Description

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

```
. use http://www.stata-press.com/data/r15/gsem_lbw
(Hosmer & Lemeshow data)
. describe
Contains data from http://www.stata-press.com/data/r15/gsem_lbw.dta
    obs:        189  Hosmer & Lemeshow data
    vars:       11  21 Mar 2016 12:28
    size:   2,646   (_dta has notes)
```

```
+------------------------------------------------------------------+
<table>
<thead>
<tr>
<th>variable</th>
<th>type</th>
<th>format</th>
<th>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>birth weight &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>birth weight (g)</td>
</tr>
</tbody>
</table>
+------------------------------------------------------------------+
```

Sorted by:

`. notes
_dta:
2. Data from a study of risk factors associated with low birth weights.

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

1
Fitting the combined model

The model we wish to fit is

\[
\begin{align*}
\text{age} & \rightarrow \text{Poisson} \\
\text{smoke} & \rightarrow \text{Poisson} \\
\text{ht} & \rightarrow \text{Poisson} \\
\text{lwt} & \rightarrow \text{Poisson} \\
\text{1b.race} & \rightarrow \text{Bernoulli} \\
\text{2.race} & \rightarrow \text{Bernoulli} \\
\text{3.race} & \rightarrow \text{Bernoulli} \\
\text{ui} & \rightarrow \text{Poisson} \\
\text{ptl} & \rightarrow \text{log} \\
\text{low} & \rightarrow \text{logit}
\end{align*}
\]

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 : ptt
Family : Poisson
Link : log
Log likelihood = -198.56178

|              | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|--------------|--------|-----------|-------|------|---------------------|
| low          |        |           |       |      |                     |
| ptt          | 0.5418 | 0.3462    | 1.56  | 0.118| -0.136799 to 1.220472 |
| age          | -0.0271| 0.0365    | -0.74 | 0.457| -0.0985418 to 0.0443412 |
| smoke        | 0.9233 | 0.4008    | 2.30  | 0.021| 0.137739 to 1.708951  |
| ht           | 1.8325 | 0.6916    | 2.65  | 0.008| 0.4769494 to 3.188086 |
| lwt          | -0.0151| 0.0069    | -2.19 | 0.029| -0.0287253 to -0.0015763 |
| race         |        |           |       |      |                     |
| black        | 1.2626 | 0.5264    | 2.40  | 0.016| 0.2309023 to 2.294392 |
| other        | 0.8621 | 0.4392    | 1.96  | 0.050| 0.0013548 to 1.722803 |
| ui           | 0.7585 | 0.4594    | 1.65  | 0.099| -0.1418484 to 1.658875 |
| _cons        | 0.4612 | 1.2046    | 0.38  | 0.702| -1.899729 to 2.822176 |
| ptl          |        |           |       |      |                     |
| age          | 0.0371 | 0.0299    | 1.24  | 0.215| -0.0214946 to 0.0956142 |
| smoke        | 0.9602 | 0.3396    | 2.83  | 0.005| 0.2944796 to 1.626027 |
| ht           | -0.1853| 0.7272    | -0.25 | 0.799| -1.610607 to 1.239906  |
| _cons        | -2.9855| 0.7842    | -3.81 | 0.000| -4.52255 to -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 http://www.stata-press.com/data/r15/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 `g` button.

4. Create the independent variables.
   
   Select the Add observed variables set tool, ```Variable Name```, 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


Also see

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