

Example 7 — Random-effects regression with continuous endogenous covariate
[Description](#)[Remarks and examples](#)[Reference](#)[Also see](#)

Description

In this example, we show how to estimate and interpret the results of an extended regression model with a continuous outcome, a continuous endogenous covariate, and random effects.

Remarks and examples

[stata.com](#)

We will use `nlswork.dta`, a subsample of the NLSY data ([Center for Human Resource Research 1989](#)) on young women aged 14–26 in 1968. These data are panel data; each individual was surveyed in multiple years ranging from 1968 to 1988.

Suppose that we want to study the relationship between the natural logarithm of wage (`ln_wage`) and the number of years at a job (`tenure`). We also model `ln_wage` with a quadratic effect of the individual's age (`age` and `c.age#c.age`), living in a metropolitan area (`not_smsa`), and whether the individual is African American (`2.race`). We suspect that the unobserved factors that influence the individual's job tenure are correlated with the unobserved factors that influence their wage, so we treat job tenure as an endogenous covariate. We use an individual's union status (`union`) and whether she lived in the southern United States (`south`) as instrumental covariates for tenure. Of course, these are not the instruments we would choose in real research, but they are useful for demonstrating how to use the commands below.

We also want to account for the within-panel correlation in our data, so we fit a random-effects model using `xtregress`. Before we can fit our model, we must use `xtset` to specify the panel identifier variable, in this case, `idcode`. Our data have already been `xtset`, so we type `xtset` to display the settings.

```
. use https://www.stata-press.com/data/r16/nlswork
(National Longitudinal Survey. Young Women 14-26 years of age in 1968)
. xtset
    panel variable:  idcode (unbalanced)
    time variable:  year, 68 to 88, but with gaps
                   delta: 1 unit
```

We are now ready to fit our model. We want to make inferences about how our covariates affect the log wage in the population, not just in our sample. Therefore, we add the `vce(robust)` option so that subsequent calls to `margins` will consider our sample as a draw from the population.

By default, `xtregress` includes random effects for both `ln_wage` and `tenure` and allows these random effects to be correlated. Because of the complexity of this model, the command may take a few minutes to run.

2 Example 7 — Random-effects regression with continuous endogenous covariate

```

. xtregress ln_wage age c.age#c.age not_smsa 2.race,
> endogenous(tenure = age c.age#c.age union 2.race south) vce(robust)
(setting technique to bhhh)
Iteration 0: log pseudolikelihood = -53610.76
Iteration 1: log pseudolikelihood = -53602.997
Iteration 2: log pseudolikelihood = -53602.576
Iteration 3: log pseudolikelihood = -53602.573
Iteration 4: log pseudolikelihood = -53601.921
Iteration 5: log pseudolikelihood = -53601.801
Iteration 6: log pseudolikelihood = -53601.753
Iteration 7: log pseudolikelihood = -53601.623
Iteration 8: log pseudolikelihood = -53601.563
Iteration 9: log pseudolikelihood = -53601.547
(switching technique to nr)
Iteration 10: log pseudolikelihood = -53601.495
Iteration 11: log pseudolikelihood = -53601.41
(switching technique to bhhh)
Iteration 12: log pseudolikelihood = -53601.41
Extended linear regression      Number of obs      =      19,007
Group variable: idcode          Number of groups   =       4,134
                                Obs. per group:
                                min =             1
                                avg =             4.6
                                max =             12
Integration method: mvaghermite      Integration pts.   =       7
                                Wald chi2(5)       =      384.25
Log pseudolikelihood = -53601.41     Prob > chi2       =       0.0000
                                (Std. Err. adjusted for 4,134 clusters in idcode)

```

	Coef.	Robust Std. Err.	z	P> z	[95% Conf. Interval]	
ln_wage						
age	.0161086	.0134428	1.20	0.231	-.0102388	.042456
c.age#c.age	-.0011178	.0002402	-4.65	0.000	-.0015887	-.000647
not_smsa	-.172498	.0122743	-14.05	0.000	-.1965552	-.1484408
race						
black	-.2374388	.0254533	-9.33	0.000	-.2873263	-.1875513
tenure	.2300781	.0277646	8.29	0.000	.1756605	.2844957
_cons	1.690136	.2077606	8.14	0.000	1.282933	2.097339
tenure						
age	.0892847	.0599348	1.49	0.136	-.0281852	.2067547
c.age#c.age	.0033688	.0009943	3.39	0.001	.0014199	.0053176
union	.5584566	.0740956	7.54	0.000	.4132318	.7036814
race						
black	.4691202	.1101411	4.26	0.000	.2532476	.6849929
south	-.4024058	.0628545	-6.40	0.000	-.5255983	-.2792132
_cons	-2.929734	.8800349	-3.33	0.001	-4.65457	-1.204897
var(e.ln_w~e)	.3654205	.0786259			.2396866	.5571114
var(e.tenure)	6.656475	.1285168			6.409292	6.913189
corr(e.ten~e, e.ln_wage)	-.9055589	.0213219	-42.47	0.000	-.9395846	-.8538145

