Example 31g — Two-factor measurement model (generalized response)

Description

We demonstrate a two-factor generalized linear measurement model with the same data used in [SEM] Example 29g:

```
. use https://www.stata-press.com/data/r16/gsem_cfa
(Fictional math abilities data)
. describe
Contains data from https://www.stata-press.com/data/r16/gsem_cfa.dta
obs: 500 Fictional math abilities data
vars: 19 21 Mar 2018 10:38
(_dta has notes)

       storage  display value variable name type format label variable label
school   byte    %9.0g School id
id       long    %9.0g Student id
q1       byte    %9.0g result q1 correct
q2       byte    %9.0g result q2 correct
q3       byte    %9.0g result q3 correct
q4       byte    %9.0g result q4 correct
q5       byte    %9.0g result q5 correct
q6       byte    %9.0g result q6 correct
q7       byte    %9.0g result q7 correct
q8       byte    %9.0g result q8 correct
attr1    float   %26.0g agree Skills taught in math class will help me get a better job.
att2     float   %26.0g agree Math is important in everyday life
att3     float   %26.0g agree Working math problems makes me anxious.
att4     float   %26.0g agree Math has always been my worst subject.
att5     float   %26.0g agree I am able to learn new math concepts easily.
test1    byte    %9.0g Score, math test 1
test2    byte    %9.0g Score, math test 2
test3    byte    %9.0g Score, math test 3
test4    byte    %9.0g Score, math test 4
```

Sorted by:
```
. notes
_dta:
1. Fictional data on math ability and attitudes of 500 students from 20 schools.
2. Variables q1-q8 are incorrect/correct (0/1) on individual math questions.
3. Variables att1-att5 are items from a Likert scale measuring each student's attitude toward math.
4. Variables test1-test4 are test scores from tests of four different aspects of mathematical abilities. Range of scores: 0-100.
These data record results from a fictional instrument measuring mathematical ability. Variables \( q_1 \) through \( q_8 \) are the items from the instrument.

In this example, we will also be using variables \( att_1 \) through \( att_5 \). These are five Likert-scale questions measuring each student’s attitude toward math.

See *Multiple-factor measurement models* in [SEM] *Intro 5* for background.

**Remarks and examples**

Remarks are presented under the following headings:

- Fitting the two-factor model
- Fitting the model with the Builder

**Fitting the two-factor model**

We extend the measurement model fit in [SEM] *Example 29g* from one factor, math ability, to two factors, math ability and attitude. The model we wish to fit is:

In this model, mathematical ability affects the correctness of the answers to the items just as previously. The new component, attitude toward mathematics, is correlated with math ability. We expect this correlation to be positive, but that is yet to be determined.

What is important about the attitudinal questions is that the responses are ordinal, that is, the ordering of the possible answers is significant. In other cases, we might have a categorical variable taking on, say, five values; even if the values are 1, 2, 3, 4, and 5, there is no case in which answer 5 is greater than answer 4, answer 4 is greater than answer 3, and so on.
For our attitude measures, however, response 5 signifies strong agreement with a statement and 1 signifies strong disagreement. We handle the ordinal property by specifying that the attitudinal responses are family ordinal, link logit, also known as ordered logit or ordinal logistic regression, and also known in Stata circles as ologit.

In the command language, to fit a one-factor measurement model with math ability, we would type

```
gsem (MathAb -> q1-q8), logit
```

To include the second factor, attitude correlated with math ability, we would type

```
gsem (MathAb -> q1-q8, logit) ///
(MathAtt -> att1-att5, ologit)
```

The covariance between MathAtt and MathAb does not even appear in the command! That is because latent exogenous variables are assumed to be correlated in the command language unless you specify otherwise; in path diagrams, such variables are correlated only if a curved path is drawn between them.

There is another, minor difference in syntax between the one-factor and two-factor models that is worth your attention. Notice that the logit was outside the parentheses in the command to fit the one-factor model, but it is inside the parentheses in the command to fit the two-factor model. Actually, logit could have appeared inside the parentheses to fit the one-factor model. When options appear inside parentheses, they affect only what is specified inside the parentheses. When they appear outside parentheses, they affect all parenthetical specifications.

To obtain the estimates of the two-factor model, we type

