EC6322 Advanced Industrial Organization
Demand Estimation

Hsin-Tien Tiffany Tsai

National University of Singapore

Today's Learning Goal

- Demand Estimation: Introduction, theory
- Discrete Choice Demand Model: Berry (1994)
- Estimation, Endogeneity
  - Empirical Example: Reimers and Waldfogel (2021)

Data Structure

Aggregate Market Level Data

- Aggregate Market Level Data
    - Longitudinal: one market/store across time
    - Cross-sections: multiple markets/stores
    - Panel: multiple markets/stores across time
- Typical variables used in estimation
    - Aggregate quantity
    - Prices/attributes/instruments
    - Definition of market size
    - Distribution of demographics (sometimes)

Typical Data Structure

- Market/Time 1
- Market/Time 2
- Market size (M) assumption needed to get shares.
- **Other type of data: Problem set!**

Logit Model

- Let d_i indicate the choice individual i makes where $d_i \in \{1, \cdots, J\}$. Individuals choose d to maximize their utility U, which is generally written as:

$$U_{ij} = \delta_j + \epsilon_{ij}$$

- where:
  - δ_j relates observed factors to the utility individual i receives from choosing option j.
  - ϵ_{ij} are unobserved to the econometrician but observed to the individual.

Logit Model

- Therefore, $d_{ij} = 1$ if $U_{ij} > U_{ij'}, \forall j' \neq j$, which is equivalent to $\delta_j + \epsilon_{ij} > \delta_{j'} + \epsilon_{ij'}, \forall j' \neq j$.
- With the $\epsilon' s$ unobserved, the probability of individual i making choice j is given by:

$$
\begin{aligned}
\mathrm{Pr}_{ij} &= \mathrm{Prob}\left(\delta_j + \epsilon_{ij} > \delta_{j'} + \epsilon_{ij'}, \forall j' \neq j\right) \\
&= \mathrm{Prob}\left(\epsilon_{ij'} - \epsilon_{ij} < \delta_j - \delta_{j'}, \forall j' \neq j\right) \\
&= \int_{\epsilon} \mathbb{1}\left(\epsilon_{ij'} - \epsilon_{ij} < \delta_j - \delta_{j'}, \forall j' \neq j\right) f(\epsilon) d\epsilon
\end{aligned}
$$

- where $\mathbb{1}(\cdot)$ is the indicator function. Note that, regardless of what distributional assumptions are made on the $\epsilon' s$, the probability of choosing a particular option does not change when we:
  - Add a constant to the utility of all options (utility is relative to one of the options, only differences in utility matter).
  - Multiply by a positive number (need to scale something, generally the variance of the ϵ 's).

Logit Model

- This means that in modeling utility we need to make some normalization - that is we need to bolt down a zero to measure things against. Normally we do the following:
  - Normalize the mean utility of the outside good to zero.
  - Normalize the coefficient on the idiosyncratic error term to 1.
- This allows us to interpret our coefficients and do estimation.

- Consider the case when the choice is binary and there are two choices $\{1, 2\}$. The type 1 extreme value (EVT1) c.d.f. is [1] $F(\epsilon) = \exp(-\exp(-\epsilon))$.
- To get the probability of choosing 1 , we have:

$$
\begin{aligned}
\Pr\left(d_i = 1\right) &= \mathrm{Prob}\left(\epsilon_{i2} - \epsilon_{i1} < \delta_1 - \delta_2\right) \\
&= \mathrm{Prob}\left(\epsilon_{i2} < \epsilon_{i1} + \delta_1 - \delta_2\right) \\
&= \exp\left(-\exp\left(-\left(\epsilon_{i1} + \delta_1 - \delta_2\right)\right)\right)
\end{aligned}
$$

Logit Model

- But ϵ_{i1} is unobserved and distributed as a EVT1 distribution, so we need to integrate it out:

$$\begin{aligned}
\Pr\left(d_i = 1\right) &= \int_{-\infty}^{\infty} \left[\exp\left(-\exp\left(-\left(\epsilon_{i1} + \delta_1 - \delta_2\right)\right)\right)\right] \cdot f\left(\epsilon_{i1}\right) d\epsilon_{i1} \\
&= \int_{-\infty}^{\infty} \exp\left(-\exp\left(-\epsilon_{i1}\right)\left[1 + \exp\left(\delta_2 - \delta_1\right)\right]\right) \exp\left(-\epsilon_1\right) d\epsilon_{i1}
\end{aligned}$$

