Last lecture slides covered case where dependent variable reflected choice between 2 options.

Multinomial choice models are used when more than two options.

Much of the theory and intuition is similar to bivariate choice, but some new issues arise.

Most commonly used models are extensions of probit and logit.

Multinomial probit, multinomial logit and conditional logit.

Reading: Koop (2008, pages 287-296) and Gujarati (chapter 9)
Example taken from marketing literature
Data from $N = 136$ households in Rome, Georgia.
Data taken from supermarket optical scanners.
Which of four brands of crackers (Sunshine, Keebler, Nabisco and Private label) did each household purchase
As explanatory variable: the price of all four brands of crackers in the store at time of purchase as explanatory variables.
For future reference, note the following properties of this example:

- Note 1: The explanatory variable is a characteristic of the crackers NOT the households
- Attribute of the choice NOT an attribute of the individual
- Note 2: The options the household faces are unordered
- In contrast, some applications involve an ordered choice set
  - E.g. Consumer survey: Is a product Bad, Fair, Good, Excellent?
  - E.g. Rating agencies rate bonds as B, B++, A, A++, etc.
- Ordered (also called Ordinal) choice models will be covered in next topic
- Multinomial logit and probit are for unordered choices
Example 2

- Example taken from transportation economics literature
- Survey $N$ individuals on whether they take car, public transport or bicycle to work
- Explanatory variables:
  - Age
  - Income
  - Commuting time
- Note: age and income are characteristics of the individual
- Commuting time varies over both individuals and choices
- Commuting time using car is different from using bicycle
- Also individual 1 may live in city centre (commuting time by bicycle short), individual 2 may live in suburbs (commuting time by bicycle long)
Multinomial Choice Models

- Binary choice models had $Y_i$ dummy variable (takes on values 0 or 1).
- Multinomial choice $Y_i$ can take on values 0, 1, $\ldots$, $J$.
- $J + 1$ options individual can choose between.
- Subscripts $j = 0, \ldots, J$ will indicate options.
- Subscripts $i + 1, \ldots, N$ will indicate individuals who make choices.
Random Utility Framework with Multiple Choices

- Individual will make the choice which yields the highest utility.
- $U_{ji}$: utility of individual $i$ when choosing option $j$
- Remember: with binomial choice we said: choice 1 is made if $U_1 \geq U_0$ which is same as $U_1 - U_0 \geq 0$
- Difference in utility between choices 1 and 0 that matter
- With multinomial choice: choose one option as benchmark (option 0).
- Utility of every other option compared to benchmark
- It does not matter which choice you make as the benchmark.
Notation for utility differences between making a choice and a benchmark choice:

\[ Y_{ji}^* = U_{ji} - U_{0i} \]

- \( Y_{ji}^* \) is unobservable
- We observe the choice made:

\[ Y_i = j \] if individual \( i \) makes choice \( j \).

Relationship between \( Y_{ji}^* \) and \( Y_i \):

- if \( Y_{ji}^* < 0 \) for \( j = 1, \ldots, J \), then individual \( i \) chooses the benchmark alternative and \( Y_i = 0 \).
- Otherwise, individual \( i \) makes choice which yields the highest value for \( Y_{ji}^* \) and \( Y_i = j \).
Multinomial probit and logit models begin with:

\[ Y_{ji}^* = \alpha_j + \beta_{j1} X_{1ij} + \beta_{j2} X_{2ij} + \ldots + \beta_{jk} X_{kij} + \epsilon_{ji} \]

Logit and probit arise out of different assumptions about the errors.

Before we get to this models, look carefully at the subscripts in equation.

This is actually \( J \) different regressions: one for comparing each option to benchmark.

In each regression we have different coefficients.

\( \alpha_j \) is the intercept in regression involving the difference in utility between option \( j \) and option 0.

