Limited Dependent Variable Models II

Fall 2008

Models with Multiple Choices

- The binary response model was dealing with a decision problem with two alternatives. This can be generalized to one with more than two alternatives.

Models with Multiple Choices

- The binary response model was dealing with a decision problem with two alternatives. This can be generalized to one with more than two alternatives.
  - choice of transportation: car, bus, bicycle...

Models with Multiple Choices

- The binary response model was dealing with a decision problem with two alternatives. This can be generalized to one with more than two alternatives.
  - choice of transportation: car, bus, bicycle...
  - occupational choice: blue/white collar, professional, craft...

Models with Multiple Choices

- The binary response model was dealing with a decision problem with two alternatives. This can be generalized to one with more than two alternatives.
  - choice of transportation: car, bus, bicycle...
  - occupational choice: blue/white collar, professional, craft...
  - assignments of credit ratings to corporate bonds

Models with Multiple Choices

- The binary response model was dealing with a decision problem with two alternatives. This can be generalized to one with more than two alternatives.
  - choice of transportation: car, bus, bicycle...
  - occupational choice: blue/white collar, professional, craft...
  - assignments of credit ratings to corporate bonds
- We will examine two broad types of choice sets, **ordered** and **unordered**.

Models with Multiple Choices

- The binary response model was dealing with a decision problem with two alternatives. This can be generalized to one with more than two alternatives.
  - choice of transportation: car, bus, bicycle...
  - occupational choice: blue/white collar, professional, craft...
  - assignments of credit ratings to corporate bonds
- We will examine two broad types of choice sets, **ordered** and **unordered**.
  - (unordered) choice of transportation

Models with Multiple Choices

- The binary response model was dealing with a decision problem with two alternatives. This can be generalized to one with more than two alternatives.
  - choice of transportation: car, bus, bicycle...
  - occupational choice: blue/white collar, professional, craft...
  - assignments of credit ratings to corporate bonds
- We will examine two broad types of choice sets, **ordered** and **unordered**.
  - (unordered) choice of transportation
  - (ordered) credit rating to corporate bonds

Models with Multiple Choices

- The binary response model was dealing with a decision problem with two alternatives. This can be generalized to one with more than two alternatives.
  - choice of transportation: car, bus, bicycle...
  - occupational choice: blue/white collar, professional, craft...
  - assignments of credit ratings to corporate bonds
- We will examine two broad types of choice sets, **ordered** and **unordered**.
  - (unordered) choice of transportation
  - (ordered) credit rating to corporate bonds
- As we shall see, quite different econometric techniques are used for the two types of models.

Multinomial Logit Model

- We first consider *unordered*-choice models. Two models are common again, logit and probit. Due to the need to evaluate multiple integrals of the normal distribution, the logit model becomes more popular.
- Let $Y_i \in \{0, 1, 2, ..., K\}$. Then, the multinomial logit model specifies the following probabilities for alternatives: for $j = 0, 1, ..., K$,

$$\Pr(Y_i = j) = \frac{\exp\left(\beta_{j0} + \beta_{j1} X_i\right)}{\sum_{k=0}^{K} \exp\left(\beta_{k0} + \beta_{k1} X_i\right)}.$$

Multinomial Logit Model

- The parameters in the model are *identifiable* up to normalization. To see this, multiply all the coefficients by a factor λ. Do the probabilities change?
- A convenient normalization is setting the coefficients of one alternative, say $j = 0$, to zero. Thus,

$$
\begin{aligned}
\Pr(Y_i = 0) &= \frac{1}{1 + \sum_{k=1}^{K} \exp(\beta_{k0} + \beta_{k1} X_i)}, \\
\Pr(Y_i = j) &= \frac{\exp\left(\beta_{j0} + \beta_{j1} X_i\right)}{1 + \sum_{k=1}^{K} \exp(\beta_{k0} + \beta_{k1} X_i)}, \quad \text{for } j \neq 0.
\end{aligned}
$$

Independence of Irrelevant Alternatives

- Note that the log odds-ratios between two alternatives are only expressed as a function of the parameters of the two alternatives, but not of those for any other alternatives.

