Empirical Models in Marketing

Marketing Models
& Models of Choice

Oded Netzer - Columbia Business School
Marketing Models

Class Structure

- Lectures – mainly the needed econometric tools
- Basic coding – hands-on experience with the tools (using R)
- Discussing articles – pouring marketing into the topic (led by you)
- Course project

Oded Netzer - Columbia Business School
Marketing Models

Class Syllabus

- Session 1 – Marketing models
- Session 2 – Logit, Probit, Nested Logit
- Session 3 – Logit, Probit, Nested Logit
- Session 4 – Heterogeneity in choice models
- Session 5 – Dynamics in choice models
- Session 6 – Hidden Markov models

Class Syllabus

- Session 7 – Customer base analysis \
- Session 8 – Fields experiments in marketing
- Session 9 – Text mining in marketing (Alain Lemaire)
- Session 10 – Networked marketing
- Session 11 – Bridging behavioral Research and quantitative Modeling
- Session 12 – Presentations

Mathematical Models

- Models are simplified representation of a complex system. The can be built to further our understanding of the working of a complex system; to predict future outcomes; and to help design interventions to lead the system to a particular outcome

- Mathematical models use a mathematical representation of systems.

Some Modeling Quotes

“ The choice is always the same. You can make your model more complex and more faithful to reality, or you can make it simpler and easier to handle. Only the most naïve scientists believe that the perfect model is one that perfectly represents reality. “

James Gleik, Chaos: Making a New Science, 1987, p.278

“Make it as simple as possible, not any simpler”

Albert Einstein

Some Modeling Quotes

"All models are wrong.
Some are useful."

G.E.P. Box

Type of Models

- Measurement
  - Measure response, demand, preferences…
- Theoretical
  - Explain, provide insights…
- Decision Support
  - Aid decision, politics..

Measurement Models

- Empirical models of measuring responses and demand estimation

Aggregate Models
- Sales response models
- Market share models
- Diffusion models

Individual Level Models
- Choice models
- Conjoint analysis

Empirical IO demand models

- Two main objectives: explain and\or predict

Theoretical Models

- Explain complex phenomenon
- Serve as input in specification of other models
- Built on assumptions

  Substantive richness vs. Tractable Analysis
- Examples include game-theoretic models for various marketing phenomena

Decision Support Models

- For aiding managerial decisions
- Models have empirical and optimization components
- Examples include
  - Conjoint models
  - Agent–based models
  - Sales promotion calendar
  - Advertising media scheduling
  - Targeted advertising

Empirical Methodology

- Econometrics
  - Discrete choice model; Endogeneity
  - Structural modeling
- Probability
- Statistics
  - Bayesian Statistics
- Machine Learning
  - Analysis of large data sets
  - Text mining

Analytical Methodology

- Game theory for modeling strategic interactions
  - Nash Equilibrium
  - Multi period games
  - Monopoly, duopoly, oligopoly games
  - Signaling games

Characteristics of Good Models

- Simplicity vs. substantive richness
  - Internal vs. External validity
- Robustness
  - Ability to give sensible results
  - Results not too dependent on assumptions
- Adaptive
- Flexible
- Scalability

Oded Netzer - Columbia Business School
Marketing Models

Process of Empirical Model Building

- Idea (see Varian 1997)
- Theory
- Model specification
- Data collection
- Estimation
- Model assessment
- Implementation

Model Specification

- Statistical models are built to capture variation in some substantive quantity of interest
  - Sales vary across markets
  - Market share varies over time
  - Consumer choices vary across consumers etc…
- Response/Explanatory variables
  - Response variables are variables of primary interest
  - Explanatory variables are used to the explain the variation in response

Model Specification

- Each model has two components
  - Systematic component
    - Models how explanatory variables combine to influence the variable of interest $E(sales_t) = \beta_1 + \beta_2 price_t$
  - Stochastic component
    - Captures idiosyncratic variation from one observation to the other. Modeled using probability distribution.
- Example – Sales\Advertising Response function

$$Sales_t = \underbrace{\beta_1 + \beta_2 Adv_t}_{\text{Linear systematic part}} + \varepsilon_t \leftarrow \text{Stochastic part}$$

