Description

To illustrate crossed models, we use

```
. use http://www.stata-press.com/data/r13/gsem_fifeschool
(School data from Fife, Scotland)
```

Contains data from http://www.stata-press.com/data/r13/gsem_fifeschool.dta

<table>
<thead>
<tr>
<th>variable name</th>
<th>storage type</th>
<th>display format</th>
<th>value label</th>
</tr>
</thead>
<tbody>
<tr>
<td>pid</td>
<td>int</td>
<td>%9.0g</td>
<td>Primary school ID</td>
</tr>
<tr>
<td>sid</td>
<td>byte</td>
<td>%9.0g</td>
<td>Secondary school ID</td>
</tr>
<tr>
<td>attain</td>
<td>byte</td>
<td>%9.0g</td>
<td>Attainment score at age 16</td>
</tr>
<tr>
<td>vrq</td>
<td>int</td>
<td>%9.0g</td>
<td>Verbal-reasoning score from final year of primary school</td>
</tr>
<tr>
<td>sex</td>
<td>byte</td>
<td>%9.0g</td>
<td>1: female; 0: male</td>
</tr>
</tbody>
</table>

Sorted by: pid sid

```
. notes
_dta:
2. Each observation is a different student. Each student attended a primary school (pid) and a secondary school (sid).
3. pid and sid are crossed, not nested. All combinations are possible.
```

Rabe-Hesketh and Skrondal (2012, 443–460) give an introduction to crossed-effects models and provide other examples of crossed-effects models by using the school data from Fife, Scotland.

See _Structural models 1: Linear regression_ and _Multilevel mixed-effects models_ in [SEM] intro 5 for background.

Remarks and examples

Remarks are presented under the following headings:

_The crossed model_

_Fitting the model with the Builder_
The crossed model

In these data, 3,435 students attended 148 different primary schools and 19 different secondary schools. The model we wish to fit is

\[ \text{attain} \sim \varepsilon_1 \text{.sex} \text{.sid}_2 \text{.pid}_1 \]

We include latent (random) effects for primary and secondary school because we think that school identities may have an effect.
The command syntax to fit this model is

```
gsem (attain <- i.sex M1[pid] M2[sid])
```

```
Fitting fixed-effects model:
Iteration 0:  log likelihood = -8701.1372
Iteration 1:  log likelihood = -8701.1372
```

```
Refining starting values:
Grid node 0:  log likelihood = -8587.4875
```

```
Fitting full model:
Iteration 0:  log likelihood = -8587.4875
Iteration 1:  log likelihood = -8573.5619
Iteration 2:  log likelihood = -8562.9985
Iteration 3:  log likelihood = -8561.7425
Iteration 4:  log likelihood = -8561.7354
Iteration 5:  log likelihood = -8561.7354
```

```
Generalized structural equation model               Number of obs = 3435
Log likelihood = -8561.7354
(1) [attain]M2[pid] = 1
(2) [attain]M1[pid] = 1
```

```
|       | Coef.   | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|-------|---------|-----------|-----|-----|----------------------|
| attain <- |         |           |     |     |                      |
| 1.sex    | 0.4986152 | 0.0982634 | 5.07 | 0.000 | 0.3060224 - 0.6912079 |
| M1[pid]  | 1 (constrained) |       |     |     |                      |
| M2[sid]  | 1 (constrained) |       |     |     |                      |
| _cons    | 5.257372 | 0.1813784 | 28.99 | 0.000 | 4.901876 - 5.612867  |
| var(M1[pid]) | 1.104316 | 0.2022595 |     |     | 0.7712451 - 1.581226 |
| var(M2[sid]) | 0.3457086 | 0.1608679 |     |     | 0.1388744 - 0.860937 |
| var(e.attain) | 8.053437 | 0.1990023 |     |     | 7.672694 - 8.453074  |
```

Notes:

1. These data are not nested, but the diagram above would look the same even if they were. The fact that primary and secondary schools are crossed and not nested is, however, specified when we enter the model into the SEM Builder and is implicit in the command syntax.

