Papers to be covered:

- Berry (1994)
- Berry, Levinsohn, and Pakes (1995)

1 Why demand analysis/estimation?

- Many papers in empirical IO focuses on estimation of demand models
- Why?
- IO theory mostly concerned about supply-side (firm-side). However, important determinants of firm behavior are **costs**, which are usually unobserved.
- Fundamental question in empirical IO: how much market power do firms have?

  Market power measured by markup: $\frac{p-mc}{p}$.

  Problem: mc not observed!

  much of the "new empirical industrial organization" (NEIO) motivated by this data problem.

  For example, you observe high prices in an industry. Is this due to market power, or due to high costs? Cannot answer this question directly, because we don't observe costs.
- NEIO takes *indirect approach*: obtain estimate of firms' markups by estimating firms' demand functions.
- Intuition is most easily seen in monopoly example:
  - $\max_p pq(p) - C(q(p))$, where $q(p)$ is demand curve.
  - FOC: $q(p) + pq'(p) = C'(q(p))q'(p)$

– At optimal price p^*, **Inverse Elasticity Property** holds:

$$(p^* - MC(q(p^*))) = -\frac{q(p^*)}{q'(p^*)}$$

or

$$\frac{p^* - mc\,(q(p^*))}{p^*} = -\frac{1}{\epsilon(p^*)},$$

where $\epsilon(p^*)$ is $q'(p^*)\dfrac{p^*}{q(p^*)}$, the price elasticity of demand.

– Hence, if we can estimate $\epsilon(p^*)$, we can infer what the markup $\frac{p^* - mc(q(p^*))}{p^*}$ is, even when we don't observe the marginal cost $mc\,(q(p^*))$.

– Similar exercise holds for oligopoly case (as we will show below).

– Caveat: validity of exercise depends crucially on using the right supply-side model (in this case: monopoly without entry possibility).

If costs were observed: markup could be estimated directly, and we could test for vaalidity of monopoly pricing model (ie. test whether markup= $\frac{-1}{\epsilon}$).

- Start by reviewing some standard approaches to demand estimation, and motivate why recent literature in empirical IO has developed new methodologies.

2 Review: demand estimation

- Linear demand-supply model:

$$\begin{aligned} \text{Demand:} \quad & q_t^d = \gamma_1 p_t + \mathbf{x'_{t1}}\beta_{\mathbf{1}} + u_{t1} \\ \text{Supply:} \quad & p_t = \gamma_2 q_t^s + \mathbf{x'_{t2}}\beta_{\mathbf{2}} + u_{t2} \\ \text{Equilibrium:} \quad & q_t^d = q_t^s \end{aligned}$$

- Demand function summarizes consumer preferences; supply function summarizes firms' cost structure

- Focus on estimating demand function:

$$\text{Demand:} q_t = \gamma_1 p_t + \mathbf{x'_{t1}} \boldsymbol{\beta_1} + u_{t1}$$

- If u_1 correlated with u_2, then p_t is endogenous in demand function: cannot estimate using OLS. Important problem.

- Instrumental variable (IV) methods: assume there are instruments Z's so that $E(u_1 \cdot \mathbf{Z}) = 0$.

- Properties of appropriate instrument Z for endogenous variable p:

  1. Uncorrelated with error term in demand equation: $E(u_1 Z) = 0$. **Exclusion** restriction. (order condition)
  2. Correlated with endogenous variable: $E(Zp) \neq 0$. (rank condition)

- The x's are exogenous variables which can serve as instruments:

  1. x_{t2} are *cost shifters*; affect production costs. Correlated with p_t but not with u_{t1}: use as instruments in demand function.
  2. x_{t1} are *demand shifters*; affect willingness-to-pay, but not a firm's production costs. Correlated with q_t but not with u_{2t}: use as instruments in supply function.

  The demand models used in empirical IO different in flavor from "traditional" demand specifications. Start by briefly showing traditional approach, then motivating why that approach doesn't work for many of the markets that we are interested in.

2.1 "Traditional" approach to demand estimation

- Consider modeling demand for two goods 1,2 (Example: food and clothing).

- Data on prices and quantities of these two goods across consumers, across markets, or over time.

- Consumer demand determined by utility maximization problem:

$$\max_{x_1,x_2} U(x_1,x_2) \text{ s.t. } p_1x_2 + p_2x_2 = M$$

- This yields demand functions $x_1^*(p_1,p_2,M)$, $x_2^*(p_1,p_2,M)$.
- Equivalently, start out with *indirect utility function*

$$V(p_1,p_2,M) = U(x_1^*(p_1,p_2,M), x_2^*(p_1,p_2,M))$$

- Demand functions derived via *Roy's Identity*:

$$x_1^*(p_1,p_2,M) = -\frac{\partial V}{\partial p_1} / \frac{\partial V}{\partial M}$$
$$x_2^*(p_1,p_2,M) = -\frac{\partial V}{\partial p_2} / \frac{\partial V}{\partial M}$$

  This approach is often more convenient empirically.

