Description

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

predict after sem creates new variables containing observation-by-observation values of estimated factor scores (meaning predicted values of latent variables) and predicted values for latent and observed endogenous variables. Out-of-sample prediction is allowed.

When predict is used on a model fit by sem with the group() option, results are produced with the appropriate group-specific estimates. Out-of-sample prediction is allowed; missing values are filled in for groups not included at the time the model was fit.

predict allows two syntaxes. You can type

    . predict *stub**, ...

to create variables named *stub*1, *stub*2, ..., or you can type

    . predict *var1 var2*..., ...

to create variables named *var1*, *var2*, ...

predict may not be used with summary statistics data.

Menu

Statistics > SEM (structural equation modeling) > Predictions

Syntax

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

predict [*type*] *stub** [*if*] [*in*], <u>sc</u>ores

statistic	Description
xb	linear prediction for all OEn variables; the default
xb(*varlist*)	linear prediction for specified OEn variables
<u>xbl</u>atent	linear prediction for all LEn variables
<u>xbl</u>atent(*varlist*)	linear prediction for specified LEn variables
<u>lat</u>ent	factor scores for all latent variables
<u>lat</u>ent(*varlist*)	factor scores for specified latent variables

Key: OEn = observed endogenous; LEn = latent endogenous

Options

xb calculates the linear prediction for all observed endogenous variables in the model. xb is the default if no option is specified.

xb(*varlist*) calculates the linear prediction for the variables specified, all of which must be observed endogenous variables.

xblatent and xblatent(*varlist*) calculate the linear prediction for all or the specified latent endogenous variables, respectively.

latent and latent(*varlist*) calculate the factor scores for all or the specified latent variables, respectively. The calculation method is an analog of regression scoring; namely, it produces the means of the latent variables conditional on the observed variables used in the model. If missing values are found among the observed variables, conditioning is on the variables with observed values only.

scores is for use by programmers. It provides the first derivative of the observation-level log likelihood with respect to the parameters.

Programmers: In single-group sem, each parameter that is not constrained to be 0 has an associated equation. As a consequence, the number of equations, and hence the number of score variables created by predict, may be large.

Remarks and examples

See [SEM] **Example 14**.

Factor scoring for latent variables can be interpreted as a form of missing-value imputation—think of each latent variable as an observed variable that has only missing values.

When latent variables are present in the model, linear predictions from predict, xb are computed by substituting the factor scores in place of each latent variable before computing the linear combination of coefficients. This method will lead to inconsistent coefficient estimates when the factor score contains measurement error; see Bollen (1989, 305–306).

Reference

Bollen, K. A. 1989. *Structural Equations with Latent Variables*. New York: Wiley. https://doi.org/10.1002/9781118619179.

Also see

[SEM] **Example 14** — Predicted values

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

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