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

Price "endogeneity": price and market share highest for brand C. Perhaps due to unobserved quality. (This problem occurred in the Trajtenberg (1989) analysis of demand for CAT scanners.)

■■■

In explanation, we start out with simplest setup, with most restrictive assumptions, and later describe more complicated extensions.

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

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

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: these are both "structural errors".

$\xi_1, \ldots, \xi_J$ are commonly interpreted as "unobserved quality". All else equal, consumers more willing to pay for brands for which ξ_j is high.

Important: ξ_j, as unobserved quality, is correlated with price p_j (and also potentially with characteristics X_j). It is the source of the endogeneity problem in this demand model.

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:

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

Aggregate market shares are:

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

$\tilde{s}(\cdots)$ is the "predicted share" function, for fixed values of the parameters α and β, and the unobservables $\xi_1, \ldots, \xi_J$.

■■■

- 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.

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

  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) 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 $\delta_1, \ldots, \delta_J$:

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

- 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 $\to$ have to "normalize" $\delta_0 = 0$.
- Output from this step: $\hat{\delta}_j \equiv \delta_j(\hat{s}_1, \ldots, \hat{s}_J),\ j = 1, \ldots, J$ (J numbers)

Second step: IV estimation

- Going back to definition of δ_j's:

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

- 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).

■■■

What are appropriate instruments (Berry, p. 249)?

- 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).
- 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.

■■■

One simple case of inversion step:

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:

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

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

$$(\log \hat{s}_j - \log \hat{s}_0) = X_j\beta - \alpha p_j + \xi_j. \tag{2}$$

Eq. (2) is called a "logistic regression" by bio-statisticians, who use this logistic transformation to model "grouped" data. So in the simplest MNL logit, the estimation method can be described as "logistic IV regression".

See Berry paper for additional examples (nested logit, vertical differentiation).

■■■

SUPPLY SIDE

- After we estimate demand side, can we estimate supply side also? In particular, we want to derive estimates of the firms' markups.
- From our demand estimation, we have estimated the demand function for brand j, which we denote as follows:

$$D^j\left(\underbrace{X_1, \ldots, X_J}_{\equiv \vec{X}}; \underbrace{p_1, \ldots, p_J}_{\equiv \vec{p}}; \underbrace{\xi_1, \ldots, \xi_J}_{\equiv \vec{\xi}}\right)$$

- Specify costs of producing brand j:

$$C^j\left(q_j, w_j, \omega_j\right)$$

where q_j is total production of brand j, w_j are observed cost components associated with brand j (e.g. could be characteristics of brand j), ω_j are unobserved cost components (another structural error)

Importantly, we assume that there are no (dis-)economies of scope, so that production costs are simply additive across car models, for a muultiproduct firm.

- Then profits for brand j are:

$$\Pi_j = D^j\left(\vec{X}, \vec{p}, \vec{\xi}\right) p_j - C^j\left(D^j\left(\vec{X}, \vec{p}, \vec{\xi}\right), w_j, \omega_j\right)$$

- For multiproduct firm: assume that firm K produces all brands $j \in \mathcal{K}$. Then its profits are

$$\tilde{\Pi}_K = \sum_{j \in \mathcal{K}} \Pi_j = \sum_{j \in \mathcal{K}} \left[D^j\left(\vec{X}, \vec{p}, \vec{\xi}\right) p_j - C^j\left(D^j\left(\vec{X}, \vec{p}, \vec{\xi}\right), w_j, \omega_j\right)\right].$$

- In order to proceed, we need to assume a particular model of oligopolistic competition. The most common assumption is *Bertrand (price) competition.* (Note that because firms produce differentiated products, Bertrand solution does not result in marginal cost pricing.)

- Under price competition, equilibrium prices are characterized by J equations (which are the J best-response functions for the J brands):

$$\begin{aligned} &\frac{\partial \tilde{\Pi}_K}{\partial p_j} = 0, \ \forall j \in \mathcal{K}, \ \forall K \\ \Leftrightarrow & D^j + \sum_{j' \in K} \frac{\partial D^{j'}}{\partial p_j}\left(p_{j'} - C_1^{j'}|_{q_{j'}=D^{j'}}\right) = 0 \end{aligned}$$

where C_1^j denotes the derivative of C^j with respect to first argument.

