Ordered Models in gretl

Quantitative Microeconomics

R. Mora

Department of Economics
Universidad Carlos III de Madrid

Outline

Introduction

The Ordered Probit and ML Estimation

- Consider m observed outcomes: $y = 0, 1, ..., m$.
- Consider the latent variable model without a constant:

$$y^* = x'\beta + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0,1)$$

- Define $m-1$ cut-off points: $\alpha_1 < ... < \alpha_{m-1}$
- We do not observe y^*, but we observe choices according to the following rule

$$y = 0 \text{ if } y^* \leq \alpha_1$$
$$y = 1 \text{ if } \alpha_1 < y^* \leq \alpha_2$$
$$\vdots$$
$$y = m \text{ if } \alpha_{m-1} < y^*$$

ML estimation

- Since:

$$\Pr(y=0|x)=\Pr\left(x'\beta+\varepsilon\leq\alpha_1\right)=1-\Phi\left(x'\beta-\alpha_1\right)$$
$$\Pr(y=1|x)=\Phi\left(x'\beta-\alpha_1\right)-\Phi\left(x'\beta-\alpha_2\right)$$
$$\vdots$$
$$\Pr(y=m|x)=\Phi\left(x'\beta-\alpha_{m-1}\right)$$

- Thus, for a sample of N observations, the likelihood is:

$$\mathrm{L}(b)=\prod_{i=1}^{N}\Big\{\left(1-\Phi\left(x_i'b-\alpha_1\right)\right)^{1(y_i=0)}\times$$
$$\left(\Phi\left(x_i'b-\alpha_1\right)-\Phi\left(x_i'b-\alpha_2\right)\right)^{1(y_i=1)}\times\ldots\times$$
$$\left(\Phi\left(x_i'b-\alpha_{m-1}\right)\right)^{1(y_i=m)}\Big\}$$

Basic Commands in gretl for Ordered Probit Estimation

- `probit`: computes Maximum Likelihood order probit estimation if the dependent variable is not binary but is discrete
- `omit/add`: tests joint significance
- `$yhat`: returns probability estimates
- `$lnl`: returns the log-likelihood for the last estimated model
- `logit`: computes Maximum Likelihood logit estimation if the dependent variable is not binary but is discrete

- in this Session, we are going to learn how to use `probit` for ordered probit models

Ordered Probit in gretl

probit *depvar* *indvars* --robust --verbose --p-values

- *depvar* must either take on only non-negative integer values, or be explicitly marked as discrete.
    - In case the variable has non-integer values, it will be recoded internally.

- by default, standard errors are computed using the negative inverse of the Hessian.
- options:
    1. `--robust`: covariance matrix robust to model misspecification
    2. `--verbose`: shows information from all numerical iterations

Estimates

- ML estimates for the β coefficients and for the cut-off points are obtained.
  - The latter are reported by `gretl` as `cut1`, `cut2` and so on.
- `$uhat` yields generalized residuals; `$yhat` returns $x'\widehat{\beta}$.
  - It is thus possible to compute an unbiased estimate of the latent variable U simply by adding the two together.
- Output shows χ^2_q statistic test for null that all slopes are zero

Example: Simulated Data

Activity

- $y^* = 0.07 \times educ - 1.0 \times kids + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0,1)$

$$y = 0 \text{ if } h^* \leq 0.5 \text{ (inactive)}$$
$$y = 1 \text{ if } 0.5 < y^* \leq 2.5 \text{ (part-time)}$$
$$y = 2 \text{ if } 2.5 < y^* \text{ (full-time)}$$

- education brings utility the more you work
- having a kid brings more utility the less you work

probit *work educ kids*

```
Model 1: Ordered Probit, using observations 1-5000
Dependent variable: work

             coefficient   std. error      z       p-value
  --------------------------------------------------------
  educ        0.0795682    0.00402396    19.77    5.03e-87  ***
  kids       -0.955750     0.0366670    -26.07    8.94e-150 ***

  cut1        0.656064     0.0677545     9.683    3.56e-22  ***
  cut2        2.63981      0.0764705    34.52     3.93e-261 ***

Mean dependent var   0.657000   S.D. dependent var   0.599852
Log-likelihood      -3903.096   Akaike criterion     7814.193
Schwarz criterion    7840.262   Hannan-Quinn         7823.330

Number of cases 'correctly predicted' = 3107 (62.1%)
Likelihood ratio test: Chi-square(2) = 1054.95 [0.0000]

Test for normality of residual -
  Null hypothesis: error is normally distributed
  Test statistic: Chi-square(2) = 0.108987
  with p-value = 0.946965
```

