Ordered Response and Multinomial Logit Estimation

Quantitative Microeconomics

R. Mora

Department of Economics
Universidad Carlos III de Madrid

Outline

Introduction

The Ordered Probit Model with 3 possible outcomes

$$U = x'\beta + \varepsilon, \varepsilon \sim N(0,1)$$

$$y = 0 \text{ if } U \leq \alpha_1$$
$$y = 1 \text{ if } \alpha_1 < U \leq \alpha_2$$
$$y = 2 \text{ if } \alpha_2 < U$$

- We do not observe U, but the choice of each individual among the three alternatives
- This choice is represented by y, which is an ordinal variable (i.e. it does not have cardinal interpretation)
- The aim is to obtain estimates for β, α_1, and α_2

Multinomial logit with three alternatives

$$\Pr(y=0|x) = 1 - \Pr(y=1|x) - \Pr(y=2|x)$$
$$\Pr(y=1|x) = \frac{\exp(x'\beta_1)}{1+\exp(x'\beta_1)+\exp(x'\beta_2)}$$
$$\Pr(y=2|x) = \frac{\exp(x'\beta_2)}{1+\exp(x'\beta_1)+\exp(x'\beta_2)}$$

- The choice is represented by y, which is a qualititative variable (i.e. it has neither cardinal nor ordinal interpretation)
- The aim is to obtain estimates for β_1 and β_2

Ordered Response Estimation

Example: No-work, part-time, and full-time

h^*: desired weekly time of work (in tens of hours)

$$h^* = -0.5 + 0.07 * educ - 1.0 * kids + \varepsilon, \varepsilon \sim N(0,1)$$

- Suppose there are only two possible labor contracts: part-time and full-time contracts:
  - part-time contracts have working time of at most 20 hours per week
  - full-time contracts have working time larger than 20 hours per week
- In our data, we do not observe h^*. We observe:
  - $y = 0$ for all individuals who choose not to work ($h^* \leq 0$).
  - $y = 1$ for those who work part-time ($0 < h^* \leq 2$).
  - $y = 2$ for those who work full time ($2 < h^*$).

Normalization

- Let $y^* = 0.07 * educ - 1.0 * kids + \varepsilon$ so that $h^* = -0.5 + y^*$
- Not working implies that

$$-0.5 + y^* \leq 0 \Rightarrow y^* \leq 0.5$$
$$(\alpha_1 = 0.5)$$

- Working full time implies that

$$-0.5 + y^* > 2 \Rightarrow y^* > 2.5$$

Model without constant

The model can be normalized without a constant

$$y^* = 0.07 * educ - 1.5 * kids + \varepsilon,\ \varepsilon \sim N(0,1)$$

$$y = 0 \text{ if } y^* \leq 0.5$$
$$y = 1 \text{ if } 0.5 < y^* \leq 2.5$$
$$y = 2 \text{ if } 2.5 < y^*$$

Probabilities in extremes

$$\begin{aligned}\Pr(y=0|x) &= \Pr\left(x'\beta+\varepsilon \le \alpha_1|x\right)\\ &= \Phi\left(-\left(x'\beta-\alpha_1\right)\right)\\ &= 1-\Phi\left(\left(x'\beta-\alpha_1\right)\right)\end{aligned}$$

$$\begin{aligned}\Pr(y=2|x) &= \Pr\left(\alpha_2 < x'\beta+\varepsilon|x\right)\\ &= \Pr\left(\varepsilon > -\left(x'\beta-\alpha_2\right)|x\right)\\ &= \Phi\left(\left(x'\beta-\alpha_2\right)\right)\end{aligned}$$

(Note that $\Phi(a) = 1-\Phi(-a)$, because the normal distribution is symmetric)

Intermediate Probabilities

- In the 3-alternative example, there is only one intermediate probability:

$$\begin{aligned}\Pr(y=1|x) &= \Pr\left(\alpha_1 < x'\beta+\varepsilon \le \alpha_2|x\right)\\ &= \Pr\left(\varepsilon > -\left(x'\beta-\alpha_1\right), \varepsilon \le -\left(x'\beta-\alpha_2\right)|x\right)\end{aligned}$$

