

Differentiated Products Supply and Demand

Phil Haile
Yale University

Fall 2025

Outline (2-3 class meetings)

1. Differentiated Products Supply
2. Berry-Levinsohn-Pakes (“BLP”) Demand Model
3. BLP Estimator and Instruments
4. Estimation Algorithms

Motivation

IO is mostly about market power:

- sources
- implications for profits and firm incentives
- impact on consumers
- effects of policy

Standard models of firms with market power: Oligopoly. Typically, some form of differentiated products competition in prices.

Oligopoly Supply

Baseline workhorse model:

- firms produce differentiated goods/products, selling to consumers with heterogeneous preferences
- static model, complete information (model of long run eqm)
 - ▶ set of products, their non-price characteristics already set
 - ▶ Nash eqm in simultaneous price setting game in each **market**
 - market usually defined by time or geography in practice.

Firm Cost Functions

Variable cost $C_j(Q_j, w_{jt}, \omega_{jt}, \gamma)$ for product j

- Q_j total quantity of good j sold
- w_{jt} observable cost shifters; may include product characteristics x_{jt} that will affect demand (later)
- ω_{jt} unobserved cost shifters (“cost shocks”); may be correlated with latent demand shocks (later)
- γ parameters
- for multi-product firms, we’ll assume variable cost additive across products for simplicity

We will ignore fixed costs: these affect entry/exit/innovation (later) but not pricing conditional on these things.

Demand

Notation

- J_t products/goods/choices in market t (for now $J_t = J$)
- $p_t = (p_{1t}, \dots, p_{Jt})$, prices of all goods
- $\chi_t = (\chi_{1t}, \dots, \chi_{Jt})$, other characteristics of goods affecting demand (observed and unobserved to us)

Demand system:

$$q_{jt} = q_j(p_t, \chi_t) \quad j = 1, \dots, J.$$

Equilibrium Pricing

With single-product firms and constant marginal cost (simple case)

$$\pi_{jt} = q_j(p_t, \chi_t) [p_{jt} - mc_j(w_{jt}, \omega_{jt}, \gamma)]$$

FOC wrt to p_{jt} :

$$p_{jt} = mc_{jt} - q_j(p_t, \chi_t) \left(\frac{\partial q_j}{\partial p_{jt}} \right)^{-1}$$

This is **inverse elasticity pricing** (i.e., monopoly pricing) against the “**residual demand curve**” $q_j(p_t, \chi_t)$:

$$\frac{p_{jt} - mc_{jt}}{p_{jt}} = - \frac{q_j(p_t, \chi_t)}{p_{jt}} \left(\frac{\partial q_j}{\partial p_{jt}} \right)^{-1}.$$

Equilibrium Pricing with Multi-Product Firms

With multi-product firms, firm f 's profit is

$$\begin{aligned}\Pi_{ft} &= \sum_{j \in \mathcal{J}_f} \pi_{jt} \\ &= \sum_{j \in \mathcal{J}_f} q_j(p_t, \chi_t) [p_{jt} - mc_j(w_{jt}, \omega_{jt}, \gamma)]\end{aligned}$$

FOC wrt p_j :

$$p_{jt} = mc_{jt} - \left(\frac{\partial q_j}{\partial p_{jt}} \right)^{-1} \left[q_j(p_t, \chi_t) + \sum_{k \in \mathcal{J}_f \setminus \{j\}} \frac{\partial q_k}{\partial p_{jt}} (p_{kt} - mc_{kt}) \right]$$

(firm **internalizes** effects of p_{jt} on profit from all of its products).

Supply Model

What we get from this...

1. Holding all else fixed, markups/prices depend on the own- and cross-price elasticities of residual demand.

⇒ For good quantitative predictions of firm behavior and market outcomes, **we will need good estimates of demand** (at least the own- and cross-price derivatives)

Supply Model

What we get from this...

2. If we know demand, we can also perform a **small miracle**:

► re-arrange FOC above:

$$mc_{jt} = p_{jt} + q_j(p_t, \chi_t) \left(\frac{\partial q_j}{\partial p_{jt}} \right)^{-1}$$

eqm (p, q) + supply model + estimated demand → estimates of marg costs!

► with multiproduct firms, same thing in system of equations:

$$mc_{jt} = p_{jt} + \left(\frac{\partial q_j}{\partial p_{jt}} \right)^{-1} \left[q_j(p_t, \chi_t) + \sum_{k \in J_f \setminus \{j\}} \frac{\partial q_k}{\partial p_{jt}} (p_{kt} - mc_{kt}) \right]$$

[see Rosse (1970), Bresnahan (1981, 1987), BLP (1995), Berry-Haile (2014)].

Supply Model

What we get from this...

3. If we know demand and marginal costs, we can “predict” a lot of things—i.e., give the quantitative implications of the model for **counterfactual** worlds: e.g., what prices, consumer choices, profits, consumer welfare ... if
- ▶ products were less differentiated?
 - ▶ a tax or tariff were imposed?
 - ▶ two suppliers merged?
 - ▶ a certain new product had not been introduced?
 - ▶ school vouchers were provided to poor students?
 - ▶ ...

Demand can be important on its own. But good demand estimates open a world of possibilities for answering questions about markets and competition policy.

Demand Isn't Easy

Typically we need to know levels/elasticities of demand at particular points; i.e., effects of one price change holding all else fixed. “Local average effects” don’t reveal these.

The main challenge: unobserved demand shifters (“demand shocks”) at the level of the good \times market—e.g., unobserved product characteristics or cross-market differences in tastes (important to acknowledge: not everything is observed!)

- demand shocks are among the things that must be held fixed to measure the relevant demand elasticities etc.

Explicit modeling of these demand shocks central in the applied IO literature following Berry-Levinsohn-Pakes 1995 (and often ignored outside this literature). Handling them correctly requires approaches outside the basic toolkit of applied micro!

A Key Challenge

Econ 115: the quantity demanded of a given good j depends on the prices and characteristics of all related goods (substitutes and complements). This includes the latent demand shocks associated with all of those goods.

