Example 35g — Ordered probit and ordered logit

Description

Below we demonstrate ordered probit and ordered logit in a measurement-model context. We are not going to illustrate every family/link combination. Ordered probit and logit, however, are unique in that a single equation is able to predict a set of ordered outcomes. The unordered alternative, mlogit, requires $k - 1$ equations to fit $k$ (unordered) outcomes.

To demonstrate ordered probit and ordered logit, we use the following data:

```
. use https://www.stata-press.com/data/r16/gsem_issp93
(Selection from ISSP 1993)
. describe
Contains data from https://www.stata-press.com/data/r16/gsem_issp93.dta
     obs:       871  Selection for ISSP 1993
     vars:        8  21 Mar 2018 16:03
     (_dta has notes)
storage   display  value
variable name   type   format    label
id       int     %9.0g   respondent identifier
y1      byte     %26.0g   agree5 too much science, not enough feelings & faith
y2      byte     %26.0g   agree5 science does more harm than good
y3      byte     %26.0g   agree5 any change makes nature worse
y4      byte     %26.0g   agree5 science will solve environmental problems
sex     byte     %9.0g   sex
age     byte     %9.0g   age (6 categories)
edu     byte     %20.0g   education (6 categories)
```

Sorted by:
. notes
_dta:
2. Full text of y1: We believe too often in science, and not enough in feelings and faith.
3. Full text of y2: Overall, modern science does more harm than good.
4. Full text of y3: Any change humans cause in nature, no matter how scientific, is likely to make things worse.
5. Full text of y4: Modern science will solve our environmental problems with little change to our way of life.

See _Structural models 5: Ordinal models_ in [SEM] _Intro 5_ for background.
Remarks and examples

Remarks are presented under the following headings:

- Ordered probit
- Ordered logit
- Fitting the model with the Builder

Ordered probit

For the measurement model, we focus on variables \( y_1 \) through \( y_4 \). Each variable contains 1–5, with 1 meaning strong disagreement and 5 meaning strong agreement with a statement about science.

Ordered probit produces predictions about the probabilities that a respondent gives response 1, response 2, \ldots, response \( k \). It does this by dividing up the domain of an \( N(0, 1) \) distribution into \( k \) categories defined by \( k - 1 \) cutpoints, \( c_1, c_2, \ldots, c_{k-1} \). Individual respondents are assumed to have a score \( s = X\beta + \epsilon \), where \( \epsilon \sim N(0, 1) \), and then that score is used along with the cutpoints to produce probabilities for each respondent producing response 1, 2, \ldots, \( k \).

\[
\Pr(\text{response is } i \mid X) = \Pr(c_{i-1} < X\beta + \epsilon \leq c_i)
\]

where \( c_0 = -\infty; \ c_k = +\infty \); and \( c_1, c_2, \ldots, c_{k-1} \) and \( \beta \) are parameters of the model to be fit. This ordered probit model has long been known in Stata circles as oprobit.

We have a set of four questions designed to determine the respondent’s attitude toward science, each question with \( k = 5 \) possible answers ranging on a Likert scale from 1 to 5. With ordered probit in hand, we have a way to take a continuous variable, say, a latent variable we will call SciAtt, and produce predicted categorical responses.

The measurement model we want to fit is

![Diagram of ordered probit model]
We fit the model in the command language by typing

```
.gsem (y1 y2 y3 y4 <- SciAtt), oprobit
```

Fitting fixed-effects model:

- Iteration 0: log likelihood = -5227.8743
- Iteration 1: log likelihood = -5227.8743

Refining starting values:

- Grid node 0: log likelihood = -5230.8106

Fitting full model:

- Iteration 0: log likelihood = -5230.8106 (not concave)
- Iteration 1: log likelihood = -5132.1849 (not concave)
- Iteration 2: log likelihood = -5069.5037
- Iteration 3: log likelihood = -5040.4779
- Iteration 4: log likelihood = -5040.2397
- Iteration 5: log likelihood = -5039.8242
- Iteration 6: log likelihood = -5039.823
- Iteration 7: log likelihood = -5039.823