- Since $-(x'\beta-\alpha_1) < -(x'\beta-\alpha_2)$:

$$\begin{aligned}\Pr\left(\varepsilon \le -\left(x'\beta-\alpha_2\right)|x\right) - \Pr\left(\varepsilon < -\left(x'\beta-\alpha_1\right)|x\right)\\ = \Phi\left(-\left(x'\beta-\alpha_2\right)\right) - \Phi\left(-\left(x'\beta-\alpha_1\right)\right)\end{aligned}$$

$$\Pr(y=1|x) = \Phi\left(x'\beta-\alpha_1\right) - \Phi\left(x'\beta-\alpha_2\right)$$

Conditional Expectation and OLS

$$
\begin{aligned}
\Pr\left(y=0|x\right) &= 1-\Phi\left(\left(x'\beta-\alpha_1\right)\right)\\
+\Pr\left(y=1|x\right) &= \Phi\left(x'\beta-\alpha_1\right)-\Phi\left(x'\beta-\alpha_2\right)\\
+\Pr\left(y=2|x\right) &= \Phi\left(\left(x'\beta-\alpha_2\right)\right)
\end{aligned}
$$

....

$$\sum \Pr\left(y=j|x\right)=1 \text{ (Probabilities sum to one)}$$

- In general, OLS does not work because the conditional expectation of the dependent variable is not linear any more:

$$
\begin{aligned}
E\left(y|x\right) &= 0\times\Pr\left(y=0|x\right)+1\times\Pr\left(y=1|x\right)+2\times\Pr\left(y=2|x\right)\\
&= \Pr\left(y=1|x\right)+2\times\Pr\left(y=2|x\right)\\
&= \Phi\left(x'\beta-\alpha_1\right)+\Phi\left(x'\beta-\alpha_2\right)
\end{aligned}
$$

ML Estimation

The probability of any observation can be expressed as

$$\Pr(y|x) = \left(1-\Phi\left(x'\beta-\alpha_1\right)\right)^{1(y=0)} \times \left(\Phi\left(x'\beta-\alpha_1\right)-\Phi\left(x'\beta-\alpha_2\right)\right)^{1(y=1)} \times \left(\Phi\left(x'\beta-\alpha_2\right)\right)^{1(y=2)}$$

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

$$\mathrm{L}(b) = \prod_{i=1}^{N} \left\{ \left(1-\Phi\left(x_i'b-\alpha_1\right)\right)^{1(y_i=0)} \times \left(\Phi\left(x_i'b-\alpha_1\right)-\Phi\left(x_i'b-\alpha_2\right)\right)^{1(y_i=1)} \times \left(\Phi\left(x_i'b-\alpha_2\right)\right)^{1(y_i=2)} \right\}$$

Marginal Effects

Options for reporting results:

- When the latent variable equation has a simple interpretation, this is probably a good way of reporting the model.
- Alternatively, marginal effects for the probabilities of each of the categories can be computed.
    - When the independent variable is discrete, marginal effects can be computed as in the binary case.
- Finally, we can estimate the effect on the expected value of the observed variable.

Marginal Effects when regressor is continuous

Marginal effects when regressor is continuous

$$\frac{\partial \Pr(y=0|x)}{\partial x_j} = -\phi\left(x'\beta - \alpha_1\right)\beta_j$$

$$\frac{\partial \Pr(y=1|x)}{\partial x_j} = \left(\phi\left(x'\beta - \alpha_1\right) - \phi\left(x'\beta - \alpha_2\right)\right)\beta_j$$

$$\frac{\partial \Pr(y=2|x)}{\partial x_j} = \phi\left(x'\beta - \alpha_2\right)\beta_j$$

The Multinomial Model

Random Utility Model

- Assume that there are three transport alternatives: bus, car, train:

$$U_b = x_b'\beta_b + \varepsilon_b$$
$$U_c = x_c'\beta_c + \varepsilon_c$$
$$U_t = x_t'\beta_t + \varepsilon_t$$

  where $\{\varepsilon_b, \varepsilon_c, \varepsilon_t\}$ are the effects on utility unobserved by the econometrician

- Let $y = 0$ if bus is chosen, $y = 1$ if car is chosen, and $y = 2$ if train is chosen.
  - y does not have any cardinal or ordinal meaning!

