1 Outline.

1. Motivation

2. Linear Probability Model

3. Bernoulli Dependent Variables

4. Random Utility Interpretation

5. Logit

6. Random Coefficient Logit

2 Motivation.

- In this section, we study the models and computational methods for demand estimation.

- In empirical applications, we most frequently study differentiated product markets.

- Products are characterized by a bundle of characteristics and a price.

- Frequently, the data takes the form of obsevations on individual choices.

- That is, we see a menu of products (bundles of characteristics), prices and a consumer choice.

- The dependent variable takes the form of zeros an ones.

- Demand estimates are a critical input into much applied work.

- Examples:

1. (Residual) demand elasticities are used to measure market power.

2. Demand is used to measure the welfare impacts of policy (CS,CV, EV).

3. Positive economic description of valuation of product characteristics (e.g. school quality, environmental amenities, good reputation in online auctions)

4. First stage input for strucutural model of market equilibrium.

- In this module, you will learn how to estimate the elasticity of demand using the logit and random coefficinets logit.

- In this module, you will be exposed to algorithms for minimization.

- You will apply these models to data from a marketing study of the demand for new computers.

- In the marketing study, participants are asked to rank a menu of alternatives from highest to lowest.

- Such studies are commonly issued by firms in order to forecast demand for new products.

- To protect the identity of the firm, some information is disguised.

3 Linear Probability Model.

- Discussion here follows Chapter 15 of Cameron and Trivedi (grad textbook for applied econometrics).

- Considering a simple case where agents are faced with a dichotomous choice.

- Example, individual level data on consumers $n = 1, ..., N$

- Subscribe/don't subscribe to cable television across markets.

- For each consumer, data on chanel selection, identity of local monopoly provider (e.g. Time/Warner, ATT) and price.

- Let $y_n \in \{0, 1\}$ denote the value of the dependent variable (i.e. subscribe/don't subscribe).

- Let x_n denote exogenous covariates.

$$\begin{aligned} E[y_n|x_n] &= 0 \cdot \Pr[y_n = 0|x_n] + 1 \cdot \Pr[y_n = 1|x_n] \\ &= \Pr[y_n = 1|x_n] \end{aligned}$$

- If we use a linear regression model, then:

$$\begin{aligned} E[y_n|x_n] &= x_n'\beta \\ y_n &= x_n'\beta + \varepsilon_n \\ \varepsilon_n &= (y_n - E[y_n|x_n]) \end{aligned}$$

- Assume $E[\varepsilon_n|x_n] = E\left[(y_n - E[y_n|x_n])\,|x_n\right] = 0$.

- Plausible since the error term is difference between the predicted and actual frequency.

- Given these assumptions, we can estimate the above model using regression.

- This is sometimes referred to as the "linear probability model".

- Problems with linear probability model:

1. Estimates may predict that $E[y_n|x_n] > 1$ or $E[y_n|x_n] < 0$.

2. Effiiciency loss from ignoring constraint that $0 < E[y_n|x_n] < 1$.

3. Error term does not have full support under the null.

4 Bernoulli Dependent Variables.

- Now, let's constrain dependent variable to lie between zero and one.

- Let $F : R \to R$ be a real valued function with $0 < F(z) < 1$.

- One potential model is then:

$$E[y_n|x_n] = F(x_n'\beta)$$

- This model has the advantage that it always constrains the rhs to be a well defined probability.

- Two common choices for F are the standard normal cdf (probit model) and $F(z) = \frac{1}{1+e^{-z}}$ (binomial logit).

- These models can be estimated by MLE or NLLS.

- The log-likelihood for this model is:

$$L(\beta) = \frac{1}{N} \sum \left\{ y_n F(x_n' \beta) + (1 - y_n)(1 - F(x_n' \beta)) \right\}$$

- The NLLS estimator is:

$$\widehat{\beta}_{NLLS} = \arg\min_{\beta} \frac{1}{2} E_N \left[\left(y_N - F(x_n' \beta) \right)^2 \right]$$

- Note that the FOC to the NLLS problem is:

$$E_N \left[x_n f(x_n' \beta) \left(y_N - F(x_n' \beta) \right) \right] = 0$$

- At the true parameter value, β_0, $E\left[y_N - F(x_n'\beta_0)|x_n\right] = 0$ since $y_N - F(x_n'\beta_0)$ is just a prediction error (that is, the difference between the expected and actual value of y_n.

