Multiple Choice & BLP

Hansen (2022, Sections 26.7-26.12)

Yasuyuki Matsumura (Kyoto University)

Last Updated: October 15, 2025

https://yasu0704xx.github.io

References

- Train (2009, Chapters 2-7, 11)
- Wooldridge (2010, Chapter 13)
- Handbook of Industrial Organization (Volume 4, 2021)
- 上級計量経済学 07 (Conditional Maximum Likelihood Estimation) & 08 (Nonlinear GMM Estimation) by 柳先生

Contents

Review on Sections 26.1-26.6

Mixed Logit

Simple Multinomial Probit

General Multinomial Probit

Ordered Response

Count Data

Review on Random Utility Model

Berry, Levinson and Pakes (1995)

Appendix:
Semiparametric Discrete Choice Model (Kaneko & Toyama, 2025)

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:

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

  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:

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

- 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

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

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 $F(\eta|\alpha)$
  - $\eta \sim \text{Normal}(\gamma, D)$ with diagonal covariance matrix D
  - $\eta \sim \text{Normal}(\gamma, \Sigma)$ 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:

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

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:

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

  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 Probability & 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.

Review on Random Utility Model

Berry and Haile (2021, Handbook of IO)

ELSEVIER

Handbook of Industrial Organization

Volume 4, Issue 1, 2021, Pages 1-62

Chapter 1 - Foundations of demand estimation ☆

Steven T. Berry, Philip A. Haile

Show more

+ Add to Mendeley   Share   Cite

https://doi.org/10.1016/bs.hesind.2021.11.001 ↗

Get rights and content ↗

Abstract

Demand elasticities and other features of demand are critical determinants of the answers to most positive and normative questions about market power or the functioning of markets in practice. As a result, reliable demand estimation is an essential input to many types of research in Industrial Organization and other fields of economics. This chapter presents a discussion of some foundational issues in demand estimation. We focus on the distinctive challenges of demand estimation and strategies one can use to overcome them. We cover core models, alternative data settings, common estimation approaches, the role and choice of instruments, and nonparametric identification.

Setup: Random Utility Model

- Discrete choice demand is commonly represented with a random utility model.
- Let $j = 1, \cdots, J_i$ index the inside goods available to consumer i, while $j = 0$ denotes the outside good.
- A consumer's choice set is characterized by J_i and a set χ_i, which may include observed characteristics of consumer i, observed characteristics (including prices) of the available goods, observed characteristics of the local market, and characteristics of the market or goods that are unobserved to the researcher.
- Each consumer i has a conditional indirect utility (henceforth, utility) u_{ij} for good j. Consumer i knows her utilities for all goods and chooses the good yielding her highest utility.

Heterogenous Preference

- Consumer preferences are permnitted to be heterogenous, even when conditioning on any consumer characteristics included in χ_i.
- This heterogeneity is modeled by treating utility as varying at random across consumers: Given the choice set (J_i, χ_i), each consumer's utility vector $(u_{i0}, u_{i1}, \cdots, u_{iJ_i})$ is an independent draw from a joint distribution $F_u(\cdot|J_i, \chi_i)$.
- Because a consumer's behavior depends only on her ordinal ranking of goods, below we will normalize the location and scale of each consumer's utility vector w.l.o.g.
- We assume that the distribution $F_u(\cdot|J_i, \chi_i)$ is such that "ties" $(u_{ij} = u_{ik}$ for $j \neq k)$ occur with probability zero.

Consumer-Specific Choice Probability

- Represent consumer i's choice with the vector $(q_{i1}, \cdots, q_{iJ_i})$, where

$$q_{ij} = 1\left\{u_{ij} \geq u_{ik}, \quad \forall k \in \{0, 1, \cdots, J_i\}\right\}.$$

- Then, consumer-specific choice probabilities are given by

$$\begin{aligned}
s_{ij} &= \mathbb{E}[q_{ij}|J_i, \chi_i] \\
&= \int_{\mathcal{A}_{ij}} dF_u(u_{i0}, \cdots, u_{i1}, \cdots, u_{iJ_i}|J_i, \chi_i), \\
\mathcal{A}_{ij} &= \left\{(u_{i0}, u_{i1}, \cdots, u_{iJ_i}) : u_{ij} \geq u_{ik} \, \forall k\right\}.
\end{aligned}$$

Example: 2-Product Case (Berry and Haile, 2021, Figure 1)

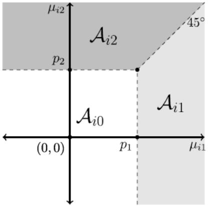

FIGURE 1

Choice regions for goods 0, 1, and 2.