How do we hold something fixed if it isn't observed?

The Demand System

With J related goods, demand for **each one** takes the form

$$q_j = D(x, p, \xi)$$

where

- p is a J -vector of **all goods'** prices.
- x is the matrix of **all** non-price observables
- ξ is a J -vector of demand shocks for **all goods**.

This is a system of J equations, all depending on all elements of (x, p, ξ) .

Demand Is Not Regression

$$q_j = D_j(x, p, \xi) \quad (1)$$

- RHS has many latent shocks;
- D_j is a standard regression function only under a strong functional form assumption: that the J components of ξ enter D_j only through a scalar index
- applying regression methods to (1) might allow one to recover certain weighted average derivatives of demand, but those have little value

⇒ not obvious how to proceed, even if prices were exogenous!

let that sink in

Price Endogeneity Adds to the Challenge

- all J endogenous prices are on RHS of demand for each good
- eqm pricing implies that each price depends on **all demand shocks and all cost shocks**
 - prices are endogenous (e.g., correlated with ξ vector)
 - control function generally is **not** a valid solution (Blundell & Matzkin, 2014)
- clear that we need sources of exogenous price variation, but
 - ▶ what exactly is required?
 - ▶ how do we proceed?

Ways Forward

A common approach starts by building up a demand system from a smaller set of parameters appearing in a specification of consumer utilities.

And we'll use a “trick” (a useful mathematical result) to deal with the fact that each good's demand is affected by J structural errors.

Notes:

- deriving demand from utilities offers parsimony: many own- and cross-price elasticities from a modest number of parameters
- it also offers a way to impose some natural exclusivity or symmetry conditions
- but is isn't essential
- the trick (or some other trick) is essential.

NEXT: RANDOM UTILITY DISCRETE CHOICE

Discrete Choice Demand

The choice setting

- consumers have unit demands, choose one of the available options
- generally, one option should be “none of the above” — what we will call the “outside good” (without this, there would be no aggregate demand elasticity!)

Note: discrete choice is more general than it seems; e.g., a single option for a consumer could be {Dodge Caravan + Porsche 911} or {four boxes of cookies + a gallon of milk}. Many key insights we'll cover here extend to models of “multiple discrete choice” or continuous choice.

Random Utility Discrete Choice

Random Utility Specification

- differentiated goods $j \in \{1, \dots, J\}$
- **conditional indirect utilities** of consumer i : u_{ij} (“utility”)
- $(u_{i1}, \dots, u_{iJ}) \sim F_u(\cdot)$ for all i
- outside good 0
 - ▶ only utility differences matter, so wlog. we could set $u_{i0} = 0 \forall i$ —or give it any distribution we want— i.e., we may (must!) **normalize** its distribution.

[Aside: correct use of the term “normalization” is always like this—for an assumption one **may** make without loss and, therefore, **must** make to avoid trivial indeterminacy. The only nuance: without loss for what purpose?]

Choice Probabilities From the Model

- consumer i 's choice (quantities)

$$q_{ij} = 1\{u_{ij} \geq u_{ik} \ \forall k = 0, \dots, J\}$$

(typical assumptions imply ties happen w/probability zero)

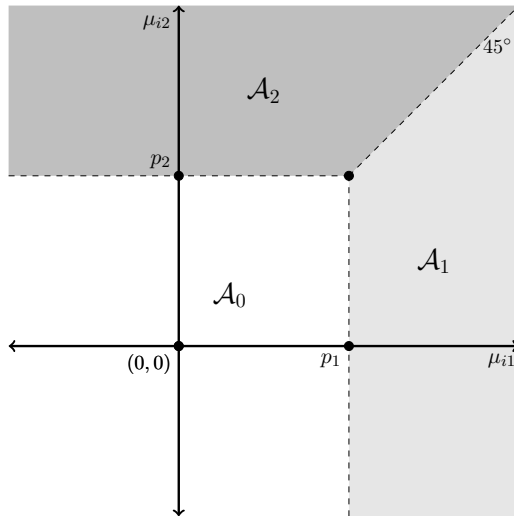
- choice probabilities

$$\begin{aligned} s_{ij} &= \Pr(q_{ij} = 1) \\ &= \int_{\mathcal{A}_j} dF_U(u_{i1}, \dots, u_{iJ}) \end{aligned}$$

where

$$\mathcal{A}_j = \left\{ (u_{i1}, \dots, u_{iJ}) \in \mathbb{R}^J : u_{ij} \geq u_{ik} \ \forall k \right\}.$$

Example: $J = 2, u_{i0} = 0, u_{ij} = \mu_{ij} - p_j \quad j = 1, 2$



Demand and Utility

Some Comments

- utility maximization is one (often convenient) way to represent how consumers make choices
- but demand is well-defined without this — consumers need not maximize utilities or have complete information, e.g.,
- indeed, utility is a notion we make up, and representing demand with utility maximization requires extra assumptions
- relatedly, randomness in the “utilities” could instead reflect noise/errors in consumer choice (e.g., Luce, 1959)
 - ▶ profit-maximizing firms don't care what the randomness represents (unless they can affect it)
 - ▶ but the interpretation will matter for welfare—one in a long list of reasons one must be especially careful about welfare analysis.

NEXT: BLP DEMAND MODEL

Berry, Levinsohn, and Pakes (1995) “BLP”

This is **the** standard empirical model of demand and supply of differentiated products in IO and beyond.

Many of the ideas also in Berry (1994), mostly for simpler models. [Read this first!](#)

Many extensions and variations, some of which we will get to.

Some Goals in BLP

1. parsimonious specification to generate the distribution $F_U(\cdot|p, \chi)$ of random utilities and, thus, demand
2. sufficiently rich heterogeneity in preferences to permit reasonable/flexible substitution patterns
3. explicit treatment of demand shocks and the resulting endogeneity “problem(s)”
4. clarify the identification challenges and possible solutions, including appropriate instruments
5. computationally feasible (in early 1990s!) algorithm for consistent estimation of the model and standard errors.