- By iterated expectations, our moment condition will be equal to zero at the true parameter value.

- After we estimate the model, one interesting thing to do is to study the marginal effects:

$$\frac{\partial E[y_n|x_n]}{\partial x_n} = \frac{\partial F(x_n'\beta)}{\partial x_n} = f(x_n'\beta)x_n$$

- The first and second derivatives for MLE are easy to compute.

- This greatly simplifies computation.

5 Random Utility Interpretation.

- One common use of the probit and binomial logit model is to model utility maximizing consumers.

- Consider a consumer faced with a dichotomouos choice $y \in \{1, 2\}$.

- Example work or don't work/vote don't vote.

- Let z denote some characteristics of the individual that impact that optimal choice.

- Suppose that the utility of individual i can be written as:

$$u_n = \begin{cases} x_n'\beta + \varepsilon_n \text{ if } y_n = 1 \\ 0 \text{ if } y_n = 2 \end{cases}$$

- In the above, assume that ε_n is a standard normally distributed error term which represents a stochastic preference shock.

- Utility maximization implies that $y_n = 1 \Leftrightarrow x_n'\beta + \varepsilon_n > 0$

- Note that we have normalized the utility of the second choice equal to zero WLOG.

- This is because we can always add or subtract a constant from utility and leave choice behavior unaltered.

- Also, we assume that the variance of ε_n is equal to one WLOG.

- This is because we can always multiply utility by a constant without changing choice behavior.

- We shall rigorously justify these statements below.

- Our model implies that:

$$\begin{aligned}\Pr[y_n &= 2|x_n] = F(x_n'\beta) \\ \Pr[y_n &= 1|x_n] = 1 - F(x_n'\beta)\end{aligned}$$

- In the above, F is the standard normal cdf.

- In the literature, two interpretations of ε_n have been proposed.

- The first is that ε_n captures unobserved heterogeneity about the individual's utility.

- That is, the econometrician only imperfectly observes the agent's utility and ε_n is a component of utility seen by an agent, but not the econometrician.

- Remark: exogeneity assumptions are potentially very strong here.

- The second is that ε_n captures the fact that utility is truly random.

6 Models with Multiple Choices.

- The above examples are for models where the number of choices is restricted to 2.

- However, in many applications of interest, 3 or more choices could be relevant (e.g. in differentiated product markets, occupation, etc...).

- Suppose that each agent n has $j = 1, ..., J$ choices available.

- Define the latent utilities u_n (J by 1) as the following SUR model:

$$\begin{aligned} u_n^* &= x_n' B_0 + \varepsilon_n' \\ Var[u_n | x_n] &= \Omega_n = [w_{0ij}] \\ B_0 &= [\beta_{0j}] \end{aligned}$$

- Define the $y_{nj} = 1\{u^*_{nj} > u^*_{nj'}\ j' \neq j\}$

- That is, $y_{nj} = 1$ if and only if the jth option is chosen.

- Then the likelihood function will have the form:

$$
\begin{aligned}
L(\theta) &= \frac{1}{N}\sum_n [\sum_{j=1}^{J} y_{nj} p_{nj}] \\
p_{nj} &= \ln(\Pr[y_{nj} = 1 | x_n])
\end{aligned}
$$

- In this model, we allow the regressors to be the same across choices but the parameters β_{0j} to vary.

- In some applications, we will want the regressors to be choice specific and the covariates to be the same.

- An example is differentiated product markets.

- We might want to model the utility of a car (j) as:

$$u^*_{nj} = \beta_1 \text{horsepower} + \beta_2 \text{size} + \beta_3 1\{\text{sedan}\} + \varepsilon_{nj}$$

- The regressors are then choice specific, x_{nj}.

