Multiple Choice

Hansen (2022, Sections 26.7-26.12)

Yasuyuki Matsumura (Kyoto University)

Last Updated: October 13, 2025

https://yasu0704xx.github.io

Textbooks

- Train (2009, Chapters 2-7, 11)
- Wooldridge (2010, Chapter 13)
- Hansen (2022, Chapter 26)
- 上級計量経済学 07 (Conditional Maximum Likelihood Estimation) by 柳先生

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.

Simple 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 $\mathsf{AME}_j = \mathbb{E}\left[\delta_j(X)\right]$ can be estimated by $\widehat{\mathsf{AME}}_j = \frac{1}{n} \sum_{i=1}^{n} \widehat{\delta}_j(X_i)$.

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

$$P_j(w, x) = \frac{\exp\left(x_j^\top \gamma\right)}{\sum_{l=1}^J \exp\left(x_l^\top \gamma\right)}. \tag{3}$$

- 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:

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

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:

$$\frac{P_j(X|\theta)}{P_l(X|\theta)} = \frac{\exp(X^\top \beta_j)}{\exp(X^\top \beta_l)}. \tag{4}$$

- This restriction (4) 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

- 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.

- Under the above structure, the nested logit response probabilities are given by

$$P_{jk} = P_{k|j} P_j,$$

where

$$P_{k|j} = \frac{\exp\left(\frac{W^\top \beta_{jk} + X_{jk}^\top \gamma}{\tau_j}\right)}{\sum_{m=1}^{K_j} \exp\left(\frac{W^\top \beta_{jk} + X_{jk}^\top \gamma}{\tau_j}\right)}, \quad P_j = \frac{\left(\sum_{m=1}^{K_j} \exp\left(\frac{W^\top \beta_{jm} + X_{jm}^\top \gamma}{\tau_j}\right)\right)^{\tau_j}}{\sum_{l=1}^{J} \left(\sum_{m=1}^{K_l} \exp\left(\frac{W^\top \beta_{lm} + X_{lm}^\top \gamma}{\tau_l}\right)\right)^{\tau_l}}.$$

- Letting θ be the parameters, the log-likelihood function is given by

$$l_n(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{k=1}^{K_j} 1\{Y_i = jk\} \left(\log P_{k|j}(W_i, X_i|\theta) + \log P_j(W_i, X_i|\theta)\right).$$

- Marginal effects can (in principle) be caluculated but are complicated functions of the coefficients.

Mixed Logit

Mixed Logit (Random Coefficient Logit)

- Another solution to the IIA issue is the mixed logit (random coefficient logit) model.
- Suppose that there exist J alternatives. Assume that the utility of the j-th alternative is given by

$$U_j^* = W^\top \beta_j + X_j^\top \eta + \epsilon_j, \tag{5}$$

  where η is an individual-specific random variable with distribution $F(\eta|\alpha)$ with parameters α, and ϵ_j is an error with IID extreme value.
- Typical choices for $F(\eta|\alpha)$
  - $\eta \sim$ Normal(γ, D) with diagonal covariance matrix D
  - $\eta \sim$ Normal(γ, Σ) with unconstrained covariance matrix Σ
  - Log-normally distributed η (to enforce $\eta \geq 0$)
- For computational simplicity, it is common to partition X_j so that some variables have random coefficients and others have fixed coefficients.

How mixed logit mitigate IIA?

- Letting $\gamma = \mathbb{E}[\eta]$ and $V_j = X_j^\top(\eta - \gamma) + \epsilon_j$, the model in (5) can be rewritten as

$$U_j^* = W^\top \beta_j + X_j^\top \gamma + V_j,$$

which is the conventional random utility framework with errors V_j.

- Notice that the errors V_j are conditionally heteroscedastic and correlated across alternatives:

$$\mathbb{E}[V_j V_l | X_j X_l] = X_j^\top \mathsf{var}(\eta) X_l.$$

Here, the non-zero correlation means that the IIA property is partially broken, providing the mixed logit model with more flexibility than the conditional logit model.