BLP Random Utility Specification

(slightly simplified)

$$u_{ijt} = x_{jt}\beta_{it} - \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

- consumer i , good/product j , market t
(imagine setting with many markets, each with many consumers)
- price p_{jt}
- $x_{jt} \in \mathbb{R}^K$ non-price observables (product or market characteristics)
- ξ_{jt} *unobserved* demand shock at level of product \times market
- ϵ_{ijt} idiosyncratic (and latent) “taste for product j ”

...

Preference Heterogeneity

$$u_{ijt} = x_{jt}\beta_{it} - \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt}$$

Sources of **consumer** heterogeneity: $\epsilon_{ijt}, \beta_{it} = (\beta_{it}^1, \dots, \beta_{it}^K)$

- $\beta_{it}^k = \beta_0^k + \sigma_k \zeta_{it}^k$ (“random coefficient” = taste for $x_{jt}^{(k)}$)
- $(\epsilon_{i0t}, \dots, \epsilon_{iJt}, \zeta_{it}^1, \dots, \zeta_{it}^K)$ i.i.d. across csrs and mkts
- typically:
 - ▶ $\epsilon_{ijt} \sim$ i.i.d. type 1 extreme value (like multinomial logit)
 - ▶ $\zeta_{it}^k \sim$ i.i.d. standard normal, or drawn from actual distribution of demographics (e.g., income) in market t . Or could have both.

Exogenous and Endogenous Product Characteristics

Recall

$$u_{ijt} = x_{jt}\beta_{it} - \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt} \quad (2)$$

- **exogenous** characteristics: x_t satisfy $E[\xi_t | x_t] = 0$
- **endogenous** characteristics: p_{jt} (usually a scalar, price)
 - ▶ typically each p_{jt} will depend on *whole vector* $\xi_t = (\xi_{1t}, \dots, \xi_{Jt})$ (also on own and others' product characteristics and costs)
 - ▶ we need to distinguish true effects of prices on demand from the effects of ξ_t ; this will require instruments
- of course (2) is not an estimating equation (u_{ijt} is not observed)
- because prices and quantities are all endogenous, you may suspect (correctly) that instruments for prices alone may not suffice. More on this to come.

Utility Specification, Rewritten

Rewrite

$$\begin{aligned}u_{ijt} &= x_{jt}\beta_{it} - \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt} \\ &= \delta_{jt} + \nu_{ijt}\end{aligned}$$

where

$$\begin{aligned}\delta_{jt} &= x_{jt}\beta_0 - \alpha p_{jt} + \xi_{jt} \quad (\text{“mean utility” of good } j \text{ in market } t) \\ \nu_{ijt} &= \sum_k x_{jt}^k \sigma^k \zeta_{it}^k + \epsilon_{ijt} \\ &\equiv x_{jt}\tilde{\beta}_{it} + \epsilon_{ijt} \quad (\text{defining } \tilde{\beta}_{it} \text{—the random part of } \beta_{it}).\end{aligned}$$

Market Shares

- recall $u_{ijt} = \delta_{jt} + \nu_{ijt}$
- let $\delta_{0t} = 0$ (the normalization); define $\delta_t = (\delta_{1t}, \dots, \delta_{Jt})$
- \approx continuum of consumers in each market*
 \implies market shares = choice probabilities =

$$s_{jt} = \Pr(y_{it} = j) = \int_{\mathcal{A}_j(\delta_t)} dF_\nu(\nu_{i0t}, \nu_{i1t}, \dots, \nu_{iJt})$$

where

$$\mathcal{A}_j(\delta_t) = \left\{ (\nu_{i0t}, \nu_{i1t}, \dots, \nu_{iJt}) \in \mathbb{R}^{J+1} : \delta_{jt} + \nu_{ijt} \geq \delta_{kt} + \nu_{ikt} \quad \forall k \right\}$$

* really, enough that sampling error on choice probabilities is negligible compared to that of moments based on variation across products/markets.

Demand

- market shares again

$$s_{jt} = \int_{\mathcal{A}_j(\delta_t)} dF_{\nu}(\nu_{i0t}, \nu_{i1t}, \dots, \nu_{iJt})$$

- with random coefficients, $F_{\nu}(\cdot)$ is really $F_{\nu}(\cdot | x_t, \sigma)$ where
 - ▶ $x_t = (x_{1t}, \dots, x_{Jt}) \in \mathbb{R}^{K \times J}$
 - ▶ $\sigma = (\sigma^1, \dots, \sigma^K)$
- so $s_{jt} = s_j(\delta_t, x_t, \sigma)$
- if M_t is the total measure of consumers in market t , quantities demanded are

$$q_{jt} = M_t \times s_j(\delta_t, x_t, \sigma).$$

Discussion

Key features of the BLP model

- explicit modeling of demand shocks
- consumer heterogeneity through random coefficients

We discussed the need to be explicit about demand shocks—only then can we (a) define “ceteris paribus,” (b) define the endogeneity problem, (c) determine what are valid solutions.

Why random coefficients? A parsimonious way to allow “reasonable” “substitution patterns.”

Why Random Coefficients?

Without random coefficients:

$$\begin{aligned} u_{ijt} &= \underbrace{x_{jt}\beta_0 - \alpha p_{jt} + \xi_{jt}}_{\delta_{jt}} + \epsilon_{ijt} \\ &= \delta_{jt} + \epsilon_{ijt} \end{aligned}$$

If ϵ_{ijt} are iid and independent of (x, p) , e.g. as in the multinomial logit or probit models, products differ only in δ_{jt}

- \implies market shares depend only on the mean utilities;
- \implies price elasticities (own and cross) depend only on mean utilities too.

Example

Two autos with (virtually) identical market shares in 2024:

	MSRP (base)	Mkt Share
Automobile 1		1.2%
Automobile 2		1.2%

Without random coefficients, model implies same mean utility for each and therefore: same own-price demand elasticity, same cross-price elasticity wrt price of any third automobile, say Ford F-series pickup (#1 market share) or Toyota Camry (2nd best-selling sedan).