Model Specification

- Example – Sales\Advertising Response function

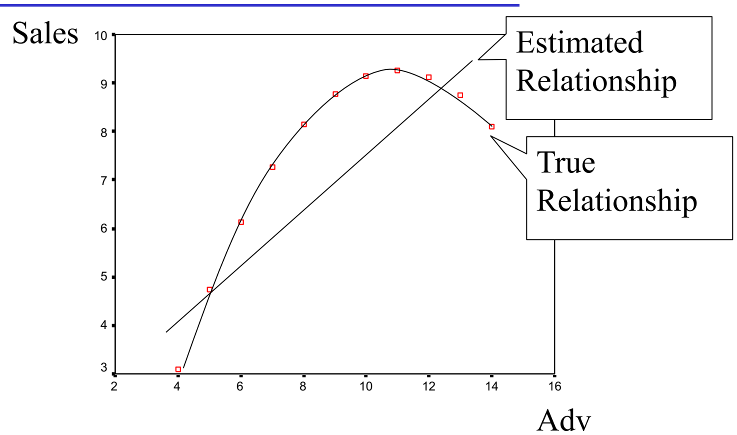
$$Sales_t = \underbrace{\beta_1 + \beta_2 Adv_t}_{\text{Linear systematic part}} + \varepsilon_t \leftarrow \text{Stochastic part}$$

- Why do we need ε_t ?

Measurement Error

- True relationship is $Y^* = \beta X + e$
- We observe Y which is noise measure of Y^* such that $Y = Y^* + u$
- If we assume that u and e are independent of each other and u is independent of Y^*, substituting $Y^* = \beta X + e$ into $Y = Y^* + u$ we get $Y = \beta X + u + e = Y = \beta X + \varepsilon$

Error in Functional Form



Specification of the Systematic Part

- Choice of Dependent\Independent variables
  - Subject matter theory
  - Data availability
  - Exploratory analysis
- Independent variables can be
  - Exogenous
  - Endogenous
  - Manipulated
  - Trends\Seasonality
  - Qualitative (dummy variables)
  - Lags\leads

Specification of the Stochastic Part

- Choice of probability distribution based on the nature of the response variables and nature of the systematic part
  - Continuous (restricted/unrestricted)
  - Binary/multinomial/pick any
  - Ordered
  - Counts
- Robustness considerations
- Tractability, i.e., whether analytical closed functional from can result for quantities of interest

Choice Models

Type of Choice Decision

- Discrete choice – one among the choice set
- Multinomial discrete – can buy more than one
- Continuous – choice + quantity

Some definitions:
- Choice set – Finite number of alternatives – MECE
- Alternatives – bundle of characteristics
- Consumers – 1) preferences for attributes => for alternatives
  2) characteristics

The Luce Model (1959)

The probability that individual i chooses alternative j and time t:

$$P(choice_{it} = j) = \frac{W_{ijt}}{\sum_{k=1}^{J} W_{ikt}}$$

The IIA property

Oded Netzer - Columbia Business School
Choice Models in Marketing

Random Utility Models

- The indirect utility of alternative j is U_j
- $U_{ijt} = V_{ijt} + \varepsilon_{ijt}$, $V_0 = 0$

$$U_{ijt} = \alpha_{ij} - \beta_1 price_{ijt} + \beta_2 adv_{ijt} + \beta_3 display_{ijt} + \varepsilon_{ijt}$$

- Alternative with largest utility is chosen

$$P(choice_{it} = j) = prob(V_{ijt} + \varepsilon_{ijt} > V_{ikt} + \varepsilon_{ikt}, \forall k \neq j) =$$

$$prob(\varepsilon_{ijt} - \varepsilon_{ikt} > V_{ikt} - V_{ijt}, \forall k \neq j)$$

- Utilities are relative - "only differences matter"
- Probit $\varepsilon_{ijt} : N(0, \sigma^2)$ $\boldsymbol{\varepsilon}_{\mathbf{ijt}} : N(0, \Sigma)$
- Logit $\varepsilon_{ijt} :$ Extreme Value

Oded Netzer - Columbia Business School
Choice Models in Marketing

Only Difference in Utilities Matters

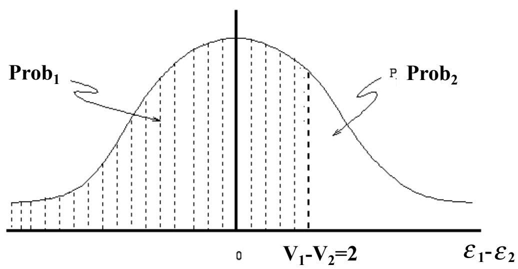
- Suppose $U_{i1t} = 5 + \varepsilon_{i1t}$ and $U_{i2t} = 3 + \varepsilon_{i2t}$