Response Probability & Log-Likelihood

- The conditional response probabilities given η is

$$P_j(w, x|\eta) = \frac{\exp\left(w^\top \beta_j + x_j^\top \eta\right)}{\sum_{l=1}^{J} \exp\left(w^\top \beta_l + x_l^\top \eta\right)},$$

following from (3).

- The unconditional response probabilities are given by

$$P_j(w, x) = \int P_j(w, x|\eta) dF(\eta|\alpha). \tag{6}$$

- Letting θ be the list of all parameters, the log-likelihood function is given by

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

Monte Carlo Integration

- The integral in (6) is not available in closed form.
- A standard implementation is Monte Carlo integration (estimation by simulation).
- Let $\{\eta_1, \cdots, \eta_G\}$ be a set of IID pseudo-random draws from $F(\eta|\alpha)$.
- The simulation estimator of (6) is

$$\tilde{P}_j(w, x) = \frac{1}{G} \sum_{g=1}^{G} P_j(w, x|\eta_g),$$

which converges in probability to $P_j(w, x)$ in (6) as G increases.

Simple Multinomial Probit

Simple Multinomial Probit

- The simple multinomial probit model assumes that the utility from alternative j is given by

$$U_j^* = W^\top \beta_j + \epsilon_j, \quad \epsilon_j \sim \text{IID, Normal}(0, 1), \tag{7}$$

which is identical to the simple logit model, except for the error structure.

- As it assumes that the errors are independent, the simple multinomial probit model does not allow two alternatives to be close substitutes, which is not, but similar to, the IIA.

Conditional Multinomial Probit

- The conditional multinomial probit model assumes that the utility from alternative j is given by

$$U_j^* = W^\top \beta_j + X_j^\top \gamma + \epsilon_j, \quad \epsilon_j \sim \text{IID, Normal}(0,1), \tag{8}$$

  which is identical to the conditional logit model, except for the error structure.
- Note that the simple multinomial probit structure (7) is a special case of the conditional multinomial models.

Response Probabilitiy

- Under the structure (8), the response probabilities are given by the following one-dimensional normal integral over the $J-1$ fold product of normal distribution functions:

$$P_j(W,X) = \int_{-\infty}^{\infty} \prod_{l \neq j} \Phi\left(W^\top(\beta_j - \beta_l) + (X_j - X_l)^\top \gamma + \nu\right) \phi(\nu) d\nu, \tag{9}$$

  where $\Phi(\cdot)$ and $\phi(\cdot)$ are the CDF and PDf of Normal$(0,1)$, respectively.
- The response probabilities (9) are not available in closed form.
- Letting θ denote the parameters, The log-likelihood function is given by

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

General Multinomial Probit

General Multinomial Probit

- The general multinomial probit model assumes that the utility from alternative j is given by

$$U_j^* = W^\top \beta_j + X_j^\top \gamma + \epsilon_j, \quad \epsilon \sim \text{Normal}(0, \Sigma),$$

where Σ is an unconstrained variance-covariance matrix.

- Identification:
    - The coefficients β_j and γ are only identified up to scale.
    - The coefficients β_j are only identified relative to a baseline alternative J.
    - Σ requires some normalization.[1]
- The response probabilities is typically estimated by simulated maximum likelihood (SML), which is developed by Geweke, Hajivassiliou and Keane (1997, GHK).

[1] Note that the scale of differenced utility $U_j^* - U_J^*$ cannot be identified.

Ordered Response

Ordered Response

- Let $Y = \{1, \cdots, J\}$ be an ordered discrete variable and X be a vector of regressors. Here we assume that X does not include the intercept.
- The standard approach to ordered response assumes the latent utility framework:

$$U^* = X^\top \beta + \epsilon, \quad \epsilon \sim G.$$

Under the above structure, the ordered response model specifies that the response Y is determined by the following set of threshold crossing rules:

$$\begin{aligned} Y &= 1 \quad \text{if} \quad U^* \leq \alpha_1, \\ Y &= 2 \quad \text{if} \quad \alpha_1 < U^* \leq \alpha_2, \\ &\vdots \\ Y &= J \quad \text{if} \quad \alpha_{J-1} < U^*. \end{aligned}$$

where $\alpha_1 < \cdots < \alpha_{J-1}$ are unknown, non-stochastic, parameters.

