

predict after gsem — Generalized linear predictions, etc.

Syntax Remarks and examples	Menu Reference	Description Also see	Options
--------------------------------	-------------------	-------------------------	---------

Syntax

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

```
predict [type] {stub*|newvarlist} [if] [in] [, options_obs_endog options_latent]
```

The default is to predict observed endogenous variables with empirical Bayes means predictions of the latent variables.

<i>options_obs_endog</i>	Description
outcome(<i>depvar</i> [#]) ¹	specify observed response variable (default all)
mu	calculate expected value of <i>depvar</i>
pr	calculate probability (synonym for mu when μ is a probability)
eta	calculate expected value of linear prediction of <i>depvar</i>
nooffset	make calculation ignoring Poisson or nbreg offset or exposure
fixedonly	make calculation setting latent variables to 0 —or—
means	use empirical Bayes means of latent variables; the default
modes	use empirical Bayes modes of latent variables
intpoints(#)	# of integration points; default is intpoints(7) (means only)
tolerance(#)	tolerance; default is tolerance(1.0e-8) (means and modes)
iterate(#)	maximum # of iterations; default is iterate(1001) (means and modes)

¹ outcome(*depvar* #) is allowed only after mlogit, ologit, and oprobit. 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.

<i>options_latent</i>	Description
* <code>latent</code>	calculate empirical Bayes prediction of all latent variables
* <code>latent(varlist)</code>	calculate empirical Bayes prediction of specified latent variables
<code>se(stub* newvarlist)</code>	calculate standard errors
<code>means</code>	use empirical Bayes means of latent variables; the default
<code>modes</code>	use empirical Bayes modes of latent variables
<code>intpoints(#)</code>	# of integration points; default is <code>intpoints(7)</code> (means only)
<code>tolerance(#)</code>	tolerance; default is <code>tolerance(1.0e-8)</code> (means and modes)
<code>iterate(#)</code>	maximum # of iterations; default is <code>iterate(1001)</code> (means and modes)

* Either `latent` or `latent()` must be specified to obtain predictions of latent variables.

Menu

Statistics > SEM (structural equation modeling) > Predictions

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, or endogenous or exogenous latent variables.

Out-of-sample prediction is allowed in three cases:

1. if the prediction does not involve latent variables, or
2. if the prediction involves latent variables, directly or indirectly, option `fixedonly` is specified, or
3. if the prediction involves latent variables, directly or indirectly, the model is multilevel and no observational-level latent variables are involved.

`predict` has two ways of specifying the name(s) of the variable(s) to be created:

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

or

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

The first creates variables named `stub1`, `stub2`, ... The second creates variables named as 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 creating one variable and you specify either `outcome()` or `latent()`.

Options

`outcome(depvar [#])` and `latent[(varlist)]` determine what is to be calculated.

neither specified	predict all observed response variables
<code>outcome(<i>depvar</i> [#])</code> specified	predict specified observed response variable
<code>latent</code> specified	predict all latent variables
<code>latent(<i>varlist</i>)</code> specified	predict specified latent variables

If you are predicting latent variables, both empirical Bayes means and modes are available; see options `means`, `modes`, `intpoints(#)`, `tolerance(#)`, and `iterate(#)` below.

If you are predicting observed response variables, you can obtain $g^{-1}(\mathbf{x}\hat{\beta})$ or $\mathbf{x}\hat{\beta}$; see options `mu` and `eta` below. Predictions can include latent variables or treat them as 0; see option `fixedonly`. If predictions include latent variables, then just as when predicting latent variables, both means and modes are available; see options `means`, `modes`, `intpoints(#)`, `tolerance(#)`, and `iterate(#)`.

`mu` and `pr` specify that $g^{-1}(\mathbf{x}\hat{\beta})$ be calculated, the inverse-link of the expected value of the linear predictions. `x` by default contains predictions of latent variables. `pr` is a synonym for `mu` if response variables are multinomial, ordinal, or Bernoulli. Otherwise, `pr` is not allowed.

`eta` specifies that $\mathbf{x}\hat{\beta}$ be calculated, the expected value of the linear prediction. `x` by default contains predictions of latent variables.

`fixedonly` and `nooffset` are relevant only if observed response variables are being predicted.

`fixedonly` concerns predictions of latent variables used in the prediction of observed response variables. `fixedonly` specifies latent variables be treated as 0, and thus only the fixed-effects part of the model is used to produce the predictions.

`nooffset` is relevant only if option `offset()` or `exposure()` were specified at estimation time. `nooffset` specifies that `offset()` or `exposure()` be ignored, thus producing predictions as if all subjects had equal exposure.

`means`, `modes`, `intpoints(#)`, `tolerance(#)`, and `iterate(#)` specify what predictions of the latent variables are to be calculated.

`means` and `modes` specify that empirical Bayes means or modes be used. Means are the default.

`intpoints(#)` specifies the number of numerical integration points and is relevant only in the calculation of empirical Bayes means. `intpoints()` defaults to the number of integration points specified at estimation time or to `intpoints(7)`.

`tolerance(#)` is relevant for the calculation of empirical Bayes means and modes. It specifies the convergence tolerance. It defaults to the value specified at estimation time with `gsem`'s `adaptopts()` or to `tolerance(1e-8)`.

`iterate(#)` is relevant for the calculation of empirical Bayes means and modes. It specifies the maximum number of iterations to be performed in the calculation of each integral. It defaults to the value specified at estimation time with `gsem`'s `adaptopts()` or to `tolerance(1e-8)`.

Remarks and examples

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

See [\[SEM\] intro 7](#), [\[SEM\] example 28g](#), and [\[SEM\] example 29g](#).

Reference

Skrondal, A., and S. Rabe-Hesketh. 2009. Prediction in multilevel generalized linear models. *JRSSA* 172: 659–687.

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] [methods and formulas for gsem](#) — Methods and formulas