The LPM may give bad results

```
Model 2: OLS, using observations 1-5000
Dependent variable: work
Heteroskedasticity-robust standard errors, variant HC1

             coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const       0.251884     0.0292683     8.606    1.00e-17  ***
  educ        0.0354561    0.00169676   20.90     4.97e-93  ***
  kids       -0.431542     0.0152135   -28.37     3.02e-164 ***

Mean dependent var    0.657000   S.D. dependent var   0.599852
Sum squared resid     1457.253   S.E. of regression   0.540024
R-squared             0.189855   Adjusted R-squared   0.189531
F(2, 4997)            651.6896   P-value(F)           3.2e-252
Log-likelihood       -4012.480   Akaike criterion     8030.960
Schwarz criterion     8050.512   Hannan-Quinn         8037.813
```

To test joint significance, omit does not work

- you cannot use omit to test joint significance using the Wald test

omit educ kids –wald

```
-

No independent variables left after omissions

Error executing script: halting
 > omit educ kids --wald
```

- the result of the LR test is available in the output, and we can still conduct it "by hand"

LR test for the significance of all parameters

```
# estimating unrestricted model and storing loglikelihood
probit work educ kids --quiet
scalar lur= $lnl

# estimating restricted model and storing loglikelihood
probit work const --quiet
scalar lr= $lnl

# computing the LR statistic and p-value
scalar LR=2*(lur-lr)
scalar pval = pvalue(X, 1, LR)
```

- The result coincides with the output in `probit`:

```
? printf " LR = %.8g\n p-value = %.8g\n", LR,pval
 LR = 1054.9521
 p-value = 2.0415848e-231
```

Marginal Effects

Partial effects on predicted probabilities

- How best to interpret results from ordered models?
  - latent variable equation
  - $\frac{\partial \mathrm{E}(y|x)}{\partial x_j} = \frac{\partial \Pr(y=0|x)}{\partial x_j} \times 0 + \frac{\partial \Pr(y=1|x)}{\partial x_j} \times 1 + \ldots + \frac{\partial \Pr(y=m|x)}{\partial x_j} \times m$
- Alternatively, you might just want to report the effect on the probability of observing the ordered categories
  - if x_j is discrete we compute as in the binary case the discrete change in the predicted probabilities associated with changing x_j.
  - typically, partial effects for intermediate probabilities are quantitatively small and often statistically insignificant.

Marginal Effects: Discrete Change

Discrete change from x_0 to x_1

- estimate the model and store $\widehat{\beta}$ and $\widehat{\alpha}$
- predict index functions $x_0'\widehat{\beta}$ and $x_1'\widehat{\beta}$
- generate the individual marginal effects

$$\Delta\widehat{\Pr}\left(y=0|x\right)=\Phi\left(x_0'\widehat{\beta}-\widehat{\alpha}_1\right)-\Phi\left(x_1'\widehat{\beta}-\widehat{\alpha}_1\right)=-\Delta\Phi\left(x'\widehat{\beta}-\widehat{\alpha}_1\right)$$

$$\Delta\widehat{\Pr}\left(y=1|x\right)=\Delta\Phi\left(x'\widehat{\beta}-\widehat{\alpha}_1\right)-\Delta\Phi\left(x'\widehat{\beta}-\widehat{\alpha}_2\right)$$

$$\vdots$$

$$\Delta\widehat{\Pr}\left(y=m|x\right)=\Delta\Phi\left(x'\widehat{\beta}-\widehat{\alpha}_{m-1}\right)$$