  Customer *i* chooses brand 1 if:

  1) $\varepsilon_{i1t} > \varepsilon_{i2t}$
  2) ε_{i2t} exceeds ε_{i1t} by less than 2

That is,

$$\text{Prob}(\text{choice}_{it} = 1) = \text{Prob}(U_{it1} - U_{it2} > 2) = \text{Prob}(\varepsilon_{it2} - \varepsilon_{it1} < 2)$$



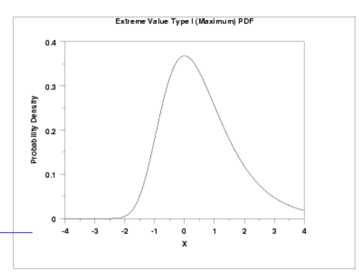
The Logit Model (McFadden 1974)

- $U_{ijt} = V_{ijt} + \varepsilon_{ijt}$ $\quad V_0 = 0$
- ε_{ijt} are IID Extreme Value Distributed
- μ is scale parameter (commonly =1)
- η is location parameter (commonly=0)
- $V(\varepsilon) = \dfrac{\pi^2}{6}$
- Difference between two EV is logistic

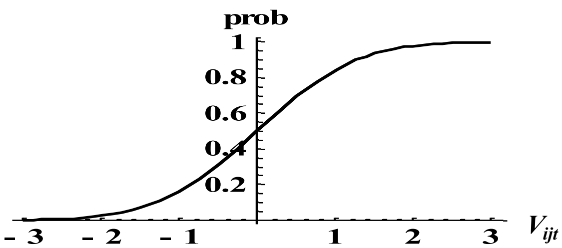
$$f(\varepsilon_j) = e^{-\varepsilon_j} e^{-e^{-\mu(\varepsilon_j - \eta)}}$$

$$F(\varepsilon_j) = e^{-e^{-\mu(\varepsilon_j - \eta)}}$$

$$P(choice_{it} = j) = \frac{e^{V_{ijt}}}{\sum_{k=1}^{J} e^{V_{ikt}}}$$



The Logit Model - CDF



$$P(choice_{it} = j) = \frac{e^{V_{ijt}}}{\sum_{k=1}^{J} e^{V_{ikt}}}$$

The Logit Model Properties

- IIA

Assume equal utility for Coke, Pepsi and Sprite

Prob		Prob		Logit
Coke	0.5	Coke	0.25	0.33
Sprite	0.5	Pepsi	0.25	0.33
		Sprite	0.5	0.33

•Distribution of max-utilities

The Logit Model Elasticities

Own Elasticity

$$E_{ip_{jt}} \frac{\partial \text{Prob}_{ijt}}{p_{ijt}} \frac{p_{ijt}}{\text{Prob}_{ijt}} = \beta p_{ijt}(1 - \text{Prob}_{ijt})$$

Cross Elasticity

$$E_{ip_{kt}} \frac{\partial \text{Prob}_{ijt}}{p_{ikt}} \frac{p_{ikt}}{\text{Prob}_{ijt}} = -\beta p_{ikt} \text{Prob}_{ikt}$$

Same cross elasticities for all j

Heterogeneity Solution for IIA

100 Consumers

	50 strongly prefer Colas	50 strongly prefer uncolas	
Coke	$P_{coke} = 1$	$P_{coke} = 0$	$MS_{coke} = 1/2$
7-UP	$P_{7\text{-}UP} = 0$	$P_{7\text{-}UP} = 1$	$MS_{7\text{-}UP} = 1/2$

Coke	$P_{coke} = 1/2$	$P_{coke} = 0$	$MS_{coke} = 1/4$
Pepsi	$P_{pepsi} = 1/2$	$P_{pepsi} = 0$	$MS_{pepsi} = 1/4$
7-UP	$P_{7\text{-}UP} = 0$	$P_{7\text{-}UP} = 1$	$MS_{7\text{-}UP} = 1/2$

Thus, modeling preference and choice at the individual level minimizes the IIA problem

Identification Issues

- Utilities are relative - "only differences matter"

$$P(choice_{it} = j) = prob(V_{ijt} + \varepsilon_{ijt} > V_{ikt} + \varepsilon_{ikt}, \forall k \neq j) =$$
$$prob(\varepsilon_{ijt} - \varepsilon_{ikt} > V_{ikt} - V_{ijt}, \forall k \neq j)$$

- Brand intercepts
- Demographics
- Logit scale parameter (Jordan Louviere)

