Demand Estimation - Discrete Choice Models

Li Zhao, SJTU

Spring, 2017
Demand Estimation

- **Traditional models**: Product space models.
- **Discrete Choice models (Characteristics Space)**
  - 1st generation: Logit / Probit, Nested Logit, Mixed Logit
  - 2nd generation: Logit with endogenous variable
  - 3rd generation: Random coefficient + price endogeneity
Outline

1. Discrete Choice Models
2. Logit
3. Nested Logit
4. Mixed Logit
Consumer chooses at most one good from finite set of goods.
We will define products as bundles of characteristics (includes unobservables).
We will assume that preferences are defined on these characteristics.
Each consumer chooses the bundle that maximizes her utility.
Consumers have different preferences for different characteristics.
Aggregate demand is the sum over individual demands.
Discrete Choice Framework

- Products are bundles of characteristics: \( j = 0, 1, ..., J \) where 0 is the outside good.
- Consumer preferences are defined on space of characteristics.
- The set of alternatives, called the choice set, needs to exhibit three characteristics.
  - First, the alternatives must be mutually exclusive.
  - Second, the choice set must be exhaustive.
  - Third, the number of alternatives must be finite.
- Discrete choice models are usually derived under an assumption of utility-maximizing behavior by the decision maker.
The number of parameters is primarily determined by the dimensional of the characteristics and independent of the number of products.

We give micro foundation of the formation of elasticity.

We are estimating the joint distribution of preferences over characteristics.

If a new good is introduced, we can value that good since it is simply a bundle of characteristics.

We can model consumer heterogeneity.
A decision maker, labeled $i$, faces a choice among $J$ alternatives.

The utility that decision maker $i$ obtains from alternative $j$ is $U_{nj}$, $j = 1, \ldots, J$.

The researcher observes some attributes of the alternatives as faced by the decision maker and can specify a function that relates these observed factors to the decision maker’s utility. Call this representative utility $V_{ij}$.

Utility is decomposed as $U_{ij} = V_{ij} + \epsilon_{ij}$.

The characteristics of $\epsilon_{ij}$, such as its distribution, is critical!

The probability that decision maker $i$ chooses alternative $j$ is

$$p_{ij} = \text{Prob}(U_{ij} > U_{ik}, \forall k \neq j).$$
Example

- Consider a person who can take either a car or a bus to work.
- We observe the time and cost that the person would incur under each mode.
- There are factors other than time and cost that affect the person’s utility and hence his choice.

\[ U_{\text{car}} = \alpha T_{\text{car}} + \beta M_{\text{car}} + \varepsilon_{\text{car}}; \]
\[ U_{\text{bus}} = \alpha T_{\text{bus}} + \beta M_{\text{bus}} + \varepsilon_{\text{bus}}. \]
Choice Probability

- Define the subset of “consumers” that lead to choice $j$ as
  \[ A_j(\theta) = \{ \varepsilon | u_{ij} > u_{ik}, \forall k \} \]

- The probability that consumer $i$ chooses product $j$ is
  \[ s_j = \int_{\varepsilon \in A_j(\theta)} f(\varepsilon) d\varepsilon. \]

- Under the assumption that “market size” $M$ is very large and $\varepsilon_i$’s are i.i.d. across consumers, the Law of Large Numbers implies that market demand converges to $M \cdot s_j$. 
Normalization: choices of individual consumers are invariant to any transformation of utilities.

Invariant to additive shifts implies normalizing mean utility of outside good to zero. Deduct $u_{i0}$ from each $u_{ij}$ for $j = 1, \ldots, J$.

Invariant to scale leads to normalizing one of the other parameters (typically variance of $F$) to one.
Types of Discrete Choice Models

- **Logit / Probit**
  - $\varepsilon \sim$ Type 1 extreme value distribution.
  - $\varepsilon \sim$ Multivariate Normal.

- **Nested Logit**
  - $\varepsilon \sim$ Generalized extreme value distribution.

