The Ordered and Multinomial Models

Quantitative Microeconomics

R. Mora

Department of Economics
Universidad Carlos III de Madrid

Outline

Motivation

We consider the following two extensions from binary dependent models:

- Ordered response models: The dependent variable takes a number of **finite** and **discrete** values that contain **ordinal information**.
- Multinomial response models: The dependent variable takes a number of **finite** and **discrete** values that **DO NOT** contain **ordinal information**.

As in the probit and logit cases, the dependent variable is not strictly continuous. Estimation will be carried out using the ML estimator.

Examples of ordered models

- Credit rating, using seven categories, from "absolutely not credit worthy" to "credit worthy".
- Decision to remain inactive, to work part-time, or to work full-time.
- In an income regression, income levels are coded in intervals: $[0,1000)$, $[1000,1500)[1500,2000),[2000,\infty)$
- On value statements, several answers with ordinal content: "completely disagree", "disagree", "somewhat agree", "completely agree"

Examples of multinomial models

- Choice of transport mode: train, bus, car
- Economic status: inactive, unemployed, self-employed, employee
- Education field choice: hard science, health sciences, social sciences, humanities

Ordered Response Models

- The two standard models are the ordered probit and the ordered logit.
- The approach is equivalent: we simply use for the ordered probit the normal CDF $\Phi()$ and for the ordered logit the logistic CDf $\Lambda()$.
- OLS does not work because the dependent variable does not have cardinal meaning:
  - credit worthiness: 0,1,2,3,4,5: the change from 0 to 1 does not have to be "equivalent" to the change from 4 to 5.
  - activity: inactive=0, part-time=1, full-time=2: While inactive is zero hours of work, in practice code 1 reflects any hours of work between 1 and (usually) 30 hours of work, and code 2 reflects more 30 hours of work. This implies that there is no proportionality in going from 0 to 1 and going from 1 to 2.

Simplification

- Binary choice models (LPM, probit, logit) could potentially be used by grouping all categories into two major ones,
- This is the case when the sample is small and the ordinal categories can be logically be grouped in two major categories.
- In some cases, this is probably a very bad idea (income intervals).

- Consider three observed outcomes: $y = 0, 1, 2$.
- Consider the latent variable model without a constant:

$$y^* = \beta_1 x_1 + ... + \beta_k x_k + \varepsilon$$

where $\varepsilon \sim \mathcal{N}(0,1)$.

- Define two cut-off points: $\alpha_1 < \alpha_2$
- We do not observe y^*, but we observe choices according to the following rule

$$\begin{aligned} y &= 0 \text{ if } y^* \leq \alpha_1 \\ y &= 1 \text{ if } \alpha_1 < y^* \leq \alpha_2 \\ y &= 2 \text{ if } \alpha_2 < y^* \end{aligned}$$

Example: activity

- $y = 0$ if inactive, $y = 1$ if part-time, $y = 2$ if full-time
- $y^* = \beta_e \times educ + \beta_k \times kids + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0,1)$
- Then

$$
\begin{aligned}
y &= 0 \text{ if } \beta_e \times educ + \beta_k \times kids + \varepsilon \leq \alpha_1 \\
y &= 1 \text{ if } \alpha_1 < \beta_e \times educ + \beta_k \times kids + \varepsilon \leq \alpha_2 \\
y &= 2 \text{ if } \alpha_2 < \beta_e \times educ + \beta_k \times kids + \varepsilon
\end{aligned}
$$

- Note that we could alternatively introduce a constant β_0 and assume that $\alpha_1 = 0$.

Interpretation

- As in other nonlinear models, marginal effects can be computed to learn about the partial effects of a small change in explanatory variable x_j.
- For ordered models we can compute marginal effects on the predicted probabilities along the same principles.

Partial effects on predicted probabilities

- For binary choice models, we focused on the effects on the probability that y is equal to one.
- In the ordered models, things are not so simple: we now have more than two outcomes:

$$\frac{\partial \Pr(y=0|x)}{\partial x_j} = -\phi\left(x'\beta-\alpha_1\right)\beta_j$$
$$\frac{\partial \Pr(y=1|x)}{\partial x_j} = \left(\phi\left(x'\beta-\alpha_1\right)-\phi\left(x'\beta-\alpha_2\right)\right)\beta_j$$
$$\frac{\partial \Pr(y=2|x)}{\partial x_j} = \phi\left(x'\beta-\alpha_2\right)\beta_j$$

- if x_j is discrete we compute as in the binary case the discrete change in the predicted probabilities associated with changing x_j.

Partial Effects

- The partial effect of x_j on the predicted probability of
  - the highest outcome has the same sign as β_j.
  - the lowest outcome has the opposite sign to β_j
  - intermediate outcomes cannot, in general, be inferred from the sign of β_j.
- The last results is due to two offsetting effects. Suppose $\beta_j > 0$ and you increase x_j. The intermediate category
  - may become more likely since the probability of the lowest category falls.
  - may also become less likely because the probability of the highest category increases.
- Typically, partial effects for intermediate probabilities are quantitatively small and often statistically insignificant.

Discussion

How best to interpret results from ordered models?

