Title

> **Example 2b —** Linear regression with exogenous treatment

Description

In this example, we show how to estimate and interpret the results of an extended regression model with a continuous outcome and exogenous binary treatment.

Remarks and examples

In [ERM] **Example 2a**, we analyzed the effect of having a college degree on wages as a binary endogenous covariate. Now suppose that we approach our research question instead in the potential-outcomes framework. With this approach, we consider the expected wage for each individual without a college degree versus the expected wage for each individual with a college degree. Specifically, we might like to know the average expected change in wages for those who complete college, the average treatment effect on the treated (ATET).

As before, we use `wageed.dta` with educational attainment data on 6,000 adults. We control for differences in job tenure (`tenure`) and age (`age`) by specifying them in the main equation. For the time being, we consider the treatment (`college`) to be exogenous. We want to make inferences about the average effect of a college degree on the wages of all individuals who complete college, not just the subjects in our study sample, so we specify `vce(robust)`. This will allow us to estimate the standard errors of the ATET accounting for the fact the variables in our sample represent just one draw from the population.

```
. use https://www.stata-press.com/data/r18/wageed
(Wages for 20 to 74 year olds, 2015)

. eregress wage c.age##c.age tenure, extreat(college) vce(robust)

Iteration 0:  Log pseudolikelihood = -13989.589
Iteration 1:  Log pseudolikelihood = -13989.586
Iteration 2:  Log pseudolikelihood = -13989.586
```

Extended linear regression					Number of obs	=	6,000
					Wald chi2(8)	=	439363.91
Log pseudolikelihood = -13989.586					Prob > chi2	=	0.0000

wage	Coefficient	Robust std. err.	z	P>|z|	[95% conf. interval]	
college# c.age						
No	.2454534	.0180052	13.63	0.000	.2101638	.280743
Yes	.7042756	.0225386	31.25	0.000	.6601007	.7484505
college# c.age#c.age						
No	-.0018998	.0001935	-9.82	0.000	-.002279	-.0015206
Yes	-.0054223	.000243	-22.31	0.000	-.0058986	-.0049459
college# c.tenure						
No	.3206065	.0207164	15.48	0.000	.2800031	.36121
Yes	.4935213	.0257599	19.16	0.000	.4430329	.5440097
college						
No	9.851871	.3701276	26.62	0.000	9.126435	10.57731
Yes	4.384709	.4654545	9.42	0.000	3.472435	5.296983
var(e.wage)	6.20477	.1152627			5.982922	6.434843

Because we specified the command as a treatment-effects model, `eregress` automatically interacts the `college` variable with all other covariates in the model, thus essentially creating separate models for those who graduate from college and those who do not. There is nothing wrong with interpreting the coefficients. This is, after all, just a regression. The coefficients labeled `no` are the estimates of the parameters of the wage model for those who are not college graduates. The coefficients labeled `yes` are the estimates of the parameters for the model of those who are college graduates. Tenure in the company has a larger effect for college graduates than nongraduates. It is 49 cents an hour per tenure year for college graduates and 32 cents for nongraduates. The effect of age is more difficult to interpret because of the quadratic term. The effect of age is clearly different between the groups, but the pattern of that difference is not obvious. See [ERM] **Example 2c** for some tools you could apply to this model that would make that pattern obvious. The effect of college graduation is harder still to see. For any person, it would be the difference of the values predicted by the two models. Again, see [ERM] **Example 2c** for ways to visualize the effect.

If we are interested only in the average effect, we can estimate that using the estat teffects command.

```
. estat teffects, atet
```

Predictive margins

Number of obs   = 6,000
Subpop. no. obs = 2,234

	Margin	Unconditional std. err.	z	P>|z|	[95% conf. interval]	
ATET						
college						
(Yes vs No)	7.62719	.0863465	88.33	0.000	7.457954	7.796426

The average wage is estimated to be $7.63 higher per hour for the population of college graduates than the wage would have been if those same individuals had not completed college.

We have ignored several potential complications in this example. One of which is that unobserved factors such as ability that influence whether individuals complete college could also influence their wage. In that case, the treatment assignment (obtaining a college degree) would be endogenous. If the treatment were endogenous, we would model its coefficients and the correlation between the treatment assignment errors and the outcome errors. See [ERM] **Example 2c** for an example with an endogenous treatment.

Also see

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

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

[ERM] **estat teffects** — Average treatment effects for extended regression models

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