- **Random Coefficient Logit**
  - Heterogeneous consumer preference.
Outline

1. Discrete Choice Models
2. Logit
3. Nested Logit
4. Mixed Logit
Logit - Introduction

- Logit is the most widely used discrete choice model.

\[ u_{ij} = X_j \beta - \alpha p_j + \varepsilon_{ij}. \]

- It is derived under the assumption that \( \varepsilon_{ij} \) is iid type I extreme value for all \( i \).

- The critical part of the assumption is that the unobserved factors are uncorrelated over alternatives, as well as having the same variance for all alternatives.

- This assumption, while restrictive, provides a very convenient form for the choice probability.
Type I Extreme Value Distribution

- A Type I extreme value distribution has CDF
  \[ F(v) = \exp(-\exp(-v)) \]
on \( v \in \mathbb{R} \).

- Motivation for extreme value distribution: Suppose that \((X_1, X_2, \ldots)\) is a sequence of independent random variables, each with the standard exponential distribution \( G(x) = 1 - e^{-x} \). The distribution of \( Y = \max\{X_1, X_2, \ldots X_n\} - \ln(n) \) converges to the (Type I) standard extreme value distribution as \( n \to \infty \).

- Distribution of the difference between two extreme value variables:
  \[ \epsilon^*_jk = \epsilon_{ij} - \epsilon_{ik}. \]
  \[ F(\epsilon^*_jk) = \frac{\exp(\epsilon^*_jk)}{1 + \exp(\epsilon^*_jk)}. \]
Logit Choice Probability

- The utility function
  \[ u_{ij} = X_j \beta - \alpha p_j + \varepsilon_{ij} \]
  where \( \varepsilon_{ij} \) is distributed i.i.d. Type I extreme value.
- The probability that decision maker \( i \) chooses alternative \( j \) is
  \[ P_{ij}(\theta) = \frac{\exp(X_j \beta - \alpha p_j)}{\sum_{k=0}^{J} \exp(X_k \beta - \alpha p_k)}. \]
- The market share of product \( j = 0, 1, \ldots, J \) is
  \[ s_j(\theta) = \frac{\exp(X_j \beta - \alpha p_j)}{\sum_{k=0}^{J} \exp(X_k \beta - \alpha p_k)}. \]
- Normalize the mean utility of the outside good to zero implies
  \[ s_0(\theta) = \frac{1}{\sum_{k=0}^{J} \exp(X_k \beta - \alpha p_k)}. \]
Estimation

- A sample of $N$ decision makers is obtained for the purpose of estimation.
- Since the Logit probabilities take a closed form, the traditional maximum-likelihood procedures can be applied.

\[ L = \prod_{i=1}^{N} \prod_{j=0}^{J} (P_{ij})^{y_{ij}}, \]

where $y_{ij} = 1$ if person $i$ chooses alternative $j$.
- The log-likelihood function is

\[ LL = \sum_{i=1}^{N} \sum_{j=0}^{J} y_{ij} \ln P_{ij}. \]
Consumer Welfare

- Consumer surplus is the utility get for obtaining project.

\[ CS_i = \frac{1}{\alpha} \max_j (U_j \text{ for all } j). \]

  - Parameter \( \alpha \) is the marginal utility of money (coefficient on price).

- Expected consumer surplus is

\[ E(CS_i) = \frac{1}{\alpha} E[\max_j (V_j + \epsilon_j \text{ for all } j)] \]

\[ = \frac{1}{\alpha} \ln(\sum_{j=1}^{J} \exp(V_j) + C). \]

  - Taking average over the unobserved component of utility.

- Welfare evaluation for policy change

  - Changes in observed characteristics.
  - Changes in equilibrium price from alternative market structures.
  - Removal / Introduction of new goods.
Example: Guadagni and Little (Mkt Sci 1983)

- A multinomial Logit model of brand choices.
- 32 weeks of purchases of regular ground coffee by 100 household.
- The study shows high statistical significance of brand loyalty, size loyalty, store promotion, and promotional price cut.
Guadagni and Little (1983) - Model

- Model:
  \[ u_{ij} = X_{ij} \beta - \alpha p_j + \epsilon_{ij}. \]

- \( y_{ijt} = 1 \) if consumer \( i \) chooses product \( j \) at time \( t \).