Example

Two autos with (virtually) identical market shares in 2024:

	MSRP (base)	Mkt Share
Automobile 1	\$31,500	1.2%
Automobile 2	\$40,400	1.2%

Without random coefficients, model implies same mean utility for each and therefore: same own-price demand elasticity, same cross-price elasticity wrt price of any third automobile, say Ford F-series pickup (#1 market share) or Toyota Camry (2nd best-selling sedan).

Example

Two autos with (virtually) identical market shares in 2024:

	MSRP (base)	Mkt Share
Toyota Tacoma	\$31,500	1.2%
Tesla 3	\$40,400	1.2%

Without random coefficients, model implies same mean utility for each and therefore: same own-price demand elasticity, same cross-price elasticity wrt price of any third automobile, say Ford F-series pickup (#1 market share) or Toyota Camry (2nd best-selling sedan).

Logit and IIA

What I've said so far doesn't depend on the distribution ϵ_{ijt} are drawn from. In the logit model, market shares, elasticities, and substitution patterns have simple closed forms in which, e.g., substitution is always proportional to market shares. These particular closed forms are features of the “IIA” property.

But although the exact IIA property is unique to the logit, other specifications of iid epsilons have the same types of problems.

Does this matter?

Yes! Remember: the main reason to estimate demand is to quantify the own- and cross-price demand responses!

- these determine responses to counterfactual changes in market structure
- these used with a model of the supply side to infer firm markups, market power, implications of mergers, entry incentives, etc.
- these are what determine welfare effects of price changes, tax, tariff, merger, collusion, entry, etc.

⇒ if we impose a prior restrictions on substitution patterns, we are restricting the model's ability to give us useful guidance on the questions of interest.

The Science and Art of Modeling

Models that have only iid additive taste shocks impose **very restrictive** relationships between the *levels* of market shares and the matrix of own and cross-price *derivatives* (and therefore on counterfactual predictions). These restrictions do not come from economics but from an assumption **chosen for analytical convenience**.

Models must abstract from reality. And in a finite sample some kind of functional form restrictions are always necessary for estimation. But a good model/approximation should try to avoid placing strong arbitrary restrictions on the key quantities of interest.

How do random coefficients help?

Real goods differ in multiple dimensions; real consumers have (heterogeneous) preferences over these differences

- random coefficients on product characteristics can capture this
 - ▶ large $\beta_i^k \iff$ strong taste for characteristic x^k (e.g., fuel efficiency, pickup dummy)
 - ▶ i 's first choice likely to have high value of x^k
 - ▶ i 's second choice too! (note: cross elasticities are always about 1st vs. 2nd choices)
- incorporating this allows more sensible substitution patterns: competition is mostly “local” – i.e., between firms offering products appealing to the same consumers.

TABLE 1
DESCRIPTIVE STATISTICS

Year	No. of Models	Quantity	Price	Domestic	Japan	European	HP/Wt	Size	Air	MPG	MP\$
1971	92	86.892	7.868	0.866	0.057	0.077	0.490	1.496	0.000	1.662	1.850
1972	89	91.763	7.979	0.892	0.042	0.066	0.391	1.510	0.014	1.619	1.875
1973	86	92.785	7.535	0.932	0.040	0.028	0.364	1.529	0.022	1.589	1.819
1974	72	105.119	7.506	0.887	0.050	0.064	0.347	1.510	0.026	1.568	1.453
1975	93	84.775	7.821	0.853	0.083	0.064	0.337	1.479	0.054	1.584	1.503
1976	99	93.382	7.787	0.876	0.081	0.043	0.338	1.508	0.059	1.759	1.696
1977	95	97.727	7.651	0.837	0.112	0.051	0.340	1.467	0.032	1.947	1.835
1978	95	99.444	7.645	0.855	0.107	0.039	0.346	1.405	0.034	1.982	1.929
1979	102	82.742	7.599	0.803	0.158	0.038	0.348	1.343	0.047	2.061	1.657
1980	103	71.567	7.718	0.773	0.191	0.036	0.350	1.296	0.078	2.215	1.466
1981	116	62.030	8.349	0.741	0.213	0.046	0.349	1.286	0.094	2.363	1.559
1982	110	61.893	8.831	0.714	0.235	0.051	0.347	1.277	0.134	2.440	1.817
1983	115	67.878	8.821	0.734	0.215	0.051	0.351	1.276	0.126	2.601	2.087
1984	113	85.933	8.870	0.783	0.179	0.038	0.361	1.293	0.129	2.469	2.117
1985	136	78.143	8.938	0.761	0.191	0.048	0.372	1.265	0.140	2.261	2.024
1986	130	83.756	9.382	0.733	0.216	0.050	0.379	1.249	0.176	2.416	2.856
1987	143	67.667	9.965	0.702	0.245	0.052	0.395	1.246	0.229	2.327	2.789
1988	150	67.078	10.069	0.717	0.237	0.045	0.396	1.251	0.237	2.334	2.919
1989	147	62.914	10.321	0.690	0.261	0.049	0.406	1.259	0.289	2.310	2.806
1990	131	66.377	10.337	0.682	0.276	0.043	0.419	1.270	0.308	2.270	2.852
All	2217	78.804	8.604	0.790	0.161	0.049	0.372	1.357	0.116	2.099	2.086

Note: The entry in each cell of the last nine columns is the sales weighted mean.

TABLE VII
SUBSTITUTION TO THE OUTSIDE GOOD

Given a price increase, the percentage who substitute to the outside good (as a percentage of all who substitute away.)		
Model	Logit	BLP
Mazda 323	90.870	27.123
Nissan Sentra	90.843	26.133
Ford Escort	90.592	27.996
Chevy Cavalier	90.585	26.389
Honda Accord	90.458	21.839
Ford Taurus	90.566	25.214
Buick Century	90.777	25.402
Nissan Maxima	90.790	21.738
Acura Legend	90.838	20.786
Lincoln Town Car	90.739	20.309
Cadillac Seville	90.860	16.734
Lexus LS400	90.851	10.090
BMW 735i	90.883	10.101

Which random coefficients?

