

**example 3b** — Probit regression with endogenous covariate and treatment[Description](#)[Remarks and examples](#)[Also see](#)

## Description

We model a binary outcome that depends on a continuous endogenous covariate and has an endogenous treatment by using `eprobit` with the `endogenous()` and `entreat()` options.

## Remarks and examples

[stata.com](#)

Continuing from [\[ERM\] example 3a](#), State U administrators have implemented a voluntary program to increase retention freshman year. Whether a student chose to participate is stored in the indicator variable `program`. They are concerned that unobservable factors that influence a student's decision to participate in the college retention program also influence the probability of graduation. For example, students who have higher self-motivation may be more likely to join and also more likely to graduate without the program. Thus, they are concerned that participation in the program may be an endogenously chosen treatment.

The researchers believe the program was easier to access for students who lived on campus freshman year. They also think students who had scholarships may have been more motivated to attend the program. However, they do not believe either of these variables independently affects the probability of graduation after controlling for other covariates in the model. They use an indicator for on-campus residence during the freshman year (`campus`), having a scholarship of any kind (`scholar`), and parents' income in the treatment assignment model.

## 2 example 3b — Probit regression with endogenous covariate and treatment

```
. eprobit graduate income i.roommate, endogenous(hsgpa = income i.hscomp)
> entreat(program = i.campus i.scholar income) vce(robust)

Iteration 0:  log pseudolikelihood = -2793.4696
Iteration 1:  log pseudolikelihood = -2792.9298
Iteration 2:  log pseudolikelihood = -2792.9017
Iteration 3:  log pseudolikelihood = -2792.9016
```

```
Extended probit regression          Number of obs   =      2,500
                                   Wald chi2(8)       =      404.26
Log pseudolikelihood = -2792.9016  Prob > chi2     =      0.0000
```

	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
<b>graduate</b>						
program#						
c.income						
0	.1760785	.0201833	8.72	0.000	.13652	.2156371
1	.1925761	.021294	9.04	0.000	.1508405	.2343116
roommate#						
program						
yes#0	.3110885	.0814304	3.82	0.000	.1514878	.4706892
yes#1	.2475942	.0756877	3.27	0.001	.0992491	.3959394
program#						
c.hsgpa						
0	1.160053	.4590276	2.53	0.011	.2603759	2.059731
1	.9379774	.4450455	2.11	0.035	.0657043	1.81025
program						
0	-4.350156	1.312558	-3.31	0.001	-6.922721	-1.77759
1	-3.393398	1.242536	-2.73	0.006	-5.828725	- .9580717
<b>program</b>						
campus						
yes	.7433155	.0735249	10.11	0.000	.5992092	.8874217
scholar						
yes	.8970451	.0585469	15.32	0.000	.7822952	1.011795
income	-.0799274	.0088987	-8.98	0.000	-.0973686	-.0624862
_cons	-.3810042	.0860131	-4.43	0.000	-.5495867	-.2124217
<b>hsgpa</b>						
income	.0478626	.0016461	29.08	0.000	.0446363	.0510889
hscomp						
moderate	-.1350116	.0115013	-11.74	0.000	-.1575538	-.1124694
high	-.2269435	.019326	-11.74	0.000	-.2648218	-.1890652
_cons	2.79442	.0128088	218.16	0.000	2.769315	2.819525
var(e.hsgpa)	.0685874	.0019597			.064852	.0725379
<b>corr(e.pro-m, e.graduate)</b>						
corr(e.hsgpa, e.graduate)	.3791651	.1035775	3.66	0.000	.1605878	.5622919
corr(e.hsgpa, e.graduate)	.4001679	.089854	4.45	0.000	.2109447	.5604834
corr(e.hsgpa, e.program)	-.0201748	.02637	-0.77	0.444	-.0717594	.0315174

The main equation output is slightly different from that in [ERM] example 3a. Because program was specified as a treatment, it was automatically interacted with each of the other covariates in the graduate equation.

The correlation between the errors from the graduation equation and those from the program participation equation is estimated to be 0.38 and is significantly different from zero. The researchers conclude that unobservable factors that increase the chance of participating in the program also increase the chance of graduating.

Now, we use `estat teffects` to estimate the ATE of program participation on college graduation. We specified `vce(robust)` when we fit the model, so `estat teffects` reports standard errors and tests for the population ATE.

```
. estat teffects
Predictive margins                                Number of obs    =      2,500
```

	Unconditional		z	P> z	[95% Conf. Interval]	
	Margin	Std. Err.				
ATE program (1 vs 0)	.1017846	.0488905	2.08	0.037	.005961	.1976081

We estimate that the ATE is 0.10. In other words, the average probability of graduating increases by 0.10 when all students participate in the program versus when no students participate in the program.