- If we make the further assumption that marginal costs are constant, and linear in cost components:

$$C_1^j = c^j \equiv w_j \gamma + \omega_j$$

(where γ are parameters in the marginal cost function) then the best-response equations become

$$D^j + \sum_{j' \in K} \frac{\partial D^{j'}}{\partial p_j} \left(p_{j'} - c^j \right) = 0. \tag{3}$$

Note that because we have already estimated the demand side, the demand functions $D^j,\ j = 1, \ldots, J$ and full set of demand slopes $\frac{\partial D^{j'}}{\partial p_j},\ \forall j, j' = 1, \ldots, J$ can be calculated.

- This suggest a two-step approach to estimating cost parameters γ (analogous to two-step demand estimation):

Inversion: the system of best-response equation (3) is J equation in the J unknowns $c^j,\ j = 1, \ldots, J$.

Note that once you solve for the marginal costs $c^1, \ldots, c^J$, you can already compute the markups $\left(p_j - c^j\right)/p_j$ for each brand j. This is the oligopolistic equivalent of using the "inverse-elasticity" condition to calculate a monopolist's market power.

IV estimation: Estimate the regression $c^j = w_j \gamma + \omega_j$. Allow for endogeneity of observed cost components w_j by using demand shifters as instruments. Assume you have instruments U_j such that $E(\omega U) = 0$, then find γ to minimize sample analogue $\frac{1}{J} \sum_{j=1}^{J} \left(c_j - w_j \gamma\right) U_j$.

- Naturally, you can also estimate the demand and supply side jointly: estimate (α, β, γ) all at once by jointly imposing the moment conditions $E(\xi Z) = 0$ and $E(\omega U) = 0$.

This is not entirely straightforward, since the "dependent variables" on the supply side, the marginal costs $c^1, \ldots, c^J$, are themselves function of the demand parameters α, β. So in order to estimate jointly, we have to employ a more complicated "nested" estimation procedure which we will describe below.

■■■

RANDOM COEFFICIENTS LOGIT MODEL

Return to the demand side. Next we discuss the random coefficients logit model, which is the main topic of Berry, Levinsohn, and Pakes (1995).

- Assume that utility function is:

  $$u_{ij} = X_j \beta_i - \alpha_i p_j + \xi_j + \epsilon_{ij}$$

  The difference here is that the slope coefficients (α_i, β_i) are allowed to vary across households i.

- We assume that, across the population of households, the slope coefficients (α_i, β_i) are i.i.d. random variables. The most common assumption is that these random variables are jointly normally distributed:

  $$(\alpha_i, \beta_i)' \sim N\left((\bar{\alpha}, \bar{\beta})', \Sigma \right).$$

  For this reason, α_i and β_i are called "random coefficients".

  Hence, $\bar{\alpha}$, $\bar{\beta}$, and Σ are additional parameters to be estimated.

- Given these assumptions, the mean utility δ_j is $X_j \bar{\beta} - \bar{\alpha} p_j + \xi_j$, and

  $$u_{ij} = \delta_j + \epsilon_{ij} + (\beta_i - \bar{\beta}) X_j - (\alpha_i - \bar{\alpha}) p_j$$

  so that, even if the ϵ_{ij}'s are still i.i.d. TIEV, the composite error is not. Here, the simple MNL inversion method will not work.

- The estimation methodology for this case is developed in Berry, Levinsohn, and Pakes (1995).