We must choose which characteristics have random coefficients

- dummies for subsets of products?
 - ▶ this yields something equivalent (up to different distributional assumptions) to the *nested logit*: see Berry (1994)
- certain horizontal or vertical characteristics (parts of X, P)?

In practice, the choice depends on the application and data set, including instruments. Too many RC's for the data available will often yield imprecise estimates of σ .

NEXT: ESTIMATION OF THE BLP DEMAND MODEL

Recall: Utility Specification, Rewritten

$$\begin{aligned}u_{ijt} &= x_{jt}\beta_{it} - \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt} \\ &= \delta_{jt} + \nu_{ijt}\end{aligned}$$

where

$$\begin{aligned}\delta_{jt} &= x_{jt}\beta_0 - \alpha p_{jt} + \xi_{jt} \quad (\text{“mean utility” of good } j \text{ in market } t) \\ \nu_{ijt} &= \sum_k x_{jt}^k \sigma^k \zeta_{it}^k + \epsilon_{ijt} \\ &\equiv x_{jt}\tilde{\beta}_{it} + \epsilon_{ijt} \quad (\text{defining } \tilde{\beta}_{it} \text{—the random part of } \beta_{it}).\end{aligned}$$

Estimation of Demand with Market-Level Data

Highest-Level View

- we observe x_t , p_t , sh_t , w_t , and $\tilde{z}_t \leftarrow$ excluded IV
(for clarity, sh_{jt} and sh_t denote the *observed* market shares)
- given a choice of parameters, the model predicts market shares as a function of observables
- choose the parameters so that the predictions match the observed shares as well as possible, while obeying (as well as possible) the assumed exogeneity conditions on x_t and w_t .

Estimation with Market-Level Data: A Partial Sketch

1. start with demand model alone
2. suppose $F_\nu(\cdot|x_t, \sigma)$ is known (i.e., σ known)
3. for each market t , find $\delta_t \in \mathbb{R}^J$ such that $s_j(\delta_t, x_t, \sigma) = sh_{jt} \forall j$
i.e., “invert” model at observed market shares to find mean utilities δ_t
4. using IV ($E[\xi_{jt}|z_{jt}] = 0$), estimate the equation

$$\delta_{jt} = x_{jt}\beta_0 - \alpha p_{jt} + \xi_{jt}.$$

Some Details to Fill In

ok ... a lot of details

1. What instruments?
2. Will the “inversion” step actually work?
3. What about σ ??
4. Formally define estimator, computational algorithm(s)
5. Standard errors?
6. Add Supply Side
 - ▶ additional restrictions aid estimation of demand
 - ▶ estimate parameters γ of marginal cost function
(why? may care directly; and needed for counterfactuals that change equilibrium quantities unless mc is constant).

NEXT: INSTRUMENTS

Instruments for Estimating Demand

Broadly speaking, we need variables that exogenously shift all endogenous variables—*prices and quantities—independently*.

This may be counterintuitive: to estimate demand, we might think instruments for prices were all we needed. As we discussed earlier, however, exogenous variation in prices generally doesn't suffice: we need prices to move exogenously, but also to hold the demand shocks fixed. More below.

Typical (Excluded) Instruments for Estimating Demand

1. Excluded cost shifters w_t

- ▶ classic demand instrument, e.g., wages, material costs, shipping cost to market t , taxes/tariffs, demand shifters from other markets

2. Proxies for excluded costs shifters

- ▶ typical: price of same good in another mkt (“Hausman instruments”); properly excluded if demand shocks in one market not correlated with those in others

Typical (Excluded) Instruments for Estimating Demand

3. Markup shifters

- ▶ exogenous changes in ownership (e.g., merger)
- ▶ characteristics of “nearby” consumers (“Waldfoegel instruments”)
 - e.g., firms in some industries may use same price for all markets in a region (e.g., CT)
 - e.g., age/income/education in Greenwich may affect prices (markups) in New Haven, but may be independent of New Haven preferences (including New Haven demand shocks) conditional on New Haven observables

4. “BLP Instruments” x_{-jt}

- ▶ by assumption, $E[\xi_{jt}|x_t] = E[\xi_{jt}]$
- ▶ affect quantities directly; affect prices (markups) via FOC

Later: “optimal instruments” — i.e., optimal functions of the excluded instruments for estimating the unconditional moments.

NEXT: INVERSION

Will the Inversion Step Work?

Given x, σ and any positive shares sh , define $\Phi : \mathbb{R}^J \rightarrow \mathbb{R}^J$ by

$$\Phi(\delta) = \delta + \ln(sh) - \ln(s(\delta, x, \sigma))$$

BLP show (under mild conditions on the linear random coefficients random utility model) that for any nonzero shares sh , Φ is a **contraction**, i.e.,

- it has a unique fixed point in δ
 - ▶ $s(\delta_t, x_t, \sigma)$ has an inverse: we can write $\delta_t = \delta(s_t; x_t, \sigma)$ (see also Berry, 1994)
- \exists convergent algorithm: start with guess δ^0 , set $c = 1$
 1. let $\delta^c = \Phi(\delta^{c-1})$, $c = 1, 2, \dots$,
 2. repeat to convergence

(invertibility for more general models: Berry, Gandhi, & Haile, 2013).

NEXT: Identification?

What About sigma??

- inversion result \implies for any market shares s_t and any σ , we can find a vector δ_t that rationalizes the data (fits market shares perfectly)
- a non-identification result? there is NO information about σ from market shares?

What are we forgetting?

- variation across markets (and products): the structural errors
 $\xi_{jt} = \delta_{jt} - x_{jt}\beta_0 - \alpha p_{jt}$ implied by candidate (α, β, σ) must be mean-independent of exogenous observables across markets and products
- (this is just like linear regression: for any $(x, y, \beta) \exists e$ such that $y = x\beta + e$, but $x \perp e$ is what ensures identification of β).