To illustrate, consider an example with $J_i = 2$. Let p_j denote the price of good j and let

$$u_{ij} = \mu_{ij} - p_j$$

for $j > 0$, where (μ_{i1}, μ_{i2}) are drawn from a joint distribution $F_\mu(\cdot)$. Set $u_{i0} = 0$, normalizing the location of utilities. Fig. 1 then illustrates the regions in (μ_{i1}, μ_{i2})-space leading consumer i to choose goods 0, 1, and 2. For example, only consumers for whom $\mu_{i2} - p_2 > 0$ prefer good 2 to the outside option. The dark gray region is the set of (μ_{i1}, μ_{i2}) combinations such that this holds and $\mu_{i2} - p_2 > \mu_{i1} - p_1$, i.e., the set $\mathcal{A}_{i2}$. Similarly, the light gray region corresponds to $\mathcal{A}_{i1}$. The choice probabilities for consumer i then correspond to the probability measure assigned to each region by $F_\mu(\cdot)$.

Discrete Choice Demand Model

- A parametric random utility specification:

$$\text{For } j > 0, \quad u_{ijt} = x_{jt}\beta_{it} - \alpha_{it}p_{jt} + \xi_{jt} + \epsilon_{ijt}, \tag{10}$$

$$\text{For } j = 0, \quad u_{i0t} = \epsilon_{i0t}, \tag{11}$$

ϵ_{ijt} is an IID draw from a standard type-1 extreme value distribution,

yielding a mixed[2] multinomial logit[3] model.

- The notion of a market t is central to this formulation and will allow a precise characterization of the endogeneity problems inherent to demand estimation.

[2] The term "mixed" reflects the heterogeneity across consumers in the parameters α_{it} and β_{it} that characterize their marginal rates of substitution between the various observed and unobserved characteristics.

[3] A normal distribution will yield a mixed multinomial probit.

Market

- Let $\mathcal{J}_t$ denote the set of products (inside goods) available to consumers in market t, and let $J_t = |\mathcal{J}_t|$.
- p_{ijt} represents the price of good j in market t.
- $x_{ijt} \in \mathbb{R}^K$ represents other observable characteristics of good j in market t.
- ξ_{jt} is an unobserved factor (demand shock) associated with good j and market t.
    - The demand shock $\xi_j t$ is often described as a measure of good j's unobserved characteristics.
    - But this is more restrictive than necessary: ξ_{jt} can represent any combination of latent taste variation and latent product characteristics common to consumers in market t.
    - For example, a high value of ξ_{jt} may simply indicate that consumers in market t have a high mean taste for good j.
- Let $x_t = (x_{1t}, \cdots, x_{J_t,t})$, $p_t = (p_{1t}, \cdots, p_{J_t,t})$, $\xi_t = (\xi_{1t}, \cdots, \xi_{J_t,t})$, and $\chi_t = (x_t, p_t, \xi_t)$.

Endogeneity

- Typically, one allows prices p_t to be correlated with ξ_t.
    1. Standard models of oligopoly competition imply that prices are endogenous.
    2. The equilibrium price of any good j in market t will depend on all components of x_t and ξ_t, as these alter the residual demand for good j.
    3. Equilibrium prices are affected by latent shocks to marginal costs, which we typically expect to be correlated with demand unobservables.
    4. When marginal costs are upward-sloping, this will imply dependence of equilibrium marginal costs (and thus, prices) on demand shocks.
- Exogeneity of the remaining product characteristics x_t is often assumed, and we will do so in what follows.
    - This is not essential.
    - On the demand side, allowing endogeneity of additional characteristics is conceptually straightforward but lead to more demanding instrumental variables requirements.