Generalized structural equation model

<table>
<thead>
<tr>
<th>Response</th>
<th>Family</th>
<th>Link</th>
<th>Number of obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>: y1</td>
<td>: ordinal</td>
<td>: probit</td>
<td>= 871</td>
</tr>
<tr>
<td>y2</td>
<td>SciAtt</td>
<td>1.424366 ,2126574</td>
<td>6.70</td>
</tr>
<tr>
<td>y3</td>
<td>SciAtt</td>
<td>1.283359 ,1797557</td>
<td>7.14</td>
</tr>
<tr>
<td>y4</td>
<td>SciAtt</td>
<td>-.0322354,0612282</td>
<td>-0.53</td>
</tr>
</tbody>
</table>

Log likelihood = -5039.823
( 1) [y1]SciAtt = 1

<table>
<thead>
<tr>
<th>y1</th>
<th>y2</th>
<th>y3</th>
<th>y4</th>
<th>/y1</th>
</tr>
</thead>
<tbody>
<tr>
<td>SciAtt</td>
<td>1 (constrained)</td>
<td>SciAtt</td>
<td>1.424366 ,2126574</td>
<td>6.70</td>
</tr>
<tr>
<td>SciAtt</td>
<td>1.283359 ,1797557</td>
<td>7.14</td>
<td>0.000</td>
<td>.931044 1.635674</td>
</tr>
<tr>
<td>SciAtt</td>
<td>-.0322354,0612282</td>
<td>-0.53</td>
<td>0.599</td>
<td>-.1522405 .0877697</td>
</tr>
<tr>
<td>cut1</td>
<td>-1.343148 ,0726927</td>
<td>-1.485623 -1.200673</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cut2</td>
<td>.0084719 ,0521512</td>
<td>-.0937426 .1106863</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cut3</td>
<td>.7876538 ,0595266</td>
<td>.6709837 .9043238</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cut4</td>
<td>1.989985 ,0999181</td>
<td>1.794149 2.18582</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example 35g — Ordered probit and ordered logit

<table>
<thead>
<tr>
<th>( y_2 )</th>
<th>cut1</th>
<th>-1.997245</th>
<th>.1311972</th>
<th>-2.254387</th>
<th>-1.740104</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cut2</td>
<td>-.8240241</td>
<td>.0753839</td>
<td>-.9717738</td>
<td>-.6762743</td>
</tr>
<tr>
<td></td>
<td>cut3</td>
<td>.0547025</td>
<td>.0606036</td>
<td>-.0640784</td>
<td>.1734834</td>
</tr>
<tr>
<td></td>
<td>cut4</td>
<td>1.419923</td>
<td>.1001258</td>
<td>1.22368</td>
<td>1.616166</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( y_3 )</th>
<th>cut1</th>
<th>-1.271915</th>
<th>.0847483</th>
<th>-1.438019</th>
<th>-1.105812</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cut2</td>
<td>.1249493</td>
<td>.0579103</td>
<td>.0114472</td>
<td>.2384515</td>
</tr>
<tr>
<td></td>
<td>cut3</td>
<td>.9752553</td>
<td>.0745052</td>
<td>.8292277</td>
<td>1.121283</td>
</tr>
<tr>
<td></td>
<td>cut4</td>
<td>2.130661</td>
<td>.1257447</td>
<td>1.884206</td>
<td>2.377116</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( y_4 )</th>
<th>cut1</th>
<th>-1.484063</th>
<th>.0646856</th>
<th>-1.610844</th>
<th>-1.357281</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cut2</td>
<td>-.4259356</td>
<td>.0439145</td>
<td>-.5120065</td>
<td>-.3398647</td>
</tr>
<tr>
<td></td>
<td>cut3</td>
<td>.1688777</td>
<td>.0427052</td>
<td>.0851771</td>
<td>.2525782</td>
</tr>
<tr>
<td></td>
<td>cut4</td>
<td>.9413113</td>
<td>.0500906</td>
<td>.8431356</td>
<td>1.039487</td>
</tr>
</tbody>
</table>

var(SciAtt) | .5265523 | .0979611 | .3656637 | .7582305

Notes:

1. The cutpoints \( c_1, \ldots, c_4 \) are labeled \( \text{cut1}, \ldots, \text{cut4} \) in the output. We have a separate cutpoint for each of the four questions \( y_1, \ldots, y_4 \). Look at the estimated cutpoints for \( y_1 \), which are \(-1.343, 0.008, 0.788, \) and 1.99. The probabilities that a person with \( \text{SciAtt} = 0 \) (its mean) would give the various responses are

\[
\begin{align*}
\text{Pr}(\text{response 1}) &= \text{normal}(-1.343) = 0.090 \\
\text{Pr}(\text{response 2}) &= \text{normal}(0.008) - \text{normal}(-1.343) = 0.414 \\
\text{Pr}(\text{response 3}) &= \text{normal}(0.788) - \text{normal}(0.008) = 0.281 \\
\text{Pr}(\text{response 4}) &= \text{normal}(1.99) - \text{normal}(0.788) = 0.192 \\
\text{Pr}(\text{response 5}) &= 1 - \text{normal}(1.99) = 0.023
\end{align*}
\]

2. The path coefficients \( (y_1 \ y_2 \ y_3 \ y_4 \leftarrow \text{SciAtt}) \) measure the effect of the latent variable we called science attitude on each of the responses.

3. The estimated path coefficients are 1, 1.42, 1.28, and \(-0.03 \) for the four questions.

4. If you read the questions—they are listed above—you will find that in all but the fourth question, agreement signifies a negative attitude toward science. Thus \( \text{SciAtt} \) measures a negative attitude toward science because the loadings on negative questions are positive and the loading on the single positive question is negative.

5. The direction of the meanings of latent variables is always a priori indeterminate and is set by the identifying restrictions we apply. We applied—or more correctly, \( \text{gsem} \) applied for us—the constraint that \( y_1 \leftarrow \text{SciAtt} \) has path coefficient 1. Because statement 1 was a negative statement about science, that was sufficient to set the direction of \( \text{SciAtt} \) to be the opposite of what we hoped for.

The direction does not matter. You simply must remember to interpret the latent variable correctly when reading results based on it. In the models we fit, including more complicated models, the signs of the coefficients will work themselves out to adjust for the direction of the variable.
Ordered logit

The description of the ordered logit model is identical to that of the ordered probit model except that where we assumed a normal distribution in our explanation above, we now assume a logit distribution. The distributions are similar.

To fit an ordered logit (ologit) model, the link function shown in the boxes merely changes from probit to logit:

We can fit the model in the command language by typing:

```
. gsem (y1 y2 y3 y4 <- SciAtt), ologit
Fitting fixed-effects model:
Iteration 0:  log likelihood = -5227.8743
Iteration 1:  log likelihood = -5227.8743
Refining starting values:
Grid node 0:  log likelihood = -5127.9026
Fitting full model:
Iteration 0:  log likelihood = -5127.9026 (not concave)
Iteration 1:  log likelihood = -5065.4679
Iteration 2:  log likelihood = -5035.9766
Iteration 3:  log likelihood = -5035.0943
Iteration 4:  log likelihood = -5035.0353
Iteration 5:  log likelihood = -5035.0352
Generalized structural equation model Number of obs = 871
Response : y1
Family : ordinal
Link : logit
Response : y2
Family : ordinal
Link : logit
Response : y3
Family : ordinal
Link : logit
Response : y4
Family : ordinal
Link : logit
Log likelihood = -5035.0352
( 1) [y1]SciAtt = 1
```
|       | Coef.  | Std. Err. | z     | P>|z| | [95% Conf. Interval] |
|-------|--------|-----------|-------|------|----------------------|
| y1    | SciAtt | 1 (constrained) |      |      |                      |
| y2    | SciAtt | 1.394767  | .2065479 | 6.75 | 0.000 | .9899406  | 1.799593 |
| y3    | SciAtt | 1.29383   | .1845113 | 7.01 | 0.000 | .9321939  | 1.655465 |
| y4    | SciAtt | -.0412446 | .0619936 | -0.67| 0.506 | -.1627498 | .0802606 |
| y1    | cut1   | -2.38274  | .1394292  |      |      | -2.656016  | -2.109464 |
|       | cut2   | -.0088393 | .08197918 |      |      | -.1832207  | .1655422  |
|       | cut3   | 1.326292  | .106275   |      |      | 1.117997   | 1.534587  |
|       | cut4   | 3.522017  | .1955535  |      |      | 3.138739   | 3.905295  |
| y2    | cut1   | -3.51417  | .2426595  |      |      | -3.989774  | -3.038566 |
|       | cut2   | -1.421711 | .135695   |      |      | -1.687669  | -1.155754 |
|       | cut3   | .0963154  | .1046839  |      |      | -.1086612  | .3014921  |
|       | cut4   | 2.491459  | .1840433  |      |      | 2.130741   | 2.852178  |
| y3    | cut1   | -2.263557 | .1618806  |      |      | -2.580838  | -1.946277 |
|       | cut2   | .2024798  | .1012122  |      |      | .0041075   | .400852   |
|       | cut3   | 1.695997  | .1393606  |      |      | 1.422855   | 1.969138  |
|       | cut4   | 3.828154  | .2464566  |      |      | 3.345108   | 4.3112    |
| y4    | cut1   | -2.606013 | .1338801  |      |      | -2.868413  | -2.343613 |
|       | cut2   | -.6866199 | .0718998  |      |      | -.8275369  | -.5456949 |
|       | cut3   | .268862   | .0684577  |      |      | .1346874   | .4030366  |
|       | cut4   | 1.561921  | .0895438  |      |      | 1.386419   | 1.737424  |
| var(SciAtt) | 1.715641 | .3207998 |      |      | 1.189226  | 2.475077  |

Note:
1. Results are nearly identical to those reported for ordered probit.

### Fitting the model with the Builder

Use the diagram in *Ordered probit* above for reference.

1. Open the dataset.
   
   In the Command window, type
   
   ```
   . use https://www.stata-press.com/data/r16/gsem_issp93
   ```

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 SciAtt.

Select the Add measurement component tool, 😮, 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 Latent variable name to SciAtt;
   b. select y1, y2, y3, and y4 by using the Measurement variables control;
   c. check Make measurements generalized;
   d. select Ordinal, Probit 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. Estimate.

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

6. To fit the model in Ordered logit, change the type of generalized response for each of the measurement variables.

   a. Choose the Select tool, 🖍.
   b. Click on the y1 rectangle. In the Contextual Toolbar, select Ordinal, Logit in the Family/Link control.
   c. Repeat this process to change the family and link to Ordinal, Logit for y2, y3, and y4.

7. Estimate again.

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 for the ordered probit model in the Builder by typing

    . webgetsem gsem_oprobit

You can open a completed diagram for the ordered logit model in the Builder by typing

    . webgetsem gsem_ologit

Reference

Also see

[SEM] **Example 1** — Single-factor measurement model

[SEM] **Example 27g** — Single-factor measurement model (generalized response)

[SEM] **Example 33g** — Logistic regression

[SEM] **Example 36g** — MIMIC model (generalized response)

[SEM] **Example 37g** — Multinomial logistic regression

[SEM] **Intro 5** — Tour of models

[SEM] **gsem** — Generalized structural equation model estimation command