- The latent utilities will then be of the form:

$$u^*_{nj} = x'_{nj}\beta + \varepsilon_{nj}$$

- Models of this form are the most common for applied work in demand estimation.

- This is called the conditional logit model.

6.1 Identification.

- Observe that making affine transformations of utilities leaves choice behavior unchanged.

- That is, for scalars, μ and σ, $(\sigma > 0)$ it follows that:

$$
\begin{aligned}
y_{nj} &= 1 \;\Leftrightarrow 1\{u^*_{nj} > u^*_{nj'} \quad j' \neq j\} \\
&\Leftrightarrow 1\{\sigma u^*_{nj} + \mu > \sigma u^*_{nj'} + \mu \;\; j' \neq j\} \\
&\Leftrightarrow 1\{0 = \max_{j=1,\ldots,J} \sigma \left(u^*_{nj} - u^*_{nj'} \right)\}
\end{aligned}
$$

- Thus, our model is only identified up to first differences in y^*_n and scalar multiplication of y^*_n.

- Since only first differences matter, we can set $\varepsilon_{n1} = 0$ as a normalizing assumption.

- This is because the model below will generate the same behavior as our original model:

$$
\begin{aligned}
u_{n1}^* &= x_n'\beta_{01} \\
u_{nj}^* &= x_n'\beta_{01} + \varepsilon_{nj} - \varepsilon_{n1}
\end{aligned}
$$

- Because the scale, we could multiply through by a constant and not change behavior.

- We could therefore normalize the diagonal of the variance matrix equal to one.

6.2 Estimation.

- In general, estimation of models with multiple choices is hard because computing $p_{nj} = \Pr[y_{nj} = 1|x_n]$ is difficult.

- Let $F(\varepsilon_n)$ denote the joint distribution of the random preference shocks. Then:

$$p_{nj} = \int 1\left\{x'_{nj}\beta + \varepsilon_{nj} > x'_{nj'}\beta + \varepsilon_{nj'} \text{ for } j' \neq j\right\}$$

- In general, it is not possible express the solution to this integral in a closed form.

- In fact, for general F, evaluating this probability by simulation may be very difficult beyond just a few choices.

- If there are 100 choices, this will be a 100 dimensional integral!

- The most common solution to this problem is to assume that the random preference shocks are distributed as in the logit model above.

- In this case, it can be shown that:

$$
\begin{aligned}
p_{nj} &= \Pr\left\{\varepsilon_{nj'} - \varepsilon_{nj} < \left(x_{nj} - x_{nj}\right)' \beta_0 \text{ for } j' \neq j\right\} \\
&= \frac{\exp(x'_{nj}\beta_0)}{1 + \sum_{j'=1}^{J} \exp(x'_{nj'}\beta_0)}
\end{aligned}
$$

- In the above, we have normalized the utiltiy of one good (the outside good) to zero since we are only identified up to first differences.

- Also, we have implicity assumed that the variance of the error term is $\frac{\pi^2}{6}$

- Thus, we have imposed the identifying assumptions discussed in the previous section.

- This model is extremely convenient computationally.

- Note that the likelihood will take the simple form:

$$L(\beta) = \frac{1}{N}\sum_n \left[\left(\sum_j y_{nj} \left(x'_{nj}\beta - \log\left(1 + \sum_{j'=1}^{J} \exp(x'_{nj'}\beta) \right) \right) \right) \right.$$

- The derivatives take the form:

$$\begin{aligned}
\frac{\partial L(\beta)}{\partial \beta} &= \frac{1}{N}\sum_n \sum_j p_{nj}(x_{nj} - Ex_i) \\
\frac{\partial L(\beta)}{\partial \beta \partial \beta'} &= -\frac{1}{N}\sum_n \sum_j p_{nj}(x_{nj} - Ex_i)(x_{nj} - Ex_i)' \\
Ex_i &= \frac{1}{N}\sum_n p_{nj} x_{nj}
\end{aligned}$$

- Note that this will greatly simplify the numerical analysis of the logit model.