\( \beta_{j1} \) is the coefficient on first explanatory variable in this regression.

etc.
Warning of Issues Concerning Explanatory Variables

- Note: $i$ and $j$ subscripts appear on the explanatory variables.
- Explanatory variables can be characteristics of individuals or choices.
- E.g. characteristics of choices (e.g. price of each cracker brand) could explain why individual makes a particular choice.
- Or characteristics of an individual (e.g. age of individual) could explain choice.
- E.g. older people tend to choose old-fashioned Nabisco brand.
- Multinomial logit and multinomial probit can handle explanatory variables both types of explanatory variables.
Usually explanatory variables will either be attributes of choice or of individuals, but not both
So usually it will either be $X_{1i}$ or $X_{1j}$
However, there are exceptions (see commuting time in Example 2)
But I write $X_{1ij}, \ldots, X_{kij}$ as general notation to cover both options
More confusion: some textbooks use multinomial probit or multinomial logit only with individual characteristics (so $X_{1i}, \ldots, X_{ki}$)

Koop (2008) sometimes uses this notation, but empirical example uses explanatory variables which are attributes of choices (price of cracker brands)

Sorry for this confusion

Gujarati (2010) has multinomial logit model as having explanatory variables which are characteristics of individuals

Gretl User Guide defines multinomial logit with explanatory variables which are characteristics of individuals

But Gretl User Guide defines multinomial probit with explanatory variables which are characteristics of choices or individuals
A common practice in many textbooks is to use multinomial probit/logit when you have explanatory variables which are attributes of individuals.

Conditional logit and conditional probit used when explanatory variables are attributes of choices (so $X_{1j}, \ldots, X_{kj}$).

For reasons to be explained later, conditional logit/probit have trouble when explanatory variables are attributes of individuals.

To attempt to minimize confusion, notes below follow this division.

Bottom line: multinomial probit/multinomial logit can handle explanatory variables which are attributes of either individuals or choices.

Conditional probit/logit can only handle explanatory variables which are attributes of choices.
To simplify, work with one explanatory variable which contains attributes of individual:

\[ Y_{ji}^* = \beta_j X_i + \varepsilon_{ji} \]

Multinomial logit and probit based on this set of \( J \) regressions, but differ in the assumptions made about the errors.

Multinomial probit model: multivariate Normal distribution

\[ \Pr(Y_i = j) \] can be obtained using properties of Normal distribution.

Multinomial logit: type-1 extreme value distribution (do not worry about what this is)

Key thing: for multinomial logit model probability of individual \( i \) making choice \( j \) is:

\[ \Pr(Y_i = j) = \frac{\exp(\beta_j X_i)}{1 + \sum_{s=1}^{J} \exp(\beta_s X_i)} \]
Pros and Cons of Multinomial Probit

- Remember: different error in each of $J$ regressions involving each utility difference.
- Errors in different equations could be correlated with one another.
- E.g. If choice between 3 options then have two regressions involving utility differences $Y_{1i}^*$ and $Y_{2i}^*$.
- Could have $\text{corr}(\varepsilon_{1i}, \varepsilon_{2i}) \neq 0$ and multivariate Normal distribution allows for this.
- This gives multinomial probit nice flexible properties.
• Number of correlations is \( \frac{J(J-1)}{2} \)
• If \( J \) large, number of such correlations grows can grow huge
• Multinomial probit has to estimate all these correlations
• For this reason, multinomial probit is typically used only if the number of options is relatively small.
• Gretl only does 3 options (\( J = 2 \)) and calls it *bivariate probit*
Pros and Cons of Multinomial Logit

- When $J$ is large, multinomial logit model is more popular
- Assumes errors in the different equations are uncorrelated with one another.
- Easier to estimate, but uncorrelatedness may be undesirable
- It implies choice probabilities must satisfy an *independence of irrelevant alternatives* (or IIA) property.
Suppose that, initially, commuter has a choice between car \((Y = 0)\) or public transport \((Y = 1)\).

IIA property relates to odds ratio: \(\frac{Pr(Y=0)}{Pr(Y=1)}\)

IIA says that this odds ratio will be the same, regardless of what the other options are.

E.g. suppose, initially, \(\frac{Pr(Y=0)}{Pr(Y=1)} = 1\)

Commuter equally likely to take the car or public transport

Must be \(Pr(Y = 0) = Pr(Y = 1) = 0.5\)
Now suppose bicycle lane is constructed so commuters can now bicycle to work \((Y = 2)\) is now an option.