We might be interested if those students who self-selected into the program increased their graduation probability by more than 0.10. We estimate the average treatment effect on the treated (ATET).

```
. estat teffects, atet
Predictive margins                                Number of obs    =      2,500
                                                Subpop. no. obs  =      1,352
```

	Unconditional		z	P> z	[95% Conf. Interval]	
	Margin	Std. Err.				
ATET program (1 vs 0)	.1015967	.0493961	2.06	0.040	.0047821	.1984113

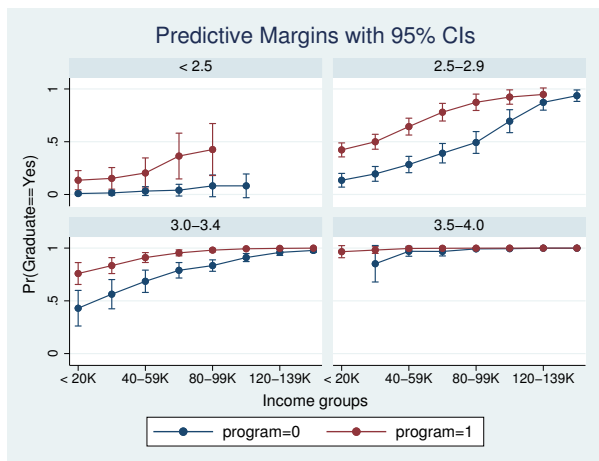
In this case, the students who chose the program did no better on average than choosing by flipping a coin. Both the ATE and ATET are 0.10.

Those are the overall averages. Do graduation rates for participants and nonparticipants differ by high school GPA and parents' income? Our dataset has grouping variables, so we can let `margins` estimate graduation rates subpopulations defined by all three covariates.

```
. margins, over(program incomegrp hsgpagrp) vce(unconditional)
```

The output is copious. You can type the command and see it if you like. The patterns are easier to see on a `marginsplot`.

```
. marginsplot, plot(program) xlabel(0 4 8 12)
```



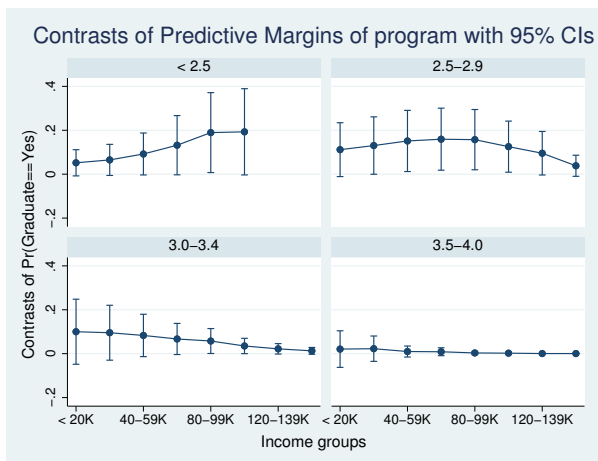
The red line shows expected graduation rates for those who participated in the program. The blue line shows rates for nonparticipants. Clearly, the differences between the groups in the program and those out of the program differ dramatically across GPA and family income. For GPAs at or above 3.5, the graduation rates are so high that there was no room for differences. For most other groups, the graduation rates are estimated to be substantially higher among those who participated. The only exceptions are extremely low-income students with GPAs below 2.5 and extremely high-income students with GPAs at or above 3.5.

We were careful not to call the comparisons above effects or attribute them directly to the program. They are indeed expected rates for the groups, but the students self-selected into program participation groups. If we want to compare graduation rates assuming all students don't participate and then assuming all students do participate, we need to instruct `margins` to `fix()` the values for program participation and also add the `r_` to `program`.

```
. margins r_program, over(incomegrp hsgpagrp) vce(unconditional)
> predict(fix(program)) contrast(nowald)
(output omitted)
```

The output is again long, so we leave you to see it for yourself. The graphs reveal the patterns across groups.

```
. marginsplot, by(hsgpagrp) xlabel(0 4 8 12)
```



These differences are close to what we would have seen had we differenced the red and blue lines of the first graph. In this graph, each point is an estimate of the average treatment effect for a subpopulation defined by a range of GPAs and a range of family income. We note that the confidence intervals, as represented by the capped lines, are fairly wide.

## Also see

[ERM] [eprobit](#) — Extended probit regression

[ERM] [eprobit postestimation](#) — Postestimation tools for eprobit

[ERM] [estat teffects](#) — Average treatment effects for extended regression models

[ERM] [intro 3](#) — Endogenous covariates features

[ERM] [intro 5](#) — Treatment assignment features

[ERM] [intro 8](#) — Conceptual introduction via worked example