Multiple Choice

Hansen (2022, Sections 26.7-26.12)

Yasuyuki Matsumura (Kyoto University)

Last Updated: October 9, 2025

https://yasu0704xx.github.io

Textbooks

- Train (2009, Chapters 2-7, 11)
- Wooldridge (2010, Chapter 13)
- Hansen (2022, Chapter 26)

Contents

Review on Sections 26.1-26.6

Multinomial Response

- Suppose that we have a multinomial random variable $Y \in \{1, 2, \cdots, J\}$, and k-dimensional regressors $X \in \mathbb{R}^k$.
- The conditional distributions of Y given X is summarized by the response probability

$$P_j(x) = \mathbb{P}\left(Y = j | X = x\right).$$

- The response probabilities $P_1(x), \cdots, P_J(x)$ are nonparametrically identified and can be arbitrary functions of x.

Latent Utility

- Multinomial response is typically motivated and derived from a latent utility model. Assume that the utility of choosing alternative j is expressed as

$$U_j^* = X^\top \beta_j + \epsilon_j \tag{1}$$

  where β_j are coefficients and ϵ_j is an idiosyncratic error of individual- and product-level.

- An individual is assumed to select the alternative with the highest utility:

$$Y = j \iff U_j^* \geq U_l^* \text{ for all } l.$$

- To identify β_j in (1), researchers are required to impose a normalization. The standard choice is to set $\beta_j = 0$ for a base alternative j and interpret the reported coeeficients $\widehat{\beta}_j$ as differences relative to the base alternative.

Canonical Multinomial Logit

- Assume that the utility of alternative j is given by (1), and the error vector $(\epsilon_1, \cdots, \epsilon_J)$ has the generalized extreme value (GEV) joint distribution:

$$F(\epsilon_1, \cdots, \epsilon_J) = \exp\left(-\left[\sum_{j=1}^{J} \exp\left(-\frac{\epsilon_j}{\tau}\right)\right]^{\tau}\right).$$

- Then, the response probabilities equal

$$P_j(X) = \frac{\exp\left(\frac{X^\top \beta_j}{\tau}\right)}{\sum_{l=1}^{J} \exp\left(\frac{X^\top \beta_l}{\tau}\right)}. \tag{2}$$

- The canonical multinomial logit model is an extension of the binary logit model to the case with an unordered multinomial dependent variable and heterogenous coefficients β_j's.
- The explanatory vector X is common across the choices in the multinomial logit model.
- Note that the scale of the coefficients τ is not identified.

Maximum Likelihood Estimation

- Noting that the response probabilities in (2) are functions of the parameter vector $\beta = (\beta_1, \cdots, \beta_J)$. we can express the probability mass function for Y as

$$\pi(Y|X,\beta) = \prod_{j=1}^{J} P_j(X|\beta)^{1\{Y=j\}}.$$

- Thus, the log-likelihood function is given by

$$l_n(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{J} 1\{Y_i = j\} \log P_j(X_i|\beta).$$

- Then, the maximum likelihood estimator (MLE) is $\widehat{\beta} = \arg\max_{\beta} l_n(\beta)$, which has no algebraic solution and so needs to be found numerically.

Marginal Effect

- The coefficients themselves are difficult to interpret.
- In applications, it is common to examine and report marginal effects:

$$\delta_j(x) = \frac{\partial}{\partial x} P_j(x) = P_j(x) \left(\beta_j - \sum_{l=1}^{J} \beta_l P_l(x) \right),$$

  which can be estimated by

$$\widehat{\delta}_j(x) = \widehat{P}_j(x) \left(\widehat{\beta}_j - \sum_{l=1}^{J} \widehat{\beta}_l \widehat{P}_l(x) \right).$$

- The average marginal effect $\text{AME}_j = \mathbb{E}\left[\delta_j(X)\right]$ can be estimated by $\widehat{\text{AME}}_j = \frac{1}{n} \sum_{i=1}^{n} \widehat{\delta}_j(X_i)$.

Conditional Logit

- Assume that the utility of alternative j is given by

$$U_j^* = X_j^\top \gamma + \epsilon_j,$$

and the error vector $(\epsilon_1, \cdots, \epsilon_J)$ are distributed IID Type 1 extreme value:

$$F(\epsilon_j) = \exp\left(-\exp(-\epsilon_j)\right), \quad j = 1, \cdots, J.$$

- Then, the response probabilities equal

$$P_j(w, x) = \frac{\exp\left(x_j^\top \gamma\right)}{\sum_{l=1}^J \exp\left(x_l^\top \gamma\right)}.$$

- The conditional logit model is an extension of the binary logit model to the case with heterogenous explanatory vectors X_j's across choices.
- The coefficient γ does not depend on the chioces in the conditional logit mnodel.
- Let $\theta = (\beta_1, \cdots, \beta_J, \gamma)$. Given the observations $\{Y_i, X_i\}$ where $X_i = (X_{1i}, \cdots, X_{Ji})$, the log-likelihood function is

$$l_n(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{J} 1\{Y_i = j\} \log P_j(X_i|\theta).$$

The MLE is given by $\widehat{\theta} = \arg\max_{\theta} l_n(\theta)$, which has no algebraic solution and needs to be found numerically.

