Economics 217 - Multinomial Choice Models

- So far, most extensions of the linear model have centered on either a binary choice between two options (work or don't work) or censoring options.
- Many questions in economics involve consumers making choices between more than two varieties of goods
    - Ready-to-eat cereal
    - Vacation destinations
    - Type of car to buy
- Firms also have such multinomial choices
    - In which country to operate
    - Where to locate a store
    - Which CEO to hire
- Techniques to evaluate these questions are complex, but widely used in practice. Generally, they are referred to as **Discrete Choice Models**, or **Multinomial Choice Models**

Multinomial Choice - The basic framework

- Suppose there are individuals, indexed by i
- They choose from J options of a good, and may only choose one option.
- If they choose option k, then individual i receives U_{ik} in utility, where

$$U_{ik} = V_{ik} + \epsilon_{ik}$$

  - V_{ik} is observable utility (to the econometrician). This can be linked to things like product characteristics, demographics, etc..
  - ϵ_{ik} is random utility. The econometrician doesn't see this, but knows its distribution. This actually makes the problem a bit more reasonable to characterize empirically

- Utility maximization - individual i chooses option k if

$$U_{ik} > U_{ij} \;\; \forall \;\; j \neq k$$

- This maximization problem involves comparing observable utility for each option, while accounting for random utility.

Multinomial Choice - The basic framework

- From here, there are a variety of techniques that one can use to estimate multinomial choice models
- Multinomial Logit is the easiest, and will be derived below
  - Assumes a particular functional form that has questionable properties, but produces closed form solutions
  - There are two ways to derive the multinomial logit - we will go over the easier approach, though I have also derived the second approach in the notes.
- Nested Logit is more realistic:
  - Consumers choose between larger groups (car vs. truck) before making more refined choices (two-door vs. four-door)
  - Also yields closed form solutions, but results can depend on choices over "nests"
- Additional extensions to multinomial choice are beyond this course, but can be used if you understand the basic assumptions
  - Multinomial Probit (requires heavy computation)
  - Random coefficients logit (variation in how agents value attributes of choices)

Multinomial Distribution

- Recall the binomial distribution:

$$f(y;p) = \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}$$

- Remember that p is the probability some event (eg. unemployment) occurs, and y is the number of times the event occurs after n attempts.
  - $n-y$ is the number of times the event does not occur.
- When there are more than two choices, the distribution is generalized as **multinomial**
- Defining π_j as the probability that option j is chosen, the multinomial distribution is written as:

$$\begin{aligned} f(y;p) &= n! \prod_{j=1}^{J} \frac{1}{y_j!} \pi_j^{y_j} \\ &= \frac{n!}{y_1! y_2! \cdots y_{J-1}! y_J!} \pi_1^{y_1} \pi_2^{y_2} \cdots \pi_{J-1}^{y_{J-1}} \pi_J^{y_J} \end{aligned}$$

- This is the PDF that is used for maximum likelihood. We wish to estimate J π_j's.
  - Do you think we can? Or do you think that we need to?
- Let's now take the next step and link the likelihood function to data.

Multinomial Logit - Derivation

- Recall for the Logit model we link the log odds ratio to data

$$\log\left(\frac{p_i}{1-p_i}\right) = \mathbf{x}_i^T \beta$$

- We exponentiate and rearrange to get:

$$p_i = \frac{\exp\left(\mathbf{x}_i^T \beta\right)}{1+\exp\left(\mathbf{x}_i^T \beta\right)}$$

- We must extend this link to having multiple options in the multinomial model.
  - Since there is a linear dependency in our probabilities (ie. they sum to one), we must choose a **reference group**
- We write the log odds ratio relative to the reference group ($j = 1$) as:

$$\log\left(\frac{\pi_{ij}}{\pi_{i1}}\right) = \mathbf{x}_{ij}^T \beta_j$$

- Note that the relative probability is specific to j: β_j.
  - β_j: The effect of some covariate on the choice between j and 1 *may* vary by j.
  - $\mathbf{x}_{ij}^T$ is a vector covariates for i that *may* vary by j.
  - Eg. Price matters for choice between compact cars, but not between compact and luxury.