Choice Probabilities in the Population/Market Shares

- Choice probabilities in the population reflect a mixture of the choice probabilities conditional on each possible combination of $(\alpha_{it}, \beta_{it})$.
- In the mixed multinomial logit model, choice probabilities in the population (i.e., market shares) are given by

$$s_{jt} = \int \frac{\exp\left(x_{jt}\beta_{it} - \alpha_{it}p_{jt} + \xi_{jt}\right)}{\sum_{k=0}^{J_t} \exp\left(x_{kt}\beta_{it} - \alpha_{it}p_{kt} + \xi_{kt}\right)} dF(\alpha_{it}, \beta_{it}; t), \tag{12}$$

  where $F(\cdot; t)$ denotes the joint distribution of $(\alpha_{it}, \beta_{it})$ in market t.
- The latent taste parameters $(\alpha_{it}, \beta_{it})$ are often referred to as random coefficients.

Specification for $F(\cdot\,; t)$

- Each component k of the random coefficient vector β_{it} is commonly specified as taking the form

$$\beta_{it}^{(k)} = \beta_0^{(k)} + \beta_\nu^{(k)} \nu_{it}^{(k)} + \sum_{l=1}^{L} \beta_d^{(l,k)} d_{ilt}. \tag{13}$$

  - $\beta_0^{(k)}$is a parameter shifting all consumers' tastes for $x_{jt}^{(k)}$.
  - Each d_{ilt} represents a characteristics (e.g., demographic measure) of individual i.
  - Each $\nu_{it}^{(k)}$ is a random variable with a pre-specified distribution (e.g., a standard normal).
  - The parametrs $\beta_d^{(l,k)}$ and $\beta_\nu^{(k)}$ govern the extent of variation in tastes for $x_{jt}^{(k)}$ across consumers with different demographic characteristics d_{it} or taste shocks $\nu_{it}^{(k)}$.
  - The distinction between d_{ilt} and $\nu_{it}^{(k)}$ reflects the fact that each d_{ilt} (or at least its distribution in the population) is assumed to be known.

- The treatment of the coefficient on price, α_{it}, is similar. A typical specification takes the form

$$\log(\alpha_{it}) = \alpha_0 + \alpha_y y_{it} + \alpha_\nu \nu_{it}^{(0)}$$

  - y_{it} represents consumer-specific measures that are posited to affect price sensitivity.
  - The variables included in y_{it} might overlap partially or entirely with d_{it}.

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)

- 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 characterized by random utility 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.

Recommended Readings

- Berry (1994)
- Berry, Levinson and Pakes (1999)
- 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）」[4]
- RA Bootcamp 講義資料 [2024] [2025] by 遠山先生

[4]近日中に書籍化されるようです.

Market-Level Data

- In many applications, the key data are observed at the market level. In such cases, one typically observes
  - the number of goods J_t available to consumers in each market t;
  - their prices and other observable characteristics p_t, x_t;
  - their observed market shares $\tilde{s}_{jt}$, typically measured as the total quantity of good j sold in market t divided by the number of consumers (e.g., households) in that market;
  - the distribution of consumer characteristics (d_{it}, y_{it}) in each market; and
  - (possibly) additional variables w_t (e.g., cost shifters) that might serve as appropriate instruments.

Estimation with Aggregate Market-Level Data

- The standard approach to estimation of discrete choice demand from market-level data was developed in BLP (1995), with many subsequent variations and extensions.
- Here we consider the slightly simplified version of their model with a non-random coefficient on price. Thus, the random utility specification (10) becomes

$$u_{ijt} = x_{jt}\beta_{it} - \alpha_0 p_{jt} + \xi_{jt} + \epsilon_{ijt}, \tag{14}$$

  for $j > 0$, with $u_{i0t} = \epsilon_{i0t}$.
- Following BLP (1995), we assume that each ϵ_{ijt} is an IId drawfrom a standard type-1 extreme value distribution, and that each $\nu_{it}^{(k)}$ in (13) is an IID draw from a standard normal distribution.

- Observe that (14) can be rewritten as

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \epsilon_{ijt}, \tag{15}$$

where we have defined

$$\delta_{jt} = x_{jt}\beta_0 - \alpha_0 p_{jt} + \xi_{jt}, \tag{16}$$

$$\mu_{ijt} = \sum_{k=1}^{K} x_{jt}^{(k)} \left(\sum_{l=1}^{L} \beta_d^{(l,k)} d_{ilt} + \beta_\nu^{(k)} \nu_{it}^{(k)} \right). \tag{17}$$

- Let $F_\mu(\cdot|x_t, \beta_d, \beta_\nu)$ denote the conditional joint distribution of the stochastic terms $(\mu_{i1t}, \cdots, \mu_{iJ_tt})$ given $(x_t, \beta_d, \beta_\nu)$. Given the assumptions above, this distribution is known up tp the parameters (β_d, β_ν).