Model re-parametrization

$$\varepsilon_{01} \equiv \varepsilon_b - \varepsilon_c, \qquad \varepsilon_{02} \equiv \varepsilon_b - \varepsilon_t$$
$$x'\beta_{01} \equiv x_b'\beta_b - x_c'\beta_c, \qquad x'\beta_{02} \equiv x_b'\beta_b - x_t'\beta_t$$

- Assumption: $\{\varepsilon_{01}, \varepsilon_{02}\} \sim F$, where F is symmetric.

$$\Pr(y=0|x) = F\left(x'\beta_{01}, x'\beta_{02}\right)$$
$$\Pr(y=1|x) = F\left(-x'\beta_{01}, x'(\beta_{02}-\beta_{01})\right)$$
$$\Pr(y=2|x) = F\left(-x'\beta_{02}, -x'(\beta_{02}-\beta_{01})\right)$$

The Multinomial Logit

When vector $\{\varepsilon_b, \varepsilon_c, \varepsilon_t\}$ has a extreme value distribution, then we have the **Multinomial Logit**:

$$\Pr(y=0|x) = 1 - \Pr(y=1|x) - \Pr(y=2|x)$$

$$\Pr(y=1|x) = \frac{\exp(x'\beta_1)}{1+\exp(x'\beta_1)+\exp(x'\beta_2)}$$

$$\Pr(y=2|x) = \frac{\exp(x'\beta_2)}{1+\exp(x'\beta_1)+\exp(x'\beta_2)}$$

- OLS does not work because the conditional expectation is not linear
- ML estimation gives consistent and asymptotically normal estimates
  - In `gretl`, the `logit` command estimates the multinomial logit model when the dependent variable is not binary and is discrete if we use the `--multinomial` option

Example: Car, Bicycle, Train

$$\Pr(car|x) = 1 - \Pr(bicycle|x) - \Pr(train|x)$$
$$\Pr(bicycle|x) = \frac{\exp(x'\beta_1)}{1+\exp(x'\beta_1)+\exp(x'\beta_2)}$$
$$\Pr(train|x) = \frac{\exp(x'\beta_2)}{1+\exp(x'\beta_1)+\exp(x'\beta_2)}$$

where

$$x'\beta_1 = 1 - 0.1 * age - .1 * income - 3 * kids + 5 * center$$
$$x'\beta_2 = 1 - .1 * income + 1.5 * kids + 2 * center$$

Note that $\beta_1 = \beta_{bicycle} - \beta_{car}$ and that $\beta_2 = \beta_{train} - \beta_{car}$.
What is the relative value of *train* against *bicycle*?

$$x'\beta_1 = 1 - 0.1 * age - .1 * income - 3 * kids + 5 * center$$
$$x'\beta_2 = 1 - .1 * income + 1.5 * kids + 2 * center$$

- $\beta_{1,age} = -.1 < 0$: As people age, it brings more value to use the car than the bicycle.
- $\beta_{1,income} = \beta_{2,income} = -.1 < 0$: The higher the income, , the more likely is car over train and bicycle.
- $\beta_{1,kids} = -3, \beta_{2,kids} = 1.5$: then $\beta_{car,kids} - \beta_{train,kids} = -1.5 < 0$ y $\beta_{bicycle,kids} - \beta_{train,kids} = -3 - 1.5 = -4.5 < 0$: The more kids, the more likely is train over car and bicycle.
- If the journey goes through the city center, train and bicycle are more likely than car (and how is train valued versus bicycle?)

logit transport const age income kids center --multinomial

Model 3: Multinomial Logit, using observations 1–5000
Dependent variable: transport, Standard errors based on Hessian

	Coefficient	Std. Error	z	p-value
transport = 2				
const	0.522376	0.501274	1.0421	0.2974
age	−0.0969857	0.00890363	−10.8928	0.0000
income	−0.0957303	0.00589953	−16.2268	0.0000
kids	−2.78035	0.222090	−12.5190	0.0000
center	5.29508	0.422875	12.5216	0.0000
transport = 3				
const	0.867895	0.197401	4.3966	0.0000
age	0.00590025	0.00495462	1.1909	0.2337
income	−0.100911	0.00351216	−28.7320	0.0000
kids	1.42980	0.0869553	16.4429	0.0000
center	1.89417	0.0899837	21.0502	0.0000

Mean dependent var	2.253800	S.D. dependent var	0.915616
Log-likelihood	−2687.761	Akaike criterion	5395.521
Schwarz criterion	5460.693	Hannan–Quinn	5418.363

Number of cases 'correctly predicted' = 3787 (75.7 percent)
Likelihood ratio test: $\chi^2(8) = 3712.464$ [0.0000]

Summary

- We cannot estimate ordered or multinomial logit by OLS.
- Maximum Likelihood estimation gives consistent and asymptotically normal results.
- Marginal effects can be computed as in the simpler binary cases.