Multinomial Logit - Derivation

- Exponentiate and solve for π_{ij}

$$\pi_{ij} = \pi_{i1} \exp\left(\mathbf{x}_{ij}^T \beta_j\right)$$

- Next, use the requirement that all probabilities sum to 1

$$\pi_{i1} + \sum_{j=2}^{J} \pi_{ij} = 1$$

- Substituting for π_{ij}, we get:

$$\pi_{i1} + \sum_{j=2}^{J} \pi_{i1} \exp\left(\mathbf{x}_{ij}^T \beta_j\right) = 1$$

- Solving for π_{i1}

$$\pi_{i1} = \frac{1}{1 + \sum_{j=2}^{J} \exp\left(\mathbf{x}_{ij}^T \beta_j\right)}$$

- Thus, the probability of option j, π_{ij}, is

$$\pi_{ij} = \frac{\exp\left(\mathbf{x}_{ij}^T \beta_j\right)}{1 + \sum_{s=2}^{J} \exp\left(\mathbf{x}_{is}^T \beta_s\right)}$$

Multinomial Logit - Assumptions

- The multinomial logit formula is pretty simple

$$\pi_{ij} = \frac{\exp\left(\mathbf{x}_{ij}^T \beta_j\right)}{1 + \sum_{s=2}^{J} \exp\left(\mathbf{x}_{is}^T \beta_s\right)}$$

- The multinomial logit has a pretty sharp property that is usually not good in practice: Independence of Irrelevant Alternatives (IIA)
- Precisely, when choosing between two goods, substitution with other goods does not matter
- To see IIA in practice, take the ratio of probabilities between some good j and another k

$$\frac{\pi_{ij}}{\pi_{ik}} = \frac{\exp\left(\mathbf{x}_{ij}^T \beta_j\right)}{\exp\left(\mathbf{x}_{ik}^T \beta_k\right)}$$

- Thus, the relative probabilities of two outcomes do not depend on the other $J-2$ outcomes.
- Techniques such as multinomial probit, and nested logit, avoid this strong prediction.

Multinomial Logit - Estimation in R

- There are a few packages in R to estimate the multinomial logit.
  - **mlogit** is the best.
- The package also includes a number of datasets that we can use to demonstrate the model. Since it is pretty simple, we will use the dataset "Cracker".
- After loading mlogit, you can call the data internal to the package via the following command:

```
data("Cracker", package = "mlogit")
str(Cracker)
```

- Each row represents an individual, and "choice" represents the chosen brand. This will be the outcome variable.
- For each brand of cracker, the dataset contains the following information
  - **price** observed for individual i
  - Whether or not there was an in-store display observed by individual i, **disp**.
  - Whether or not there was a newspaper ad observed by individual i, **feat**.

Multinomial Logit - Estimation in R

- To setup the data.frame for estimation, you must create an mlogit data object.

```
data_c<-mlogit.data(Cracker, shape="wide", choice="choice",
varying=c(2:13))
```

- "data_c" is the mlogit data object in "wide" format
  - "Cracker" is the original data frame
  - ' shape="wide" ' tells us to list the data in a format that I will describe with R.
  - "varying=c(2:13)" indicates the variables from the dataset that vary by individual (prices they observe, advertisements they see
- To estimate the model, run:

```
m <- mlogit(choice~price+disp+feat,data_c)
summary(m)
```

- Can estimate the model with product specific coefficients using

```
m2 <- mlogit(choice~0|price+disp+feat,data_c)
summary(m2)
```

Extra: Multinomial Logit from Extreme Value Distribution

- Choices are independent of one another, and ϵ_{ik} follows an extreme value I distribution (also known as the Gumbel distribution).

$$\begin{aligned} f(\epsilon_{ik}) &= \exp(-\epsilon_{ik})\exp(-\exp(-\epsilon_{ik})) \\ \Pr(\epsilon < \epsilon_{ik}) = F(\epsilon_{ik}) &= \exp(-\exp(-\epsilon_{ik})) \end{aligned}$$

- Recall that from utility maximization - individual i chooses option k if

$$U_{ik} > U_{ij} \;\; \forall \;\; j \neq k$$

- We now seek the probability that this outcome occurs, which can then be compared empirically to the share of agents that choose option k over all other j.