$$\log\left(\frac{\Pr(Y_i = k)}{\Pr(Y_i = 0)}\right) = \beta_{k0} + \beta_{k1}X_i.$$

Independence of Irrelevant Alternatives

- Note that the log odds-ratios between two alternatives are only expressed as a function of the parameters of the two alternatives, but not of those for any other alternatives.

$$\log\left(\frac{\Pr\left(Y_i = k\right)}{\Pr\left(Y_i = 0\right)}\right) = \beta_{k0} + \beta_{k1} X_i .$$

- This is called the **Independence of Irrelevant Alternatives (IIA)**. This is a convenient property as regards estimation. From a behavioral viewpoint, this is not so attractive.

Independence of Irrelevant Alternatives

- Note that the log odds-ratios between two alternatives are only expressed as a function of the parameters of the two alternatives, but not of those for any other alternatives.

$$\log\left(\frac{\Pr(Y_i = k)}{\Pr(Y_i = 0)}\right) = \beta_{k0} + \beta_{k1} X_i.$$

- This is called the **Independence of Irrelevant Alternatives (IIA)**. This is a convenient property as regards estimation. From a behavioral viewpoint, this is not so attractive.
- This property follows from the independence and homoskedasticity of errors in the original structural model.

$$Y_{ki}^* = \beta_{k0} + \beta_{k1} X_i + u_{ki}$$

$$Y_i = k \ \text{ if } Y_{ki}^* > Y_{ji}^*, \text{ for all } j \neq k,$$

where u_{ki} follows the Type I extreme value distribution

$$F(u_{ki}) = \exp(-\exp(-u_{ki})).$$

Example: Violation of IIA

- Consider a choice of transportation between car and *red* bus, which are only currently available transportation in a city. Suppose that the choice probabilities are equal:

$$\Pr(car) = \Pr(red\ \ bus) = 0.5 \Longrightarrow \frac{\Pr(car)}{\Pr(red\ \ bus)} = 1.$$

Example: Violation of IIA

- Consider a choice of transportation between car and *red* bus, which are only currently available transportation in a city. Suppose that the choice probabilities are equal:

$$\Pr(car) = \Pr(red\ \ bus) = 0.5 \Longrightarrow \frac{\Pr(car)}{\Pr(red\ \ bus)} = 1.$$

- Suppose a city government introduces *blue* bus that is identical to red bus except for color. It is reasonable that the behavior of car drivers will not be affected at all by the introduction of blue bus. And people using bus are spilt evenly between blue and red bus. Thus,

$$\Pr(car) = 0.5,\ \Pr(red\ \ bus) = \Pr(blue\ \ bus) = 0.25,$$

$$\frac{\Pr(car)}{\Pr(red\ \ bus)} = 2.$$

Example: Violation of IIA

- Consider a choice of transportation between car and *red* bus, which are only currently available transportation in a city. Suppose that the choice probabilities are equal:

$$\Pr(car) = \Pr(red\ \ bus) = 0.5 \Longrightarrow \frac{\Pr(car)}{\Pr(red\ \ bus)} = 1.$$

- Suppose a city government introduces *blue* bus that is identical to red bus except for color. It is reasonable that the behavior of car drivers will not be affected at all by the introduction of blue bus. And people using bus are spilt evenly between blue and red bus. Thus,

$$\Pr(car) = 0.5,\ \Pr(red\ \ bus) = \Pr(blue\ \ bus) = 0.25,$$

$$\frac{\Pr(car)}{\Pr(red\ \ bus)} = 2.$$

- However, the IIA implies that the odd ratios should be the same wehther another alternative exists or not, which is obviously violated in this example.

Marginal Effects in Multinomial Logit Model

- β_{k1} can be interpreted as the marginal effect of X on the log odds-ratio of alternative k to the baseline alternative, 0.
- The marginal effect of X on the probability of choosing alternative k can be expressed as

$$\frac{\partial \Pr(Y_i = k)}{\partial X_i} = \Pr(Y_i = k) \left[\beta_{k1} - \sum_{j=0}^{K} \Pr(Y_i = j) \beta_{j1} \right]$$

  Hence, the marginal effect of X on alternative k involves not only the parameters of k but also the ones of all other alternatives.
- Note that the marginal effect need not have the same sign of β_{k1}.