2. We typed `attain <- i.sex M1[pid] M2[sid]`. We would have typed `attain <- i.sex M1[pid] M2[pid]<sid>` had secondary school been nested within primary school.

3. `gsem` produced the following note when it began estimation: “crossed random effects detected; option intmethod(laplace) assumed”. `gsem` provides four integration methods. The default is `mvaghermite`, which stands for mean-variance adaptive Gauss–Hermite quadrature. The others are `mcaghermite` (mode-curvature adaptive Gauss–Hermite quadrature); `ghermite` (nonadaptive Gauss–Hermite quadrature); and `laplace` (Laplacian approximation).

In general, the adaptive methods `mvaghermite` and `mcaghermite` are considered superior in terms of accuracy to the nonadaptive method `ghermite`, which is considered superior to the approximation method `laplace`. They also take longer.

Fitting crossed models can be difficult. You may specify `intmethod()` with one of the superior methods, but be aware, convergence may not be achieved in finite time.
Fitting the model with the Builder

Use the diagram in *The crossed model* above for reference.

1. Open the dataset.
   
   In the Command window, type
   
   . use http://www.stata-press.com/data/r13/gsem_fifeschool

2. Open a new Builder diagram.
   
   Select menu item Statistics > SEM (structural equation modeling) > Model building and estimation.

3. Put the Builder in gsem mode by clicking on the button.

4. Create the endogenous variable.
   
   a. Select the Add Observed Variable tool, and then click in the diagram about one-third of the way in from the right and halfway down from the top. After adding it, you can click inside the rectangle and move the variable if you wish.
   
   b. In the Contextual Toolbar, select attain with the Variable control.

5. Create the observed exogenous variable.
   
   a. Select the Add Observed Variable tool, and then click in the diagram about one-third of the way in from the left and halfway down from the top.
   
   b. In the Contextual Toolbar, type 1.sex in the Variable control (typing 1.sex rather than using the button to create i.sex prevents the rectangle corresponding to the base category for this binary variable from being created);

6. Create the pid-level latent variable.
   
   a. Select the Add Multilevel Latent Variable tool, and click below the rectangle for attain.
   
   b. In the Contextual Toolbar, click on the button.
   
   c. Select the nesting level and nesting variable by selecting 2 from the Nesting depth control and selecting pid > Observations in the next control.
   
   d. Specify M1 as the Base name.
   
   e. Click on OK.

7. Create the sid-level latent variable.
   
   a. Select the Add Multilevel Latent Variable tool, and click above the rectangle for attain.
   
   b. In the Contextual Toolbar, click on the button.
   
   c. Select the nesting level and nesting variable by selecting 2 from the Nesting depth control and selecting sid > Observations in the next control.
   
   d. Specify M2 as the Base name.
   
   e. Click on OK.
8. Create the paths from the exogenous variables to attain.
   a. Select the Add Path tool, 
   b. Click in the right side of the 1.sex rectangle (it will highlight when you hover over it), and drag a path to the left side of the attain rectangle (it will highlight when you can release to connect the path).
   c. Continuing with the tool, draw paths from bottom of the sid2 double oval to the top of the attain rectangle and from the top of the pid1 double oval to the bottom of the attain rectangle.

9. Clean up the location of the paths.
   If you do not like where the paths have been connected to the rectangles or oval, use the Select tool, , to click on the path, and then simply click on where it connects to a rectangle or oval and drag the endpoint.

10. Estimate.
    Click on the Estimate button, , in the Standard Toolbar, and then click on OK in the resulting GSEM estimation options dialog box.

You can open a completed diagram in the Builder by typing

```
    . wgetsem gsem_cross
```

Reference

Also see

- [SEM] example 38g — Random-intercept and random-slope models (multilevel)
- [SEM] gsem — Generalized structural equation model estimation command
- [SEM] intro 5 — Tour of models