Logit Scale Parameter

$U_{ijt}^{*} = X_{ijt}\beta^{*} + \varepsilon_{ijt}^{*}$ Suppose $\text{Var}(\varepsilon_{ijt}^{*}) = k^2$

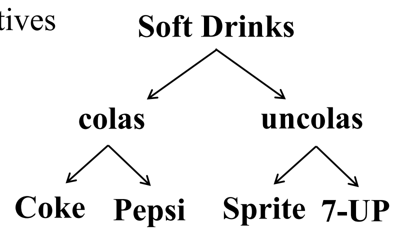
Express $k^2 = \dfrac{\sigma^2\pi^2}{6}$

$$\frac{U_{ijt}^{*}}{\sigma} = X_{ijt}\frac{\beta}{\sigma} + \frac{\varepsilon_{ijt}^{*}}{\sigma} \Rightarrow \quad U_{ijt} = X_{ijt}\beta + \varepsilon_{ijt} \quad \text{Var}(\varepsilon_{ijt}) = \frac{\pi^2}{6}$$

Only the ratio β^{*}/σ can be estimated

Nested Logit

- Allow for correlation between alternatives
- Within branch IIA holds



$$U_{ijt\in branch} = \beta' X_{ijt} + \gamma' Z_{i,branch,t} + \varepsilon_{i,branch,} + \varepsilon_{ijt}$$

Attributes of brands in branch

Attribute of the branch

$\varepsilon_{i,branch,t}$ and ε_{ijt} are distributed independently

$$P_{ijt} = P_{ijt|branch} P_{i,branch,t}$$

Nested Logit does not imply sequential choice.

Nested Logit

$$P(choice_{it} = j \mid branch = k) = \frac{e^{\beta' X_{ijt}}}{\sum_{l\in k} e^{\beta' X_{ilt}}}$$

Simple MNL

Inclusive value of branch k $\quad \mathrm{IV}_k = \log[\sum_{j\in k} \exp(\beta' X_{ijt})]$

$$IV_k = E\left(\max_{j\in k} U_k\right) = E\left(\max_{j\in k} V_j + \varepsilon_j\right)$$

$$\mathrm{Prob}(branch = k) = \frac{e^{(\gamma' Z_{ikt} + \mu IV_k)}}{\sum_{b=1}^{B} e^{(\gamma' Z_{ibt} + \mu IV_b)}}$$

The probability of choosing a branch is determined by the expected maximum utility of the brands within the branch (plus branch specific benefits)

For μ=1 we are back to the MNL

Nested Logit

What is μ

$1 - \mu$ is measure of correlation within each nest

for $\mu = 1$ we get back the MNL

for $\mu = 0$ we get perfect correlation

The Probit Model

- The indirect utility of alternative j is U_j
- $U_{ijt} = V_{ijt} + \varepsilon_{ijt},\ V_0 = 0$
- ε_{ijt} are IID Normal

- Consider three alternatives:

$$\boldsymbol{\varepsilon}_{\mathbf{it}} = MVN(0, \boldsymbol{\Sigma}) \qquad \boldsymbol{\Sigma} = \begin{matrix} \sigma_{11}^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22}^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33}^2 \end{matrix}$$

$\text{Prob}(U_1 > U_2 \text{ and } U_1 > U_3) =$

$\text{Prob}(V_1 + \varepsilon_1 > V_2 + \varepsilon_2 \text{ and } V_1 + \varepsilon_1 > V_3 + \varepsilon_3) =$

$\text{Prob}(\varepsilon_2 - \varepsilon_1 < V_1 - V_2 \text{ and } \varepsilon_3 - \varepsilon_1 < V_1 - V_3)$

The Probit Model

Let $\mu_{12} = \varepsilon_2 - \varepsilon_1$ and $\mu_{13} = \varepsilon_3 - \varepsilon_1$

$$\operatorname{cov}(\mu_{12}, \mu_{13}) = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} & \sigma_1^2 - \sigma_{13} - \sigma_{12} + \sigma_{23} \\ \sigma_1^2 - \sigma_{13} - \sigma_{12} + \sigma_{23} & \sigma_1^2 + \sigma_3^2 - 2\sigma_{13} \end{bmatrix}$$

$$\text{Prob (Choice}_{it} = 1) = \int_{-\infty}^{V_1 - V_2} \int_{-\infty}^{V_1 - V_3} f(\mu_{12}, \mu_{13}) d\mu_{12} d\mu_{13}$$