- First note: for a given α_i, β_i, the choice probabilities for household i take MNL form:

  $$Pr(i,j) = \frac{\exp\left(X_j \beta_i - \alpha_i p_j + \xi_j\right)}{1 + \sum_{j'=1}^{J} \exp\left(X_{j'} \beta_i - \alpha_i p_{j'} + \xi_{j'}\right)}.$$

- In the whole population, the aggregate market share is just

$$\begin{aligned}\tilde{s}_j &= \int\int Pr(i,j,)dG(\alpha_i,\beta_i) \\ &= \int\int \frac{\exp\left(X_j\beta_i - \alpha_i p_j + \xi_j\right)}{1+\sum_{j'=1}^{J}\exp\left(X_{j'}\beta_i - \alpha_i p_{j'} + \xi_{j'}\right)} dG(\alpha_i,\beta_i) \\ &= \int\int \frac{\exp\left(\delta_j + (\beta_i - \bar{\beta})X_j - (\alpha_i - \bar{\alpha})p_j\right)}{1+\sum_{j'=1}^{J}\exp\left(\delta_{j'} + (\beta_i - \bar{\beta})X_{j'} - (\alpha_i - \bar{\alpha})p_{j'}\right)} dG(\alpha_i,\beta_i) \\ &\equiv \tilde{s}_j^{RC}\left(\delta_1,\ldots,\delta_J;\bar{\alpha},\bar{\beta},\Sigma\right)\end{aligned} \tag{4}$$

  that is, roughly speaking, the weighted sum (where the weights are given by the probability distribution of (α,β)) of $Pr(i,j)$ across all households.

  The last equation in the display above makes explicit that the predicted market share is not only a function of the mean utilities $\delta_1,\ldots,\delta_J$ (as before), but also functions of the parameters $\bar{\alpha},\bar{\beta},\Sigma$. Hence, the inversion step described before will not work, because the J equations matching observed to predicted shares have more than J unknowns (i.e. $\delta_1,\ldots,\delta_J;\bar{\alpha},\bar{\beta},\Sigma$).

  Moreover, the expression in Eq. (4) is difficult to compute, because it is a multidimensional integral. BLP propose *simulation methods* to compute this integral. We will discuss simulation methods later. For the rest of these notes, we assume that we can compute $\tilde{s}_j^{RC}$ for every set of parameters $\bar{\alpha},\bar{\beta},\Sigma$.

■■■

ESTIMATING THE RC DEMAND MODEL

We would like to proceed, as before, to estimate via GMM, exploiting the population moment restriction $E(\xi Z)=0$. We would like to estimate the parameters by minimizing the sample analogue of the moment condition:

$$\min_{\bar{\beta},\bar{\alpha},\Sigma}\left|\left|\frac{1}{J}\sum_{j=1}^{J}\left(\delta_j - X_j\bar{\beta} + \bar{\alpha}p_j\right)Z_j\right|\right| \equiv Q\left(\bar{\alpha},\bar{\beta},\Sigma\right).$$

But problem is that we cannot perform inversion step as before, so that we cannot derive $\delta_1,\ldots,\delta_J$.

So BLP propose a "nested" estimation algorithm, with an "inner loop" nested within an "outer loop"

- In the **outer loop**, we iterate over different values of the parameters $\theta \equiv (\bar{\alpha}, \bar{\beta}, \Sigma)$. Let $\hat{\theta}$ be the current values of the parameters being considered.
- In the **inner loop**, for the given parameter values $\hat{\theta}$, we wish to evaluate the objective function $Q(\hat{\theta})$. In order to do this we must:
  1. At current $\hat{\theta}$, we solve for the mean utilities $\delta_1(\hat{\theta}), \ldots, \delta_J(\hat{\theta})$ to solve the system of equations

     $$\begin{aligned} \hat{s}_1 &= \tilde{s}_1^{RC}\left(\delta_1, \ldots, \delta_J; \hat{\theta}\right) \\ \vdots & \quad \vdots \\ \hat{s}_J &= \tilde{s}_J^{RC}\left(\delta_1, \ldots, \delta_J; \hat{\theta}\right). \end{aligned}$$

     Note that, since we take the parameters $\hat{\theta}$ as given, this system is J equations in the J unknowns $\delta_1(\hat{\theta}), \ldots, \delta_J(\hat{\theta})$.
  2. For the resulting $\delta_1(\hat{\theta}), \ldots, \delta_J(\hat{\theta})$, calculate

     $$Q(\hat{\theta}) \equiv \left\| \frac{1}{J} \sum_{j=1}^{J} \left(\delta_j(\hat{\theta}) - X_j \hat{\bar{\beta}} + \hat{\bar{\alpha}} p_j\right) Z_j \right\|. \tag{5}$$
- Then we return to the outer loop, which searches until it finds parameter values $\hat{\theta}$ which minimize Eq. (5).
- Essentially, the original inversion step is now nested inside of the estimation routine.