IIA property says that addition of new alternative does not alter the fact that \(\frac{\Pr(Y=0)}{\Pr(Y=1)} = 1\).

Is IIA property is reasonable?

It is possible in this case

E.g. if 20% of commuters start bicycling and these cyclists drawn equally from other two options

Could end up with \(\Pr(Y = 2) = 0.2\) and \(\Pr(Y = 0) = \Pr(Y = 1) = 0.40\)

Note this still implies \(\frac{p(y=0)}{p(y=1)} = 1\), so IIA is satisfied
Red Bus-Blue Bus Problem (an example where IIA is not satisfied)

- Commuter originally has choice between car \((Y = 0)\) or Red Bus \((Y = 1)\)
- Suppose \(p(Y = 0) = p(Y = 1) = \frac{1}{2}\), thus \(\frac{Pr(Y=0)}{Pr(Y=1)} = 1\)
- Now bus company paints half of their buses Blue
- Treat Blue Bus as a new option
- Probably not reasonable to assume IIA holds
E.g. suppose initially $p(Y = 0) = p(Y = 1) = \frac{1}{2}$ and, thus, $\frac{Pr(Y = 0)}{Pr(Y = 1)} = 1$.

Equally likely to take car or bus

Since Blue Bus is virtually identical to Red Bus, adding third option likely leave commuter just as likely to take the car to work

This would imply $p(Y = 0) = 0.50$ and $p(Y = 1) = p(Y = 2) = 0.25$.

Hence, introduction of Blue Bus option implies $\frac{p(Y = 0)}{p(Y = 1)} = 2$.

Odds ratio is changed by adding new option:

This violates IIA and is not allowed for by multinomial logit model.
Is IIA reasonable or not?
- Depends on your empirical application.
- Sometimes it is reasonable, but other times not.
- Note: there are extensions of multinomial logit models (not covered in this course) that do not have restrictive IIA property:
- *Nested logit model* and *mixed logit model* are two such extensions.
Estimation of Multinomial Choice Models

- We will not describe in detail estimation and testing with multinomial choice models.
- Maximum likelihood methods are used.
- Gretl will estimate them.
- Provide point estimates of coefficients, P-values for telling whether coefficients are significant.
- Various predicted probabilities and marginal effects.
- We will see how this is done in the computer session.
Use Empirical Example 1 (cracker data set)

$N = 136$ households

Each chooses between four brands of crackers: Sunshine, Keebler, Nabisco and Private label

We select "Private label" as the benchmark alternative

Remember: model utility difference relative to choice 0
Hence, multinomial logit model has three equations: the first based on utility difference between Sunshine and Private label crackers, etc.

All equations include intercept and price of all four brands of crackers as explanatory variables.

Note: be careful with respect points made on: “Warning of Issues Concerning Explanatory Variables”

This example has explanatory variables which differ across alternatives and three equations with different coefficients.

This is not quite the same as the models in Gujarati (chapter 9).
Interpretation of Multinomial Logit Results