Identification of sigma: loose nonparametric intuition

Changes in choice sets

- recall $s_{jt} = s(\delta_t, x_t, \sigma)$
- consider two markets, same p_t, x_t, ξ_t in each
- remove 1 car in one of them... where does its mkt share “go”?
 - ▶ to cars with large mkt shares?
 - ▶ to cars similar to the one removed in some dimension(s)?
- similar idea with cts variation across/within markets

One source of the looseness: how can we fix $\xi_t = (\xi_{1t}, \dots, \xi_{Jt})$ in two markets when ξ_t is latent??
 Instruments will shift things independently of ξ_t , but this isn't the same as fixing them! In reality, instruments for quantities are what enable us to hold ξ_t fixed. More on this later.

Identification of sigma: parametric intuition

Counting parameters and moments:

- trial value of $\sigma \implies \delta_t(sh_t; x, \sigma)$ by inversion
- with trial value of $\alpha, \beta \implies \xi_t(\sigma, \alpha, \beta)$
- IV orthogonality condition: $E[\xi_{jt}(\sigma, \alpha, \beta) Z_{jt}] = 0 \forall j, t$
- what kind of Z do we need? we need at least as many moment conditions as parameters
 - ▶ $\implies x_{jt}$ plus excluded IV for price is not enough: (α, β) are not all the parameters!
 - ▶ to identify σ , we need additional excluded instruments.

A minimal requirement: as many exogenous variables as we have parameters in the model.

Nonparametric Identification of Demand

Even if considerations lead to reliance on parsimonious parametric specifications, we'd like to know what does (or does not) permit identification without such restrictions.

Berry and Haile (2014) examine a nonparametric generalization of the BLP demand model using market-level data. The main requirement: instruments creating independent exogenous variation in all $2J$ endogenous variables: **prices and quantities**:

- intuitively: **move 1** price; **hold fixed $J - 1$** prices and J demand shocks $\rightarrow 2J$
- instruments: “BLP instruments” (exogenous characteristics of other products) plus J others (shifters of costs or markups)

Details later in the course, including more clarity about why the BLP instruments are valid and necessary.

NEXT: ESTIMATION DETAIL

Basic Idea for BLP Estimator (demand alone)

idea: method of moments estimator

- any guess at the parameters (σ, α, β) implies (after inversion) values $\xi_{jt}(\sigma, \alpha, \beta)$ for the latent demand shocks rationalizing the data
- moments conditions: $E[\xi_{jt}(\sigma, \alpha, \beta) z_{jt}] = 0$
- sample analog $E[\xi_{jt}(\sigma, \alpha, \beta) z_{jt}] \approx \frac{1}{JT} \xi_{jt}(\sigma, \alpha, \beta) z_{jt}$
- GMM estimator
 - ▶ $(\hat{\sigma}, \hat{\alpha}, \hat{\beta})$ chosen to make sample analog close to zero
 - ▶ optimal weighting of moments for efficiency

Note: this means fitting market shares exactly, orthogonality as close as possible, based on large ($\rightarrow \infty$ faster than J) number of consumers per market.

Some Complications

1. model predictions $s_j(\delta_t, x_t, \sigma)$ involve high-dimensional integrals (recall 2-D picture)
 - ▶ use simulation to approximate
 - ▶ \implies “method of simulated moments”
2. moment conditions involve $\xi_t(\sigma, \alpha, \beta)$, which has no closed form \rightarrow two options for computation of the estimator:
 - (i) solve contraction at each trial value of (σ, α, β)
 \implies “nested fixed point” algorithm (BLP; or)
 - (ii) forget about contraction, solve the BLP constrained optimization problem directly using specialized algorithms adapted to the BLP details (Dube-Fox-Su, 2012)

(we'll discuss both options soon).

Defining the Estimator

Notation

- let $\theta = (\theta_1, \theta_2) = ([\alpha, \beta_0], \sigma)$
- let Z_{jt} denote the exogenous variables $(x_{jt}, w_{jt}, \tilde{z}_{jt})$
- let $\delta_{jt}(\theta_2)$ be shorthand for $\delta_j(sh_t; x_t, \sigma)$.

The BLP Estimator

GMM estimator of θ defined as solution to mathematical program:

$$\min_{\theta} \quad g(\xi(\theta))' \Omega g(\xi(\theta)) \quad \text{s.t.}$$

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

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

$$\log(sh_{jt}) = \log(s_j(\delta_t, x_t, \theta_2))$$

$$s_j(\delta_t, x_t, \theta_2) = \int \frac{\exp[\delta_{jt}(\theta_2) + x_{jt}\tilde{\beta}]}{1 + \sum_k \exp[\delta_{kt}(\theta_2) + x_{kt}\tilde{\beta}]} f_{\tilde{\beta}}(\tilde{\beta}|\theta_2) d\tilde{\beta}_i$$

$$\Omega = \text{standard GMM weight matrix.}$$

NEXT: COMPUTATION OF THE BLP ESTIMATOR

BLP Estimation Algorithm (Sketch)

“Nested Fixed Point” algorithm (used for other things too)

- Outer Loop: search over trial values of θ
- Inner Loop: given θ , find solution for $\xi(\theta)$
 - ▶ given θ_2 , solve for $\delta(\theta_2)$ as fixed point of contraction mapping
 - ▶ then $\xi_{jt}(\theta) = \delta(\theta_2) + \alpha p_{jt} - x_{jt}\beta_0$

```
begin outer loop
  try new  $\theta$ 
  begin inner loop
    solve contraction
  end inner loop
  calculate GMM criterion
end outer loop
```

BLP Estimation Algorithm: More Detail

1. set up approximation of predicted shares (given θ) via simulation

▶ draw values of ζ_{it}^k from normal distribution $\rightarrow \tilde{\beta}_{it}$

▶ take many draws to simulate population of csrs:

$$s_j(x_t, \delta_t, \theta_2) = \int \frac{\exp[x_{jt}\tilde{\beta}_{it} + \delta_{jt}]}{1 + \sum_{j'} \exp[x_{j't}\tilde{\beta}_{it} + \delta_{j't}]} f(\tilde{\beta}_{it}|\theta) d\tilde{\beta}_{it}$$

$$\approx \frac{1}{NS} \sum_{i=1}^{NS} \frac{\exp[x_{jt}\tilde{\beta}_{it} + \delta_{jt}]}{1 + \sum_{j'} \exp[x_{j't}\tilde{\beta}_{it} + \delta_{j't}]}$$

▶ more sophisticated sampling (“importance sampling”) possible

▶ use this approximation routine every time $s_j(x_t, \delta_t, \theta_2)$ appears below

▶ use the same set of draws of ζ_{it}^k each time to avoid “jittering”.