- Control variable \( X \):
  - Product dummy.
  - Regular price.
  - Promotion: presence of promotion, amount of price-cut (possible zero) during promotion, lagged promotion variable (if the previous purchase of coffee was a promotional purchase of an alternative of the same brand as brand-size \( j \)).
  - Loyalty: weighted average of past purchases of the same brand / size.

- By bringing in past customer purchase behavior, we conveniently model purchase probability heterogeneity while treating customers homogeneously.
Guadagni and Little (1983) - Data

- Data being collected by optical scanning of the Universal Product Code (UPC).
  - Store data vs. panel data.
- Definition of set of alternative (products set).
  - Exact level of aggregation and which ones to include in the relevant set are not necessarily obvious.
  - Ground / instant, caffeinated / decaf, within instant, freeze dried / non-freeze.
  - Size: small and large.
- Ground coffee store and panel records from four Kansas City supermarkets for the 78-week period in late 1970s.
- 2000 households, each of which has indicated it makes 90% or more of its purchases at one of the stores collecting the data.
Guadagni and Little (1983) - Counterfactual

- Now we wish to study how the market responds to the retailers’ actions.
- Analytically, determination of aggregate market response calls for an integration of the customer response function over a joint distribution of Xs.
- One standard approach to such a high dimensional integration is Monte Carlo sampling.
- A simpler and easier method to implement is to use the actual customers, time periods and attribute variables of the data.
- Many different response calculations are possible. Some examples are
  - (1) regular price elasticity of share,
  - (2) the cross elasticity of one brand-size’s share to another’s price,
  - (3) share response to promotion,
  - (4) the cannibalization of one size of a brand by the promotion of another and
  - (5) the elasticity of price during a promotion.
Limitation of Logit - Substitution Patterns

IIA (Independence from Irrelevant Alternatives)

- For any two alternatives i and k, the ratio of the logit probabilities is

\[
\frac{P_{ij}}{P_{ik}} = \frac{\exp(V_{ij})}{\sum} = \frac{\exp(V_{ij})}{\sum} = \exp(V_{ij} - V_{ik})
\]

- This ratio does not depend on any alternatives other than i and k.
- “Red-bus–blue-bus problem”
  - Assume that city has two transportation schemes: walk, and red bus, with shares 50%, 50%.
  - Now consider introduction of third option: train. IIA implies that odds ratio between walk/red bus is still 1.
  - What if third option were blue bus? IIA implies that odds ratio between walk/red bus would still be 1.
  - This is especially troubling if you want to use Logit model to predict penetration of new products.

- Overcome IIA: Probit, Nested Logit, Mixed Logit, Random Coefficient Logit.
Recall the random utility model takes the form

\[ u_{ij} = X_j \beta - \alpha p_j + \varepsilon_{ij}. \]

\[ p_{ij} = \text{Prob}(U_{ij} > U_{ik}, \forall k \neq j). \]

- In Logit, we assume \( \varepsilon_{ij} \sim \text{i.i.d. extreme value distribution.} \)
- In Probit, we assume \( \varepsilon_i = [\varepsilon_{i1}, ..., \varepsilon_{iJ}] \sim \text{multivariate normal distribution.} \)
  - Probit allows for correlation between unobserved utility among choices and does not reply on IIA assumptions.
Outline

1. Discrete Choice Models
2. Logit
3. Nested Logit
4. Mixed Logit
A nested Logit model is appropriate when the set of alternatives faced by a decision maker can be partitioned into subsets, called nests, in such a way that the following properties hold:

- For any two alternatives that are in the same nest, the ratio of probabilities is independent of the attributes or existence of all other alternatives. That is, \( IIA \) holds within each nest.
- For any two alternatives in different nests, the ratio of probabilities can depend on the attributes of other alternatives in the two nests. \( IIA \) does not hold in general for alternatives in different nests.
Nested Logit - Illustration

- Auto, (Red Bus, Blue Bus).
- $\varepsilon_{nj}$ have a joint cumulative distribution of

$$F(\varepsilon_{n1}, \varepsilon_{n2}, \ldots) = \exp\left(- \sum_{s=1}^{S} \left( \sum_{j \in J_g} \exp\left(- \frac{\varepsilon_{nj}}{\lambda_s}\right) \right) \lambda_s \right).$$

- Within the nest the correlation coefficient of is approximately $1 - \lambda_s$, Between the nests choices are independent.