No closed form

Estimating the Logit

- Estimation is typically done via MLE

$$Y_{ijt} = \begin{cases} 1 & \text{if alternative j was chosen by customer i at time t} \\ 0 & \text{if alternative j was not chosen by customer i at time t} \end{cases}$$

- The likelihood of the chosen alternative for customer i is:

$$L_{it} = \prod_{j=1}^{J} \text{Prob}(Choice_{it} = j)^{Y_{ijt}}$$

- Aggregating over customers and time:

$$L = \prod_{i=1}^{N} \prod_{t=1}^{T_i} \prod_{j=1}^{J} \text{Prob}(Choice_{it} = j)^{Y_{ijt}}$$

Maximum Likelihood Estimation

- We are interested in the value of θ, defined by θ^*, that maximizes the likelihood.
- We typically maximize the log-likelihood, why?
- The maximum likelihood is consistent and asymptotically unbiased. Can be biased in small samples.
- θ^* is asymptotically normally distributed. That is, in large samples: $\theta^* \sim N(\theta, I(\theta)^{-1})$
- Where the Fisher Information matrix is $I(\theta) = E\left\{-\dfrac{d^2 LL(\theta)}{d^2\theta}\right\}$

Estimating the Logit

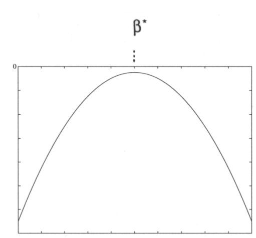
- Log-Likelihood

$$LL = \log(L) = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{j=1}^{J} Y_{ijt} \log\left(\text{Prob}(Choice_{it} = j)\right) =$$

Which could be further simplified to

$$\sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{j=1}^{J} Y_{ijt} \log\left(\frac{e^{V_{ijt}}}{\sum_{l=1}^{J} e^{V_{ilt}}}\right) = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{j=1}^{J} Y_{ijt} \left[V_{ijt} - \log\left(\sum_{l=1}^{J} e^{V_{ilt}}\right)\right] =$$

$$\sum_{i=1}^{N} \sum_{t=1}^{T_i} \left[\sum_{j=1}^{J} Y_{ijt} V_{ijt} - \sum_{j=1}^{J} Y_{ijt} \log\left(\sum_{l=1}^{J} e^{V_{ilt}}\right)\right]$$



Oded Netzer - Columbia Business School
Choice Models in Marketing

Other Examples of Likelihood Functions

- Count - Poisson
- Let $y = \{y_1, ..., y_n\}$ represent the purchases made by n individuals in a given time period.
- We can model y_i using Poisson distribution
- Under Poisson $pr(y_i \mid \theta) = \theta^{y_i} \mathrm{e}^{-\theta} / y_i!$
- Write the likelihood of observing $y = \{y_1, ..., y_n\}$
- consider the following for data: 1 0 1 1 0 0 0 1 0 0
  - What is the maximum likelihood estimate of θ?

Oded Netzer - Columbia Business School
Choice Models in Marketing

Other Examples of Likelihood Functions

- Continuous - Normal
- Let $y = \{y_1, ..., y_n\}$ represent the sales of a brand in n markets
- Assuming $y_i = \boldsymbol{\beta}\mathbf{x_i} + e_i$ and $e_i \sim N(0, \sigma^2)$
- The likelihood can be written as:

$$L(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}) = \prod_{i=1...n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(y_i - \boldsymbol{\beta}\mathbf{x}_i)^2) =$$

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp(-\frac{1}{2\sigma^2} \sum_{i=1...n} (y_i - \boldsymbol{\beta}\mathbf{x}_i)^2)$$

Oded Netzer - Columbia Business School
Choice Models in Marketing

Model Assessment

- How well is model doing in terms of:
  - Substantive insights
  - Predictive performance
- Model adequacy
  - Does the model fit the data well?
  - Does it recover key aspect of the data well?
  - Is the systematic component adequate?
  - Is the stochastic component adequate?
- Model comparison
- Predictive performance

Model Fit

- Unfortunately LL is not intuitive to interpret

$$U^2 = \rho^2 = 1 - \frac{L(X)}{L_0}$$

L_0 is the likelihood of the LL of the null model (only intercept)

- Hypothesis Testing – Likelihood Ratio Test

Model A – restricted model (smaller LL)
Model B – unrestricted model (bigger LL)

$$LL \text{ statistic} = 2(LL_B - LL_A) : x^2_{d.f.=\text{difference in the \# of parameters}}$$