BLP Estimation Algorithm: More Detail

1. set up approximation of shares via simulation
2. take a trial value of the parameters θ
 - ▶ selected by search algorithm (e.g., Nelder-Meade in BLP).

BLP Estimation Algorithm: More Detail

1. set up approximation of shares via simulation
2. take a trial value of the parameters θ
3. start with initial guess at $\delta_t^0 \forall t$
 - ▶ e.g., from MNL or nested logit.

BLP Estimation Algorithm: More Detail

1. set up approximation of shares via simulation

2. take a trial value of the parameters θ

3. start with initial guess at $\delta_t^0 \forall t$

4. solve fixed point problem by iterating on the contraction

$$\delta_{jt}^{h+1} = \delta_{jt}^h + \log(sh_{jt}) - \log(s_j(\delta^h, x_t, \theta_2)) \quad \forall j \rightarrow \delta_{jt}(\theta_2) \quad \forall jt$$

▶ convergence tolerance τ_{inner}

▶ tempting to make tolerance loose, at least in beginning, but this can lead to big trouble

▶ better than simple iteration on contraction: “SQUAREM” acceleration method of Reynaerts, Nash, and Varadahn (2012).

BLP Estimation Algorithm: More Detail

1. set up approximation of shares via simulation
2. take a trial value of the parameters θ
3. start with initial guess at $\delta_t^0 \forall t$
4. solve fixed point problem by iterating on the contraction
$$\delta_{jt}^{h+1} = \delta_{jt}^h + \log(sh_{jt}) - \log(s_j(\delta^h, x_t, \theta_2)) \forall j \rightarrow \delta_{jt}(\theta_2) \forall jt$$
5. $\delta_{jt}(\theta_2) = x_{jt}\beta_0 - \alpha p_{jt} + \xi_{jt} \implies \hat{\xi}_t(\theta)$
6. sample analog of $g(\xi(\theta)) = \frac{1}{JT} Z_{jt} \hat{\xi}_{jt}(\theta)$
7. plug into GMM objective function
8. iterate (from 2) to convergence

Standard errors: standard MSM (e.g., Pakes-Pollard, 1989); formulas in paper based on $J \rightarrow \infty$. See Freyberger (2015) for version when $T \rightarrow \infty$.

“Nonlinear” vs. “Linear” Parameters

Important Simplification:

- $\theta_1 = (\alpha, \beta_0)$ enter *objective function* linearly
- so given θ_2 and Ω , we have closed-form expression for optimal θ_1
- so outer loop *search* only involves $\theta_2 = \sigma$
- (this also makes it natural to interpret the inner loop as solving for all ξ_{jt} rather than all δ_{jt}).

NEXT: AN ALTERNATIVE ALGORITHM

The NFP Algorithm

Challenges

- contraction rate (“inner loop”) can be slow
(although recent progress on this, noted already)
- “outer loop” optimization is hard
 - ▶ existence of inner loop often yields highly non-convex optimization problem, search algo can fail
 - ▶ user-defined tolerances important; can get wrong answer without realizing it
 - ▶ NFP algorithm *seems* unnecessarily “expensive:” it forces constraints on market shares and linear parameters hold exactly at every guess of θ_2 when we care only about constraints’ holding at the final solution (but shortcuts turn out to be dangerous!).

Constrained Optimization (“MPEC”) Algorithm

(mathematical programming with equilibrium constraints)

- Dube, Fox, and Su (2012): alternative computation approach for the **same estimator**
- the general idea:
 - ▶ NFP is just a routine to solve a constrained optimization problem: **minimize GMM objective function over parameters, subject to constraint that the inner loop fixed point equations hold**
 - ▶ so try off-the-shelf constrained optimization solvers (e.g., Knitro) that work well for “sufficiently nice” problems
 - ▶ DFS show that certain tricks can (sometimes) make the BLP-MPEC problem “sufficiently nice.”

BLP by MPEC: A Simple Version

- constrained optimization formulation of BLP estimator

$$\begin{aligned} \min_{\theta, \xi} \quad & \xi' Z' \Omega Z' \xi \quad \text{s.t.} \\ \log(sh_{jt}) &= \log(s_j(x_t, \xi_t, \theta)) \quad \forall j, t \\ \text{where } s_j(x_t, \xi_t, \theta) &\equiv \frac{1}{NS} \sum_{i=1}^{NS} \frac{\exp[x_{jt}\beta_0 + \xi_{jt} + \sum_k x_{jkt}\zeta_{it}^k]}{1 + \sum_{\ell} \exp[x_{\ell t}\beta_0 + \xi_{\ell t} + \sum_k x_{\ell kt}\zeta_{it}^k]} \end{aligned}$$

Note: ξ_{jt} treated as parameters (a lot of them!) whose values (along with θ) must—at the final solution—equate predicted and observed shares.

