `mu(`*outcome*`)`: The `mu()` function specifies the calculation of the mean prediction for *outcome*. If `mu()` is specified without *outcome*, the mean prediction for the first outcome is implied.

   `pr(`*outcome*`)`: The `pr()` function is a synonym for `mu(`*outcome*`)` when *outcome* identifies a multinomial, ordinal, or Bernoulli outcome.

2. `eta(`*outcome*`)`: The `eta()` function specifies the calculation of the linear prediction for *outcome*. If `eta()` is specified without *outcome*, the linear predictor for the first outcome is implied.

   When you specify *exp*, both of these functions may be used repeatedly, in combination, and in combination with other Stata functions and expressions.

`classpr` calculates predicted probabilities for each latent class.

`classposteriorpr` calculates predicted posterior probabilities for each latent class. The posterior probabilities are a function of the latent class predictors and the fitted outcome densities.

`conditional(`*ctype*`)`, `marginal`, and `pmarginal` specify how latent variables are handled in computing *statistic*.

`conditional()` specifies that *statistic* will be computed conditional on specified or estimated continuous latent variables.

conditional(ebmeans), the default, specifies that empirical Bayes means be used as the estimates of the latent variables. These estimates are also known as posterior mean estimates of the latent variables.

conditional(ebmodes) specifies that empirical Bayes modes be used as the estimates of the latent variables. These estimates are also known as posterior mode estimates of the latent variables.

conditional(fixedonly) specifies that all latent variables be set to zero, equivalent to using only the fixed portion of the model.

marginal specifies that the predicted *statistic* be computed marginally with respect to the latent variables.

Although this is not the default, marginal predictions are often very useful in applied analysis. They produce what are commonly called population-averaged estimates. They are also required by margins for models with continuous latent variables.

For models with continuous latent variables, the *statistic* is calculated by integrating the prediction function with respect to all the latent variables over their entire support.

For models with categorical latent variables, mu is the only supported *statistic*. The overall expected value of each outcome is predicted by combining the class-specific expected values using the latent class probabilities.

pmarginal specifies that the overall expected value of each outcome be predicted by combining the class-specific expected values using the posterior latent class probabilities. This option is allowed only with the default *statistic*, mu.

nooffset is relevant only if option offset() or exposure() was specified at estimation time. nooffset specifies that offset() or exposure() be ignored, which produces predictions as if all subjects had equal exposure.

outcome(*depvar* [*#*]) specifies that predictions for *depvar* be calculated. Predictions for all observed response variables are computed by default. If *depvar* is a multinomial or an ordinal outcome, then *#* optionally specifies which outcome level to predict.

class(*lclspec*) specifies that predictions for latent class *lclspec* be calculated. Predictions for all latent classes are computed by default. For models with one categorical latent variable, such as C, *lclspec* can be a class value, such as class(2) or its equivalent factor-variable notation, class(2.C). For models with two or more categorical latent variables, such as C and D, *lclspec* may only be in factor-variable notation, such as class(2.C) or class(2.C#1.D).

latent and latent(*varlist*) specify that the continuous latent variables be estimated using empirical Bayes predictions. By default or if the ebmeans option is specified, empirical Bayes means are computed. With the ebmodes option, empirical Bayes modes are computed.

latent requests empirical Bayes estimates for all latent variables.

latent(*varlist*) requests empirical Bayes estimates for the specified latent variables.

ebmeans specifies that empirical Bayes means be used to predict the latent variables.

ebmodes specifies that empirical Bayes modes be used to predict the latent variables.

se(*stub** | *newvarlist*) calculates standard errors of the empirical Bayes estimators and stores the result in *newvarlist*. This option requires the latent or latent() option.

scores calculates the scores for each coefficient in e(b). This option requires a new variable list of length equal to the number of columns in e(b). Otherwise, use *stub** to have predict generate enumerated variables with prefix *stub*.

Integration

intpoints(*#*) specifies the number of quadrature points used to compute marginal predictions and the empirical Bayes means; the default is the value from estimation.

iterate(*#*) specifies the maximum number of iterations when computing statistics involving empirical Bayes estimators; the default is the value from estimation.

tolerance(*#*) specifies convergence tolerance when computing statistics involving empirical Bayes estimators; the default is the value from estimation.

Remarks and examples

stata.com

Out-of-sample prediction is allowed for all predict options except scores.

predict has two ways of specifying the names of the variables to be created:

```
. predict stub*, ...
```

or

```
. predict firstname secondname ..., ...
```

The first creates variables named *stub*1, *stub*2, ... The second creates variables with names that you specify. We strongly recommend using the *stub** syntax when creating multiple variables because you have no way of knowing the order in which to specify the individual variable names to correspond to the order in which predict will make the calculations. If you use *stub**, the variables will be labeled and you can rename them.

The second syntax is useful when you create one variable and specify outcome(), expression(), class(), or latent().

See [SEM] **Intro 7**, [SEM] **Example 28g**, [SEM] **Example 29g**, [SEM] **Example 50g**, and [SEM] **Example 52g**.

Reference

Skrondal, A., and S. Rabe-Hesketh. 2009. Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society, Series A* 172: 659–687. https://doi.org/10.1111/j.1467-985X.2009.00587.x.

Also see

[SEM] **gsem** — Generalized structural equation model estimation command

[SEM] **gsem postestimation** — Postestimation tools for gsem

[SEM] **Intro 7** — Postestimation tests and predictions

[SEM] **Example 28g** — One-parameter logistic IRT (Rasch) model

[SEM] **Example 29g** — Two-parameter logistic IRT model

[SEM] **Example 50g** — Latent class model

[SEM] **Example 52g** — Latent profile model

[SEM] **Methods and formulas for gsem** — Methods and formulas for gsem

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