- Conditional probability of choice $j$ given $g$ is $P_{j|g} = \frac{\exp\left(\frac{\delta_j}{\lambda_s}\right)}{\sum_{j' \in J_g} \exp\left(\frac{\delta_{j'}}{\lambda_s}\right)}$.

- The probability of choosing group $g$ is $P_g = \frac{(\sum_{j \in J_g} \exp\left(\frac{\delta_j}{\lambda_s}\right))^{\lambda_s}}{\sum_{g'} (\sum_{j' \in J_{g'}} \exp\left(\frac{\delta_{j'}}{\lambda_{s'}}\right))^{\lambda_{s'}}}$.

- $P_j = P_{j|g} \cdot P_g$. 
Nested Logit - Estimation

- ML joint estimation.
- Sequential estimation using nesting structure.
  - Estimate lower model: Within the nest we have a conditional MNL with coefficients $\beta / \lambda_s$.
  - Compute Inclusive value, $\sum_{j \in J_g} \exp\left(\frac{\delta_j}{\lambda_s}\right)$ using estimates from step 1.
  - Estimate upper model with inclusive value as explanatory variable, recover $\lambda_s$. 
Example: Morey et. al. (1993)


- Participation and site choice for Atlantic salmon fishing are modeled in the context of a repeated three-level nested Logit model.

- Consumer’s surplus measures are derived for different levels of species availability in the Penobscot River, the most important salmon river in New England.

- For comparison, six other travel-cost models are estimated.

- Estimating these models indicate the importance of modeling the participation decision, including income effects, and of adopting a nested Logit structure rather than a single-level Logit structure.
Morey et. al. (1993) - Model

- Fishing season is divided into $T$ periods such that in each period the individual takes, at most, one fishing trip.

\[ U_{jt} = V_j + \varepsilon_{jt}. \]

- $j = 0$ (nonfishing alternative), $j = 1 – 5$ are Maine sites, and $j = 6 – 8$ are Canadian sites.

\[ V_0 = V_0(ppy - p_j, C + ppy, yrs, club, age). \]

\[ V_j = V_j(ppy - p_j, C + ppy - p_j, Catch_j). \]

- $\varepsilon_{jt}$ are drawn from generalized extreme distribution

\[ F(\varepsilon) = \exp[-e^{-s\varepsilon_0} - [e^{-s\varepsilon_1} + ... - e^{-s\varepsilon_5}]^{\lambda/s} + (e^{-s\varepsilon_6} + ... + e^{-s\varepsilon_8})^{\lambda/s}]^{1/\lambda}. \]
Morey et. al. (1993) - Counterfactual

- Individual’s expected maximum utility in a single period is

\[ V = V(ppy, yrs, club, age, P, Catch). \]

- Per-period compensation variation is

\[ V(ppy, yrs, club, age, P^0, Catch^0) = V(ppy - PPCV, yrs, club, age, P^1, Catch^1). \]

- Compare utilities under 3 scenarios
  - Elimination of the Atlantic salmon fishery at the Penobscot River.
  - Double catch rates at the Penobscot River.
  - Halve catch rates at the Penobscot River.
The paper tries six other travel-cost models to compare the results.

- Repeated Logit with /without income effects.
- Standard Logit treating non-participation is not one of the alternative.
- A partial demand (share) model of the site choices.
- Single-site linear demand function.
- Single-site log-linear demand function.
Mixed Logit (1)

- Mixed Logit obviates the three limitations of standard Logit by allowing for random taste variation, unrestricted substitution patterns, and correlation in unobserved factors over time.
- Like Probit, the mixed Logit model has been known for many years but has only become fully applicable since the advent of simulation.
- A mixed logit model is any model whose choice probabilities can be expressed in the form

\[ P_{ij} = \int L_{ij}(\beta) \cdot f(\beta) \cdot d\beta \]

where \( L_{ij}(\beta) \) is the Logit probability evaluated at parameters \( \beta \):

\[ L_{ij}(\beta) = \frac{\exp(V_{ij}(\beta))}{\sum_j \exp(V_{ij}(\beta))} \]