- Take a particular functional form for V (for example, Translog):

$$\log V_k(p_1,p_2,M) = \alpha_0 + \sum_{i=1,2} \alpha_i \log\left(\frac{p_i}{M_k}\right) +$$
$$\frac{1}{2}\sum_{i=1,2}\sum_{j=1,2} \beta_{ij} \log\left(\frac{p_i}{M_k}\right) \log\left(\frac{p_j}{M_k}\right) + \eta \log\left(\frac{p_i}{M_k}\right) x_k$$

  α's and β's are parameters to be estimated.

- Log-version of Roy's Identity yields the *expenditure shares* (i.e., $w_{ik} = \frac{p_i x_{ik}}{M_k}$):

$$w_{ik}(p_1,p_2,M_k) = \frac{\partial V_k}{\partial \log\left(\frac{p_i}{M_k}\right)} / \sum_{i=1,2} \frac{\partial V_k}{\partial \log\left(\frac{p_i}{M_k}\right)}$$

  - Only estimate equation for first good (since shares sum to 1)
  - To estimate: add error term to share equation. (Literature on interpretation of this error.)

    If prices endogenous: use cost shifters (which could include weather, transportation costs, fabric costs) which vary across time, across geographic areas.

- This "standard" approach not convenient for many markets which we are interested in: automobile, airlines, cereals, toothpaste, etc. These markets characterized by:
  - Many alternatives: too many parameters to estimate using traditional approach
  - At individual level, usually only choose one of the available options (discrete choices). Consumer demand function not characterized by FOC of utility maximization problem.

  These problems have been addressed by

  - Modeling demand for a product as demand for the characteristics of that product: **Hedonic** analysis (Rosen (1974), Bajari and Benkard (2005)).

    This can be difficult in practice when there are many characteristics, and characteristics not continuous.
  - Discrete choice: assume each consumer can choose at most *one* of the available alternatives on each purchase occasion. This is the approach taken in the moden empirical IO literature.

3 Discrete-choice approach to modeling demand

- There are N alternatives in market. Each purchase occasion, each consumer i divides her income y_i on (at most) one of the alternatives, and on an "outside good":

$$\max_{n,z} U_i(x_n, z) \text{ s.t. } p_n + p_z z = y_i$$

  where

  - x_n are chars of brand n, and p_n the price
  - z is quantity of outside good, and p_z its price
  - outside good ($n = 0$) denotes the non-purchase of any alternative (that is, spending entire income on other types of goods).

- Substitute in the budget constraint ($z = \frac{y-p_n}{p_z}$) to derive *conditional indirect utility functions* for each brand:

$$U^*_{in}(p_n, p_z, y) = U_i(x_n, \frac{y_i - p_n}{p_z}).$$

If outside good is bought:

$$U^*_{i0}(p_z, y) = U_i(0, \frac{y_i - p_n}{p_z}).$$

- Consumer chooses the brand yielding the highest cond. indirect utility:

$$\max_n U^*_{in}(p_n, p_z, y_i)$$

- U^*_{in} usually specified as sum of deterministic and stochastic part:

$$U^*_{in}(p_n, p_z, y_i) = V_{in}(p_n, p_z, y_i) + \epsilon_{in}$$

ϵ_{in} observed by agent i, not by econometrician (this is a **structural error**). From agent's point of view, utility and choice are *deterministic*.

- Distributional assumptions on ϵ_{in}, $n = 0 \dots N$ determine the form of consumer i's choice probabilities. Probability that consumer i buys brand n is:

$$D_{in}(p_1 \dots p_N, p_z, y_i) = \text{Prob}\left\{\epsilon_{i0}, \dots, \epsilon_{iN} : U^*_{in} > U^*_{ij} \text{ for } j \neq n\right\}$$

- If consumers are identical, and $\{\epsilon_{i0}, \dots, \epsilon_{iN}\}$ is *iid* across agents i (and there are a very large number of agents), then $D_{in}(p_1 \dots p_N, p_z, y)$ is also the *aggregate market share*.

- Common assumptions:

  – $(\epsilon_{i0}, \dots, \epsilon_{iN})$ distributed multivariate normal: **multinomial probit**. Choice probabilities do not have closed form, but they can be simulated (Keane (1994), McFadden (1989)). (cf. GHK simulator, described below.)

    But difficult when there are large number of choices, because number of parameters in the variance matrix Σ also grows very large.

– $(\epsilon_{in},\ n = 0, \ldots, N)$ distributed *i.i.d.* type I extreme value across i:

$$F(\epsilon) = \exp\left[-\exp\left(-\frac{\epsilon - \eta}{\mu}\right)\right]$$

with the location parameter $\eta = 0.577$ (Euler's constant), and the scale parameter (usually) $\mu = 1$.