- Then we do the substitution rule where $t = \exp\left(-\epsilon_{i1}\right)$ and $dt = -\exp\left(-\epsilon_{i1}\right) d\epsilon_{i1}$. Also note that $\exp(-\infty) = 0$ and $\exp(\infty) = \infty$. So, we get:

$$\begin{aligned}
\Pr\left(d_i = 1\right) &= \int_{\infty}^{0} \exp\left(-t\left[1 + \exp\left(\delta_2 - \delta_1\right)\right]\right) - dt \\
&= \left.\frac{\exp\left(-t\left[1 + \exp\left(\delta_2 - \delta_1\right)\right]\right)}{-\left[1 + \exp\left(\delta_2 - \delta_1\right)\right]}\right|_0^{\infty} \\
&= \frac{\exp\left(\delta_1\right)}{\exp\left(\delta_1\right) + \exp\left(\delta_2\right)}
\end{aligned}$$

Logit Model

- After we obtain the choice probabilities, we can estimate the parameters by MLE. Multinomial logit model. Basically, adding more choices with i.i.d. EVT1 errors yields the multinomial logit. The probability (likelihood) of choosing j is:

$$\Pr(d_i = j) = \frac{\exp(\delta_j)}{\sum_{k=0}^{J} \exp(\delta_k)}$$

- We normalize the utility of outside good to zero, then this yields the closed form expressions for the share of consumers who purchase inside goods j and outside good 0 :

$$s_j = \frac{\exp(\delta_j)}{1 + \sum_{k=1}^{J} \exp(\delta_k)}$$
$$s_0 = \frac{1}{1 + \sum_{k=1}^{J} \exp(\delta_k)}$$

- In the literature, when people talk about logit model in demand estimation, they usually mean "multinomial logit model".

Summary: Logit Model

- This model assumes everyone has the same taste for quality but have different idiosyncratic taste for the product.
- Utility is given by

$$U_{ij} = \delta_j + \epsilon_{ij}$$

- $\epsilon_{ij} \overset{iid}{\sim}$ extreme value type I $\left[F(\epsilon) = e^{-e^{-\epsilon}}\right]$. This is a very helpful assumption as it allows for the aggregate shares to have an analytical form.
  I.e.:

$$\Pr\left(U_{ij} \geq U_{ik} \forall k\right) = \frac{\exp\left(\delta_j\right)}{\sum_{k=0,\ldots,J} \exp\left(\delta_k\right)} \tag{2}$$

Summary: Logit Model

- This ease in aggregation comes at a cost, the embedded assumption on the distribution on tastes creates more structure than we would like on the aggregate substitution matrix.

- Independence of Irrelevant Alternatives (IIA): Ratio of choice probabilities between two options j and k doesn't depend on utilities of any other product. I.e.,:

$$\frac{P_{ij}}{P_{ik}} = \frac{e^{\delta_{ij}}}{e^{\delta_{ik}}}$$

  (Red bus-Blue bus issue)

- See McFadden 1972 for details on the construction.

Independence of Irrelevant Alternatives (IIA)

- One of the properties of the multinomial logit model is that $\frac{Pr_{ij}}{Pr_{ik}}$ does not depend upon what other alternatives are available:

$$\frac{Pr_{ij}}{Pr_{ik}} = \frac{\exp\left(\delta_j\right) / \sum_{q=0}^{J} \exp\left(\delta_q\right)}{\exp\left(\delta_k\right) / \sum_{q=0}^{J} \exp\left(\delta_q\right)}$$

$$= \frac{\exp\left(\delta_j\right)}{\exp\left(\delta_k\right)}$$

- The disadvantage of the IIA assumption is best illustrated by the red bus/blue bus problem.
  - Consider an individual who is deciding between riding a blue bus and taking a car. Adding a red bus, assuming that the only difference in utility between a red bus and a blue bus is in the error, will double the probability of taking a bus relative to taking a car. For example, say before the red bus is introduced, the probability of riding a blue bus versus taking a car is 50% versus 50%. The logit model predicts that after the red bus is introduced, the probability of choosing (red bus, blue bus, car) will be ($33.3\%, 33.3\%, 33.3\%$) respectively. But in reality, a more reasonable prediction would be $(25\%, 25\%, 50\%)$.