Extra: Multinomial Logit - Derivation

- First, let's consider option k against some other option j. The probability the consumer purchases k:

$$\Pr\left(U_{ik} > U_{ij}\right) = \Pr\left(V_{ik} + \epsilon_{ik} > V_{ij} + \epsilon_{ij}\right)$$

- Rearranging to isolate ϵ_{ij}

$$\Pr\left(U_{ik} > U_{ij}\right) = \Pr\left(V_{ik} - V_{ij} + \epsilon_{ik} > \epsilon_{ij}\right)$$

- This simply says that the difference in observable utility plus ϵ_{ik} is greater than ϵ_{ij}. Put differently, unobserved utility in option j is not sufficient to make-up for the other factors influencing the decision between k and j.

- Imposing the CDF of the Gumbel distribution, and treating ϵ_{ik} as a conditioning variable, we have:

$$\Pr\left(U_{ik} > U_{ij} \middle| \epsilon_{ik}\right) = F\left(V_{ik} - V_{ij} + \epsilon_{ik}\right)$$

- Given ϵ_{ik}, what is the probability that this occurs for all $j \neq k$?

Extra: Multinomial Logit - Derivation

- Since unobserved utility is independent across goods, the intersection of these events is just their probabilities multiplied together
- So, the probability that k is chosen over j for all $j \neq k$, conditional on ϵ_{ik}, is:

$$\Pr\left(U_{ik} > U_{ij} \ \forall\, j \neq k \,\middle|\, \epsilon_{ik}\right) = \prod_{j \neq k} F\left(V_{ik} - V_{ij} + \epsilon_{ik}\right)$$

- For the final step before some algebra, recall that this is a *conditional probability*. We still need to account for the possible values of ϵ_{ik}
- Formally, the unconditional probability that k is chosen, P_{ik}, is written as:

$$P_{ik} = \Pr\left(U_{ik} > U_{ij} \ \forall\, j \neq k\right) = \int \Pr\left(U_{ik} > U_{ij} \ \forall\, j \neq k \,\middle|\, \epsilon_{ik}\right) f\left(\epsilon_{ik}\right) d\epsilon_{ik}$$

- Basically, what we're doing is taking each $\int \Pr\left(U_{ik} > U_{ij} \ \forall\, j \neq k \,\middle|\, \epsilon_{ik}\right)$, and then weighting by the pdf $f\left(\epsilon_{ik}\right)$.

Extra: Multinomial Logit - Derivation

- Imposing the solution for the choice of k conditional on ϵ_{ik}:

$$P_{ik} = \int_{-\infty}^{\infty} \prod_{j \neq k} F\left(V_{ik} - V_{ij} + \epsilon_{ik}\right) f\left(\epsilon_{ik}\right) d\epsilon_{ik}$$

- Imposing the parameterization of the extreme value distribution, we have:

$$P_{ik} = \int_{-\infty}^{\infty} \prod_{j \neq k} \exp\left(-\exp\left(-\left(V_{ik} - V_{ij} + \epsilon_{ik}\right)\right)\right) \exp\left(-\epsilon_{ik}\right) \exp\left(-\exp\left(-\epsilon_{ik}\right)\right) d\epsilon_{ik}$$

- Note that since $\exp\left(-\exp\left(-\epsilon_{ik}\right)\right) = \exp\left(-\exp\left(-\left(V_{ik} - V_{ik} + \epsilon_{ik}\right)\right)\right)$, we can simply as:

$$P_{ik} = \int_{-\infty}^{\infty} \prod_{j} \exp\left(-\exp\left(-\left(V_{ik} - V_{ij} + \epsilon_{ik}\right)\right)\right) \exp\left(-\epsilon_{ik}\right) d\epsilon_{ik}$$

- Simplifying this is not too hard, once you note a few convenient features of the extreme value distribution.