- compute the averages

Example: Having an *Additional* Kid

```
# marginal effects of having an additional kid
probit work educ kids --quiet
genr beta=$coeff[1:2]
genr alpha=$coeff[3:4]
series kids0=kids
matrix x0={educ,kids0}
series kids1=kids0+1
matrix x1={educ,kids1}
series x1b = x1*beta
series x0b = x0*beta

series Mg_kid0 = (cdf(N,x0b-alpha[1])-cdf(N,x1b-alpha[1]))
series Mg_kid1 = (cdf(N,x1b-alpha[1])-cdf(N,x0b-alpha[1])) \
                -(cdf(N,x1b-alpha[2])-cdf(N,x0b-alpha[2]))
series Mg_kid2 = (cdf(N,x1b-alpha[2])-cdf(N,x0b-alpha[2]))
```

summary *Mg_kid** --simple

```
summary Mg_kid* --simple
Summary statistics, using the observations 1 - 5000

                    Mean        Minimum        Maximum      Std. Dev.
Mg_kid0          0.31409        0.13797        0.36398       0.057924
Mg_kid1         -0.25621       -0.31999       -0.13648       0.058980
Mg_kid2        -0.057872       -0.13917     -0.0014979       0.050060
```

- having an extra child increases the probability of not working by over 30 percentage points
- this increase comes from reductions of 26.6 percentage points in the probability of working part-time and of 5.8 percentage points in the probability of working full time

Marginal Effects: Infinitesimal Change

Calculus approximation

- estimate the model and store $\widehat{\beta}$ and $\widehat{\alpha}$
- predict index function $x'\widehat{\beta}$ and compute its average $\overline{x'\widehat{\beta}}$
- generate the calculus approximation:

$$\frac{\partial \widehat{\Pr}(y=0|x)}{\partial x_j} = -\phi\left(\overline{x'\widehat{\beta}} - \widehat{\alpha}_1\right)\widehat{\beta}_j$$

$$\frac{\partial \widehat{\Pr}(y=1|x)}{\partial x_j} = \left(\phi\left(\overline{x'\widehat{\beta}} - \widehat{\alpha}_1\right) - \phi\left(\overline{x'\widehat{\beta}} - \widehat{\alpha}_2\right)\right)\widehat{\beta}_j$$

$$\frac{\partial \widehat{\Pr}(y=2|x)}{\partial x_j} = \phi\left(\overline{x'\widehat{\beta}} - \widehat{\alpha}_2\right)\widehat{\beta}_j$$

Example of Calculus Approximation

```
# one extra year of education: calculus approximation
genr beta=$coeff[1:2]
genr alpha=$coeff[3:4]
matrix X={educ,kids}
series Xb=X*beta
scalar meanXb=mean(Xb)
scalar Mg_educ0=-pdf(N,meanXb-alpha[1])*$coeff(educ)
scalar Mg_educ1= (pdf(N,meanXb-alpha[1]) \
         -pdf(N,meanXb-alpha[2]))*$coeff(educ)
scalar Mg_educ2= pdf(N,meanXb-alpha[2])*$coeff(educ)
```

```
? printf " Mg_educ0 = %.8g\n Mg_educ1 = %.8g\n Mg_educ2 = %.8g\n", \
          Mg_educ0, Mg_educ1, Mg_educ2

 Mg_educ0 = -0.030702828
 Mg_educ1 = 0.023540482
 Mg_educ2 = 0.0071623465
```

Effects on the ordered response variable

Effects on y

- once we have the marginal effects for the probabilities, obtaining the marginal effect on the ordered response variable y is very simple
- For the continuous case:

$$\frac{\partial \widehat{\mathrm{E}}(y|x)}{\partial x_j} = \frac{\partial \widehat{\Pr}(y=1|x)}{\partial x_j} \times 1 + \frac{\partial \widehat{\Pr}(y=2|x)}{\partial x_j} \times 2$$

- For the discrete case:

$$\Delta \widehat{\mathrm{E}}(y|x) = \Delta \widehat{\Pr}(y=1|x) + 2 \times \Delta \widehat{\Pr}(y=2|x)$$

```
# marginal effects on work
scalar EMg_kid=mean(Mg_kid1)+2*mean(Mg_kid2)
scalar EMg_educ=Mg_educ1+2*Mg_educ2
```

```
? printf " EMg_kid = %.8g\n EMg_educ = %.8g\n", EMg_kid, EMg_educ
 EMg_kid = -0.371958
 EMg_educ = 0.037865175
```

Summary

- gretl allows for ML estimation of the ordered probit and ordered logit model
- marginal effects can be easily computed