Example 34g — Combined models (generalized responses)

Description

We demonstrate how to fit a combined model with one Poisson regression and one logit regression by using the following data:

```
. use https://www.stata-press.com/data/r17/gsem_lbw
(Hosmer & Lemeshow data)
. describe
Contains data from https://www.stata-press.com/data/r17/gsem_lbw.dta
Observations: 189 Hosmer & Lemeshow data
Variables: 11 21 Mar 2020 12:28
=_dta has notes
```

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Storage type</th>
<th>Display format</th>
<th>Value label</th>
</tr>
</thead>
<tbody>
<tr>
<td>id</td>
<td>int</td>
<td>%8.0g</td>
<td>Subject ID</td>
</tr>
<tr>
<td>low</td>
<td>byte</td>
<td>%8.0g</td>
<td>Birthweight &lt; 2500g</td>
</tr>
<tr>
<td>age</td>
<td>byte</td>
<td>%8.0g</td>
<td>Age of mother</td>
</tr>
<tr>
<td>lwt</td>
<td>int</td>
<td>%8.0g</td>
<td>Weight, last menstrual period</td>
</tr>
<tr>
<td>race</td>
<td>byte</td>
<td>%8.0g</td>
<td>Race</td>
</tr>
<tr>
<td>smoke</td>
<td>byte</td>
<td>%9.0g</td>
<td>Smoked during pregnancy</td>
</tr>
<tr>
<td>ptl</td>
<td>byte</td>
<td>%8.0g</td>
<td>Premature labor history (count)</td>
</tr>
<tr>
<td>ht</td>
<td>byte</td>
<td>%8.0g</td>
<td>Has history of hypertension</td>
</tr>
<tr>
<td>ui</td>
<td>byte</td>
<td>%8.0g</td>
<td>Presence, uterine irritability</td>
</tr>
<tr>
<td>ftv</td>
<td>byte</td>
<td>%8.0g</td>
<td># physician visits, 1st trimester</td>
</tr>
<tr>
<td>bwt</td>
<td>int</td>
<td>%8.0g</td>
<td>Birthweight (g)</td>
</tr>
</tbody>
</table>

Sorted by:
```
. notes
_dta:
2. Data from a study of risk factors associated with low birthweights.
```

See *Structural models 8: Dependencies between response variables* in [SEM] *Intro 5* for background.

Remarks and examples

Remarks are presented under the following headings:

- **Fitting the combined model**
- **Obtaining odds ratios and incidence-rate ratios**
- **Fitting the model with the Builder**
Fitting the combined model

The model we wish to fit is

This model has one logit equation and one Poisson regression equation, with the Poisson response variable also being an explanatory variable in the logit equation.

Because the two equations are recursive, it is not necessary to fit these models together. We could draw separate diagrams for each equation and fit each separately. Even so, many researchers often do fit recursive models together, and sometimes, it is just the first step before placing constraints across models or introducing a common latent variable. The latter might be likely in this case because neither generalized linear response has an error that could be correlated and so the only way to correlate these two responses in gsem is to add a shared latent variable affecting each.

Our purpose here is to show that you can mix models with generalized response variables of different types.
To fit the model in the command language, we type

```
.gsem (low <- ptl age smoke ht lwt i.race ui, logit)
> (ptl <- age smoke ht, poisson)
```

Iteration 0:  log likelihood = -322.96738  
Iteration 1:  log likelihood = -200.5818  
Iteration 2:  log likelihood = -198.58086
Iteration 3:  log likelihood = -198.56179
Iteration 4:  log likelihood = -198.56178

Generalized structural equation model  
Number of obs = 189

Response:  low
Family:  Bernoulli
Link:  Logit

Response:  ptl
Family:  Poisson
Link:  Log

Log likelihood = -198.56178

| Coefficient  | Std. err.  | z   | P>|z|   | [95% conf. interval] |
|--------------|------------|-----|-------|---------------------|
| **low**      |            |     |       |                     |
| ptl          | .5418366   | .346249 | 1.56 | 0.118   | -1.136799 - 1.220472  |
| age          | -.0271003  | .0364504 | -0.74 | 0.457   | 0.137739 | 1.708951         |
| smoke        | .9233448   | .4008266 | 2.30 | 0.021   | .137739 | 1.708951         |
| ht           | 1.832518   | .6916292 | 2.65 | 0.008   | .4769494 | 3.188086         |
| lwt          | -.0151508  | .0069259 | -2.19 | 0.029   | -.0287253 |-0.0015763       |
| race Black   | 1.262647   | .5264101 | 2.40 | 0.016   | .2309023 | 2.294392         |
| Other        | .8620791   | .4391532 | 1.96 | 0.050   | .0013548 | 1.722803         |
| ui           | .7585135   | .4593768 | 1.65 | 0.099   | -1.418484 | 1.658875         |
| _cons        | .4612238   | 1.20459  | 0.38 | 0.702   | -1.899729 | 2.822176         |
| **ptl**      |            |     |       |                     |
| age          | .0370598   | .0298752 | 1.24 | 0.215   | 0.0214946 | .0956142         |
| smoke        | .9602534   | .3396867 | 2.83 | 0.005   | .2944796 | 1.626027         |
| ht           | -.1853501  | .7271851 | -0.25 | 0.799   | 1.610607 | 1.239906         |
| _cons        | -2.985512  | .7842174 | -3.81 | 0.000   | -4.52255 | -1.448474         |