Leads to multinomial logit choice probabilities:

$$D_{in}(\cdots) = \frac{\exp(V_{in})}{\sum_{n'=1,\ldots,N} \exp(V_{in'})}$$

Normalize $V_0 = 0$. (Because $\sum_{n=1}^{N} D_{in} = 1$ by construction.)

Convenient, tractable form for choice probabilities. Logit model is basis for many demand papers in empirical IO.

Problems with multinomial logit

- Restrictiveness of multinomial logit: Odds ratio between any two brands n, n' doesn't depend on number of alternatives available

$$\frac{D_n}{D_{n'}} = \frac{\exp(V_n)}{\exp(V_{n'})}$$

Example: Red bus/blue bus problem:

– Assume that city has two transportation schemes: walk, and red bus, with shares 50%, 50%. So odds ratio of walk/RB= 1.

– Now consider introduction of third option: train. IIA implies that odds ratio between walk/red bus is still 1. Unrealistic: if train substitutes more with bus than walking, then new shares could be walk 45%, RB 30%, train 25%, then odds ratio walk/RB=1.5.

– What if third option were blue bus? IIA implies that odds ratio between walk/red bus would still be 1. Unrealistic: BB is perfect substitute for RB, so that new shares are walk 50%, RB 25%, bb 25%, and odds ratio walk/RB=2!

  - So this is especially troubling if you want to use logit model to predict penetration of new products.

Implication: invariant to introduction (or elimination) of some alternatives. **Independence of Irrelevant Alternatives**

- If interpret D_{in} as market share, IIA implies restrictive substitution patterns:

$$\varepsilon_{a,c} = \varepsilon_{b,c}, \text{ for all brands } a, b \neq c.$$

If $V_n = \beta_n + \alpha(y - p_n)$, then $\varepsilon_{a,c} = -\alpha p_c D_c$, for all $c \neq a$: Price decrease in brand a attracts proportionate chunk of demand from all other brands. Unrealistic!

- Changes to logit framework to overcome IIA:
  - Nested logit: assume particular correlation structure among $(\epsilon_{i0}, \ldots, \epsilon_{iN})$. Within-nest brands are "closer substitutes" than across-nest brands (cf. Maddala (1983, chap. 2)). See Goldberg (1995) for an application of this to automobile demand.

    (Diagram of demand structure from Goldberg paper)
  - Random coefficients: assume logit model, but for agent i:

$$U^*_{in} = X'_n \beta_i - \alpha_i p_n + \epsilon_{in}$$

    (coefficients are agent-specific).

    Then aggregate market share is

$$\int D_{in}(p_1 \ldots p_N, p_z, y_i; \alpha_i, \beta_i) dF(\alpha_i, \beta_i)$$

    and differs from individual choice probability. Elasticity implication of IIA disappears.

    We will focus on this model below, because it has been much used in the recente literature.
  - Important distinction between nested logit and random coefficients: NL implies IIA disappears at the individual level, RC implies IIA disappears only at aggregate level.

4 Berry (1994) approach to estimate demand in differentiated product markets

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

Note: this is different data structure than that considered in previous contexts: here, all variation is across brands (and no variation across time or markets).

Background: Trajtenberg (1989) study of demand for CAT scanners. Disturbing finding: coefficient on price is *positive*, implying that people prefer more expensive machines!

(Tables of results from Trajtenberg paper)

Possible explanation: quality differentials across products not adequately controlled for. In equilibrium of a diff'd product market where each product is valued on the basis of its characteristics, brands with highly-desired characteristics (higher quality) command higher prices. Unobserved quality leads to price endogeneity.

■■■

Here, 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):

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

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: 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, across consumers i and brands j.

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:

$$Pr\left(y_{ij} = 1 | \beta, x_{j'}, \xi_{j'}, j' = 1, \ldots, J\right) = \frac{\exp\left(\delta_j\right)}{\sum_{j'=0}^{J} \exp\left(\delta_{j'}\right)}.$$

Aggregate market shares are:

$$s_j = \frac{1}{M}\left[M \cdot Pr\left(y_{ij} = 1 | \beta, x_{j'}, \xi_{j'}, j' = 1, \ldots, J\right)\right] = \frac{\exp\left(\delta_j\right)}{\sum_{j'=1}^{J} \exp\left(\delta_{j'}\right)}$$
$$\equiv \tilde{s}_j\left(\delta_{j'}\left(x_{j'}, \beta, \xi_{j'}\right), j' = 0, \ldots, J\right) \equiv \tilde{s}_j\left(\alpha, \beta, \xi_1, \ldots, \xi_J\right).$$

$\tilde{s}(\cdots)$ 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 $\hat{s}_j, j = 1, \ldots, J$

  (Share of outside good is just $\hat{s}_0 = 1 - \sum_{j=1}^{J} \hat{s}_j$.)