7 Some Limitations of the Logit

- While the logit model is computationally convenient, it imposes some unpleasant restrictions on the data.

- It is still widely used since there are few other computationally convenient estimators.

1. Implausible substitution patters.

- In the logit model exhibits the independence of irrelevant alternatives (IIA).

- That is, the ratio of the probability of two choices does not change depending on the set of choices that are available.

$$\frac{\text{Pr(i chooses j)}}{\text{Pr(i chooses j}')} = \text{constant}$$

for all j and $j\prime$ regardless of the set of alternatives that are available.

- A famous example is the red bus/blue bus problem

- Suppose that we are studying the mode of transportation choice.

- Choice set is take the (red) bus to work or to drive.

- Suppose that these choices are equal in probability.

- Now suppose that the bus company introduces blue buses in addition to red buses.

- Suppose that consumers are indifferent about the color of their bus and that the probability of the red bus and blue bus is equal.

- IIA implies that prob(red bus)= prob(blue bus) =prob(drive) =1/3.

- A more "intuitive" answer would be rob(red bus)= prob(blue bus)= 1/4 and prob(drive)=1/2.

- This example shows that IIA can give wierd substitution patterns.

- This can also show up in terms of price elasticities.

- Suppose that we are modeling consumer demand for a differentiated product.

- Suppose that the latent utilities are:

$$y_{nj} = x'_{nj}\beta - \alpha p_j + \varepsilon_{nj}$$

- where p_j is price.

- Calculating the own and cross price elasticites.

$$\begin{aligned} \eta_{jk} &= \frac{\partial \, \text{Pr(i chooses j)}}{\partial p_k} \frac{p_k}{\text{Pr(i chooses j)}} \\ &= \begin{cases} -\alpha p_j (1 - s_j) \text{ if } j = k \\ -\alpha p_k s_k \end{cases} \end{aligned}$$

- Since in most cases there are many products so that the market shares are typically small, $(1-s_j)$ is approximately equal to price.

- This implies that the lower the price the lower the elasticity.

- This implies that markups should be higher in cheap products.

- This is clearly not appropriate in many industries.

- A second limitation is that cross price elasticities are determined by $\alpha p_k s_k$.

- Suppose that Lucky Charms and Grape Nuts are similarly priced and have a similar market share.

- An implication of this formula is that both of these will have the same cross price elasticity with CoCo Puffs.

- This is clearly a priori implausible, yet it is an assumption that we have imposed through the functional form.

2. Biased Estimates of α.

- Another problem with estimating the logit model in practice is that we will typically get estimates of α that are biased towards zero.

- Typically don't observe all of the product characteristics that are relevant to the consumer.

- If we fail to observe some product attribute, ξ_j that is correlated with price, then our exogeneity assumption is violated.

- Berry, Levinsohn and Pakes (1995) propose a strategy for consistent estimation in this case.

3. Treatment of Heterogeneity.

- In the logit model, consumers are only heterogenous because of ε_{ij}.

- ε_{ij} can be thought of as adding additional product characteristics into the model for each j and an iid random preference shock for that characteristic.

- Caplin and Nalebuff argue that this generates too much "taste for variety".

- Applied studies of welfare, such as Petrin (JPE 2002, Quantifying the Benefits of New Products: The Case of the Minivan), argue that to much of the utility comes from implausbly large draws of the ε_{ij}.

- Leads to pathological implications (e.g. markups in Bertrand may not converge to zero as market becomes thick).

8 Random coefficients.

- Since the logit model imposes fairly restrictive assumptions, researchers have searched for more general and less restrictive models.

- These models are discussed in some detail in Cameron and Travedi, Chapter 15.

- A first generalization allows for the coefficients in the conditional logit model to vary across households.

- In the notation of Cameron and Trivedi, $j = 1, ..., J$ indexes choices and $i = 1, ..., I$ indexes households.