BLP by MPEC: Better Version

More Parameters Still

- introduce superfluous parameters g

$$\begin{aligned}
 \min_{\theta, \xi, g} \quad & g' \Omega g \quad \text{s.t.} \\
 g \quad &= Z' \xi \\
 \log(sh_{jt}) \quad &= \log(s_j(x_t, \xi_t, \theta)) \quad \forall j, t_i \\
 \text{where } s_j(x_t, \xi_t, \theta) \quad &\equiv \frac{1}{NS} \sum_{i=1}^{NS} \frac{\exp[x_{jt}\beta_0 + \xi_{jt} + \sum_k x_{jkt}\zeta_{it}^k]}{1 + \sum_{\ell} \exp[x_{\ell t}\beta_0 + \xi_{\ell t} + \sum_k x_{\ell kt}\zeta_{it}^k]}
 \end{aligned}$$

- even more parameters now, but often faster computation due to *sparsity* of Hessian matrix for objective function (see Dube, Fox, Su, 2012).

MPEC Algorithm

1. single-step optimization (no inner/outer loop)
2. “canned” optimization algorithms from engineering
 - ▶ some (KNITRO is most popular) also available for Matlab, other matrix languages
 - ▶ Matlab `fmincon` can work in some cases
3. advantages
 - ▶ no user-defined tolerances or optimization parameters to mess up
 - ▶ solvers designed for constrained optimization problems, so likely more reliable than ad hoc routines written/adapted by economists.

MPEC Algorithm

Disadvantages

1. need sparsity to get speed gains (or for solvers to work at all), and this often vanishes when J large and T small.
2. generally need to code 1st and 2nd derivatives, and may need to work hard to make problem "nice" (further tricks in the formulation of the problem to induce useful sparsity); this can take a long time and introduces heavy problem-specific coding

⇒ claimed avoidance of dependence on error-prone inputs less clear

[however, "automatic differentiation" is increasingly available].

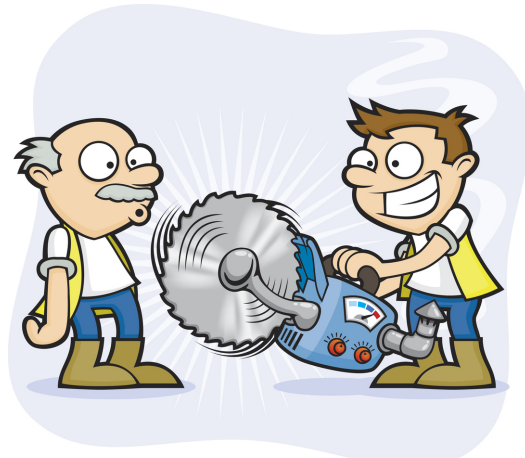
Optimization: NFP vs. MPEC?

Naive implementation of either approach can easily fail. But both can work well when one follows now-established best practices. Both have publicly available implementations.

The open source **pyBLP** implementation—discussed in some detail in Conlon and Gortmaker (2020)—offers a frontier NFP approach incorporating multiple advances and options. They also have useful online tutorials (google pyBLP) that complement the published paper. [I recommend starting here.](#)

Computation: Bottom Line

BLP is a power tool. Careless use can result in cutting off a finger, or worse—crazy elasticities. But a few now-standard practices make it possible to work safely, and on a level far beyond what is possible with simpler tools.



NEXT: “OPTIMAL INSTRUMENTS”

Optimal Instruments: An Important Digression

Loosely

- many possible functions of the exogenous variables X, Z could serve as instruments
- what is the best choice?
- particularly relevant to BLP IV: many subsets/combinations/functions of huge x_{jt} could be used

Formally: which unconditional moment conditions yield asymptotic efficiency?

Optimal Instruments

- recall $\theta = (\alpha, \beta, \sigma)$; from Chamberlain (1986) the optimal (but infeasible) demand-side instruments are

$$D_{jt}(z_t) = E \left[\frac{\partial \xi_{jt}(\theta^0)}{\partial \theta} \middle| z_t \right] \quad (3)$$

- feasible approximations: must not depend on realized demand shocks
 - ▶ sieve basis or direct approximation
 - ▶ initially explored in BLP 1995
 - ▶ better alternative in BLP 1999; simpler version in Reynaert-Verboven (2014)
 - ▶ alternative basis in Gandhi-Houde (2019) (“differentiation instruments”)
 - ▶ in practice: approximate optimal IV often help substantially
 - ▶ Conlon-Gortmaker (2020): more detail, more options, all available in pyBLP and well documented.

NEXT: BRINGING SUPPLY BACK TO THE MODEL

Adding the Supply Side Moments

- suppose

$$mc_{jt}(w_{jt}, \omega_{jt}, \gamma) = w_{jt}\gamma + \omega_{jt}$$

- recall firm FOC:

$$p_{jt} - w_{jt}\gamma + \omega_{jt} - \frac{s_j(\delta_t, x_t, \sigma)}{\alpha} \left(\frac{\partial s_j}{\partial \delta_{jt}} \right)^{-1} = 0$$

- so for any $(\sigma, \alpha, \beta, \gamma)$, we have an implied ω_{jt}
- additional moments for estimation:

$$E[\omega_{jt}(\sigma, \alpha, \beta, \gamma) \tilde{z}_{jt}] = 0$$

- Note: supply moments depend on demand parameters too; in practice, these often help precision of demand estimates— σ in particular.

Some Important Extensions and Active Topics

1. micro data, consumer panels, etc. (e.g., Petrin 2002; BLP 2004)
2. asymptotics in T (Freyberger, 2015)
3. endogenous non-price observables (e.g., Fan, 2013, Berry-Haile 2024)
4. incomplete “consideration sets” (e.g., Goeree, 2008)
5. “multiple discrete choice” (e.g., Hendel, 1999)
6. LASSO selection of covars/instruments (Gillen, Montero, Moon & Shum, 2015)
7. EL estimator (Conlon, 2013)
8. role of functional forms, NP identification (Berry & Haile, 2014,2024)
9. nonparametric estimation/inference (Compiani, 2022)
10. discriminating between models of supply (Berry & Haile, 2014; Backus, Conlon, & Sinkinson, 2024; Duarte, Magnolfi, Solvsten and Sullivan (2024))