- 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? Note that you cannot do **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 you need to know the ξ's to compute the predicted share, and they are not observed.

■■■

Berry (1994) suggests a clever IV-based estimation approach.

Assume there exist instruments Z so that $E\left(\xi Z\right) = 0$

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. We wish then to estimate (α, β) by minimizing the sample moment conditions.

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 $\delta_1, \ldots, \delta_J$:
$$\begin{aligned} \hat{s}_1 &= \tilde{s}_J\left(\delta_1(\alpha, \beta, \xi_1), \ldots, \delta_J(\alpha, \beta, \xi_J)\right) \\ \vdots &\quad \vdots \\ \hat{s}_J &= \tilde{s}_J\left(\delta_1(\alpha, \beta, \xi_1), \ldots, \delta_J(\alpha, \beta, \xi_J)\right) \end{aligned}$$
- 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 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:
$$\begin{aligned} \delta_1 &= X_1\beta - \alpha p_1 + \xi_1 \\ \vdots &\quad \vdots \\ \delta_J &= X_J\beta - \alpha p_J + \xi_J \end{aligned}$$
- 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. For example, in 2SLS, you regress p_j on Z_j in first stage, to obtain fitted prices $\hat{p}(Z_j)$. Then in second stage, you regression δ_j on X_j and $\hat{p}(Z_j)$.

  Below, we will consider the substantially more complicated case in Berry, Levinsohn, and Pakes (1995).

■■■

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

- 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. In traditional approach, 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 exploit competition within market to derive instruments. They use IV's 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.
- With panel dataset, where prices and market shares for same products are observed across many markets, could also use prices of product j in other markets as instrument for price of product j in market t (eg. Nevo (2001), Hausman (1996)).

■■■

One simple case of inversion step:

MNL case: predicted share $\tilde{s}_j(\delta_1, \ldots, \delta_J) = \frac{\exp(\delta_j)}{1+\sum_{j'=1}^{J} \exp(\delta_{j'})}$

The system of equations from matching actual to predicted shares is:

$$\begin{aligned}
\hat{s}_0 &= \frac{1}{1+\sum_{j=1}^{J} \exp(\delta_j)} \\
\hat{s}_1 &= \frac{\exp(\delta_1)}{1+\sum_{j=1}^{J} \exp(\delta_j)} \\
\vdots & \qquad \vdots \\
\hat{s}_J &= \frac{\exp(\delta_J)}{1+\sum_{j=1}^{J} \exp(\delta_j)}.
\end{aligned}$$

Taking logs, we get system of linear equations for δ_j's:

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

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).

■■■

4.1 Measuring market power: recovering markups

- Next, we show how demand estimates can be used to derive estimates of firms' markups (as in monopoly example from the beginning).
- 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)

- 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

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

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

- In order to proceed, we need to assume a particular model of oligopolistic competition.

  One 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 pricing first-order conditions for the J brands):

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

  where C_1^j denotes the derivative of C^j with respect to first argument (which is the marginal cost function).

- 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.

  Hence, from these J equations, we can solve for the J margins $p_j - C_1^j$. In fact, the system of equations is linear, so the solution of the marginal costs C_1^j is just

$$\vec{c} = \vec{p} + (\Delta D)^{-1} \vec{D}$$

  where c and D denote the J-vector of marginal costs and demands, and the derivative matrix ΔD is a $J \times J$ matrix where

$$\Delta D_{(i,j)} = \begin{cases} \frac{\partial D^i}{\partial p_j} & \text{if models } (i,j) \text{ produced by the same firm} \\ 0 & \textit{otherwise.} \end{cases}$$

  The markups measures can then be obtained as $\frac{p_j - C_1^j}{p_j}$.

  This is the oligopolistic equivalent of using the "inverse-elasticity" condition to calculate a monopolist's market power.

4.2 Estimating cost function parameters

- However, we may also be interested in estimating the coefficient in the cost function.

  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

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

- 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 $c^j,\ j = 1, \ldots, J$.

  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 $\frac{1}{J}\sum_{j=1}^{J}(c_j - w_j\gamma)U_j$.

- 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.

■■■

5 Berry, Levinsohn, and Pakes (1995): Demand estimation using 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 $\bar{\alpha},\bar{\beta},\Sigma$.