Berry Inversion

- Letting $\delta_t = (\delta_{1t}, \cdots, \delta_{J_t t})$, the market shares implied by the model take the form

$$\sigma_j(\delta_t, x_t, \beta_d, \beta_\nu, J_t) = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{\sum_{k=0}^{J_t} \exp(\delta_{kt} + \mu_{ikt})} dF_\mu(\mu_{it} | x_t, \beta_d, \beta_\nu, J_t) \tag{18}$$

  for each good j.
- Berry (1994) demonstrated that the demand system

$$\sigma(\delta_t, x_t, \beta_d, \beta_\nu, J_t) = (\sigma_1(\delta_t, x_t, \beta_d, \beta_\nu, J_t), \cdots, \sigma_{J_t}(\delta_t, x_t, \beta_d, \beta_\nu, J_t))$$

  is invertible. Given x_t, β_d, β_ν and any vector of nonzero market shares $s = (s_1, \cdots, s_{J_t})$ in market t such that $1 - \sum_{j>0} s_{jt} > 0$, there exists a unique vector δ for market t such that

$$\sigma(\delta, x_t, \beta_d, \beta_\nu, J_t) = s.$$

The BLP Estimator

- At the broadest level, an estimation strategy involves searching (or solving) for the parameters of the model that allow it to best fit the data.
- Let $\theta \equiv (\alpha_0, \beta_0, \beta_d, \beta_\nu)$ represent all the parameters of the model.
  - It will be useful to partition these as $\theta_1 \equiv (\alpha_0, \beta_0)$ and $\theta_2 \equiv (\beta_d, \beta_\nu)$.
  - In the literature, the elements of θ_1 are often referred to as the linear parameters and with θ_2 referred to as nonlinear parameters.
  - Note that we can rewrite the model's prediction of market shares (18) as

$$s_{jt} = \sigma_j(\delta_t, x_t, \theta_2, J_t).$$

- Because the identification of the model will rely on instrumental variables, it is natural to formulate an estimator using moment consditions. BLP (1995) proposed a generalized method of moments (GMM) estimation approach.

- Let T denote the number of markets in the sample and let $N = \sum_{t=1}^{T}$.
- The BLP estimator $\widehat{\theta}$ is defined as the solution to a mathematical program:

$$\min_{\theta} g\left(\xi(\theta)\right)^{\top} \Omega\, g\left(\xi(\theta)\right)$$

subject to

$$g\left(\xi(\theta)\right) = \frac{1}{N} \sum_{\forall j,t} \xi_{jt}(\theta) z_{jt}$$

$$\xi_{jt}(\theta) = \delta_{jt}(\theta_2) - x_{jt}\beta + \alpha p_{jt}$$

$$\log(\tilde{s}_{jt}) = \log\left(\sigma_j(\delta_t, x_t, \theta_2, J_t)\right)$$

$$\sigma_j(\delta_t, x_t, \theta_2, J_t) = \int \frac{\exp\left(\delta_{jt}(\theta_2) + x_{jt}\tilde{\beta}\right)}{1 + \sum_k \exp\left(\delta_{kt}(\theta_2) + x_{kt}\tilde{\beta}\right)} f_{\tilde{\beta}}(\tilde{\beta}|\theta_2),$$

where Ω denotes the standard GMM weight matrix and $f_{\tilde{\beta}}(\cdot|\theta_2)$ denotes the joint density of $\tilde{\beta}_{it} = \beta_{it} - \beta_0$, i.e., the consumer-specific components of the coefficients β_{it}.

Sketch of GMM Estimation Approach (Berry and Haile, 2021, Section 4.1)

1. take a trial value of the parameters θ;
2. for each market t, "invert" the demand model at the observed market shares $\tilde{s}_t$ to find the unique vector $\xi_t \in \mathbb{R}^{J_t}$ such that, given the definition (4.3), $\sigma_j(\delta_t, x_t, \theta_2, J_t) = \tilde{s}_{jt}$ for all j;
3. evaluate the trial value θ using a GMM criterion function based on moment conditions of the form

$$E[\xi_{jt}(\theta) z_{jt}] = 0,$$

   where $z_{jt} \supset x_{jt}$ is a vector of appropriate instrumental variables;
4. repeat from step 1 until a minimum is found.

