Title

example 46g — Endogenous treatment-effects model

Description   Remarks and examples   References   Also see

Description

To illustrate the treatment-effects model, we use the following data:

```
. use http://www.stata-press.com/data/r13/gsem_union3
(NLSY 1972)

. describe

Contains data from http://www.stata-press.com/data/r13/gsem_union3.dta
  obs:         1,693                          NLSY 1972
 vars:            24                          29 Mar 2013 11:39
 size:        79,571                          (_dta has notes)
```

variable name	storage type	display format	value label	variable label
idcode	int	%8.0g		NLS ID
year	int	%8.0g		interview year
birth_yr	byte	%8.0g		birth year
age	byte	%8.0g		age in current year
race	byte	%8.0g	racelbl	race
msp	byte	%8.0g		1 if married, spouse present
nev_mar	byte	%8.0g		1 if never married
grade	byte	%8.0g		current grade completed
collgrad	byte	%8.0g		1 if college graduate
not_smsa	byte	%8.0g		1 if not SMSA
c_city	byte	%8.0g		1 if central city
south	byte	%8.0g		1 if south
ind_code	byte	%8.0g		industry of employment
occ_code	byte	%8.0g		occupation
union	byte	%8.0g		1 if union
wks_ue	byte	%8.0g		weeks unemployed last year
ttl_exp	float	%9.0g		total work experience
tenure	float	%9.0g		job tenure, in years
hours	int	%8.0g		usual hours worked
wks_work	int	%8.0g		weeks worked last year
ln_wage	float	%9.0g		ln(wage/GNP deflator)
wage	double	%10.0g		real wage
black	float	%9.0g		race black
smsa	byte	%8.0g		1 if SMSA

```
Sorted by:  idcode

. notes

_dta:
  1. Data from National Longitudinal Survey of Young Women 14-27 years of age
     (NLSY) in 1968, Center for Human Resource Research, Ohio State
     University, first released in 1989.
  2. This data was subsetted for purposes of demonstration. Among other
     things, it contains data only for 1972.
```

See *Structural models 7: Dependencies between response variables* and *Structural models 8: Unobserved inputs, outputs, or both* in [SEM] **intro 5** for background.
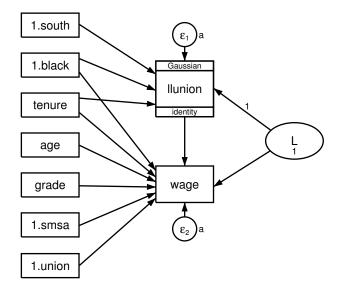
Remarks and examples

Remarks are presented under the following headings:

> *Fitting the treatment-effects model*
> *Fitting the model with the Builder*

Fitting the treatment-effects model

We wish to fit the following model:

We wish to estimate the "treatment effect" of being a union member. That is, we speculate that union membership has an effect on wages, and we want to measure that effect. The problem would be easy if we had data on the same workers from two different but nearly identical universes, one in which the workers were not union members and another in which they were.

The model above is similar to the Heckman selection model we fit in [SEM] **example 45g**. The differences are that the continuous variable (wage) is observed in all cases and that we have a path from the treatment indicator (previously selection, now treatment) to the continuous variable. Just as with the Heckman selection model, we allow for correlation by introducing a latent variable with model identification constraints.

Before we can fit this model, we need to create new variables llunion and ulunion. llunion will equal 0 if union is 1 and missing otherwise. ulunion is the complement of llunion: it equals 0 if union is 0 and missing otherwise. llunion and ulunion will be used as the dependent variables in the treatment equation, providing the equivalent of a scaled probit regression.

```
. gen llunion = 0 if union == 1
(1433 missing values generated)

. gen ulunion = 0 if union == 0
(709 missing values generated)
```

