Description

In [ERM] **Example 1a** and [ERM] **Example 1b**, we ignored the observations that were dropped because of missing data on GPA. In this example, we show you how to fit a model that includes a continuous endogenous covariate, a censored outcome, and endogenous sample selection.

Remarks and examples

In the previous two examples, the researchers excluded students who dropped out of college because they are missing college GPA data on these students. So they were estimating parameters for the population of students who graduate from college. Let's suppose they are interested in expected college GPA for all students who enroll, even those who drop out. They suspect that unobserved ability affects both the decision to stay in school and college GPA and thus that they have an endogenously selected sample.

To model the selection, they need a covariate that affects the probability that they observe a student's GPA but does not affect the level of the student's GPA. They include an indicator for whether the student participated in a retention program and whether the student had a roommate who also went to State U. They expect that students with a roommate who went to the same college were more likely to remain in school because they felt more included in the college environment.

```
. eintreg gpal gpau income, endogenous(hsgpa = income i.hscomp)
> select(graduate = hsgpa income i.roommate i.program)
  (iteration log omitted)
```

Extended interval regression

```
                                          Number of obs    =   2,500
                                               Selected =    1,528
                                            Nonselected =      972
                                             Uncensored =        0
                                          Left-censored =      150
                                         Right-censored =        0
                                         Interval-cens. =    1,378

                                          Wald chi2(2)     =  734.96
Log likelihood = -2851.3222               Prob > chi2      =  0.0000
```

	Coefficient	Std. err.	z	P>|z|	[95% conf. interval]	
income						
income	.0338548	.0075484	4.49	0.000	.0190602	.0486495
hsgpa	1.19378	.1443563	8.27	0.000	.9108467	1.476713
_cons	-.7895643	.3908796	-2.02	0.043	-1.555674	-.0234543
graduate						
hsgpa	2.215481	.4411331	5.02	0.000	1.350876	3.080086
income	.1920393	.0162334	11.83	0.000	.1602224	.2238563
roommate						
Yes	.1547087	.0455906	3.39	0.001	.0653528	.2440645
1.program	.4858749	.0523443	9.28	0.000	.383282	.5884678
_cons	-7.524521	1.237529	-6.08	0.000	-9.950034	-5.099008
hsgpa						
income	.047866	.0016981	28.19	0.000	.0445377	.0511942
hscomp						
Moderate	-.1337635	.0115749	-11.56	0.000	-.1564499	-.1110771
High	-.2284481	.0190089	-12.02	0.000	-.2657049	-.1911914
_cons	2.793802	.0132125	211.45	0.000	2.767906	2.819698
var(e.gpal)	.1753568	.0085604			.1593564	.1929636
var(e.hsgpa)	.0685863	.0019399			.0648876	.0724958
corr(e.gra~e, e.gpal)	-.9124422	.0327448	-27.87	0.000	-.9583429	-.8205981
corr(e.hsgpa, e.gpal)	.0534114	.0937195	0.57	0.569	-.1300101	.2332982
corr(e.hsgpa, e.graduate)	.2747613	.0955172	2.88	0.004	.079342	.4498437

The coefficients from the main equation for hsgpa continue to be interpreted as in [ERM] **Example 1b**. Now, however, they are estimates for the population of all admitted students, not the population of all graduates. The estimated effect of high school GPA for this population is slightly higher, 1.19 compared with 1.11.

As with [ERM] **Example 1b**, we will not further interpret this model here. Instead we refer you to the interpretation performed in [ERM] **Example 1a**. The addition of endogenous sample selection makes no difference in what commands you would type to answer questions or to how you would interpret the

results of those commands. In fact, we encourage you to run the commands discussed in [ERM] **Example 1a** on this model and compare the results. The only thing to keep in mind is that now the population we are making inferences about is all students admitted to school, not just those who graduate.

Also see

[ERM] **eintreg** — Extended interval regression

[ERM] **eintreg postestimation** — Postestimation tools for eintreg and xteintreg

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

[ERM] **Intro 4** — Endogenous sample-selection features

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