Computation & Inference

- See Section 4.4 of Berry and Haile (2021) for details.
- Open-source software: PyBLP
    - Conlon and Gortmaker (2020) serve as an introduction.

The RAND Journal of Economics

Original Article

Best practices for differentiated products demand estimation with PyBLP

Christopher Conlon, Jeff Gortmaker

First published: 26 November 2020 | https://doi.org/10.1111/1756-2171.12352 | Citations: 91

Thanks to Steve Berry, Jeremy Fox, Phil Haile, Mathias Reynaert, and Frank Verboven and seminar participants at NYU, Rochester, and the 2019 IIOC conference. Thanks to the editor Marc Rysman and to three anonymous referees. Daniel Stackman provided excellent research assistance. Any remaining errors are our own.

Choice of Instruments

- For the use of proxies of cost shifters, see e.g., Hausman, Leonard and Zona (1994); Hausman (1996); and Nevo (2001).
- For the use of exogenous characteristics of competing goods (BLP instruments), see e.g., BLP (1995); and Berry and Haile (2021, Sections 4.2.5 and 5.4).
- For the use of characteristics nearby markets (Waldfogel-Fan instruments), see e.g., Waldfogel (2003); Fan (2013); Williams and Adams (2019); and DellaVigna and Gentzkow (2019).
- For the use of exogenous changes in market structure, see e.g., Miller and Weinberg (2017).
- For the notion of "optimal" instruments, see e.g., BLP (1999); Reynaert and Verboven (2014); Gandhi and Houde (2020); and Conlon and Gortmaker (2020).

Appendix: Semiparametric Discrete Choice Model (Kaneko & Toyama, 2025)

THE JOURNAL OF INDUSTRIAL ECONOMICS 0022-1821
Volume LXXIII March 2025 No. 1

DEMAND ESTIMATION WITH FLEXIBLE INCOME EFFECT: AN APPLICATION TO PASS-THROUGH AND MERGER ANALYSIS*

Shuhei Kaneko[†]
Yuta Toyama[‡]

This article proposes a semiparametric discrete choice model that incorporates a nonparametric specification for income effects. The model allows for the flexible estimation of demand curvature, which has significant implications for pricing and policy analysis in oligopolistic markets. Our estimation algorithm adopts a method of sieve approximation with shape restrictions in a nested fixed-point algorithm. Applying this framework to the Japanese automobile market, we conduct a pass-through analysis of feebates and merger simulations. Our model predicts a higher pass-through rate and more significant merger effects than parametric demand models, highlighting the importance of flexibly estimating demand curvature.

Kaneko & Toyama (2025)

- A semiparametric discrete choice model
  - Proposing nonparametric sieve approximation of income effect
  - Resulting in more accurate estimation of demand curvature, price elasiticity, and welfare changes.
- Empirical application
  - A feebate policy in the Japanese automobile industry[5]
  - High pass-through rate
  - More significant merger effects (Toyota & Honda)

[5] Subsidy for eco-friendly cars.

Consumer Demand

- Accurate measurement of consumer demand is critical.
  - Price elasticity and substitution patterns are often what firms must consider.
  - Decision-making on pricing in oligopolistic markets
  - Evaluating the welfare consequences

Specification for Income Effects

- When estimating consumer demand for differentiated products, it is common to rely on parametric specifications.
- However, such parametrizing often imposes strict restrictions on the shape of demand curve.
- A semiparametric discrete choice model can adress this concern:
- This allows for the flexible estimation of demand curvature and price elasticity patterns.

- Combine a method of sieve approximation and nested fixed point algorithm.
- First, approximate the income effect by nonparametric sieve methods.
- Then, their model is closely aligns with the standard parametric framework of BLP.
- Second, implement a nested fixed-point algorithm to run sieve GMM estimation.

