sem — Structural equation model estimation command

Description

sem fits structural equation models. Even when you use the SEM Builder, you are using the sem command.

Menu

Statistics > SEM (structural equation modeling) > Model building and estimation

Syntax

```
sem paths [if] [in] [weight] [ , options ]
```

where `paths` are the paths of the model in command-language path notation; see [SEM] sem and gsem path notation.

**options**

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>model_description_options</code> fully define, along with <code>paths</code>, the model to be fit</td>
</tr>
<tr>
<td><code>group_options</code> fit model for different groups</td>
</tr>
<tr>
<td><code>ssd_options</code> for use with summary statistics data</td>
</tr>
<tr>
<td><code>estimation_options</code> method used to obtain estimation results</td>
</tr>
<tr>
<td><code>reporting_options</code> reporting of estimation results</td>
</tr>
<tr>
<td><code>syntax_options</code> controlling interpretation of syntax</td>
</tr>
</tbody>
</table>

Time-series operators are allowed.

bootstrap, by, collect, jackknife, permute, statsby, and svy are allowed; see [U] 11.1.10 Prefix commands. 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.

Also see [SEM] sem postestimation for features available after estimation.

Options

`model_description_options` describe the model to be fit. The model to be fit is fully specified by `paths`—which appear immediately after `sem`—and the options covariance(), variance(), and means(). See [SEM] sem model description options and [SEM] sem and gsem path notation.
_group_options_ allow the specified model to be fit for different subgroups of the data, with some parameters free to vary across groups and other parameters constrained to be equal across groups. See [SEM] _sem group options_.

:ssd_options allow models to be fit using summary statistics data (SSD), meaning data on means, variances (standard deviations), and covariances (correlations). See [SEM] _sem ssd options_.

_estimation_options_ control how the estimation results are obtained. These options control how the standard errors (VCE) are obtained and control technical issues such as choice of estimation method. See [SEM] _sem estimation options_.

_reporting_options_ control how the results of estimation are displayed. See [SEM] _sem reporting options_.

_syntax_options_ control how the syntax that you type is interpreted. See [SEM] _sem and gsem syntax options_.

### Remarks and examples

_For a readable explanation of what _sem_ can do and how to use it, see any of the intro sections. You might start with [SEM] _Intro 1_.

For examples of _sem_ in action, see any of the example sections. You might start with [SEM] _Example 1_.

For detailed syntax and descriptions, see the references below.

Remarks on three advanced topics are presented under the following headings:

* Default normalization constraints
* Default covariance assumptions
* How to solve convergence problems

#### Default normalization constraints

_sem_ applies the following rules as necessary to identify the model:

1. _means(1: LatentExogenous@0)_
   
   _sem_ constrains all latent exogenous variables to have mean 0. When the _group()_ option is specified, this rule is applied to the first group only.

2. (_LatentEndogenous <- _cons@0)_
   
   _sem_ sets all latent endogenous variables to have intercept 0.

3. (_first <- Latent@1)_
   
   If latent variable _Latent_ is measured by observed endogenous variables, then _sem_ sets the path coefficient of (_first<-Latent_) to be 1; _first_ is the first observed endogenous variable.

4. (_First<-Latent@1)_
   
   If item 3 does not apply—if latent variable _Latent_ is measured by other latent endogenous variables only—then _sem_ sets the path coefficient of _First<-Latent_ to be 1; _First_ is the first latent variable.

The above constraints are applied as needed. Here are the available overrides:

1. To override the normalization constraints, specify your own constraints. Most normalization constraints are added by _sem_ as needed. See _How sem (gsem) solves the problem for you_ under _Identification 2: Normalization constraints (anchoring)_ in [SEM] _Intro 4_.

### Remarks on three advanced topics

1. **Default normalization constraints**
2. **Default covariance assumptions**
3. **How to solve convergence problems**
2. To override means() constraints, you must use the means() option to free the parameter. To override that the mean of latent exogenous variable MyLatent has mean 0, specify the means(MyLatent) option. See [SEM] sem and gsem path notation.

3. To override constrained path coefficients from _cons, such as (LatentEndogenous <- _cons@0), you must explicitly specify the path without a constraint (LatentEndogenous <- _cons). See [SEM] sem and gsem path notation.

Default covariance assumptions

**sem** assumes the following covariance structure:

1. **covstructure(_Ex, unstructured)**
   All exogenous variables (observed and latent) are assumed to be correlated with each other.

2. **covstructure(e._En, diagonal)**
   The error variables associated with all endogenous variables are assumed to be uncorrelated with each other.

You can override these assumptions by

1. Constraining or specifying the relevant covariance to allow e.y and e.Ly to be correlated (specify the covariance(e.y*e.Ly) option); see [SEM] sem model description options.
2. Using the covstructure() option; see [SEM] sem and gsem option covstructure().

How to solve convergence problems

See [SEM] Intro 12.