- That is:

$$\begin{aligned} U_{ij} &= x_{ij}'\beta_i + \varepsilon_{ij} \\ \beta_i &\sim N(\beta, \mathbf{\Sigma}_\beta) \end{aligned}$$

- In the above, ε_{ij} comes from the Weibull distribution as before.

- Each household i is allowed to have a unique set of marginal utilities which come from a normal distribution with unknown mean and variance.

- In this model, the probability that houshold i chooses product j, p_{ij} is therefore:

$$p_{ij} = \frac{\exp(x'_{nj}\beta_i)}{1 + \sum_{j'=1}^{J} \exp(x'_{nj'}\beta_i)}$$

- The probability of choice j, p_j is therefore:

$$p_j(\beta, \Sigma_\beta) = \int \frac{\exp(x'_{nj}\beta_i)}{1 + \sum_{j'=1}^{J} \exp(x'_{nj'}\beta_i)} \phi(\beta_i|\beta, \Sigma_\beta) d\beta_i$$

- where $\phi(\beta_i|\beta, \Sigma_\beta)$ is the normal density.

- We could in principal estimate the model using MLE since our model generates a likelihood for the choice probabilities.

- If x'_{nj} has a large dimension (e.g. there are many characteristics), then evaluation of the above integral is difficult.

- Therefore, we need to estimate these models using simulation.

- We will study the theory of simulation in detail next week, however, we will sketch how to form a simulated likelihood function.

- Suppose that the β_i can be written as:

$$\begin{aligned} \beta_{i,k} &= \beta_k + \eta_i \sigma_k \; k = 1, ..., K \\ & \eta_i \text{ standard normal} \end{aligned}$$

- In this specification, we are assuming that the random coefficients are independently distributed across k with a normal distribution of mean β_k and standard deviation σ_k.

- In the simplest simulation based estimator, we could make $s = 1, ..., S$ monte carlo draws $\eta_i^{(s)}$ of the random coefficients for each household i.

- A monte carlo estimator of $\widehat{p}_j(\beta, \boldsymbol{\Sigma}_\beta)$ is then:

$$\widehat{p}_j(\beta, \boldsymbol{\Sigma}_\beta) = \frac{1}{S} \sum \frac{\exp(x'_{nj}\beta_k + \sum \eta_i \sigma_k x_{njk})}{1 + \sum_{j'=1}^{J} \exp(x'_{nj'}\beta_k + \sum \eta_i \sigma_k x_{nj'k})}$$

- Very important, remember to hold the simulation draws fixed as you vary the parameters $\beta, \boldsymbol{\Sigma}_\beta$.

- The "simulated" likelihood function would then be:

$$\ln \widehat{L}(\beta, \mathbf{\Sigma}_\beta) = \sum_{i=1}^{N} \sum_{j=1}^{J} y_{nj} \log \left(\widehat{p}_j(\beta, \mathbf{\Sigma}_\beta) \right)$$

- If we let the number of simulations become infinite (at an appropriate rate) as the sample size $N \to \infty$, this will yield a consistent estimator of our model parameters.

- It is also possible to derive the asymptotic variance matrix in a reasonably straightforward way.

- In practice, computing the standard errors using a resampling procedure (e.g. the bootstrap) may be most practical.

- There are some limitations, however.

- A first limitation is that this estimator is in general biased.

- An alternative, unbiased estimator is based on a NLLS approach:

$$\sum_{n=1}^{N} x_{n,r}\left(y_{n,j} - \widehat{p}_j(\beta, \Sigma_\beta)\right) = 0$$
$$r = 1, .., K \text{ and } j = 1, ..., J$$

- Unfortunately, this estimator is not efficient in general and may not even be smooth without using some fairly sophisticated numerical approaches.

- A second limitation is the variance of our estimates may be high if the distribution of random coefficients is flexibly specified.

- Hence, tightly parameterized models are required.

- Computational burden increases considerably in number of choices.

- An alternative approach is to use Gibbs sampling.