Title

stata.com

example 6a — Ordered probit regression with endogenous treatment

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 an ordinal outcome and endogenous treatment.

Remarks and examples

stata.com

We are studying the effect of having health insurance on women's health status, which we measure with a health score from 1 (poor) to 5 (excellent). We want to estimate the average treatment effect (ATE) of insurance on the probability of having each of the five statuses. We suspect that our model needs to account for the health insurance being an endogenous treatment.

In our fictional study, we collect data on a sample of 6,000 women between the ages of 25 and 30. In addition to the insurance indicator, we include an indicator for whether the woman exercises regularly and the number of years of schooling she completed (`grade`) as exogenous covariates. For our treatment model, we use `grade` and an indicator for whether the woman is currently working or attending school (`workschool`), which is excluded from the outcome model.

```
. use http://www.stata-press.com/data/r15/womenhlth
(Women's health status)

. eoprobit health i.exercise grade, entreat(insured = grade i.workschool)
> vce(robust)
```

(iteration log omitted)

```
Extended ordered probit regression              Number of obs     =      6,000
                                                Wald chi2(4)      =     516.93
Log pseudolikelihood = -9105.4376               Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |               Robust
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
health       |
   exercise# |
     insured |
      yes#no |   .5296149   .0619049     8.56   0.000     .4082835    .6509463
     yes#yes |   .5190249    .033872    15.32   0.000     .4526371    .5854127
             |
    insured# |
     c.grade |
          no |   .1079014   .0250326     4.31   0.000     .0588383    .1569645
         yes |   .1296456   .0107428    12.07   0.000       .10859    .1507012
-------------+----------------------------------------------------------------
insured      |
       grade |   .3060024   .0100506    30.45   0.000     .2863036    .3257012
             |
  workschool |
         yes |   .5387767   .0446794    12.06   0.000     .4512067    .6263466
       _cons |  -3.592452   .1348431   -26.64   0.000     -3.85674   -3.328165
-------------+----------------------------------------------------------------
/health      |
    insured# |
      c.cut1 |
          no |   .6282326   .2393499                      .1591154     1.09735
         yes |  -.7255086   .2470598                     -1.209737   -.2412803
    insured# |
      c.cut2 |
          no |   1.594089   .2300159                      1.143266    2.044912
         yes |   .4404531   .1986825                      .0510426    .8298636
    insured# |
      c.cut3 |
          no |   2.526424   .2241048                      2.087186    2.965661
         yes |   1.332514   .1845713                      .9707608    1.694267
    insured# |
      c.cut4 |
          no |    3.41748   .2356708                      2.955574    3.879386
         yes |   2.292828   .1760594                      1.947758    2.637899
-------------+----------------------------------------------------------------
corr(e.ins~d,|
    e.health)|   .3414241   .0940374     3.63   0.000     .1460223    .5111858
------------------------------------------------------------------------------
```

The estimated correlation between the errors from the health status equation and the errors from the health insurance equation is 0.34. This is significantly different from zero, so the treatment choice of being insured is endogenous. Because it is positive, we conclude that unobserved factors that increase the chance of having health insurance tend to also increase the chance of being in a high health status.

We see estimates of both the coefficients and the cutpoints for two equations, one for insured women (`yes`) and one for uninsured (`no`). For both insured and uninsured, exercise and education have positive effects on health status.

We could use `estat teffects` to estimate the ATE of insurance on the probabilities of each health category.

```
. estat teffects
```

Feel free to run that command and see the results. We estimate and interpret other estimates of these ATEs in [ERM] **example 6b** after adjusting for endogenous sample selection that is introduced in that example. The ATE estimates there are slightly different, but they estimate the same thing. Given a sufficiently large sample, the two sets of estimates would converge to the same values.

Also see

[ERM] **eoprobit** — Extended ordered probit regression

[ERM] **eoprobit postestimation** — Postestimation tools for eoprobit

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

[ERM] **intro 5** — Treatment assignment features

[ERM] **intro 8** — Conceptual introduction via worked example