We can now fit this model with command syntax by typing

```
. gsem (wage <- age grade i.smsa i.black tenure 1.union L)
>      (llunion <- i.black tenure i.south L@1,
>        family(gaussian, udepvar(ulunion))),
>      var(L@1 e.wage@a e.llunion@a)
Fitting fixed-effects model:

Iteration 0:   log likelihood = -3308.0497
Iteration 1:   log likelihood = -3062.8995  (not concave)
Iteration 2:   log likelihood = -3062.2307
Iteration 3:   log likelihood = -3061.5084  (not concave)
Iteration 4:   log likelihood = -3061.5029
Iteration 5:   log likelihood =  -3061.497  (not concave)
Iteration 6:   log likelihood =  -3061.497  (backed up)

Refining starting values:

Grid node 0:   log likelihood = -3104.2036

Fitting full model:

Iteration 0:   log likelihood =   -3178.91  (not concave)
Iteration 1:   log likelihood = -3112.9538
Iteration 2:   log likelihood = -3059.5126
Iteration 3:   log likelihood = -3051.6997
Iteration 4:   log likelihood =  -3051.576
Iteration 5:   log likelihood =  -3051.575
Iteration 6:   log likelihood =  -3051.575

Generalized structural equation model            Number of obs   =        1210
Log likelihood =  -3051.575

 ( 1)  [llunion]L = 1
 ( 2)  [var(e.llunion)]_cons - [var(e.wage)]_cons = 0
 ( 3)  [var(L)]_cons = 1
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
wage <-						
age	.1487409	.0193291	7.70	0.000	.1108566	.1866252
grade	.4205658	.0293577	14.33	0.000	.3630258	.4781057
1.smsa	.9117044	.1249041	7.30	0.000	.6668969	1.156512
1.black	-.7882472	.1367077	-5.77	0.000	-1.056189	-.520305
tenure	.1524015	.0369595	4.12	0.000	.0799621	.2248408
1.union	2.945816	.2749549	10.71	0.000	2.406915	3.484718
L	-1.706795	.1288024	-13.25	0.000	-1.959243	-1.454347
_cons	-4.351572	.5283952	-8.24	0.000	-5.387207	-3.315936
llunion <-						
1.black	.6704049	.148057	4.53	0.000	.3802185	.9605913
tenure	.1282024	.0357986	3.58	0.000	.0580384	.1983664
1.south	-.8542673	.136439	-6.26	0.000	-1.121683	-.5868518
L	1	(constrained)				
_cons	-1.302676	.1407538	-9.25	0.000	-1.578548	-1.026804
var(L)	1	(constrained)				
var(e.llun~n)	1.163821	.2433321			.7725324	1.753298
var(e.wage)	1.163821	.2433321			.7725324	1.753298

Notes:

1. The treatment effect is measured by the coefficient on the path *treatment_variable->continuous_variable* or, in our case, 1.union->wage. It is estimated to be 2.9458, which is practically large and statistically significant.

2. The interpretation formulas are the same as for the Heckman selection model in [SEM] **example 45g**, namely,

$$\beta = \beta^*$$

$$\gamma = \gamma^*/\sqrt{\sigma^{2*} + 1}$$

$$\sigma^2 = \sigma^{2*} + \kappa^2$$

$$\rho = \kappa/\sqrt{(\sigma^{2*} + \kappa^2)(\sigma^{2*} + 1)}$$

To remind you, β are the coefficients in the continuous-outcome equation, γ are the coefficients in the treatment equation, σ^2 is the variance of the error in the continuous-outcome equation, and ρ is the correlation between the errors in the treatment and continuous-outcome equations.

3. In the output above, σ^{2*} (var(e.wage)) is 1.1638 and κ (the path coefficient on wage<-L) is -1.7068. In [SEM] **example 45g**, we calculated ρ by hand and then showed how to use the software to obtain the value and its standard error. This time, we will go right to the software.

After obtaining symbolic names by typing gsem, coeflegend, we type the following to obtain ρ:

```
. nlcom (rho: _b[wage:L]/(sqrt(_b[var(e.wage):_cons] + 1)*sqrt(_b[var(e.wage):_
> cons] + _b[wage:L]^2)))

        rho:  _b[wage:L]/(sqrt(_b[var(e.wage):_cons] + 1)*sqrt(_b[var(e.wage):
> _cons] + _b[wage:L]^2))
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]
rho	-.574648	.060969	-9.43	0.000	-.6941451 -.455151

4. We can obtain the untransformed treatment coefficients just as we did in [SEM] **example 45g**.

5. Just as with the Heckman selection model, with gsem, the treatment-effects model can be applied to generalized outcomes and include multilevel effects. See Skrondal and Rabe-Hesketh (2004, chap. 14.5) for an example with a Poisson response function.

Fitting the model with the Builder

Use the diagram in *Fitting the treatment-effects model* above for reference.

1. Open the dataset.

   In the Command window, type

```
. use http://www.stata-press.com/data/r13/gsem_union3
. gen llunion = 0 if union == 1
. gen ulunion = 0 if union == 0
```

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 $\begin{smallmatrix}G\\SEM\end{smallmatrix}$ button.

