predict after gsem — Generalized linear predictions, etc.

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.

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

1 outcome( depvar #) is allowed only after mlogit, ologit, andoprobit. 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.
### predict after gsem — Generalized linear predictions, etc.

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

* 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()`.
predict after gsem — Generalized linear predictions, etc. 3

Options

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

neither specified                         predict all observed response variables
outcome( depvar [#] ) specified           predict specified observed response variable
latent specified                         predict all latent variables
latent( varlist ) 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 \( g^{-1}(\hat{x}\beta) \) or \( x\hat{\beta} \); 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 \( g^{-1}(\hat{x}\beta) \) 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 \( x\hat{\beta} \) 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

See [SEM] intro 7, [SEM] example 28g, and [SEM] example 29g.
Reference


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