Title

stata.com

example 34g — Combined models (generalized responses)

Description Remarks and examples Reference Also see

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/r14/gsem_lbw
(Hosmer & Lemeshow data)
. describe
Contains data from http://www.stata-press.com/data/r14/gsem_lbw.dta
  obs:           189                          Hosmer & Lemeshow data
 vars:            11                          21 Mar 2014 12:28
 size:         2,646                          (_dta has notes)
-------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
-------------------------------------------------------------------------------
id              int     %8.0g                 subject id
low             byte    %8.0g                 birth weight < 2500g
age             byte    %8.0g                 age of mother
lwt             int     %8.0g                 weight, last menstrual period
race            byte    %8.0g      race       race
smoke           byte    %9.0g      smoke      smoked during pregnancy
ptl             byte    %8.0g                 premature labor history (count)
ht              byte    %8.0g                 has history of hypertension
ui              byte    %8.0g                 presence, uterine irritability
ftv             byte    %8.0g                 # physician visits, 1st trimester
bwt             int     %8.0g                 birth weight (g)
-------------------------------------------------------------------------------
Sorted by:
. notes
_dta:
  1.  Data from Hosmer, D. W., Jr., S. A. Lemeshow, and R. X. Sturdivant. 2013.
      "Applied Logistic Regression". 3rd ed. Hoboken, NJ: Wiley.
  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.

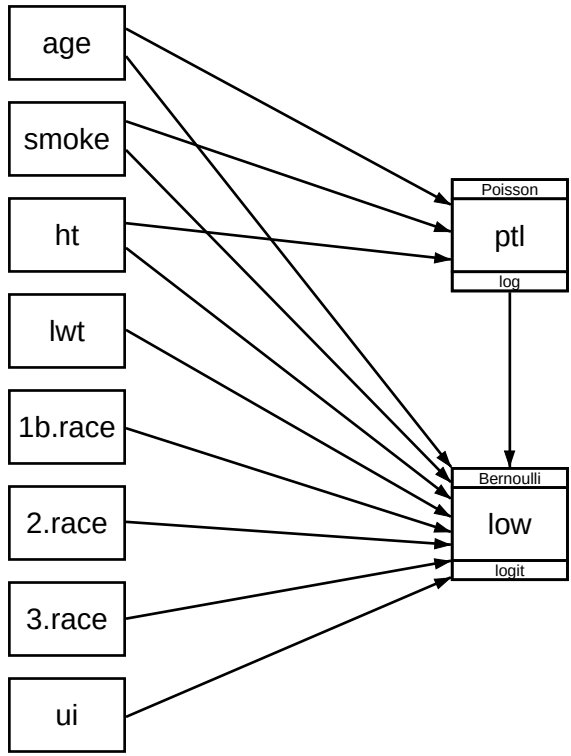
Remarks and examples

stata.com

Remarks are presented under the following headings:

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
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
low <-						
ptl	.5418366	.346249	1.56	0.118	-.136799	1.220472
age	-.0271003	.0364504	-0.74	0.457	-.0985418	.0443412
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	-.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	-.1418484	1.658875
_cons	.4612238	1.20459	0.38	0.702	-1.899729	2.822176
ptl <-						
age	.0370598	.0298752	1.24	0.215	-.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

```
. estat eform low ptl
```

	exp(b)	Std. Err.	z	P>|z|	[95% Conf.	Interval]
low						
ptl	1.719161	.5952579	1.56	0.118	.8721455	3.388787
age	.9732636	.0354759	-0.74	0.457	.9061578	1.045339
smoke	2.517698	1.00916	2.30	0.021	1.147676	5.523162
ht	6.249602	4.322407	2.65	0.008	1.611152	24.24199
lwt	.9849634	.0068217	-2.19	0.029	.9716834	.9984249
race						
white	1	(empty)				
black	3.534767	1.860737	2.40	0.016	1.259736	9.918406
other	2.368079	1.039949	1.96	0.050	1.001356	5.600207
ui	2.1351	.9808153	1.65	0.099	.8677528	5.2534
_cons	1.586014	1.910496	0.38	0.702	.1496092	16.8134
ptl						
age	1.037755	.0310032	1.24	0.215	.9787348	1.100334
smoke	2.612358	.8873835	2.83	0.005	1.342428	5.083638
ht	.8308134	.6041551	-0.25	0.799	.1997664	3.45529
_cons	.0505137	.0396137	-3.81	0.000	.0108613	.2349286

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/r14/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 GSEM 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.

6. Create the generalized response for low birthweight.

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:

```
smoke -> ptl
ht -> ptl
age -> low
smoke -> low
ht -> low
lwt -> low
1b.race -> low
2.race -> low
3.race -> low
ui -> low
```

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.

9. 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

```
. webgetsem gsem_comb
```

Reference

Hosmer, D. W., Jr., S. A. Lemeshow, and R. X. Sturdivant. 2013. *Applied Logistic Regression*. 3rd ed. Hoboken, NJ: Wiley.

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