■■■

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:

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

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 $\theta \equiv \left(\bar{\alpha}, \bar{\beta}, \Sigma\right)$. 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
  $$\begin{aligned} \hat{s}_1 &= \tilde{s}_1^{RC}\left(\delta_1, \ldots, \delta_J; \hat{\theta}\right) \\ \vdots \quad & \quad \vdots \\ \hat{s}_J &= \tilde{s}_J^{RC}\left(\delta_1, \ldots, \delta_J; \hat{\theta}\right). \end{aligned}$$
  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
  $$Q(\hat{\theta}) \equiv \frac{1}{J}\sum_{j=1}^{J}\left(\delta_j(\hat{\theta}) - X_j\hat{\bar{\beta}} + \hat{\bar{\alpha}}p_j\right)Z_j. \quad (5)$$
- 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.

■■■

Within this nested estimation procedure, we can also add a supply side to the RC model. With both demand and supply-side moment conditions, the objective function becomes:

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

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. (Assuming $M+N \geq dim(\theta)+dim(\gamma)$)

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 $G(\hat{\theta}, \hat{\gamma})$.

■■■

5.1 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 $\hat{\Sigma}$. The Cholesky factorization of a square symmetric matrix Γ is the triangular matrix $\mathbf{G}$ such that $\mathbf{G}'\mathbf{G} = \Gamma$, so roughly it can be thought of a matrix-analogue of "square root". We use the *lower triangular* version of $\hat{\Sigma}^{1/2}$.

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

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

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

(Results: marginal costs and markups from BLP paper)

6 Applications

Applications of this methodology have been voluminous. Here discuss just a few.

1. evaluation of VERs In Berry, Levinsohn, and Pakes (1999), this methodology is applied to evaluate the effects of voluntary export restraints (VERs). These were voluntary quotas that the Japanese auto manufacturers abided by which restricted their exports to the United States during the 1980's.

The VERs do not affect the demand-side, but only the supply-side. Namely, firm profits are given by:

$$\pi_k = \sum_{j \in \mathcal{K}} (p_j - c_j - \lambda VER_k) D^j.$$

In the above, VER_k are dummy variables for whether firm k is subject to VER (so whether firm k is Japanese firm). VER is modelled as an "implicit tax", with $\lambda \geq 0$ functioning as a per-unit tax: if $\lambda = 0$, then the VER has no effect on behavior, while

$\lambda > 0$ implies that VER is having an effect similar to increase in marginal cost c_j. The coefficient λ is an additional parameter to be estimated, on the supply-side.

Results (effects of VER on firm profits and consumer welfare)

2. Welfare from new goods, and merger evaluation After cost function parameters γ are estimated, you can simulate equilibrium prices under alternative market structures, such as mergers, or entry (or exit) of firms or goods. These counterfactual prices are valid assuming that consumer preferences and firms' cost functions don't change as market structures change. Petrin (2002) presents consumer welfare benefits from introduction of the minivan, and Nevo (2001) presents merger simulation results for the ready-to-eat cereal industry.

3. Geographic differentiation In our description of BLP model, we assume that all consumer heterogeneity is unobserved. Some models have considered types of consumer heterogeneity where the marginal distribution of the heterogeneity in the population is observed. In BLP's original paper, they include household income in the utility functions, and integrate out over the population income distribution (from the Current Population Survey) in simulating the predicted market shares.

Another important example of this type of obbserved consumer heterogeneity is consumers' location. The idea is that the products are geographically differentiated, so that consumers might prefer choices which are located closer to their home. Assume you want to model competition among movie theaters, as in Davis (2006). The utility of consumer i from theater j is:

$$U_{ij} = -\alpha p_j + \beta(L_i - L_j) + \xi_j + \epsilon_{ij}$$

where $(L_i - L_j)$ denotes the geographic distance between the locations of consumer I and theater j. The predicted market shares for each theater can be calculated by integrating out over the marginal empirical population density (ie. integrating over the distribution of L_i). See also Thomadsen (2005) for a model of the fast-food industry, and Houde (2006) for retail gasoline markets. The latter paper is noteworthy because instead of integrating over the marginal distribution of where people live, Houde integrates over the distribution of commuting routes. He argues that consumers are probably more sensitive to a gasoline station's location relative to their driving routes, rather than relative to their homes.

EXTRA TOPIC

7 GHK simulator: get draws from truncated multivariate normal distribution

Introduce idea of **importance sampling**, and an important case of this.

You want draws from

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \sim TN\left(\vec{\mu}, \Sigma; \vec{a}, \vec{b}\right) \equiv N\left(\vec{\mu}, \Sigma\right) \text{ s.t. } \vec{a} < \vec{x} < \vec{b} \tag{6}$$

where the difficulty is that Σ is not necessarily diagonal (i.e., elements of $\vec{x}$ are correlated).

The most obvious "brute-force" approach to simulation is an acceptance-rejection procedure, where you take draws from $N(\vec{\mu}, \Sigma)$ (the untruncated distribution), but reject all the draws which lie outside the desired truncation region. If the region is small, this procedure can be very inefficient, in the sense that you might end up rejecting very many draws.