- One option is to look at the estimated β-parameters, emphasizing the underlying latent variable equation with which we started.
- Another option might be to look at the effect on the expected value of the ordered response variable, e.g.

$$\frac{\partial \mathrm{E}(y|x)}{\partial x_j} = \frac{\partial \Pr(y=0|x)}{\partial x_j} \times 0 + \frac{\partial \Pr(y=1|x)}{\partial x_j} \times 1 + \frac{\partial \Pr(y=2|x)}{\partial x_j} \times 2$$

  This may make a lot of sense if y is a numerical variable, as in the income variable.
- Alternatively, you might just want to report the effect on the probability of observing the ordered categories.

Multinomial Response

- The dependent variable is such that
  - more than two outcomes are possible
  - the outcomes cannot be ordered in any natural way.
- Again, we could bunch two or more categories and so construct a binary outcome variable from the raw data, but in doing so, we throw away potentially interesting information.
- OLS is also not a good model in this context.
- However, the logit model for binary choice can be extended to model more than two outcomes.

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

If $x_b'\beta_b + \varepsilon_b \geq \max\ \{x_c'\beta_c + \varepsilon_c, x_t'\beta_t + \varepsilon_t\}$ then $y = 0$

If $x_c'\beta_c + \varepsilon_c > \max\ \left\{x_b'\beta_b + \varepsilon_b, x_t'\beta_t + \varepsilon_t\right\}$ then $y = 1$

If $x_t'\beta_t + \varepsilon_t > \max\ \{x_c'\beta_c + \varepsilon_c, x_t'\beta_t + \varepsilon_t\}$ then $y = 2$

Notation

- We have two unobserved independent effects

$$\varepsilon_{01} = \varepsilon_b - \varepsilon_c$$
$$\varepsilon_{02} = \varepsilon_b - \varepsilon_t$$

  - note that $\varepsilon_{12} = \varepsilon_c - \varepsilon_t = \varepsilon_{02} - \varepsilon_{01}$

- Define

$$x_b'\beta_b - x_c'\beta_c = x'\beta_{01}$$
$$x_b'\beta_b - x_t'\beta_t = x'\beta_{02}$$

Assumption

$$\{\varepsilon_{01}, \varepsilon_{02}\} \sim F$$

where F is symmetric.

Then

$$\begin{aligned}\Pr(y=0|x) &= \Pr\left(x'\beta_{01}+\varepsilon_{01} \geq 0, x'\beta_{02}+\varepsilon_{02} \geq 0|x\right) \\ &= \Pr\left(\varepsilon_{01} \geq -\left(x'\beta_{01}\right), \varepsilon_{02} \geq -\left(x'\beta_{02}\right)|x\right)\end{aligned}$$

Given symmetry,

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

Multinomial Logit

- We must model the probability that an individual belongs to category j conditional to having characteristics x:

$$\Pr(y = j|x)$$

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

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

- The main difference compared to the binary logit is that there are now two parameter vectors, β_1 and β_2
- in the general case with J possible responses, there are $J-1$ parameter vectors.
- This makes interpretation of the coefficients more difficult than for binary choice models.

Interpretation with three alternatives

The easiest case to think about is where β_{1j} and β_{2j} have the same sign.

- If β_{1j} and β_{2j} are positive then an increase in the variable x_j it less likely that the individual belongs to category 0...
- and $\Pr(y_i = 1|x_i) + \Pr(y_i = 2|x_i)$ increases
- to know how this total increase is allocated between these two probabilities, we need to look at the marginal effects: the partial derivative is very complex and the marginal effect $\frac{\partial \Pr(y=1|x)}{\partial x_j}$ may in fact be negative even if β_{1j}!

Independence of irrelevant alternatives (IIA)

- One important limitation of the multinomial logit is that the ratio of any two probabilities l and m depends only on the parameter vectors β_l and β_m and the explanatory variables x

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

- The inclusion or exclusion of other categories is irrelevant to the ratio of the two probabilities.
- This behavior is referred to as the "independence of irrelevant alternatives", and it can lead to counter-intuitive behavior

Example: IIA can be counter-intuitive

- Individuals can commute to work by three transportation means: blue bus, red bus, or train.
- Individuals choose one of these alternatives, and the econometrician estimates a multinomial logit modeling this decision, and obtains an estimate of

$$\frac{\Pr(y = red|x)}{\Pr(y = train|x)} = \exp\left(x'\left(\beta_{red} - \beta_{train}\right)\right)$$

- Suppose that the bus company now removes the blue bus from the set of options, do you think that $\frac{\Pr(y=red|x)}{\Pr(y=train|x)}$ would be the same as before?

Other multinomial models

- There are lots of other econometric models that can be used to model multinomial response models:
  - multinomial probit,
  - conditional logit,
  - nested logit

- They are beyond the scope of the course.

Summary

- When the dependent variable has a finite number of discrete values, we can extend the probit and logit models
  - When the dependent variable entails some ordinal information, then we can use ordered probit and logit models
  - When the dependent variable does not contain any ordinal information, we can use multinomial models. One such example is the multinomial logit.

- These are all nonlinear models, and they can all be estimated by MLE.