- Marginal effects are defined as

$$\text{for } j, \quad \delta_{jj}(x) = \frac{\partial}{\partial x_j} P_j(x) = \gamma P_j(w, x)\left(1 - P_j(x)\right), \text{ and}$$

$$\text{for } j \neq l, \quad \delta_{jl}(x) = \frac{\partial}{\partial x_l} P_j(x) = -\gamma P_j(w, x) P_l(x).$$

Slight Extensions of Conditional Logit

- Under the following assumptions on utilituy:

$$U_j^* = W^\top \beta_j + X_j^\top \gamma + \epsilon_j, \text{ or}$$
$$U_j^* = W\beta_j + X_j\gamma_1 + X_jW\gamma_2 + \epsilon_j \quad \text{(for notational simplicity, } W \in \mathbb{R}, X_j \in \mathbb{R})$$

  similar results to the one above can be easily obtained.

Independence of Irrelevant Alternatives (IIA)

- In the canonical multinomial logit model, the response probabilities are given by (2), which leads to the following unrealistic restriction:

$$\frac{P_j(X|\theta)}{P_l(X|\theta)} = \frac{\exp(X^\top \beta_j)}{\exp(X^\top \beta_l)}. \tag{3}$$

- This restriction (3) is called independence of irrelevant alternatives (IIA), meaning that the choice between j and l is independent of the other alternatives and hence the latter are irrelevant to the bivariate choice between j and l.
- Where does this IIA problem arise? The IIA structure (coming from multinomial logit model) excludes differentiated substitutability among alternatives.
- In other words, (part of) the problem is due to the restrictive correlation pattern imposed on the errors by the GEV distribution.

Nested Logit

- Assuming a more flexible correlation structure can mitigate the IIA problem, which allows subsets of alternatives to have differential correlations.
- One solution is the nested logit model, which separates the alternatives into groups (nests). Alternatives within groups are allowed to be correlated, but are assumed uncorrelated across groups.
- Suppose that there exist J groups each with K_j alternatives. Assume that the utility of the jk-th alternative is given by

$$U_{jk}^{*} = W^{\top}\beta_{jk} + X_{jk}^{\top}\gamma + \epsilon_{jk},\ F(\epsilon_{11}, \cdots, \epsilon_{JK_J}) = \exp\left(-\sum_{j=1}^{J}\left[\sum_{k=1}^{K_j}\exp\left(-\frac{\epsilon_{jk}}{\tau_j}\right)\right]^{\tau_j}\right)$$

where W and X_{jk} denote individual-specific regressors and regressors varying by alternative, respectively.

Mixed Logit

Simple Multinomial Probit

General Multinomial Probit

Ordered Response

Count Data

Berry, Levinson and Pakes (1995)

Econometrica, Vol. 63, No. 4 (July, 1995), 841–890

AUTOMOBILE PRICES IN MARKET EQUILIBRIUM

BY STEVEN BERRY, JAMES LEVINSOHN, AND ARIEL PAKES[1]

This paper develops techniques for empirically analyzing demand and supply in differentiated products markets and then applies these techniques to analyze equilibrium in the U.S. automobile industry. Our primary goal is to present a framework which enables one to obtain estimates of demand and cost parameters for a class of oligopolistic differentiated products markets. These estimates can be obtained using only widely available product-level and aggregate consumer-level data, and they are consistent with a structural model of equilibrium in an oligopolistic industry. When we apply the techniques developed here to the U.S. automobile market, we obtain cost and demand parameters for (essentially) all models marketed over a twenty year period.

KEYWORDS: Demand and supply, differentiated products, discrete choice, aggregation, simultaneity, automobiles.

Berry, Levinson and Pakes (1995, Ecta)

- Here we review a method of aggregate market demand estimation developed by Berry, Levinson and Pakes (1995).
- Setup
  - Differentiated product market
  - Demand: discrete choice model
  - Supply: Multi-product Bertrand competition, in which firms set the prices of their products
- Contributions of BLP (1995)
  - Discrete choice models are usually estimated with individual-level consumer data.
  - They proposed an approach for estimating discrete-choice demand only with aggregate, market-level sales data.

Related Literature to BLP (1995, Ecta)

- Berry (1994)
- Berry, Levinson and Pakes (1999, AER)
- Gentzkow and Shapiro (2015)
- Gandhi and Houde (2020)
- Berry and Haile (2021)
- Gandhi and Nevo (2021)
- 上武，遠山，若森，渡辺（2021）「実証ビジネス・エコノミクス第３回（経済セミナー連載）」[1]
- 遠山先生の RA Bootcamp 講義資料 [2024] [2025]

[1]近日中に書籍化されるようです.