var(ln_~e[idc~e])	.3314414	.0736048			.2144748	.5121973
var(ten~e[idc~e])	7.593483	.3027546			7.022688	8.210672
corr(ten~e[idc~e], ln_~e[idc~e])	-.8299334	.0421356	-19.70	0.000	-.8963409	-.7271053

The first two sections of the output provide the estimated coefficients in the equations for `ln_wage` and `tenure`. Because this is a linear regression, we can interpret the coefficients in the usual way. For example, we expect an increase of 0.23 in log wage for an additional year of job tenure.

Next, we see the estimates of the observation-level error variances and their correlation. This is followed by estimates of the variances of the random effects and an estimate of their correlation. If at least one of these correlations is significantly different from zero, we can conclude that `tenure` is endogenous. In our case, the correlation between the observation-level errors is -0.91 , and the correlation between the random effects is -0.83 . Because both are negative and significantly different from zero, we conclude that `tenure` is endogenous and that unobserved individual-level factors that increase job tenure tend to decrease log wage. Additionally, unobserved observation-level (time-varying) factors that increase job tenure tend to also decrease log wage.

We can also answer questions about the actual wage rather than its natural logarithm. The prediction option `expmean` can be used with the `predict()` option of `margins` to estimate the mean of the exponentiated outcome. What if we could increase everyone’s job tenure by one year—from 1 year to 2 years, from 1.5 years to 2.5 years, etc? `margins` will give us the population-average expected wage leaving each individual’s tenure at its current value if we specify `at(tenure=generate(tenure))`. `margins` will also give us the population-average expected wage treating each individual as if she has one additional year of job tenure if we specify `at(tenure=generate(tenure+1))`. We want to hold each individual’s unobserved characteristics to be those that are implied by her observed data, so we also create a variable that holds the true values of `tenure` and specify `base(tenure=tenureT)` within `predict()`.

```
. generate tenureT = tenure
(433 missing values generated)

. margins, at(tenure=generate(tenure)) at(tenure=generate(tenure+1))
> predict(expmean base(tenure=tenureT)) vce(unconditional)

Predictive margins                                Number of obs      =    19,007
Expression   : mean of ln_wage, predict(expmean base(tenure=tenureT))
1._at        : tenure                               = tenure
2._at        : tenure                               = tenure+1
              (Std. Err. adjusted for 4,134 clusters in idcode)
```

	Margin	Unconditional Std. Err.	z	P> z	[95% Conf. Interval]	
_at						
1	6.391319	.2138264	29.89	0.000	5.972227	6.810411
2	8.044742	.4898009	16.42	0.000	7.08475	9.004734

We find that the expected hourly wage is \$6.39. However, when everyone is given an additional year of job tenure, the expected hourly wage rises to \$8.04.

By adding `contrast(at(r))` to our `margins` command, we can difference those two counterfactuals and estimate the average effect of giving an additional year of job tenure. We also add `effects` to request test statistics and `nowald` to remove an unnecessary table from the output.

```
. margins, at(tenure=generate(tenure)) at(tenure=generate(tenure+1))
> predict(expmean base(tenure=tenureT)) contrast(at(r) nowald effects)
> vce(unconditional)
```

```
Contrasts of predictive margins          Number of obs    =    19,007
Expression  : mean of ln_wage, predict(expmean base(tenure=tenureT))
1._at      : tenure                = tenure
2._at      : tenure                = tenure+1
                               (Std. Err. adjusted for 4,134 clusters in idcode)
```

	Unconditional					
	Contrast	Std. Err.	z	P> z	[95% Conf. Interval]	
_at (2 vs 1)	1.653423	.2776953	5.95	0.000	1.109151	2.197696

We estimate the average effect of an additional year of job tenure is a \$1.65 increase in hourly wage.

Reference

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.

Also see

- [ERM] [eregress](#) — Extended linear regression
- [ERM] [eregress postestimation](#) — Postestimation tools for `eregress` and `xtheregress`
- [ERM] [Intro 3](#) — Endogenous covariates features
- [ERM] [Intro 6](#) — Panel data and grouped data model features
- [ERM] [Intro 9](#) — Conceptual introduction via worked example