ML Estimation

- The estimation method is a direct extension of the maximum likelihood method for a binary response model.
- Suppose that we observed N_j number of $Y = j$, for $j = 0, 1, ..., K$, and $N = \sum_{j=0}^{K} N_j$.
- The log-likelihood function from this data is written

$$\log L\left(\{\beta_{k0}, \beta_{k1}\}_{k=1}^{K}\right) = \sum_{j=0}^{K} \sum_{i=1}^{N_j} \log \Pr\left(Y_i = j\right).$$

- The MLE of $\{\beta_{k0}, \beta_{k1}\}_{k=1}^{K}$ are found by maximizing the log-likelihood with respect to each of $\{\beta_{k0}, \beta_{k1}\}_{k=1}^{K}$.

Example: Choice of Dwelling

- We analyze choice of dwelling between housing (H), apartment (A) and low-cost flat (F): for $k = H, A, F$

$$U_{ki} = \beta_{k0} + \beta_{k1} Age_i + \beta_{k2} Sex_i + \beta_{k3} \log Income_i + u_{ki}.$$

Choice of House			Choice of Apartment		
	Coeff.	Std. Err.		Coeff.	Std. Err.
age	0.027	0.010	age	0.002	0.012
sex	-0.409	0.259	sex	-0.305	0.297
log income	1.358	0.186	log income	1.495	0.216
constant	-10.753	1.560	constant	-11.703	1.820

A brief Intro. of Nested Logit Model

- When IIA fails, an alternative to the multinomial logit model will be a multivariate probit model.
- A more useful alternative is **nested logit model**, which basically groups the alternatives into subgroups that allow the variance to differ across teh groups while maintaining the IIA assumption within the groups.
- For example, it is useful to think of choice of transportation as a two-level choice problem. First, a person chooses between car and bus. If a bus is to be selected, then he chooses between red bus and blue bus.
- Thus, the probability of choosing red bus is

$$\Pr\left(red\;\; bus\right) = \Pr\left(red \middle| bus\right) \Pr\left(bus\right),$$

which is one component in the likelihood function.

Ordered Models I

- In the previous multinomial logit model, the choices were not ordered. For instance, we cannot rank car, bus or bicycle in a meaningful way.

Ordered Models I

- In the previous multinomial logit model, the choices were not ordered. For instance, we cannot rank car, bus or bicycle in a meaningful way.
- In some situations, we have a natural ordering of the outcomes even if we cannot express them as a continuous variable:

Ordered Models I

- In the previous multinomial logit model, the choices were not ordered. For instance, we cannot rank car, bus or bicycle in a meaningful way.
- In some situations, we have a natural ordering of the outcomes even if we cannot express them as a continuous variable:
    - (survey response) No / Somehow / Yes; Low / Medium / High.

Ordered Models I

- In the previous multinomial logit model, the choices were not ordered. For instance, we cannot rank car, bus or bicycle in a meaningful way.
- In some situations, we have a natural ordering of the outcomes even if we cannot express them as a continuous variable:
    - (survey response) No / Somehow / Yes; Low / Medium / High.
    - (unemployment) Unemployed / Part time / Full time.

Ordered Models I

- In the previous multinomial logit model, the choices were not ordered. For instance, we cannot rank car, bus or bicycle in a meaningful way.
- In some situations, we have a natural ordering of the outcomes even if we cannot express them as a continuous variable:
  - (survey response) No / Somehow / Yes; Low / Medium / High.
  - (unemployment) Unemployed / Part time / Full time.
- We will use ordered (probit and logit) models to analyze these situations.