Problems with Estimates from Logit Model

- Own price elasticities $\eta_{jj} = \frac{\partial s_j}{\partial p_j} \frac{p_j}{s_j} = (-\alpha p_j (1 - s_j))$. If shares are close to 0 , own price elasticities are proportional to price
  - higher price goods have higher elasticities.

- Cross-price elasticities $\eta_{jk} = \frac{\partial s_j}{\partial p_k} \frac{p_k}{s_j} = \alpha p_k s_k$. This means the cross-price elasticities for a change in product k 's price is the same for all other products $j \neq k$, and is solely a function of prices and shares, but not the relative proximity of products in characteristic space.

- Note: if you run logit, and your results do not generate these results you have bad code. This is a helpful diagnostic for programming.

Expected Utility and Consumer Surplus

- Individual i is going to choose the best alternative, so expected utility from the best choice, V_i is given by:

$$V_i = \mathbb{E}\left[\max_j \delta_j + \epsilon_{ij}\right]$$

- For the multinomial logit, this has a close form:

$$V_i = \ln\left(\sum_{j=1}^{J} \exp\left(\delta_j\right)\right) + \gamma$$

where γ is the Euler's constant.

Expected Utility and Consumer Surplus

- We can then transform utility into dollars to get consumer surplus (need something in the utility function that is measured in dollars). For example, suppose:

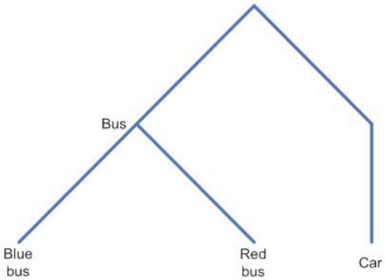
$$U_{ij} = \boldsymbol{X}_j \boldsymbol{\beta} + \gamma Z_i - \delta p_j$$

- The coefficient on price, δ gives the utility to dollar conversion:

$$\mathbb{E}\left(CS_i\right) = \frac{1}{\delta}\left[\ln\left(\sum_{j=1}^{J} \exp\left(\delta_j\right)\right) + \gamma\right]$$

Then we can calculate say the change in consumer surplus after a policy change.

Nested Logit Model



- Allows different substitution patterns within and outside nests.
- Two-level choices:
  - Choice of nest
  - Choice of alternatives within the nest
- $P(\text{Blue bus}) = P(\text{Blue bus}|\text{bus}) \times P(\text{bus})$

Nested Logit Model

- One way to solve the red bus/blue bus problem is to allow for the errors for the red bus to be correlated with the errors for the blue bus.
- Nested logit allows a nest-specific error:

$$U_{iRB} = \delta_{RB} + \theta_{iB} + \lambda v_{iRB}$$
$$U_{iBB} = \delta_{BB} + \theta_{iB} + \lambda v_{iBB}$$
$$U_{iC} = \delta_{C} + \theta_{iC} + \lambda v_{iC}$$

- where:
  - $\theta_{ik} + \lambda v_{ij}$ is distributed EVT1
  - The $\theta's$ and the $v's$ are independent
  - v_{ij} is distributed EVT1
  - The distribution of $\theta'_{ik}s$ is derived in Cardell (1997) (Econometric Theory)
  - Note: adding two extreme value errors does not give back an extreme value error.