```
. gsem (MathAb -> q1-q8, logit) > (MathAtt -> att1-att5, ologit)
```

Fitting fixed-effects model:
Iteration 0:  log likelihood =  -6629.7253
Iteration 1:  log likelihood =  -6628.7848
Iteration 2:  log likelihood =  -6628.7848
Refining starting values:
Grid node 0:  log likelihood =   -6457.4584
Fitting full model:
Iteration 0:  log likelihood =  -6457.4584
Iteration 1:  log likelihood =  -6437.9594
Iteration 2:  log likelihood =  -6400.2731
Iteration 3:  log likelihood =  -6396.3795
Iteration 4:  log likelihood =  -6394.5787
Iteration 5:  log likelihood =  -6394.4019
Iteration 6:  log likelihood =  -6394.3923
Iteration 7:  log likelihood =  -6394.3923
Generalized structural equation model      Number of obs  =      500
Response  :  q1
Family     :  Bernoulli
Link       :  logit
Response  :  q2
Family     :  Bernoulli
Link       :  logit
Response  :  q3
Family     :  Bernoulli
Link       :  logit
Response  :  q4
Family     :  Bernoulli
Link       :  logit
Example 31g — Two-factor measurement model (generalized response)

<table>
<thead>
<tr>
<th>Response</th>
<th>Family</th>
<th>Link</th>
<th>Response</th>
<th>Family</th>
<th>Link</th>
<th>Response</th>
<th>Family</th>
<th>Link</th>
<th>Response</th>
<th>Family</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bernoulli</td>
<td></td>
<td></td>
<td>Bernoulli</td>
<td></td>
<td></td>
<td>Bernoulli</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>logit</td>
<td></td>
<td></td>
<td>logit</td>
<td></td>
<td></td>
<td>logit</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Log likelihood = -6394.3923

( 1) [q1]MathAb = 1
( 2) [att1]MathAtt = 1

|               | Coef.    | Std. Err. | z      | P>|z|    | [95% Conf. Interval] |
|---------------|----------|-----------|--------|--------|----------------------|
| q1 MathAb     | 1        |           | 0.35   | 0.726  | -0.2048845 to 0.2941082 |
|               | _cons    | .0446118  | .1272964 |        |                      |
| q2 MathAb     | .3446081 | .1050264  | 3.28   | 0.001  | .1387601 to .5504562 |
|               | _cons    | -.4572215 | .0979965 | -4.67  | 0.000 to -.6492911 to -.265152 |
| q3 MathAb     | .5445245 | .1386993  | 3.93   | 0.000  | .272679 to .8163701 |
|               | _cons    | .1591406  | .1033116 | 1.54   | 0.123 to -.0433464 to .3616276 |
| q4 MathAb     | .2858874 | .0948553  | 3.01   | 0.003  | .0999743 to .4718004 |
|               | _cons    | -.3196648 | .0947684 | -3.37  | 0.001 to -.5054075 to -.1339222 |
| q5 MathAb     | .8174803 | .1867024  | 4.38   | 0.000  | .4515504 to 1.18341 |
|               | _cons    | -.04543   | .116575 | -0.39  | 0.697 to -.2739127 to .1830527 |
| q6 MathAb     | .6030448 | .1471951  | 4.10   | 0.000  | .3145478 to .8915419 |
|               | _cons    | -.309992  | .1070853 | -2.89  | 0.004 to -.5198753 to -.1001086 |
### Example 31g — Two-factor measurement model (generalized response)

<table>
<thead>
<tr>
<th>q7</th>
<th>MathAb</th>
<th>.72084</th>
<th>.1713095</th>
<th>4.21</th>
<th>0.000</th>
<th>.3850796</th>
<th>1.056601</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>_cons</td>
<td>.1047265</td>
<td>.1116494</td>
<td>0.94</td>
<td>0.348</td>
<td>-.1141023</td>
<td>.3235552</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>q8</th>
<th>MathAb</th>
<th>.5814761</th>
<th>.1426727</th>
<th>4.08</th>
<th>0.000</th>
<th>.3018428</th>
<th>.8611094</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>_cons</td>
<td>-.0250442</td>
<td>.1045134</td>
<td>-0.24</td>
<td>0.811</td>
<td>-.2298868</td>
<td>.1797983</td>
</tr>
</tbody>
</table>

### Variables

- **att1**
  - MathAtt: 1 (constrained)

- **att2**
  - MathAtt: .3788714 | .0971223 | 3.90 | 0.000 | .1885152 | .5692276 |

- **att3**
  - MathAtt: -1.592717 | .3614859 | -4.41 | 0.000 | -2.301216 | -.8842173 |

- **att4**
  - MathAtt: -.8100107 | .153064 | -5.29 | 0.000 | -1.11001 | -.5100108 |

- **att5**
  - MathAtt: .5225423 | .1170141 | 4.47 | 0.000 | .2931988 | .7518858 |

### Model Parameters

- /att1
  - cut1: -1.10254 | .1312272 |-1.359741 | -.8453396 |
  - cut2: -.2495339 | .1160385 | -.479651 | -.0221027 |
  - cut3: .2983261 | .1164414 | .0701052 | .5265471 |
  - cut4: 1.333053 | .1391907 | 1.060244 | 1.605861 |

- /att2
  - cut1: -1.055791 | .1062977 | -1.264131 | -.8474513 |
  - cut2: -.1941211 | .0941435 | -.378639 | -.0096032 |
  - cut3: .3598488 | .0952038 | .1732528 | .5464448 |
  - cut4: 1.132624 | .1082204 | .9205156 | 1.344732 |

- /att3
  - cut1: -1.053519 | .1733999 | -1.393377 | -.7136614 |
  - cut2: -.0491073 | .1442846 | -.3318999 | .2336853 |
  - cut3: .5570671 | .1538702 | .2654871 | .8586471 |
  - cut4: 1.666859 | .2135554 | 1.248298 | 2.08542 |

- /att4
  - cut1: -1.07378 | .1214071 | -1.311734 | -.8358264 |
  - cut2: -.2112462 | .1076501 | -.4222366 | -.0002559 |
  - cut3: .406347 | .1094847 | .191761 | .620933 |
  - cut4: 1.398185 | .1313327 | 1.140778 | 1.655593 |

- /att5
  - cut1: -1.244051 | .1148443 | -1.469142 | -1.018961 |
  - cut2: -.336135 | .0986678 | -.5295203 | -.1427498 |
  - cut3: .2137776 | .0978943 | .0219084 | .4056468 |
  - cut4: .9286849 | .107172 | .7186316 | 1.138738 |

- var(MathAb): 2.300652 | .7479513 | 1.216527 | 4.350909 |
- var(MathAtt): 1.520854 | .4077674 | .8992196 | 2.572228 |
- cov(MathAb, MathAtt): .8837681 | .2204606 | 4.01 | 0.000 | .4516733 | 1.315863 |
Notes:

1. The estimated covariance between math attitude and ability is 0.88.

2. There is something new in the output, namely, things labeled cut1, ..., cut4. These appear for each of the five attitudinal measures. These are the ordered logit’s cutpoints, the values on the logit’s distribution that separate attitude 1 from attitude 2, attitude 2 from attitude 3, and so on. The four cutpoints map the continuous distribution into five ordered, categorical groups.

3. There’s something interesting hiding in the MathAtt coefficients: the coefficients for two of the paths, att3 att4 <- MathAtt, are negative! If you look back to the description of the data, you will find that the sense of these two questions was reversed from those of the other questions. Strong agreement on these two questions was agreement with a negative feeling about mathematics.

`estat sd` displays the fitted variance components as standard deviations and correlations. From the following, we see that the estimated correlation between attitude and ability is 0.4725.

