Empirical Demand–Supply analysis

- Most of empirical IO focuses on estimation of demand–supply models
- Goal: say something about firm behavior (pricing, advertising, R&D, output decisions)
- While IO theory mostly concerned about supply, data constraints (lack of cost data) focus attention on demand. Fundamental tension: high prices resulting from market power ("bad") or high costs?
- On the other hand, demand characteristics (slopes) crucial to firm behavior.
- Research agenda: Estimate demand model, then simulate firm behavior of interest under alternate assumptions about behavior. Examples.
- Brief review of linear simultaneous equations
- Discrete-choice demand modeling: appropriate for differentiated product markets

Review of Simultaneous Equations

- Review linear simultaneous equations theory in specific context: estimating demand and supply
- Price endogeneity: fundamental problem in much applied IO work
- Appropriate instruments for price in supply ("demand shifters") and demand functions ("cost shifters")
- Estimation methods: IV methods (2SLS, 3SLS); Maximum likelihood

Linear Supply–demand model

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

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

- If u_1 correlated with u_2, then p_t is endogenous in demand function, and q_t is endogenous in supply relation: cannot estimate using OLS.

- Graph

- x's are exogenous variables:
  1. 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. Graph.
  2. x_{t2} are *cost shifters*; affect production costs. Correlated with p_t but not with u_{t1}: use as instruments in demand function. Graph.

Estimation methods (brief) 1

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

$$\Rightarrow \begin{pmatrix} 1 & -\gamma_1 \\ -\gamma_2 & 1 \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix} \begin{pmatrix} x_{t1} \\ x_{t2} \end{pmatrix} + \begin{pmatrix} u_{t1} \\ u_{t2} \end{pmatrix}$$

$$\Rightarrow Y\Gamma = XB + U$$

- IV methods (GMM): population moment conditions $E(u_1 \cdot \mathbf{Z}) = 0$, $E(u_2 \cdot \mathbf{Z}) = 0$ hold at true parameter values.
- 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)
  3. Is rainfall in India a good IV?
- Different IV estimators:
  1. Two-stage least squares
  2. Three-stage least squares
  3. Generally: GMM

Estimation methods (brief) 2

- Maximum likelihood:
  1. Make distributional assumptions about $\{(u_{t1}, u_{t2})_{t=1}^{T}\}$. Example: $(u_{t1}, u_{t2}) \sim$ i.i.d $N(0, \Sigma)$
  2. Likelihood function of the data is joint density of the endogenous variables $(\mathbf{q_t}, \mathbf{p_t})$ conditional on exogenous variables $(\mathbf{x_{t1}}, \mathbf{x_{t2}})$:

$$f(Y) = g(Y\Gamma - XB) * \mid \Gamma \mid \Rightarrow$$

$$\log L(Y \mid X) \sim T \log \mid \Gamma \mid - \frac{T}{2} \log \mid \Sigma \mid$$

$$-\frac{1}{2} (Y\Gamma - XB)' \, \Sigma^{-1} \, (Y\Gamma - XB)$$

  3. Maximize this with respect to Γ, B, Σ.

Empirical Demand analyses 1

Next discuss how we derive estimating equations for demand side.

- In static setting, for two goods 1,2:

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

- 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 (illustrate this next).

Empirical demand analysis 2

Empirical implementation:

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

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

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

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

  - Only estimate equation for first good (since shares sum to 1)
  - Utility max'n behavior places constraints on parameters (H^0 in prices, symmetry Slutsky matrix): can test
  - To estimate: add error term to share equation. Literature on interpretation of this error.

Empirical demand analysis 3

Usual approach not convenient for diff. product markets

- Many alternatives (autos, cereals): too many parameters to estimate
- At individual level, usually only choose one of the available options (discrete choices)

These problems have been approached by

- Demand for a product as demand for a characteristic of that product: **Hedonic** analysis (brief sidetrack to Rosen (1971, *JPE*)
- Discrete choice: assume each consumer can choose at most *one* of the available alternatives on each purchase occasion

Discrete-choice approach 1

- 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

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

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

  Note: if none of the brands are 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:

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

  Sidetrack: conditional, indirect utility

Discrete-choice approach 2

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

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

  Important: 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.*

- Sidetrack: Deterministic vs. stochastic

- Distributional assumptions on ϵ_{in}, $n = 0 \ldots N$ determine the likelihood function for agent i's purchase:

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

- If consumers are identical, and $\{\epsilon_{i0}, \ldots, \epsilon_{iN}\}$ is *iid* across agents i, then $D_n(p_1 \ldots p_N, p_z, y)$ is also the *aggregate market share.*

- Sidetrack: *i.i.d.*

Discrete-choice approach 3

Common assumptions:

- $(\epsilon_{i0}, \ldots, \epsilon_{iN})$ distributed multivariate normal: **multinomial probit**. Computationally burdensome (Keane, McFadden)

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

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

  with (usually) $\mu = 1$, and $\eta = 0.577$ (Euler's constant). $E\epsilon = 0$.

  Leads to multinomial logit

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

  Normalize $V_0 = 0$.

  Convenient, tractable form for choice probabilities.

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

- If interpret D_{in} as market share, 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!

Discrete-choice approach 4

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, ch. 2). Diagram.

- 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.

- Important distinction between nested logit and random coefficients: NL implies IIA disappears at the individual level, RC implies IIA disappears only at aggregate level.

Berry-Levinsohn-Pakes methodology (intro)

Main idea: Control for price endogeneity in *aggregate* discrete-choice framework.

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

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.

This is the type of price endogeneity tackled by Berry-Levinsohn-Pakes.

- aggregate demand model
- price endogeneity problem
- supply side
- extensions