Nested Logit Model

- Composite error term for car is independent from either the red bus error or the blue bus error. If we added a yellow bus, all errors in the bus nest would be independent conditional on choosing to take a bus. Then the probability of choosing car is derived as:

$$P_{iC} = \frac{\exp(\delta_C)}{\left[\exp\left(\frac{\delta_{RB}}{\rho}\right) + \exp\left(\frac{\delta_{BB}}{\rho}\right)\right]^{\rho} + \exp(\delta_C)}$$

- The probability of choosing red bus is:

$$\begin{aligned} P_{iRB} =& P_i(RB \mid B) P_i(B) \\ =& \frac{\left[\exp\left(\frac{\delta_{RB}}{\rho}\right) + \exp\left(\frac{\delta_{BB}}{\rho}\right)\right]^{\rho}}{\left[\exp\left(\frac{\delta_{RB}}{\rho}\right) + \exp\left(\frac{\delta_{BB}}{\rho}\right)\right]^{\rho} + \exp(\delta_C)} \\ & \times \frac{\exp\left(\frac{\delta_{RB}}{\rho}\right)}{\exp\left(\frac{\delta_{RB}}{\rho}\right) + \exp\left(\frac{\delta_{BB}}{\rho}\right)} \end{aligned}$$

Logit Inversion

- Logit is the easiest inversion to do, since

$$\ln\left[s_j\right] - \ln\left[s_0\right] = \delta_j = \sum_k \beta_k x_{kj} - \alpha p_j + \xi_j$$

$$\Rightarrow \quad \xi_j = \ln\left[s_j\right] - \ln\left[s_0\right] - \left(\sum_k \beta_k x_{kj} - \alpha p_j\right)$$

- Note that as far as estimation goes, we now are in a linear world where we can run things in the same way as we run OLS or IV or whatever. The precise routine to run will depend, as always, on what we think are the properties of ξ.

- Further simple examples in Berry 1994.

More on Estimation

- Regardless of the model we now have to choose the moment restriction we are going to use for estimation.
- This is where we can now properly deal with simultaneity in our model.
- Since consumers know ξ_j we should probably assume the firms do as well. Thus in standard pricing models you will have

$$p_j = p\left(x_j, \xi_j, x_{-j}, \xi_{-j}\right)$$

- Since p is a function of the unobservable, ξ, we should not use a moment restriction which interacts p and ξ. This is the standard endogeniety problem in demand estimation.
- It implies we need some instruments.
- There is nothing special about p in this context, if $E(\xi x) \neq 0$, then we need an instruments for x as well.

Berry (1994)

Berry (1994)

- Methodology for estimating differentiated-product discrete-choice demand models, using aggregate data. Fundamental problem is price endogeneity.

- Data structure: cross-section of market shares:

j	$\hat{s}_j$	p_j	X_1	X_2
A	25%	\$1.50	red	large
B	30%	\$2.00	blue	small
C	45%	\$2.50	green	large

- Total market size: M, J brands

Berry (1994)

- Derive market-level share expression from model of discrete-choice at the individual household level (i indexes household, j is brand):

$$U_{ij} = \underbrace{X_j\beta - \alpha p_j + \xi_j}_{\equiv \delta_j} + \epsilon_{ij}$$

  where we call δ_j the "mean utility" for brand j (the part of brand j 's utility which is common across all households i).

- Econometrician observes neither ξ_j or ϵ_{ij}, but household i observes both.
- $\xi_1, \ldots, \xi_J$ are commonly interpreted as "unobserved quality". All else equal, consumers more willing to pay for brands for which ξ_j is high.

- Price Endogeneity: price and market share highest for brand C. Perhaps due to unobserved quality.

Berry (1994)

- Make logit assumption that $\epsilon_{ij} \sim$ iid TIEV.
- Define choice indicators:

$$y_{ij} = \begin{cases} 1 & \text{if } i \text{ chooses brand } j \\ 0 & \text{otherwise} \end{cases}$$

- Given these assumptions, choice probabilities take MN logit form:

$$\Pr\left(y_{ij} = 1 \mid \beta, x_{j'}, \xi_{j'}, j' = 1, \ldots, J\right) = \frac{\exp\left(\delta_j\right)}{\sum_{j'=1}^{J} \exp\left(\delta_{j'}\right)}$$