■■■

If we wanted to add a supply side to the RC model, then the objective function is

$$Q(\theta, \gamma) = G_J(\theta, \gamma)' W_J G_J(\theta, \gamma)$$

where G_J is the $(M+N)$-dimensional vector of stacked sample moment conditions:

$$G_J(\theta,\gamma) \equiv \begin{bmatrix} \frac{1}{J}\sum_{j=1}^{J}\left(\delta_j(\theta) - X_j\bar{\beta} + \bar{\alpha}p_j\right) z_{1j} \\ \vdots \\ \frac{1}{J}\sum_{j=1}^{J}\left(\delta_j(\theta) - X_j\bar{\beta} + \bar{\alpha}p_j\right) z_{Mj} \\ \frac{1}{J}\sum_{j=1}^{J}\left(c_j(\theta) - w_j\gamma\right) u_{1j} \\ \vdots \\ \frac{1}{J}\sum_{j=1}^{J}\left(c_j(\theta) - w_j\gamma\right) u_{Nj} \end{bmatrix}$$

where M is the number of demand side IV's, and N the number of supply-side IV's. W_J is a $(M+N)$-dimensional weighting matrix.

The only change in the estimation routine described in the previous section is that the inner loop is more complicated:

In the **inner loop**, for the given parameter values $\hat{\theta}$ and $\hat{\gamma}$, we wish to evaluate the objective function $Q(\hat{\theta},\hat{\gamma})$. In order to do this we must:

1. At current $\hat{\theta}$, we solve for the mean utilities $\delta_1(\hat{\theta}),\ldots,\delta_J(\hat{\theta})$ as previously.
2. For the resulting $\delta_1(\hat{\theta}),\ldots,\delta_J(\hat{\theta})$, calculate

$$\vec{\tilde{s}}_j^{RC}(\hat{\theta}) \equiv \left(\tilde{s}_1^{RC}(\delta(\hat{\theta})),\ldots,\tilde{s}_J^{RC}(\delta(\hat{\theta}))\right)'$$

and also the partial derivative matrix

$$\mathbf{D}(\hat{\theta}) = \begin{pmatrix} \frac{\partial \tilde{s}_1^{RC}(\delta(\hat{\theta}))}{\partial p_1} & \frac{\partial \tilde{s}_1^{RC}(\delta(\hat{\theta}))}{\partial p_2} & \cdots & \frac{\partial \tilde{s}_1^{RC}(\delta(\hat{\theta}))}{\partial p_J} \\ \frac{\partial \tilde{s}_2^{RC}(\delta(\hat{\theta}))}{\partial p_1} & \frac{\partial \tilde{s}_2^{RC}(\delta(\hat{\theta}))}{\partial p_2} & \cdots & \frac{\partial \tilde{s}_2^{RC}(\delta(\hat{\theta}))}{\partial p_J} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \tilde{s}_J^{RC}(\delta(\hat{\theta}))}{\partial p_1} & \frac{\partial \tilde{s}_J^{RC}(\delta(\hat{\theta}))}{\partial p_2} & \cdots & \frac{\partial \tilde{s}_J^{RC}(\delta(\hat{\theta}))}{\partial p_J} \end{pmatrix}$$

For MN logit case, these derivatives are:

$$\frac{\partial s_j}{\partial p_k} = \begin{cases} -\alpha s_j(1-s_j) & \text{for } j=k \\ -\alpha s_j s_k & \text{for } j \neq k. \end{cases}$$

3. Use the supply-side best response equations to solve for $c_1(\hat{\theta}), \ldots, c_J(\hat{\theta})$:

$$\vec{\tilde{s}}_j^{RC}(\hat{\theta}) + \mathbf{D}(\hat{\theta}) * \begin{pmatrix} p_1 - c^1 \\ \vdots \\ p_J - c^J \end{pmatrix} = 0.$$

4. So now, you can compute $G(\bar{\theta}, \bar{\gamma})$.

■■■

SIMULATING THE INTEGRAL IN EQ. (4)

The principle of simulation: approximate an expectation as a sample average. Validity is ensured by law of large numbers.