- Mixed Logit is a mixture of the Logit function evaluated at different \( \beta \)'s with \( f(\beta) \) as the mixing distribution.
Mixed Logit (2)

- A mixed Logit model is any model whose choice probabilities can be expressed in the form

\[ P_{ij} = \int L_{ij}(\beta) \cdot f(\beta) \cdot d\beta. \]

- Discrete: The mixing distribution \( f(\beta) \) can be discrete

\[ P_{ij} = \sum_{m=1}^{M} s_m \left( \frac{\exp(b_m x_{ij})}{\sum_k \exp(b_m x_{ik})} \right). \]

This specification is useful if there are \( M \) segments in the population, each of which has its own choice behavior or preferences.

- Continuous: More frequently, the mixing distribution is specified to be continuous. For example, the density of \( \beta \) can be specified to be normal with mean \( b \) and covariance \( W \).

\[ P_{ni} = \int \left( \frac{\exp(b_m x_{ij})}{\sum_k \exp(b_m x_{ik})} \right) \phi(\beta \mid b, W) d\beta. \]

- Calculating integration is difficult, but taking average is easy. We can use simulation to replace integration with average.
Random Coefficients Logit

- In a random coefficient Logit, consumer $i$’s utility for product $j$ is
  \[ u_{ij} = X_j \beta_i - \alpha_i p_j + \xi_j + \varepsilon_{ij} \]
  where
  \[ (\alpha_i, \beta_i)' \sim \mathcal{N}((\bar{\alpha}, \bar{\beta})', \Sigma). \]
- $(\alpha_i, \beta_i)$ differ across consumers. They are called “random coefficients”.
- $(\bar{\alpha}, \bar{\beta})$ and $\Sigma$ are additional parameters to be estimated.
- We can also make parametric assumption of $(\bar{\alpha}, \bar{\beta})$ to let them depend on consumer characteristics (e.g. income).
  \[ \left( \begin{array}{c} \alpha_i \\ \beta_i \end{array} \right) \sim \left( \begin{array}{c} \bar{\alpha} \\ \bar{\beta} \end{array} \right) + \Pi D_i + v_i. \]

An important aspect of marketing practice is the targeting of consumer segments for differential promotional activity.

The goal of this paper is to assess the information content of various information sets available for direct marketing purposes.

Information on the consumer is obtained from the current and past purchase history as well as demographic characteristics.

Our results indicate there exists a tremendous potential for improving the profitability of direct marketing efforts by more fully utilizing household purchase histories.
$u_{ijt} = X_{it} \beta_i + \varepsilon_{ijt}$,

where $\varepsilon_{it} \sim N(0, \Lambda)$.

- $X_{it}$ contains product dummy, prices, display, product features.
- The random coefficients takes the form
  \[ \beta_i = \Pi D_i + \nu_i. \]
- Five brands of tuna, packaged in six-ounce cans, purchased in Springfield, Missouri.
- 400 household who remains in panel at least 1.5 years.
- Demographic variables: family income, size, retired, unemployed, female headed families.
The coupon acts as a temporary price cut equal to its face value.

We ignore dynamic effects induced by consumer stock-piling in response to coupon availability.

Incremental Sale = \( \Pr(j|\beta_i, \Lambda, \text{price} - F, X) - \Pr(j|\beta_i, \Lambda, \text{price}, X) \).

The expected net revenue is \( \pi_F = \Pr(j|\beta_i, \Lambda, \text{price} - F, X) \cdot (M - F) \).

\( \beta_i \) is estimated based on various information set

- No information at all.
- Demographic.
- Demographic + purchase history.
- Demographic + one purchase and its causal effects.
- Demographic + purchase history and all the causal effects.
Summary

- Product space and characteristics space demand models.
- Discrete choice, micro foundation, normalization.
- Logit, IIA assumption, CS.
- Extension of Logit: Probit, Nested Logit, Mixed Logit.
- Counterfactual: change in prices, removing products, change features.