4. Create the independent variables.

   Select the Add Observed Variables Set tool, ▫▫▫, and then click in the diagram about one-fourth of the way in from the left and one-fourth of the way up from the bottom.

In the resulting dialog box,

    a. select the *Select variables* radio button (it may already be selected);

    b. type `1.south`, `1.black`, `tenure`, `age`, `grade`, `1.smsa`, and `1.union` in the *Variables* control (typing `1.`*varname* rather than using the ⋯ button to create `i.`*varname* prevents the rectangle corresponding to the base category for these binary variables from being created);

    c. select `Vertical` in the *Orientation* control; and

    d. 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 `llunion`.

    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 the `1.black` rectangle.

    c. In the Contextual Toolbar, select `Gaussian, Identity` in the *Family/Link* control (it may already be selected).

    d. In the Contextual Toolbar, select `llunion` in the *Variable* control.

    e. In the Contextual Toolbar, click on the **Properties** button.

    f. In the resulting *Variable properties* dialog box, click on the **Censoring** button in the **Variable** tab.

    g. In the resulting *Censoring* dialog box, select the *Interval measured, depvar is lower boundary* radio button. In the resulting *Interval measured* box below, use the *Upper bound* control to select the variable `ulunion`.

    h. Click on **OK** in the *Censoring* dialog box, and then click on **OK** in the *Variable properties* dialog box. The Details pane will now show `llunion` as the lower bound and `ulunion` as the upper bound for our interval measure.

6. Create the endogenous `wage` variable.

    a. Select the Add Observed Variable tool, ▭, and then click about one-third of the way in from the right side of the diagram, to the right of the `grade` rectangle.

    b. In the Contextual Toolbar, select `wage` with 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 `1.south` rectangle (it will highlight when you hover over it), and drag a path to the left side of the `llunion` 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:

```
1.black -> llunion
tenure -> llunion
1.black -> wage
tenure -> wage
age -> wage
grade -> wage
1.smsa -> wage
1.union -> wage
```

8. Clean up the direction of the error terms.

    We want the error for `llunion` to be above the rectangle and the error for `wage` to be below the rectangle, but it is likely they have been created in other directions.

    a. Choose the Select tool, ![select].

    b. Click in the `llunion` rectangle.

    c. Click on one of the **Error Rotation** buttons, ⟲ ⟳, in the Contextual Toolbar until the error is above the rectangle.

    d. Click in the `wage` rectangle.

    e. Click on one of the **Error Rotation** buttons, ⟲ ⟳, in the Contextual Toolbar until the error is below the rectangle.

9. Create the latent variable.

    a. Select the Add Latent Variable tool, ○, and then click at the far right of the diagram and vertically centered between the `llunion` and `wage` variables.

    b. In the Contextual Toolbar, type L in the *Name* control and press *Enter*.

10. Draw paths from the latent variable to each endogenous variable.

    a. Select the Add Path tool, →.

    b. Click in the upper-left quadrant of the L oval, and drag a path to the right side of the `llunion` rectangle.

    c. Continuing with the → tool, create another path by clicking first in the lower-left quadrant of the L oval and dragging a path to the right side of the `wage` rectangle.

11. Place constraints on the variances and on the path from L to `llunion`.

    a. Choose the Select tool, ![select].

    b. Click on the L oval. In the Contextual Toolbar, type 1 in the σ^2 box and press *Enter*.

    c. Click on the error oval attached to the `wage` rectangle. In the Contextual Toolbar, type a in the σ^2 box and press *Enter*.

    d. Click on the error oval attached to the `llunion` rectangle. In the Contextual Toolbar, type a in the σ^2 box and press *Enter*.

    e. Click on the path from L to `llunion`. In the Contextual Toolbar, type 1 in the β box and press *Enter*.

12. Clean up the location of the paths.

    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 variable and drag the endpoint.

13. 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_treat
```

References

Center for Human Resource Research. 1989. *National Longitudinal Survey of Labor Market Experience, Young Women 14–24 years of age in 1968.* Columbus, OH: Ohio State University Press.

Skrondal, A., and S. Rabe-Hesketh. 2004. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models.* Boca Raton, FL: Chapman & Hall/CRC.

Also see

[SEM] **example 34g** — Combined models (generalized responses)

[SEM] **example 45g** — Heckman selection model

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

[SEM] **intro 5** — Tour of models