Importance sampling is a more efficient approach to simulation. In essence, you take draws from an alternative distribution whose support is concentrated in the truncation region. Principle of importance sampling:

$$\int_{\mathcal{F}} s f(s) ds = \int_{\mathcal{G}} s \frac{f(s)}{g(s)} g(s) ds.$$

That is, sampling s from $f(s)$ distribution equivalent to sampling $s * w(s)$ from $g(s)$ distribution,. with importance sampling weight $w(s) \equiv \frac{f(s)}{g(s)}$. ($f$ and g should have the same support.)

Simple example You want to simulate the mean of a standard normal distribution, truncated to the unit interval [0,1]. The desired sampling density is:

$$f(x) = \frac{\phi(x)}{\int_0^1 \phi(x)dx}$$

where $\phi()$ denotes the standard normal density.

Brute force simulation: take draws x^s from $N(0,1)$, and only keep draws in [0,1]. Simulated mean is calculated as: $\frac{\sum_{s=1}^S x^s \cdot \mathbf{1}(x^s \in [0,1])}{\sum_{s=1}^S \mathbf{1}(x^s \in [0,1])}$. Inefficient if $\sum_{s=1}^S \mathbf{1}(x^s \in [0,1]) << S$.

Importance sampling: draw from U[0,1], so that $g(x) = 1$ for $x \in [0,1]$. For each draw, importance weight is $w^s = f(x^s) = \frac{\phi(x^s)}{\int_0^1 \phi(z)dz}$ Simulated mean is $\frac{1}{S}\sum_{s=1}^S x^s w^s$. Don't need to reject any draws.

Importance sampling from truncated MVN Let $(u_1, \ldots, u_n)'$ denote an n-vector of independent multivariate standard normal random variables. Let $\Sigma^{1/2}$ denote the (lower-triangular) Cholesky factorization of Σ, with elements

$$\begin{bmatrix} s_{11} & 0 & \cdots & 0 & 0 \\ s_{21} & s_{22} & \cdots & 0 & 0 \\ \vdots & \vdots & s_{ii} & 0 & 0 \\ s_{n1} & s_{n2} & \cdots & s_{nn-1} & s_{nn} \end{bmatrix}.$$

Then we can rewrite (6) as:

$$\vec{x} = \vec{\mu} + \Sigma^{1/2}\vec{u} \sim N(\vec{\mu}, \Sigma) \;\; s.t.$$

$$\begin{pmatrix} \frac{a_1-\mu_1}{s_{11}} \\ \frac{a_2-\mu_2-s_{21}u_1}{s_{22}} \\ \vdots \\ \frac{a_n-\mu_n-\sum_{i-1}^{n-1} s_{ni}u_i}{s_{nn}} \end{pmatrix} < \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} < \begin{pmatrix} \frac{b_1-\mu_1}{s_{11}} \\ \frac{b_2-\mu_2-s_{21}u_1}{s_{22}} \\ \vdots \\ \frac{b_n-\mu_n-\sum_{i-1}^{n-1} s_{ni}u_i}{s_{nn}} \end{pmatrix} \tag{7}$$

The above suggests that the answer is to draw $(u_1, \ldots, u_n)$ **recursively**. First draw

u_1^s from $N\left(0,1;\frac{a_1-\mu_1}{s_{11}},\frac{b_1-\mu_1}{s_{11}}\right)$, then u_2^s from $N\left(0,1;\frac{a_2-\mu_2-s_{21}u_1^s}{s_{22}},\frac{b_2-\mu_2-s_{21}u_1^s}{s_{22}}\right)$, and so on.

Finally we can transform $(u_1^s,\ldots,u_n^s)$ to the desired $(x_1^s,\ldots,x_n^s)$ via the transformation

$$\vec{x}^s = \vec{\mu} + \Sigma^{1/2}\vec{u}^s.$$

Remark 1: It is easy to draw an n-dimensional vector $\vec{u}$ of independent truncated standard normal random variables with rectangular truncation conditions: $\vec{c} < \vec{u} < \vec{d}$. You draw a vector of independent uniform variables $\tilde{\vec{u}} \sim \mathcal{U}[\Phi(\vec{c}),\Phi(\vec{d})]$[1] and then transform $u_i = \Phi^{-1}(\tilde{u}_i)$.