- Results in Table 9.3 (next 2 slides)
- Note: three separate sets of regression results (in the panels of the table labelled Sunshine, Keebler and Nabisco, respectively).
- Like logit, hard to interpret size of the coefficients in multinomial logit
- But sign of any coefficient can provide some information.
- Coefficients interpreted as marginal effects relating to utility differences
- Positive coefficient in equation \( j \) means explanatory variable has positive effect on utility difference
- If this utility difference increases then individual more likely to choose alternative \( j \) relative to the benchmark choice.
- Negative coefficient make you less likely to choose option \( j \)
<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>P-val for $\beta_{ji} = 0$</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sunshine</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>$-10.06$</td>
<td>0.15</td>
<td>$[-23.59, 3.46]$</td>
</tr>
<tr>
<td>$\beta_{11}$</td>
<td>$-7.98$</td>
<td>0.01</td>
<td>$[-13.78, -2.20]$</td>
</tr>
<tr>
<td>$\beta_{12}$</td>
<td>$12.39$</td>
<td>0.04</td>
<td>$[0.77, 24.02]$</td>
</tr>
<tr>
<td>$\beta_{13}$</td>
<td>$0.37$</td>
<td>0.91</td>
<td>$[-5.83, 6.57]$</td>
</tr>
<tr>
<td>$\beta_{14}$</td>
<td>$4.83$</td>
<td>0.36</td>
<td>$[-5.54, 15.20]$</td>
</tr>
<tr>
<td><strong>Keebler</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>$-2.53$</td>
<td>0.73</td>
<td>$[-16.90, 11.85]$</td>
</tr>
<tr>
<td>$\beta_{21}$</td>
<td>$-3.10$</td>
<td>0.30</td>
<td>$[-9.01, 2.81]$</td>
</tr>
<tr>
<td>$\beta_{22}$</td>
<td>$-0.60$</td>
<td>0.92</td>
<td>$[-12.99, 2.81]$</td>
</tr>
<tr>
<td>$\beta_{23}$</td>
<td>$1.15$</td>
<td>0.70</td>
<td>$[-4.67, 6.97]$</td>
</tr>
<tr>
<td>$\beta_{24}$</td>
<td>$5.33$</td>
<td>0.25</td>
<td>$[-3.66, 14.32]$</td>
</tr>
</tbody>
</table>
Table 9.3 (cont.): Multinomial Logit Results for Cracker Data

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>P-val for $\beta_{ji} = 0$</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nabisco</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>−7.01</td>
<td>0.09</td>
<td>[−15.09, 1.07]</td>
</tr>
<tr>
<td>$\beta_{31}$</td>
<td>−1.38</td>
<td>0.48</td>
<td>[−5.23, 2.48]</td>
</tr>
<tr>
<td>$\beta_{32}$</td>
<td>5.57</td>
<td>0.12</td>
<td>[−1.37, 12.50]</td>
</tr>
<tr>
<td>$\beta_{33}$</td>
<td>0.86</td>
<td>0.65</td>
<td>[−2.84, 4.56]</td>
</tr>
<tr>
<td>$\beta_{34}$</td>
<td>4.72</td>
<td>0.06</td>
<td>[−0.23, 9.67]</td>
</tr>
</tbody>
</table>
Example: estimate of $\beta_{11}$ is negative ($-7.98$)

$\beta_{11}$ is from Sunshine crackers equation

Utility difference between Sunshine and Private label crackers is the dependent variable

$\beta_{11}$ is coefficient on explanatory variable “Price of Sunshine Crackers”.

Interpretation: “if price of Sunshine crackers increases (holding the price of other crackers constant), then the probability of choosing Sunshine crackers over Private label crackers will fall”.

Sensible result (i.e. if the price of a good increases, consumers are less likely to buy it).

P-value for $\beta_{11}$ says this coefficient is significant

Many other coefficients are statistically insignificant
With logit model we used the odds ratio.

With multinomial logit, define odds ratios for each option relative to the benchmark option:

$$\frac{Pr(Y_i = j)}{Pr(Y_i = 0)}$$

With logit, we said $\beta$ can be interpreted as a marginal effect in terms of the log odds ratio.
• Another thing to do: calculate predicted probability of choosing each alternative for an individual with a specified set of characteristics

• Yet another: calculate the probability that each individual will choose each option

• I.e. calculate $\Pr(Y_i = j)$ for $i = 1, .., N$ and $j = 0, .., J$

• Note: we have $N = 136$ individuals and four options: hence 544 different $\Pr(Y_i = j)$.

• One thing you can do is present summary statistics of these 544 probabilities.

• This is done in Table 9.4.
<table>
<thead>
<tr>
<th>Probability of Choosing:</th>
<th>Mean</th>
<th>Standard deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunshine</td>
<td>0.08</td>
<td>0.11</td>
<td>0.01</td>
<td>0.64</td>
</tr>
<tr>
<td>Keebler</td>
<td>0.07</td>
<td>0.03</td>
<td>0.02</td>
<td>0.16</td>
</tr>
<tr>
<td>Nabisco</td>
<td>0.60</td>
<td>0.10</td>
<td>0.31</td>
<td>0.80</td>
</tr>
<tr>
<td>Private Label</td>
<td>0.25</td>
<td>0.11</td>
<td>0.02</td>
<td>0.49</td>
</tr>
</tbody>
</table>
Table 9.4 tells us things like:

- Nabisco is the most popular brand
- There is one individual about whom we can say “individual is 80% sure to choose Nabisco”
- But nothing higher than 80% for any choice (lots of uncertainty in this data set)
- Results for the multinomial probit model are very similar so we will not present them
- Interpretation is as in probit, only now we have 3 equations
The conditional logit model is a special case of multinomial logit.
I thought about not covering it (since in a sense we have already covered it above).
Also Gretl does not seem to do it.
But many other textbooks, research papers and computer packages use conditional logit so I will explain here.
Return to original Random Utility Model setup:

\[ Y_{ji}^* = \alpha_j + \beta_{j1} X_{1i} + \beta_{j2} X_{2i} + \ldots + \beta_{jk} X_{ki} + \varepsilon_{ji} \]

For multinomial probit and multinomial logit (as defined earlier in this set of slides) this leads to \( j \) different equations.

Conditional logit collapses this down to one equation.

Hence, special case of multinomial logit with \( \beta_j \) being the same for \( j = 1, \ldots, J \).

Conditional logit is more compact and involves estimating fewer coefficients.

But more restrictive and only appropriate with certain types of data.
Remember multinomial logit said:

$$\Pr \left( Y_i = j \right) = \frac{\exp \left( \beta_j X_i \right)}{1 + \sum_{s=1}^{J} \exp \left( \beta_s X_i \right)}$$

Conditional logit says:

$$\Pr \left( Y_i = j \right) = \frac{\exp \left( \beta X_{ji} \right)}{\sum_{s=1}^{J} \exp \left( \beta X_{si} \right)}$$

- Examine subscripts carefully
- With conditional logit, only a single $\beta$
- In conditional logit, $X_{ji}$ – explanatory variables must have a $j$ (and possibly an $i$) subscript (must vary across options)
- Explanatory variables cannot vary only across individuals.
To see why this is so consider an example:

One explanatory variable varies across choices \((X_j)\) and a second one varies across individuals \((Z_i)\)

Conditional logit model would say:

\[
\Pr(Y_i = j) = \frac{\exp(\beta_1 X_{ij} + \beta_2 Z_i)}{\sum_{s=1}^{J} \exp(\beta_1 X_{is} + \beta_2 Z_i)}
\]

But you can never estimate \(\beta_2\).

\(\beta_2\) disappears from \(\Pr(Y_i = j)\) and thus from the likelihood function — MLE not possible
Proof

\[ \Pr(Y_i = j) = \frac{\exp(\beta_1 X_{ij} + \beta_2 Z_i)}{\sum_{s=1}^{J} \exp(\beta_1 X_{is} + \beta_2 Z_i)} \]

\[ = \frac{\exp(\beta_1 X_{ij}) \exp(\beta_2 Z_i)}{\sum_{s=1}^{J} \exp(\beta_1 X_{is}) \exp(\beta_2 Z_i)} \]

\[ = \frac{\exp(\beta_1 X_{ij}) \exp(\beta_2 Z_i)}{\exp(\beta_2 Z_i) \sum_{s=1}^{J} \exp(\beta_1 X_{is})} \]

\[ = \frac{\exp(\beta_1 X_{ij})}{\sum_{s=1}^{J} \exp(\beta_1 X_{is})}. \]
Since conditional logit is a special case of multinomial logit models, we will not discuss estimation.

Many computer packages (but not Gretl) can do MLE.

As with all the qualitative choice models, it is hard directly to interpret conditional logit coefficients.

However, we can estimate various predicted probabilities or various marginal effects with respect to the probabilities.
Multinomial choice models are used when dependent variable is choice between several options.

Multinomial probit and multinomial logit are the most popular multinomial choice models.

If choice is between $J + 1$ options, both will have $J$ equations.

Conditional logit is a special case of multinomial logit with only 1 equation.

Maximum likelihood estimation is the most common way of estimating all these models.

Explanatory variables can be characteristics of the options or the individuals (or both) and can lead to different models (so pay careful attention).

Care must be taken with interpretation of coefficients/marginal effects.