Stored results

**sem** stores the following in **e()**:  

Scalars
- e(N) number of observations
- e(N_clust) number of clusters
- e(N_groups) number of groups
- e(N_missing) number of missing values in the sample for method(mlmv)
- e(ll) log likelihood of model
- e(df_m) model degrees of freedom
- e(df_b) baseline model degrees of freedom
- e(df_s) saturated model degrees of freedom
- e(chi2_ms) test of target model against saturated model
- e(df_ms) degrees of freedom for e(chi2_ms)
- e(p_ms) p-value for e(chi2_ms)
- e(chi2sb_ms) Satorra–Bentler scaled test of target model against saturated model
- e(psb_ms) p-value for e(chi2sb_ms)
- e(sbc_ms) Satorra–Bentler correction factor for e(chi2sb_ms)
- e(chi2_bs) test of baseline model against saturated model
- e(df_bs) degrees of freedom for e(chi2_bs)
- e(p_bs) p-value for e(chi2_bs)
- e(chi2sb_bs) Satorra–Bentler scaled test of baseline model against saturated model
- e(psb_bs) p-value for e(chi2sb_bs)
- e(sbc_bs) Satorra–Bentler correction factor for e(chi2_bs)
- e(rank) rank of e(V)
- e(ic) number of iterations
- e(rc) return code
sem — Structural equation model estimation command

\[
\begin{align*}
\text{e(converged)} & : 1 \text{ if target model converged, 0 otherwise} \\
\text{e(critvalue)} & : \text{log likelihood or discrepancy of fitted model} \\
\text{e(critvalue_b)} & : \text{log likelihood or discrepancy of baseline model} \\
\text{e(critvalue_s)} & : \text{log likelihood or discrepancy of saturated model} \\
\text{e(modelmeans)} & : 1 \text{ if fitting means and intercepts, 0 otherwise}
\end{align*}
\]

Macros
\[
\begin{align*}
\text{e(cmd)} & : \text{sem} \\
\text{e(cmdline)} & : \text{command as typed} \\
\text{e(data)} & : \text{raw or ssd if SSD were used} \\
\text{e(wtype)} & : \text{weight type} \\
\text{e(wexp)} & : \text{weight expression} \\
\text{e(clustvar)} & : \text{name of cluster variable} \\
\text{e(vce)} & : \text{vcetype specified in vce()} \\
\text{e(vcetype)} & : \text{title used to label Std. err.} \\
\text{e(method)} & : \text{estimation method: ml, mlmv, or adf} \\
\text{e(technique)} & : \text{maximization technique} \\
\text{e(properties)} & : b V \\
\text{e(estat_cmd)} & : \text{program used to implement estat} \\
\text{e(predict)} & : \text{program used to implement predict} \\
\text{e(lyvars)} & : \text{names of latent } y \text{ variables} \\
\text{e(oyvars)} & : \text{names of observed } y \text{ variables} \\
\text{e(lxvars)} & : \text{names of latent } x \text{ variables} \\
\text{e(oxvars)} & : \text{names of observed } x \text{ variables} \\
\text{e(groupvar)} & : \text{name of group variable} \\
\text{e(xconditional)} & : \text{empty ifnoxconditional specified, xconditional otherwise} \\
\text{e(marginsnotok)} & : \text{predictions not allowed by margins} \\
\text{e(marginsdefault)} & : \text{default predict() specification for margins}
\end{align*}
\]

Matrices
\[
\begin{align*}
\text{e(b)} & : \text{parameter vector} \\
\text{e(b_std)} & : \text{standardized parameter vector} \\
\text{e(b_pclass)} & : \text{parameter class} \\
\text{e(Cns)} & : \text{constraints matrix} \\
\text{e(admissible)} & : \text{admissibility of } \Sigma, \Psi, \Phi \\
\text{e(ilog)} & : \text{iteration log (up to 20 iterations)} \\
\text{e(gradient)} & : \text{gradient vector} \\
\text{e(V)} & : \text{covariance matrix of the estimators} \\
\text{e(V_std)} & : \text{standardized covariance matrix of the estimators} \\
\text{e(V_modelbased)} & : \text{model-based variance} \\
\text{e(nobs)} & : \text{vector with number of observations per group} \\
\text{e(groupvalue)} & : \text{vector of group values of e(groupvar)}
\end{align*}
\]

Functions
\[
\begin{align*}
\text{e(sample)} & : \text{marks estimation sample (not with SSD)}
\end{align*}
\]

In addition to the above, the following is stored in \texttt{r()}: 

Matrices
\[
\begin{align*}
\text{r(table)} & : \text{matrix containing the coefficients with their standard errors, test statistics, } p \text{-values, and confidence intervals}
\end{align*}
\]

Note that results stored in \texttt{r()} are updated when the command is replayed and will be replaced when any \texttt{r}-class command is run after the estimation command.

References


Also see

[SEM] Intro 1 — Introduction
[SEM] sem and gsem path notation — Command syntax for path diagrams
[SEM] sem path notation extensions — Command syntax for path diagrams
[SEM] sem model description options — Model description options
[SEM] sem group options — Fitting models on different groups
[SEM] sem ssd options — Options for use with summary statistics data
[SEM] sem estimation options — Options affecting estimation
[SEM] sem reporting options — Options affecting reporting of results
[SEM] sem and gsem syntax options — Options affecting interpretation of syntax
[SEM] sem postestimation — Postestimation tools for sem
[SEM] Methods and formulas for sem — Methods and formulas for sem
[SVY] svy estimation — Estimation commands for survey data
[U] 20 Estimation and postestimation commands