Remark 2: The GHK simulator is an importance sampler. The importance sampling density is the multivariate normal density $N(\vec{\mu},\Sigma)$ truncated to the region characterized in Eqs. (7). This is a recursively characterized truncation region, in that the range of, say, x_3 depends on the draw of x_1 and x_2. Note that truncation region is different for each draw. This is different than the multivariate normal density $N(\vec{\mu},\Sigma)$ truncated to the region $(\vec{a} \leq \vec{x} \leq \vec{b})$.[2]

For the GHK simulator, the truncation probability for each draw $\vec{x}^s$ is given by:

$$\tau^s \equiv \left[\Phi\left(\frac{b_1-\mu_1}{s_{11}}\right) - \Phi\left(\frac{a_1-\mu_1}{s_{11}}\right)\right]\prod_{i=2}^{m}\left[\Phi\left(\frac{b_i-\mu_i-\sum_{j=1}^{i-1}s_{ij}u_j^s}{s_{ii}}\right) - \Phi\left(\frac{a_i-\mu_i-\sum_{j=1}^{i-1}s_{ij}u_j^s}{s_{ii}}\right)\right].$$

Remark 3: τ^s (even just for one draw: cf. Gourieroux and Monfort (1996, pg. 99)) is an unbiased estimator of the truncation probability $\text{Prob}\left(\vec{a} < \vec{x} < \vec{b}\right)$. But in general, we can get a more precise estimate by averaging over w^s:

$$T_{\vec{a},\vec{b}} \equiv \text{Prob}\left(\vec{a} < \vec{x} < \vec{b}\right) \approx \frac{1}{S}\sum_s \tau^s$$

for (say) S simulation draws.

[1] Just draw $\hat{u}$ from $\mathcal{U}[0,1]$ and transform $\tilde{u} = \Phi(c) + (\Phi(d) - \Phi(c))\hat{u}$.

[2] See Hajivassiliou and Ruud (1994), pg. 2005.

Hence, the importance sampling weight for each GHK draw is the ratio of the GHK truncation probability to the original desired truncation probability: $w^s \equiv \tau^s / T_{\vec{a},\vec{b}} \approx \tau^s / \frac{1}{S}\sum_s \tau^s$.

Hence, the GHK simulator for $\int_{\vec{a}\leq\vec{x}\leq\vec{b}} \vec{x} f(\vec{x}) d\vec{x}$, where $f(\vec{x})$ denote the $N(\vec{\mu},\Sigma)$ density, is $\frac{1}{S}\sum_{s=1}^{S} \vec{x}^s w(x^s)$, or $\sum_s \vec{x}^s \tau^s / \sum_s \tau^s$.

8 Monte Carlo Integration using the GHK Simulator

Clearly, if we can get draws from truncated multivariate distributions using the GHK simulator, we can use these draws to calculate integrals of functions of $\vec{x}$. There are two important cases here, which it is crucial not to confuse.

8.1 Integrating over untruncated distribution $\mathbf{F}(\vec{\mathbf{x}})$, but $\vec{\mathbf{a}} < \vec{\mathbf{x}} < \vec{\mathbf{b}}$ defines region of integration

If we want to calculate

$$\int_{\vec{a}<\vec{x}<\vec{b}} g(\vec{x}) f(\vec{x}) d\vec{x}$$

where f denotes the $N(\vec{\mu},\Sigma)$ density, we can use the GHK draws to derive a Monte-Carlo estimate:

$$E_{\vec{a}<\vec{x}<\vec{b}} g(\vec{x}) \approx \frac{1}{S}\sum_s g(\vec{x}^s) * \tau^s.$$

Here the weight is just τ^s (not w^s), because the desired sampling distribution is the *untruncated* MVN density. The most widely-cited example of this is the likelihood function for the multinomial probit model (cf. McFadden (1989)):

Multinomial probit with K choices, and utility from choice k $U_k = X\beta_k + \epsilon_k$. Probability that choice k is chosen is probability that $\nu_i \equiv \epsilon_i - \epsilon_k < X\beta_i - X\beta_k$, for all

$i \neq k$. For each parameter vector β, use GHK to draw S (K-1)-dimensional vectors $\vec{\nu}^s$ subject to $\vec{\nu} < (\vec{x\beta})$. Likelihood function is

$$\begin{aligned}
\text{Prob}(k) &= \int_{\vec{\nu}} \mathbf{1}\left(\vec{\nu} < (\vec{x\beta})\right) f(\vec{\nu}) d\vec{\nu} \\
&= \int_{\vec{\nu}<(\vec{x\beta})} f(\vec{\nu}) d\vec{\nu} \\
&\approx \frac{1}{S}\sum_s \tau^s.
\end{aligned}$$

8.2 Integrating over truncated (conditional) distribution $\mathbf{F}(\vec{\mathbf{x}}|\vec{\mathbf{a}} < \vec{\mathbf{x}} < \vec{\mathbf{b}})$.

The most common case of this is calculating conditional expectations (note that the multinomial probit choice probability is *not* a conditional probability!)[3].