Ordered Models II

- The data will be coded by usually assinging non-negative integer values:

$$Y_i = \begin{cases} 0 & \text{if} \quad \text{No (Low, Unemployed)} \\ 1 & \text{if} \quad \text{Somehow (Medium, Part time)} \\ 2 & \text{if} \quad \text{Yes (High, Full time)} \end{cases} .$$

Ordered Models II

- The data will be coded by usually assinging non-negative integer values:

$$Y_i = \begin{cases} 0 & \text{if} \quad \text{No (Low, Unemployed)} \\ 1 & \text{if} \quad \text{Somehow (Medium, Part time)} \\ 2 & \text{if} \quad \text{Yes (High, Full time)} \end{cases} .$$

- As before, it is assumed that the outcome Y_i is governed by a latent variable Y_i^* such that

$$Y_i^* = \beta_0 + \beta_1 X_i + u_i$$

$$Y_i = \begin{cases} 0 & Y_i^* < 0 \\ 1 & 0 \leq Y_i^* < \mu \\ 2 & Y_i^* \geq \mu \end{cases} ,$$

μ is a threshold parameter that should be estimated along with β_0 and β_1.

Ordered Models II

- The data will be coded by usually assinging non-negative integer values:

$$Y_i = \begin{cases} 0 & \text{if} \quad \text{No (Low, Unemployed)} \\ 1 & \text{if} \quad \text{Somehow (Medium, Part time)} \\ 2 & \text{if} \quad \text{Yes (High, Full time)} \end{cases} .$$

- As before, it is assumed that the outcome Y_i is governed by a latent variable Y_i^* such that

$$Y_i^* = \beta_0 + \beta_1 X_i + u_i$$

$$Y_i = \begin{cases} 0 & Y_i^* < 0 \\ 1 & 0 \leq Y_i^* < \mu \\ 2 & Y_i^* \geq \mu \end{cases} ,$$

  μ is a threshold parameter that should be estimated along with β_0 and β_1.
- Depending on the assumption of distribution of error u, the model is called *ordered probit or logit model*.

Ordered Probit Model

- We assume that u_i follows independently and identically the standard normal distribution.
- Then the probability of each outcome is derived with the normal cumulative distribution function, Φ.

$$\begin{aligned}
\Pr(Y_i = 0) &= \Phi(-\beta_0 - \beta_1 X_i) \\
\Pr(Y_i = 1) &= \Phi(\mu - \beta_0 - \beta_1 X_i) - \Phi(-\beta_0 - \beta_1 X_i) \\
\Pr(Y_i = 2) &= 1 - \Phi(\mu - \beta_0 - \beta_1 X_i).
\end{aligned}$$

- And we just need to construct the likelihood function.
- In some statistical packages (STATA) $-\beta_0$ and $\mu - \beta_0$ are reported as two threshold values.

Marginal Effects

- As before, we need to be careful in interpreting the meaning of coefficients in the ordered model.

Marginal Effects

- As before, we need to be careful in interpreting the meaning of coefficients in the ordered model.
- The marginal effects of X on the choice probabilities are

$$\begin{aligned}
\frac{\partial \Pr(Y=0)}{\partial X} &= -\beta_1 \phi(-\beta_0 - \beta_1 X_i), \\
\frac{\partial \Pr(Y=1)}{\partial X} &= \beta_1 [\phi(-\beta_0 - \beta_1 X_i) - \phi(\mu - \beta_0 - \beta_1 X_i)], \\
\frac{\partial \Pr(Y=2)}{\partial X} &= \beta_1 \phi(\mu - \beta_0 - \beta_1 X_i).
\end{aligned}$$

Marginal Effects

- As before, we need to be careful in interpreting the meaning of coefficients in the ordered model.
- The marginal effects of X on the choice probabilities are

$$\begin{aligned}
\frac{\partial \Pr(Y=0)}{\partial X} &= -\beta_1 \phi(-\beta_0 - \beta_1 X_i), \\
\frac{\partial \Pr(Y=1)}{\partial X} &= \beta_1 \left[\phi(-\beta_0 - \beta_1 X_i) - \phi(\mu - \beta_0 - \beta_1 X_i)\right], \\
\frac{\partial \Pr(Y=2)}{\partial X} &= \beta_1 \phi(\mu - \beta_0 - \beta_1 X_i).
\end{aligned}$$

- Note that if $\beta_1 > 0$, then $\frac{\partial \Pr(Y=0)}{\partial X} < 0$ and $\frac{\partial \Pr(Y=2)}{\partial X} > 0$.

