Example 8a — Random effects in one equation and endogenous covariate

Description Remarks and examples Also see

Description

In this example, we show how to estimate and interpret the results of an extended regression model with a continuous outcome and a continuous endogenous covariate. We include random effects in the outcome equation but not in the equation for the continuous endogenous covariate.

Remarks and examples

In [ERM] **Example 1a**, we examined data from a fictional university that was studying the relationship between the high school grade point average (GPA) of its admitted students and their final college GPA.

Now suppose that 100 colleges have joined together in a study of the effect of high school GPA on the final college GPA of admitted students. Again, we suspect that unobserved ability affects both high school GPA and college GPA. So we treat high school GPA as an endogenous covariate. The researchers also believe that unobserved characteristics of the college are likely to affect college GPA but not high school GPA. Therefore, we allow for random effects in only the college GPA equation. Having random effects in only the main outcome equation is rare, but occasionally it corresponds to a model of interest.

Using data on the 2,000 students expected to graduate in 2010, the researchers model college GPA (`gpa`) as a function of high school GPA (`hsgpa`). In both cases, GPA is measured in 0.01 increments, and we ignore complications due to the boundary points. We also ignore that, unfortunately, the schools have a high dropout rate and that the college GPA is missing for these students, leaving the researchers with a sample of 1,372 students.

The researchers expect that the effect of high school competitiveness on college GPA is negligible once high school GPA is controlled for. So they include a ranking of the high school (`hscomp`) as an instrumental covariate for high school GPA. They include parental income measured in $10,000s, which they believe may also influence student performance, in the main model and in the model for high school GPA.

In our dataset, each observation represents one student. The variable `collegeid` uniquely identifies the 100 schools used in the study. Before we can fit a random-effects model to our data, we need to declare our grouping variable using `xtset`.

```
. use https://www.stata-press.com/data/r19/class10re
(Classes of 2010 profile)
. xtset collegeid
Panel variable: collegeid (balanced)
```

With the data xtset, we can now estimate the parameters of the model.

```
. xteregress gpa income, endogenous(hsgpa = income i.hscomp, nore)
(setting technique to bhhh)
Iteration 0:  Log likelihood =  44.332373
Iteration 1:  Log likelihood =  44.674349
Iteration 2:  Log likelihood =  44.688506
Iteration 3:  Log likelihood =  44.690548
Iteration 4:  Log likelihood =  44.691142
Iteration 5:  Log likelihood =  44.691359
Iteration 6:  Log likelihood =  44.691445
Iteration 7:  Log likelihood =  44.691483
Iteration 8:  Log likelihood =    44.6915
Iteration 9:  Log likelihood =  44.691507
(switching technique to nr)
Iteration 10: Log likelihood =  44.691511
```

```
Extended linear regression                      Number of obs     =      1,372
Group variable: collegeid                       Number of groups  =        100

                                                Obs per group:
                                                              min =          3
                                                              avg =       13.7
                                                              max =         20

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(2)      =    2916.69
Log likelihood = 44.691511                      Prob > chi2       =     0.0000
```

	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
gpa						
income	.0558709	.003765	14.84	0.000	.0484916	.0632502
hsgpa	.9390929	.0781538	12.02	0.000	.7859142	1.092272
_cons	-.5600512	.2346858	-2.39	0.017	-1.020027	-.1000755
hsgpa						
income	.0428695	.0019412	22.08	0.000	.0390649	.0466741
hscomp						
Moderate	-.1452852	.0140801	-10.32	0.000	-.1728817	-.1176887
High	-.2339232	.0235935	-9.91	0.000	-.2801656	-.1876809
_cons	3.083431	.0167567	184.01	0.000	3.050589	3.116274
var(e.gpa)	.0470832	.0024655			.0424907	.0521721
var(e.hsgpa)	.0572604	.0021862			.053132	.0617096
corr(e.hsgpa, e.gpa)	.1979973	.0870885	2.27	0.023	.0229883	.3612321
var(gpa[colle~d])	.0633532	.0095652			.0471252	.0851695

We suppressed the random effect from the equation for high school GPA by specifying nore within the endogenous() option. Therefore, no variance is reported for college random effects affecting a student's high school GPA. The variance of the random effects affecting college GPA is estimated to be 0.06.

To check for endogeneity, we need to examine only the correlation between the student-level errors in high school and college GPAs. The estimate of this correlation is 0.2, and the corresponding test finds that it is significantly different from zero. The researchers conclude that the unobserved student-level factors that increase high school GPA tend to also increase college GPA.

Because this is a linear regression model, the coefficients can be directly interpreted. For example, the researchers expect the difference in college GPA is about 0.94 points for students with a difference of 1 point in high school GPA.

Also see

[ERM] **eregress** — Extended linear regression

[ERM] **eregress postestimation** — Postestimation tools for eregress and xteregress

[ERM] **Intro 3** — Endogenous covariates features

[ERM] **Intro 6** — Panel data and grouped data model features

[ERM] **Intro 9** — Conceptual introduction via worked example