If we want to calculate

$$E_{\vec{a}<\vec{x}<\vec{b}}\, g(\vec{x}) = \int_{\vec{a}<\vec{x}<\vec{b}} g(\vec{x}) f(\vec{x}|\vec{a} < \vec{x} < \vec{b}) d\vec{x} = \frac{\int_{\vec{a}<\vec{x}<\vec{b}} g(\vec{x}) f(\vec{x}) d\vec{x}}{Prob(\vec{a} < \vec{x} < \vec{b})}.$$

As before, we can use the GHK draws to derive a Monte-Carlo estimate:

$$E_{\vec{a}<\vec{x}<\vec{b}}\, g(\vec{x}) \approx \frac{1}{T_{\vec{a},\vec{b}}} \frac{1}{S} \sum_s g(\vec{x}^s) * \tau^s.$$

The crucial difference between this case and the previous one is that we integrate over a conditional distribution by essentially integrating over the unconditional distribution over the restricted support, but then we need to divide through by the probability of the conditioning event (i.e., the truncation probability).

[3] This is a crucial point. The conditional probability of choice k conditional on choice k is trivially 1!

An example of this comes from structural common-value auction models, where:

$$v(x,x) \equiv \mathcal{E}\left(v|x_1 = x, \min_{j\neq 1} x_j = x\right) =$$
$$\underbrace{\int\cdots\int}_{x_k\geq x,\ \forall k=3,\dots,n} \mathcal{E}\left(v|x_1,\dots,x_n\right)\ dF\left(x_3,\dots,x_n|x_1 = x, x_2 = x, x_k \geq x, k=3,\dots,n;\theta\right) =$$
$$\frac{1}{T_x}\underbrace{\int\cdots\int}_{x_k\geq x,\ \forall k=3,\dots,n} \mathcal{E}\left(v|x_1,\dots,x_n\right)\ dF\left(x_3,\dots,x_n|x_1 = x, x_2 = x;\theta\right) \tag{8}$$

where F here denotes the conditional distribution of the signals $x_3,\dots,x_n$, conditional on $x_1 = x_2 = x$, and T_x denotes the probability that $(x_k \geq x, k = 3,\dots,n|x_1 = x, x_2 = x;\theta)$.

If we assume that $\vec{x} \equiv (x_1,\dots,x_n)'$ are jointly log-normal, it turns out we can use the GHK simulator to get draws of $\tilde{x} \equiv \log\vec{x}$ from a multivariate normal distribution subject to the truncation conditions $\tilde{x}_1 = \tilde{x}, \tilde{x}_2 = \tilde{x}, \tilde{x}_j \geq \tilde{x}, \forall j = 3,\dots,n$. Let $\mathcal{A}(x)$ denote the truncation region, for each given x.

Then we approximate:

$$v(x,x) \approx \frac{1}{T_{\mathcal{A}(x)}}\frac{1}{S}\sum_s \mathcal{E}\left(v|\tilde{x}^s\right) * \tau^s$$

where $T_{\mathcal{A}(x)}$ is approximated by $\frac{1}{S}\sum_s \tau^s$.

References

BAJARI, P., AND L. BENKARD (2005): "Demand Estimation With Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," *Journal of Political Economy*, 113, 1239–1276.

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.

——— (1999): "Voluntary Export Restraints on Automobiles: Evaluating a Strategic Trade Policy," *American Economic Review*, 89, 400–430.

DAVIS, P. (2006): "Spatial Competition in Retail Markets: Movie Theaters," *RAND Journal of Economics*, pp. 964–982.

GOLDBERG, P. (1995): "Product Differentiation and Oligopoly in International Markets: The Case of the US Automobile Industry," *Econometrica*, 63, 891–951.

GOURIEROUX, C., AND A. MONFORT (1996): *Simulation-Based Econometric Methods*. Oxford University Press.

HAJIVASSILIOU, V., AND P. RUUD (1994): "Classical Estimation Methods for LDV Models Using Simulation," in *Handbook of Econometrics, Vol. 4*, ed. by R. Engle, and D. McFadden. North Holland.

HAUSMAN, J. (1996): "Valuation of New Goods under Perfect and Imperfect Competition," in *The Economics of New Goods*, ed. by T. Bresnahan, and R. Gordon, pp. 209–237. University of Chicago Press.

HOUDE, J. (2006): "Spatial Differentiation in Retail Markets for Gasoline," Manuscript, University of Wisconsin.

KEANE, M. (1994): "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62, 95–116.

MADDALA, G. S. (1983): *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press.

MCFADDEN, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995–1026.

Nevo, A. (2001): "Measuring Market Power in the Ready-to-eat Cereals Industry," *Econometrica*, 69, 307–342.

Petrin, A. (2002): "Quantifying the Benefits of New Products: the Case of the Minivan," *Journal of Political Economy*, 110, 705–729.

Rosen, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82, 34–55.

Thomadsen, R. (2005): "The Effect of Ownership Structure on Prices in Geographically Differentiated Industries," *RAND Journal of Economics*, pp. 908–929.

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