Response Probabilities & Log-Likelihood

- The distribution $G(\cdot)$ of the error ϵ is typically assumed known. Common choices include $\epsilon \sim \text{Normal}(0,1)$ (ordered probit) and $\epsilon \sim \Lambda(z) = \exp(z)/[1+\exp(z)]$ (ordered logit).
- The response probabilities are given by

$$P_j(x) = \mathbb{P}[Y = j | X = x] = \cdots = G(\alpha_j - x^\top \beta) - G(\alpha_{j-1} - x^\top \beta),$$

  and the marginal effects by

$$\frac{\partial}{\partial x} P_j(x) = \beta \left(g(\alpha_{j-1} - x^\top \beta) - g(\alpha_j - x^\top \beta) \right).$$

- Letting $\theta = (\beta, \alpha_1, \cdots, \alpha_{J-1})$, the log-likelihood function is given by

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

- It may be easier to interpret the cumulative response probabilities:

$$\mathbb{P}[Y \leq j | X = x] = G(\alpha_j x^\top \beta).$$

The marginal cumulative effects are

$$\frac{\partial}{\partial x}\mathbb{P}[Y \leq j | X = x] = -\beta g(\alpha_j x^\top \beta).$$

Count Data

Count Data

- While the multinomial model and the ordered choice model assume that the maximum realization of the discrete dependent variable is known a priori, the econometrician may not have such a priori knowledge.
- Let $Y \in \{0, 1, 2, \cdots\}$ be a discrete dependent variable with unbounded support. A count data model specifies the response probabilities $P_j(x) = \mathbb{P}[Y = j|x]$ for $j = 0, 1, 2, \cdots$ with the property $\sum_{j=0}^{\infty} P_j(x) = 1$.
- Typical specifications include Poisson regression:

$$P_j(x) = \frac{\exp\left(-\lambda(x)\right) \lambda(x)^j}{j!}, \quad \lambda(x) = \exp(x^\top \beta).$$

- The Poisson distribution satisfies the following properties:

$$\mathbb{E}[Y|X] = \exp\left(X^\top \beta\right), \quad \text{var}[Y|X] = \exp\left(X^\top \beta\right).$$

- The log-likelihood function is given by

$$l_n(\beta) = \sum_{i=1}^{n} \log P_{Y_i}(X_i|\beta) = \sum_{i=1}^{n} \left(- \exp(X_i^\top \beta) + Y_i X_i^\top \beta - \log(Y_i!) \right).$$

- Its first and second derivatives are

$$\frac{\partial}{\partial \beta} l_n(\beta) = \sum_{i=1}^{n} X_i \left(Y_i - \exp(X_i^\top \beta) \right), \text{ and}$$

$$\frac{\partial^2}{\partial \beta \partial \beta^\top} l_n(\beta) = - \sum_{i=1}^{n} X_i X_i^\top \exp\left(X_i^\top \beta \right),$$

respectively. Since the second derivative is globally negative definite, the log-likelihood function is globally concave.

More Flexibility?

- Nonparametric identification: Suppose that the true conditional mean is nonparametric. Since it is non-negative, we can write

$$\mathbb{E}[Y|X] = \exp\left(m(x)\right) \iff m(x) = \log\left(\mathbb{E}[Y|X]\right).$$

  The function $m(x)$ is nonparametrically identified and can be approximated by a series $x_K^\top \beta_K$, so that $\mathbb{E}[Y|X] \simeq \exp(X_K^\top \beta_K)$.

- Random coefficients/Negative Binomial model: Specify the Poisson parameter as $\lambda(X) = V\exp(X^\top\beta)$ where V is a random variable with a Gamma distribution. Integrating out V, the resulting conditional distribution for Y is Negative Binomial. The Negative Binomial is a popular model for count data regression, and has the advantage that the conditional mean and variance are separately varying.

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）」[2]
- 遠山先生の RA Bootcamp 講義資料 [2024] [2025]

[2]近日中に書籍化されるようです.