- Aggregate market shares are:

$$\begin{aligned} s_j &= \frac{1}{M}\left[M \cdot \Pr\left(y_{ij} = 1 \mid \beta, x_{j'}, \xi_{j'}, j' = 1, \ldots, J\right)\right] = \frac{\exp\left(\delta_j\right)}{\sum_{j'=1}^{J} \exp\left(\delta_{j'}\right)} \\ &\equiv \underbrace{\tilde{s}_j}_{\text{predicted share function}}\left(\delta_{j'}\left(x_{j'}, \beta, \xi_{j'}\right), j' = 1, \ldots, J\right) \equiv \tilde{s}_j\left(\alpha, \beta, \xi_1, \ldots, \xi_J\right) \end{aligned}$$

Berry (1994)

- Data contains observed shares: denote by $\hat{s}_j, j = 1, \ldots, J$
- Model + parameters give you predicted shares:
  $\tilde{s}_j(\alpha, \beta, \xi_1, \ldots, \xi_J), j = 1, \ldots, J$
- Principle: Estimate parameters α, β by finding those values which "match" observed shares to predicted shares: find α, β so that $\tilde{s}_j(\alpha, \beta)$ is as close to $\hat{s}_j$ as possible, for $j = 1, \ldots, J$.
- How to do this? Most obvious thing could be nonlinear least squares, i.e.

$$\min_{\alpha,\beta} \sum_{j=1}^{J} \left(\hat{s}_j - \tilde{s}_j(\alpha, \beta, \xi_1, \ldots, \xi_J)\right)^2$$

  This problem doesn't fit into standard NLS framework, because the unobservables $\xi_1, \ldots, \xi_J$ do not enter linearly and additively in the predicted share $\tilde{s}_j(\cdots)$ functions.

Berry (1994)

- Berry (1994) suggests a clever IV-based estimation approach.
- Assume there exist instruments Z so that $E(\xi Z) = 0$
  Sample analog of this moment condition is

$$\frac{1}{J}\sum_{j=1}^{J}\xi_j Z_j = \frac{1}{J}\sum_{j=1}^{J}\left(\delta_j - X_j\beta + \alpha p_j\right) Z_j$$

  which converges (as $J \to \infty$) to zero at the true values α_0, β_0.
- Problem with estimating: we do not know δ_j. Berry suggest a two-step approach

First step: Inversion

- If we equate $\hat{s}_j$ to $\tilde{s}_j(\alpha, \beta, \xi_1, \ldots, \xi_J)$, for all j, we get a system of J nonlinear equations in the J unknowns $\delta_1, \ldots, \delta_J$:

$$\begin{aligned}
\hat{s}_1 \quad &= \tilde{s}_J\left(\delta_1\left(\alpha, \beta, \xi_1\right), \ldots, \delta_J\left(\alpha, \beta, \xi_J\right)\right) \\
\vdots \quad & \qquad\qquad\qquad \vdots \\
\hat{s}_J \quad &= \tilde{s}_J\left(\delta_1\left(\alpha, \beta, \xi_1\right), \ldots, \delta_J\left(\alpha, \beta, \xi_J\right)\right)
\end{aligned}$$

- You can "invert" this system of equations to solve for $\delta_1, \ldots, \delta_J$ as a function of the observed $\hat{s}_1, \ldots, \hat{s}_J$.
- Note: the outside good is $j = 0$. Since $1 = \sum_{j=0}^{J} \hat{s}_j$ by construction, you solve for $N + 1$ free variables $\rightarrow$ have to "normalize" $\delta_0 = 0$.
- Output from this step: $\hat{\delta}_j \equiv \delta_j\left(\hat{s}_1, \ldots, \hat{s}_J\right), j = 1, \ldots, J$ (J numbers)

Second step: IV estimation

- Going back to definition of δ_j 's:

$$\begin{aligned}
\delta_1 \quad &= X_1\beta - \alpha p_1 + \xi_1 \\
\vdots \quad & \qquad\quad \vdots \\
\delta_J \quad &= X_J\beta - \alpha p_J + \xi_J
\end{aligned}$$

- Now, using estimated $\hat{\delta}_j$'s, you can calculate sample moment condition:

$$\frac{1}{J}\sum_{j=1}^{J}\left(\hat{\delta}_j - X_j\beta + \alpha p_j\right) Z_j$$

  and solve for α, β which minimizes this expression.