Utility Maximization Problem

- Let $U(m, j)$ denote the direct utility function.
  - m is a d_m dimensional vector representing the consumption of continuous choice goods.
  - $j \in \mathbb{J} = \{0, 1, \cdots, J\}$ corresponds to an alternative in the discrete choice decision, with J products available in the market. The index $j = 0$ indicates the outside goods.
- The utility maximization problem is given by

$$\max_{(m,j) \in \mathbb{R}_+^{d_m} \times \mathbb{J}} \quad U(m, j) \tag{19}$$

$$\text{s.t.} \quad P_m^T m + p_j \leq y_i,$$

where P_m is a d_m dimensional vector of prices of continuous choice goods, p_j is the price of alternative j, and y_i is income.

Conditional Indirect Utility Function

- Conditional on choice j in the discrete choice, the conditional indirect utility function is defined as

$$V(P_m, y - p_j, j) \equiv \max_{m \in \mathbb{R}_+^{d_m}} U(m, j) \text{ s.t.} P_m^T m \leq y_i - p_j. \tag{20}$$

Note that we define $p_0 = 0$ as choosing the outside good incurs no costs.

- Assume that the direct utility function satisfies

$$U(m, j) = v(j) + u(m). \tag{21}$$

- The conditional indirect utility function can be rewritten as

$$V(P_m, y - p_j, j) = v(j) + \tilde{V}(P_m, y - p_j). \tag{22}$$

Income Effect

- Assume that the continuous good is a numeraire, with its price represented by P^m. Then, we obtain

$$\tilde{V}(P^m, y - p_j) = u\left(\frac{y - p_j}{P^m}\right),$$

  implying that the utility from numeraire depends on the disposal income $y - p_j$ after choosing alternative j.

- Define the income effect term by

$$f(y - p_j) \equiv \tilde{V}(P^m, y - p_j).$$

  Note that $f(y - p_j)$ should be weakly-increasing.

Conditional Indirect Utility Function

- Letting v_{ij} denote consumer i's utility from a discrete choice good j, we specify that

$$v_{ij} = \beta^T X_j + \xi_j + \epsilon_{ij} \text{ for } j = 1, \cdots J, \tag{23}$$

$$v_{i0} = \epsilon_{i0}. \tag{24}$$

  where X_j is a vector of observable characteristics of product j, ξ_j represents its unobservable characteristics, and ϵ_{ij} is an IID idiosyncratic shock that follows the type I extreme-value distribution.
- Hence, the conditonal indirect utility function of consumer i when choosing j is given by

$$V_{ij} = \begin{cases} f(y_i - p_j) + \beta^T X_j + \xi_j + \epsilon_{ij} & \text{for } j = 1, \cdots J, \\ f(y_i) + \epsilon_{i0} & \text{for } j = 0. \end{cases} \tag{25}$$

Individual Choice Probability

- Define the choice set of consumer i as

$$\mathbb{J}_{it} = \{0\} \cup \{j \in \{1, \cdots, J_t\} : y_{it} - p_{jt} \geq 0\}, \tag{26}$$

where J_t is the total number of products available in market t.

- Given the conditional indirect utility V_{ijt} (25), the discrete choice problem is described as

$$\max_{j \in \mathbb{J}_{it}} V_{ijt}. \tag{27}$$

and the choice probability for consumer i selecting alternative j is derived as

$$s_{ijt}(y_{it}) = \frac{1(y_{it} \geq p_{jt}) \cdot \exp\left(f(y_{it} - p_{jt}) + \beta^T X_{jt} + \xi_{it}\right)}{\exp\left(f(y_{it})\right) + \sum_{k=1}^{J_t} 1(y_{it} \geq p_{kt}) \cdot \exp\left(f(y_{it} - p_{kt}) + \beta^T X_{kt} + \xi_{it}\right)}. \tag{28}$$

Market Share

- Letting y_{it} follow the distribution of income $G_t(y_{it})$, the market share is given by

$$s_{jt} = \int s_{ijt} dG_t(y_{it}). \tag{29}$$

- Market demand q_{jt} is given by

$$g_{jt} = N_t \times s_{jt}$$

where N_t denote the market size.

Practical Importance of the Flexible Income Effect

- Price Elasiticity: Avoid imposing any predetermined restrictions on how own-price elasticity varies with price.
- Pass-Through Analysis: Avoid inherent restriction on the demand curvature.
- Merger Analysis: Different curvatures of the demand funtion lead different simulated merger outcomes even under the same consumer demand with identical elasticities.