Marginal Effects

- As before, we need to be careful in interpreting the meaning of coefficients in the ordered model.
- The marginal effects of X on the choice probabilities are

$$\begin{aligned}
\frac{\partial \Pr(Y=0)}{\partial X} &= -\beta_1 \phi(-\beta_0 - \beta_1 X_i), \\
\frac{\partial \Pr(Y=1)}{\partial X} &= \beta_1 [\phi(-\beta_0 - \beta_1 X_i) - \phi(\mu - \beta_0 - \beta_1 X_i)], \\
\frac{\partial \Pr(Y=2)}{\partial X} &= \beta_1 \phi(\mu - \beta_0 - \beta_1 X_i).
\end{aligned}$$

- Note that if $\beta_1 > 0$, then $\frac{\partial \Pr(Y=0)}{\partial X} < 0$ and $\frac{\partial \Pr(Y=2)}{\partial X} > 0$.
  - If X has a positive effect on the latent variable, then by increasing X, fewer individuals will choose outcome 0, $Y_i = 0$.

Marginal Effects

- As before, we need to be careful in interpreting the meaning of coefficients in the ordered model.
- The marginal effects of X on the choice probabilities are

$$\begin{aligned}
\frac{\partial \Pr(Y=0)}{\partial X} &= -\beta_1 \phi(-\beta_0 - \beta_1 X_i), \\
\frac{\partial \Pr(Y=1)}{\partial X} &= \beta_1 \left[\phi(-\beta_0 - \beta_1 X_i) - \phi(\mu - \beta_0 - \beta_1 X_i)\right], \\
\frac{\partial \Pr(Y=2)}{\partial X} &= \beta_1 \phi(\mu - \beta_0 - \beta_1 X_i).
\end{aligned}$$

- Note that if $\beta_1 > 0$, then $\frac{\partial \Pr(Y=0)}{\partial X} < 0$ and $\frac{\partial \Pr(Y=2)}{\partial X} > 0$.
  - If X has a positive effect on the latent variable, then by increasing X, fewer individuals will choose outcome 0, $Y_i = 0$.
  - Similarly, more individuals will choose outcome 2, $Y_i = 2$.

Marginal Effects

- As before, we need to be careful in interpreting the meaning of coefficients in the ordered model.
- The marginal effects of X on the choice probabilities are

$$
\begin{aligned}
\frac{\partial \Pr(Y=0)}{\partial X} &= -\beta_1 \phi(-\beta_0 - \beta_1 X_i), \\
\frac{\partial \Pr(Y=1)}{\partial X} &= \beta_1 \left[\phi(-\beta_0 - \beta_1 X_i) - \phi(\mu - \beta_0 - \beta_1 X_i)\right], \\
\frac{\partial \Pr(Y=2)}{\partial X} &= \beta_1 \phi(\mu - \beta_0 - \beta_1 X_i).
\end{aligned}
$$

- Note that if $\beta_1 > 0$, then $\frac{\partial \Pr(Y=0)}{\partial X} < 0$ and $\frac{\partial \Pr(Y=2)}{\partial X} > 0$.
  - If X has a positive effect on the latent variable, then by increasing X, fewer individuals will choose outcome 0, $Y_i = 0$.
  - Similarly, more individuals will choose outcome 2, $Y_i = 2$.
  - In the intermediate case, the fraction of individuals will either increase or decrease, depending on the relative size of the inflow from outcome 0 and the outflow to outcome 2.

Example: Environmental Concern

- The attitudes toward environments in the survey before can be coded as

$$Y_i = 0 \text{ (no concern)}, 1 \text{ (somehow)}, 2 \text{ (very concerned)}.$$

We use the ordered probit model with age, sex, log income and smell as explanatory variables.

	Age	Sex	log income	smell	Threshold I	Threshold II
Coeff.	0.021	0.023	0.274	0.363	0.096	2.984
Std. Err.	0.005	0.125	0.080	0.138	0.746	0.697

- The computation of marginal effects and richer framework will be done in the tutorial class.