Extra: Multinomial Logit - Derivation

- Remember that the product of exponentials is just the exponential of the sums of the exponents

$$\prod_j \exp x_j = \exp\left(\sum_j x_j\right)$$

- Thus,

$$\begin{aligned} P_{ik} &= \int_{-\infty}^{\infty} \prod_j \exp\left(-\exp\left(-\left(V_{ik} - V_{ij} + \epsilon_{ik}\right)\right)\right) \exp\left(-\epsilon_{ik}\right) d\epsilon_{ik} \\ &= \int_{-\infty}^{\infty} \exp\left(-\sum_j \exp\left(-\left(V_{ik} - V_{ij} + \epsilon_{ik}\right)\right)\right) \exp\left(-\epsilon_{ik}\right) d\epsilon_{ik} \end{aligned}$$

- Using a similar rule, we can we can factor out $\exp(\epsilon_{ik})$

$$P_{ik} = \int_{-\infty}^{\infty} \exp\left(-\exp\left(-\epsilon_{ik}\right) \sum_j \exp\left(-\left(V_{ik} - V_{ij}\right)\right)\right) \exp\left(-\epsilon_{ik}\right) d\epsilon_{ik}$$

- The next step is tricky. What is the relationship between $-\exp\left(-\epsilon_{ik}\right)$ and $\exp\left(-\epsilon_{ik}\right) d\epsilon_{ik}$?

Extra: Multinomial Logit - Derivation

- Time for a change of variables, where

$$\begin{aligned} t &= -\exp\left(-\epsilon_{ik}\right) \\ dt &= \exp\left(-\epsilon_{ik}\right) d\epsilon_{ik} \\ where\ \ t &\in \left(-\infty, 0\right) \end{aligned}$$

- Thus,

$$\begin{aligned} P_{ik} &= \int_{-\infty}^{\infty} \exp\left(-\exp\left(-\epsilon_{ik}\right) \sum_j \exp\left(-\left(V_{ik} - V_{ij}\right)\right)\right) \exp\left(-\epsilon_{ik}\right) d\epsilon_{ik} \\ &= \int_{-\infty}^{0} \exp\left(t \sum_j \exp\left(-\left(V_{ik} - V_{ij}\right)\right)\right) dt \end{aligned}$$

- Completing the integral:

$$P_{ik} = \left(\frac{\exp\left(t \sum_j \exp\left(-\left(V_{ik} - V_{ij}\right)\right)\right)}{\sum_j \exp\left(-\left(V_{ik} - V_{ij}\right)\right)} \right|_{-\infty}^{0}$$

Extra: Multinomial Logit - Derivation

- And finally, simplify

$$\begin{aligned} P_{ik} &= \frac{1}{\sum_j \exp\left(-\left(V_{ik} - V_{ij}\right)\right)} \\ &= \frac{1}{\sum_j \exp(-V_{ik}) \exp\left(V_{ij}\right)} && \text{use exponent rule} \\ &= \frac{1}{\exp(-V_{ik}) \sum_j \exp\left(V_{ij}\right)} && \text{factor out } \exp(-V_{ik}) \\ &= \frac{\exp(V_{ik})}{\sum_j \exp\left(V_{ij}\right)} && \text{multiply top and bottom by } \exp(V_{ik}) \end{aligned}$$

- From here, we usually assume that observed utility is a function of covariates

$$V_{ij} = X_{ij}\beta$$

- Thus,

$$\boxed{P_{ik} = \frac{\exp(X_{ik}\beta)}{\sum_j \exp\left(X_{ij}\beta\right)}}$$