In the case of Eq. (4), note that the integral there is an expectation:

$$\mathcal{E}\left(\bar{\alpha}, \bar{\beta}, \Sigma\right) \equiv E_G \left[\frac{\exp\left(\delta_j + (\beta_i - \bar{\beta})X_j - (\alpha_i - \bar{\alpha})p_j\right)}{1 + \sum_{j'=1}^{J} \exp\left(\delta_{j'} + (\beta_i - \bar{\beta})X_{j'} - (\alpha_i - \bar{\alpha})p_{j'}\right)} | \bar{\alpha}, \bar{\beta}, \Sigma \right]$$

where the random variables are α_i and β_i, which we assume to be drawn from the multivariate normal distribution $N\left((\bar{\alpha}, \bar{\beta})', \Sigma\right)$.

For $s = 1, \ldots, S$ simulation draws:

1. Draw u_1^s, u_2^s independently from N(0,1).
2. For the current parameter estimates $\hat{\bar{\alpha}}$, $\hat{\bar{\beta}}$, $\hat{\Sigma}$, transform $(u_1^s,\ u_2^s)$ into a draw from $N\left((\hat{\bar{\alpha}}, \hat{\bar{\beta}})', \hat{\Sigma}\right)$ using the transformation

$$\begin{pmatrix} \alpha^s \\ \beta^s \end{pmatrix} = \begin{pmatrix} \hat{\bar{\alpha}} \\ \hat{\bar{\beta}} \end{pmatrix} + \hat{\Sigma}^{1/2} \begin{pmatrix} u_1^s \\ u_2^s \end{pmatrix}$$

   where $\hat{\Sigma}^{1/2}$ is shorthand for the "Cholesky factorization" of the matrix $\hat{\Sigma}$ (roughly, the matrix equivalent of square root; the Cholesky factorization of a square symmetric matrix Γ is the matrix $\mathbf{G}$ such that $\mathbf{G}'\mathbf{G} = \Gamma$).

Then approximate the integral by the sample average (over all the simulation draws)

$$\mathcal{E}\left(\hat{\hat{\alpha}}, \hat{\hat{\beta}}, \hat{\Sigma}\right) \approx \frac{1}{S} \sum_{s=1}^{S} \frac{\exp\left(\delta_j + (\beta^s - \hat{\hat{\beta}})X_j - (\alpha^s - \hat{\hat{\alpha}})p_j\right)}{1 + \sum_{j'=1}^{J} \exp\left(\delta_{j'} + (\beta^s - \hat{\hat{\beta}})X_{j'} - (\alpha^s - \hat{\hat{\alpha}})p_{j'}\right)}.$$

For given $\hat{\hat{\alpha}}, \hat{\hat{\beta}}, \hat{\Sigma}$, the law of large numbers ensure that this approximation is arbitrarily accurate as $S \to \infty$.

Important: note that the *same* uniform draws u_1^s, u_2^s, for $s = 1, \ldots, S$, are used to compute the simulated α^s, β^s no matter which the current parameter values $\hat{\hat{\alpha}}, \hat{\hat{\beta}}, \hat{\Sigma}$ are. This is important for the nice stochastic convergence properties of the estimators $\hat{\hat{\alpha}}, \hat{\hat{\beta}}, \hat{\Sigma}$, as $J \to \infty$.

References

Berry, S. (1994): "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841–890.

Trajtenberg, M. (1989): "The Welfare Analysis of Product Innovations, with an Application to Computed Tomography Scanners," *Journal of Political Economy*, 97(2), 444–479.