- If δ_j is linear in X, p and ξ (as here), then linear IV methods are applicable here (i.e. 2SLS, etc.) Later, we will consider the substantially more complicated case in Berry, Levinsohn, and Pakes (1995).

Instruments

- As in usual demand case: cost shifters. But since we have cross-sectional (across brands) data, we require instruments to very across brands in a market.
- Take the example of automobiles. One natural cost shifter could be wages in Michigan.
- But here it doesn't work, because its the same across all car brands (specifically, if you ran 2SLS with wages in Michigan as the IV, first stage regression of price p_j on wage would yield the same predicted price for all brands).

Instruments

- BLP use instruments like: characteristics of cars of competing manufacturers. Intuition: oligopolistic competition makes firm j set p_j as a function of characteristics of cars produced by firms $i \neq j$ (e.g. GM's price for the Hum-Vee will depend on how closely substitutable a Jeep is with a Hum-Vee). However, characteristics of rival cars should not affect households' valuation of firm j 's car.

- In multiproduct context, similar argument for using characteristics of all other cars produced by same manufacturer as IV.

MNL case

- Predicted share $\tilde{s}_j(\delta_1, \ldots, \delta_J) = \frac{\exp(\delta_j)}{1+\sum_{j'=1}^{J} \exp(\delta_{j'})}$
- Taking logs, we get system of linear equations for δ_j 's:

$$\log \hat{s}_1 = \delta_1 - \log(\text{ denom })$$
$$\vdots$$
$$\log \hat{s}_J = \delta_J - \log(\text{ denom })$$
$$\log \hat{s}_0 = 0 - \log(\text{ denom })$$

which yield

$$\delta_j = \log \hat{s}_j - \log \hat{s}_0, \quad j = 1, \ldots, J$$

- So in second step, run IV regression of

$$(\log \hat{s}_j - \log \hat{s}_0) = X_j \beta - \alpha p_j + \xi_j$$

- In the simplest MNL logit, the estimation method can be described as logistic IV regression.

Coding Exercise: MATLAB/STATA

Empirical Application/Nested Logit: Reimers and Waldfogel (2021, AER)

Reimers and Waldfogel (2021)

- Digitization's new crowd-based pre-purchase information affects purchase behavior and welfare.
- Do professional reviews and crowd ratings causally impact demand, and how significant are these impacts?
- Books
    1. As experience goods, books benefit significantly from pre-purchase information.
    2. The number of professional reviews, especially from highly visible outlets, is small and observable.
    3. Utilizing high-frequency data on book demand from Amazon (representing 45% of the US physical book market) helps identify causal relationships.

Empirical Strategies

- Cross-sectional variation: US, Canada, and the UK.
- Reviews and star ratings are endogenous; raters and reviewers decide when to give feedback. Appealing books sell more and get positive feedback.
- High-frequency data from multiple platforms:
  - Professional reviews cause a discontinuous jump in attention.
  - Star ratings (Chevalier and Mayzlin, 2006) using book fixed effects and cross-platform comparisons.