**Obtaining odds ratios and incidence-rate ratios**

As mentioned in [SEM] Example 33g, some researchers prefer to see exponentiated coefficients. In both odds ratios and incidence-rate ratios, exponentiation is meaningful. Exponentiated logit coefficients are odds ratios, and exponentiated Poisson regression coefficients are incidence-rate ratios. To obtain exponentiated coefficients for both equations, we type
Had we merely typed `estat eform` without the two equation names, we would have obtained exponentiated coefficients for the first equation only.

Equation names are easily found on the output or the path diagrams. Equations are named after the dependent variable.

### Fitting the model with the Builder

Use the diagram in Fitting the combined model above for reference.

1. Open the dataset.
   
   In the Command window, type
   ```
   . use https://www.stata-press.com/data/r17/gsem_lbw
   ```
   
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 independent variables.
   
   Select the Add observed variables set tool, 🔍, and then click at the bottom of the diagram about one-third of the way in from the left.

   In the resulting dialog box,
   a. select the Select variables radio button (it may already be selected);
   b. use the Variables control to select the variables age, smoke, ht, and lwt in this order;
   c. include the levels of the factor variable race by clicking on the ∈ button next to the Variables control. In the resulting dialog box, select the Factor variable radio button, select Main effect in the Specification control, and select race in the Variables control for Variable 1. Click on Add to varlist, and then click on OK;
d. continue with the Variables control and select the variable ui;
e. select Vertical in the Orientation control;
f. click on OK.

If you wish, move the set of variables by clicking on any variable and dragging it.

5. Create the generalized response for premature labor history.
   a. Select the Add generalized response variable tool, □.
   b. Click about one-third of the way in from the right side of the diagram, to the right of ht.
   c. In the Contextual Toolbar, select Poisson, Log in the Family/Link control.
   d. In the Contextual Toolbar, select ptl in the Variable control.

   a. Select the Add generalized response variable tool, □.
   b. Click about one-third of the way in from the right side of the diagram, to the right of 2.race.
   c. In the Contextual Toolbar, select Bernoulli, Logit in the Family/Link control.
   d. In the Contextual Toolbar, select low in the Variable control.

7. Create paths from the independent variables to the dependent variables.
   a. Select the Add path tool, -.
   b. Click in the right side of the age rectangle (it will highlight when you hover over it), and drag a path to the left side of the ptl rectangle (it will highlight when you can release to connect the path).
   c. Continuing with the - tool, create the following paths by clicking first in the right side of the rectangle for the independent variable and dragging it to the left side of the rectangle for the dependent variable:

   \[
   \begin{align*}
   \text{smoke} & \rightarrow \text{ptl} \\
   \text{ht} & \rightarrow \text{ptl} \\
   \text{age} & \rightarrow \text{low} \\
   \text{smoke} & \rightarrow \text{low} \\
   \text{ht} & \rightarrow \text{low} \\
   \text{lwt} & \rightarrow \text{low} \\
   \text{1b.race} & \rightarrow \text{low} \\
   \text{2.race} & \rightarrow \text{low} \\
   \text{3.race} & \rightarrow \text{low} \\
   \text{ui} & \rightarrow \text{low}
   \end{align*}
   \]
   d. Continuing with the - tool, create the path from ptl to low by clicking in the bottom of the ptl rectangle and dragging the path to the top of the low rectangle.
8. Clean up.
   If you do not like where a path has been connected to its variables, use the Select tool, 
   to click on the path, and then simply click on where it connects to a rectangle and drag the 
   endpoint.

   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
   . webgetsem gsem_comb

Reference
NJ: Wiley.

Also see
[SEM] Example 33g — Logistic regression
[SEM] Example 45g — Heckman selection model
[SEM] Example 46g — Endogenous treatment-effects model
[SEM] Intro 5 — Tour of models
[SEM] gsem — Generalized structural equation model estimation command
[SEM] estat eform — Display exponentiated coefficients