Estimation

- The utility function includes the nonparametric function $f(y-p)$ and the linear parameter β.
- Employ a sieve approximation for the nonparametric function and incorporate it into the nested fixed-point (NFP) algorithm.

Sieve Approximation

- Approximate $f(\cdot)$ by the K-th order Bernstein polynomial, i.e., by a linear function of the basis function $\Psi^K(x) = (b_0^K(x), b_1^K(x), \cdots b_K^K(x))^T$ and coefficients $\Pi = (\pi_0, \pi_1, \cdots \pi_k)^T$:

$$f(x) \simeq B_K(x) = \sum_{k=0}^{K} \pi_k b_k^K(x) \equiv \Psi^K(x)^T \Pi \tag{30}$$

where

$$b_k^K(x) = \binom{K}{k} x^k (1-x)^{K-k}, \tag{31}$$

and letting x be normalized to $[0, 1]$.

Shape Restrictions & Normalization

- Select the Bernstein polynomial as a basis function.
- Recall that $f(y-p)$ is weakly increasing (monotonicity). To incorporate this restriction within estimation, we impose constraints on the coefficients Π.
- Under $\pi_k \leq \pi_{k+1}$ for all k, the derivative of $B_K(x)$ (30) satisfies that

$$B'_K(x) = K \sum_{k=0}^{K-1} (\pi_{k+1} - \pi_k) b_k^{K-1}(x) \geq 0$$

  for all k, which is the desired monotonicity.
- The level of the income effect cannot be identified. Thus, letting $\pi_0 = 0$, we normalize $f(x)$ as $f(0) = 0$.

Approximated Model

- Under the sieve approximation above, the market share defined by (28) and (29) can be rewritten as

$$s_{jt} = \int \frac{1(y_{it} \geq p_{jt}) \cdot \exp\left(\Psi^K(y_{it} - p_{jt})^T \Pi + \beta^T X_j t + \xi_{jt}\right)}{\text{denom.}} dG_t(y_{it}), \tag{32}$$

where the denominator is given by

$$\exp\left(\Psi^K(y_{it})^T \Pi\right) + \sum_{k=1}^{J_t} 1(y_{it} \geq p_{jt}) \cdot \exp\left(\Psi^K(y_{it} - p_{jt})^T \Pi + \beta^T X_{kt} + \xi_{jt}\right)$$

- Note that there emerges an endogeneity between the product proce p_{jt} and the unobserved product characteristics ξ_{jt}.

Sieve GMM

- Moment Conditions: for $b = 1, \cdots, B$,

$$\mathbb{E}\left[\xi_{jt}(\theta) p_b(X_{jt}, W_{jt})\right] = 0, \tag{33}$$

where X_{jt} is a vector of exogenous variables, W_{jt} is a vector of IVs, $\theta = (\beta, \Pi)$, $\{p_b(X_{jt}, W_{jt})\}_{b=1,\cdots,B}$ is a sequence of known functions that can approximate any real-valued square-integrable functions of X_{jt} and W_{jt} as $B \to \infty$.

- GMM Criterion:

$$\xi(\theta)^T \tilde{P} \left(\tilde{P}^T \tilde{P}\right)^{-} \tilde{P}^T \xi(\theta)^T, \tag{34}$$

where $\xi(\theta)^T$ is a vector that stacks ξ_{jt}'s. The matrix $\tilde{P} = [P, P \otimes X]$ denotes a matrix of instruments.

NFP Algorithm

- Caluculation of the objective function & numerical optimization procedures are as follows:
  - 1. Caluculate the vector of mean utility δ by applying a contraction-mapping algorithm.
  - 2. Run a linear regression of δ on X and obtain $\widehat{\beta}$ and the residual $\widehat{\xi}_{jt}$.
  - 3. Calculate the value of the objective function (34).
  - 4. Run a nonlinear optimization routine over Π. [6]
- Inference: Generalized residual bootstrap

[6] Note that β appearing in the mean utility function can be obtained by employing a linear GMM.

Conclusion

- A new framework for a differentiated product demand model with a nonparametric income effect
- Estimate the semiparametric model with endogeneity by combining the NFP algorithm and a sieve approximation.
- Monte Carlo simulations suggest significant gains in estimating the nonparametric term of the income effect by incorporating the shape restriction.
- Applying their framework to Japanese automobile data, they demonstrate the importance of a flexible income effect specification.