Professional Reviews

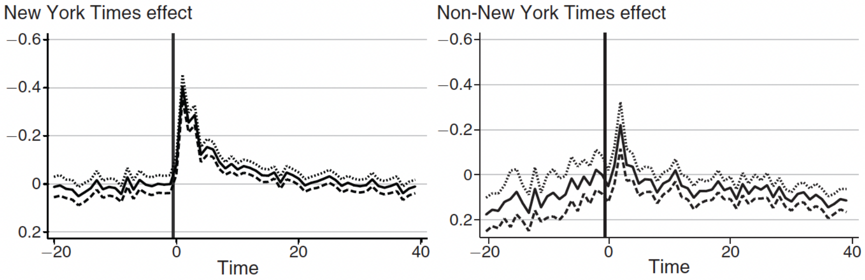


FIGURE 2. DAILY EFFECTS OF PROFESSIONAL REVIEWS ON SALES RANKS

Review Effects: Cross-Sectional Relationship

TABLE 3—CAUSAL QUANTITY EFFECTS

	Effect	SE
Price elasticity	−0.166	0.022
Amazon stars elasticity (25th percentile)	0.392	0.054
Amazon stars elasticity (50th percentile)	0.621	0.084
Amazon stars elasticity (75th percentile)	0.839	0.114
Amazon stars elasticity (mean)	0.616	0.084
NYT 0–5, not recommended	0.438	0.061
NYT 6–10, not recommended	0.165	0.031
NYT 11–20, not recommended	0.082	0.023
NYT 0–5, recommended	0.584	0.084
NYT 6–10, recommended	0.370	0.057
NYT 11–20, recommended	0.181	0.035
Other 0–10	0.114	0.037
Other 11–20	−0.021	0.034
Percent effect of review on annual q		
Other only	0.710	0.44
NYT (not recommended) only	2.189	0.43
NYT (recommended) only	4.301	0.92
NYT not recommended and other	3.833	0.94
NYT recommended and other	6.189	1.44
Average	2.641	0.52

Reimers and Waldfogel (2021): Nested Logit Model

- Consumers choose whether to purchase books (inside) or other goods (outside):

$$u_{ij} = \delta_j + \epsilon_{ij}.$$

- Let δ_j be the mean utility of product j in the status quo, when reviews and ratings are present.

$$\delta_j = \delta_j^0 + \alpha P_j + \gamma_j R_j + \psi_j.$$

- Each product's share

$$s_j = \frac{e^{\delta_j/(1-\sigma)}}{\sum_{k \in J} e^{\delta_k/(1-\sigma)}} \frac{D^{(1-\sigma)}}{1 + D^{(1-\sigma)}}.$$

  where $D = \sum_{k \in J} e^{\delta_k/(1-\sigma)}$.

- σ reflects product substitutability (1 implies full substitution).

- Identification comes from the relationship between product availability and book-buying population share increase.

Reimers and Waldfogel (2021): Nested Logit Model

- Berry (1994): $\delta_j = \ln s_j - \sigma \ln(s_{j|g}) - \ln(s_0)$.
- Product shares: $s_j = \frac{q_j}{M}$, $s_{j|g} = \frac{q_j}{Q}$, and $s_0 = 1 - \frac{Q}{M}$.
- $Q = \sum_{k \in J} q_k$: Weekly data on total physical book sales.
- Market size M: US individuals make a monthly choice (0.25×327 million for weekly observations).
- Estimation: IV regression.
  - Regress $\ln s_j - \ln s_0$ on inside share, price, and reviews. **See problem set!**

Welfare Impacts of Pre-Purchase Information

TABLE 4—WELFARE IMPACTS OF PROFESSIONAL REVIEWS AND AMAZON STAR RATINGS

	Stars	Reviews	Ratio
ΔRevenue (net)	27.51	19.98	
	(11.79)	(2.10)	
$R > \hat{R}$	92.56		
	(17.00)		
$R < \hat{R}$	−65.05		
	(6.29)		
ΔCS (baseline)	35.83	3.18	11.27
	(6.98)	(0.41)	(2.28)
ΔCS (50 categories)	35.83	3.22	11.13
	(7.13)	(0.42)	(2.23)
ΔCS (reviewed books)	1.68	3.18	0.53
	(0.30)	(0.41)	(0.13)
ΔCS ($\sigma = 0$)	36.58	3.27	11.18
	(7.43)	(0.42)	(2.34)
ΔCS ($\sigma = 0.95$)	34.67	3.03	11.42
	(6.29)	(0.39)	(2.20)
ΔCS (Marshallian: unconstrained)	37.66	4.76	7.92
	(8.01)	(0.60)	(1.80)
ΔCS (Marshallian: $\Delta q = 0$)	40.52	3.89	10.41
	(7.83)	(0.48)	(2.24)

Today's Learning Goal

- Demand Estimation: Introduction, theory
- Discrete Choice Demand Model: Berry (1994)
- Estimation, Endogeneity
  - Empirical Example: Reimers and Waldfogel (2021)