```
estat sd
```

|               | Coef.  | Std. Err. | z     | P>|z|   | [95% Conf. Interval] |
|---------------|--------|-----------|-------|-------|----------------------|
| sd(MathAb)    | 1.517  | .2465573  | 6.11  | 0.000 | 1.008236 2.025568    |
| sd(MathAtt)   | 1.233  | .1653251  | 7.50  | 0.000 | 1.069767 1.446877    |
| corr(MathAb,  | .472   | .0649541  | 7.27  | 0.000 | .3357955 .5896506    |
| MathAtt)      |        |           |       |       |                      |
```

**Fitting the model with the Builder**

Use the diagram in *Fitting the two-factor model* above for reference.

1. Open the dataset.
   In the Command window, type
   ```
   . use https://www.stata-press.com/data/r16/gsem_cfa
   ```

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 measurement component for MathAb.
   Select the Add measurement component tool, ≂, and then click in the diagram about one-third of the way up from the bottom and slightly left of the center.

In the resulting dialog box,

a. change the *Latent variable name* to MathAb;

b. select q1, q2, q3, q4, q5, q6, q7, and q8 by using the *Measurement variables* control;

c. check *Make measurements generalized*;

d. select Bernoulli, Logit in the *Family/Link* control;

e. select Down in the *Measurement direction* control;

f. click on OK.

If you wish, move the component by clicking on any variable and dragging it.
5. Create the measurement component for MathAtt.

Select the Add measurement component tool, \( \text{Add} \), and then click in the diagram about one-third of the way down from the top and slightly left of the center.

In the resulting dialog box,
   a. change the \textit{Latent variable name} to MathAtt;
   b. select att1, att2, att3, att4, and att5 by using the \textit{Measurement variables} control;
   c. check \textit{Make measurements generalized};
   d. select Ordinal, Logit in the \textit{Family/Link} control;
   e. select Up in the \textit{Measurement direction} control;
   f. click on \textit{OK}.

If you wish, move the component by clicking on any variable and dragging it.

6. Create the covariance between MathAtt and MathAb.

   a. Select the Add covariance tool, \( \text{Add} \).
   b. Click in the top-left quadrant of the MathAb oval, and drag a covariance to the bottom left of the MathAtt oval.

7. Clean up.

If you do not like where a covariance has been connected to its variable, use the Select tool, \( \text{Select} \), to simply click on the covariance, and then click on where it connects to an oval and drag the endpoint. You can also change the bow of the covariance by dragging the control point that extends from one end of the selected covariance.

8. Estimate.

Click on the \textbf{Estimate} button, \( \text{Estimate} \), in the Standard Toolbar, and then click on \textit{OK} in the resulting \textit{GSEM estimation options} dialog box.

You can open a completed diagram in the Builder by typing

\texttt{.webgetsem gsem_2fmm}

\textbf{Also see}

[SEM] \textbf{Example 27g} — Single-factor measurement model (generalized response)
[SEM] \textbf{Example 29g} — Two-parameter logistic IRT model
[SEM] \textbf{Example 32g} — Full structural equation model (generalized response)
[SEM] \textbf{Intro 5} — Tour of models
[SEM] \texttt{gsem} — Generalized structural equation model estimation command
[SEM] \texttt{estat sd} — Display variance components as standard deviations and correlations