Empirical Models of Differentiated Products

Steven T. Berry
Yale University, Cowles Foundation and NBER

June 18-19, 2015

These slides are “as presented” in the June 18-19 Masterclass at UCL; apologies for remaining typos and errors.
Differentiated Products: Introduction

While homogeneous goods is a convenient assumption for many models, it is frequently violated in practice. Many markets feature goods that are not identical, they vary in quality, features, reliability, reputation and/or geographic location. Markets of literally identical goods seem to be relatively rare, especially once differences in seller’s locations and reputations are taken into account.
Empirical modeling and estimation of differentiated products therefore seems to be important.

We might think of differentiated products as the general case and perhaps even as the typical case.
Outline of Classes

1. Introduction, Examples and Broad Ideas (this session)
2. Identification and Estimation of Differentiated Products Demand and Supply (end of this session and then next session)
3. Models with Endogenous Product Characteristics (tomorrow)
4. Example Applications to Policy Relevant Markets: Anti-trust, Heath, Media and Education (actually, largely interspersed.)
Examples

- transportation demand McFadden, Talvitie, and Associates (1977)
- market power Berry, Levinsohn, and Pakes (1995), Nevo (2001))
- mergers (e.g., Nevo (2000a) Capps, Dranove, and Satterthwaite (2003))
- environmental policy Goldberg (1998)
Examples

- equilibrium product quality Fan (2013)
- media bias Gentzkow and Shapiro (2010)
- asymmetric information and insurance Cardon and Hendel (2001), Bundorf, Levin, and Mahoney (2012), Lustig (2010)
- residential sorting Bayer, Ferreira, and McMillan (2007)
- voting Gordon and Hartmann (2013)
- school choice Hastings, Kane, and Staiger (2010), Neilson (2013)
Example: Mergers

Old “classic” merger analysis involved a discussion of “concentration,” “Herfindahl indices,” and “market definition,” which is appropriate to homogeneous goods. But “modern” merger analysis typically involves differentiated products. Key question is how close substitutes are the merging goods.

What is the level of the “diversion ratio” (of those who leave a good due to a price rise, what fraction substitute to a given set of substitute goods)?

“Structural merger analysis” is not quite standard yet, probably correctly so, but it an increasingly used tool.
School Choice

In the choice of universities or lower-level schools (school choice in US, Chile or France collège), one can think of the schools as offering differentiated products. Differentiation in location, quality, specialization. Demand is from students / parents. Price and/or admission policies are set in equilibrium.

What are the effects of different school choice programs? Of tuition increases? On student performance, student welfare?
McFadden Transportation

McFadden introduced the idea of transportation choice as a differentiated product. Products (bus, train, car) are differentiated by price, travel time and mode preference. Preferences are partly “random,” but also follow from location of home and work.
Optimal Entry

In differentiated product markets, fixed costs are often noticeably high relative to the potential market for a given product, and so the entry decision is an important component of market outcomes and consumer welfare.

Oligopoly theory says that in differentiated product markets, choice of entry on entry may not be optimal. The effect on welfare can swamp any “deadweight loss triangle” from oligopoly pricing condition on entry. Can we measure these effects empirically?
Endogenous Quality and Location

Firms choose their characteristics as function of costs, demand and competition from rival firms. Can we model this decision? What are the implications of the endogenous characteristics for empirical work? Does the market do a “good job” of generating product variety?

To what degree does my product choice set depend on my neighbors’ preferences (“preference externalities”). What are the implications of this for urban economics?
There is now a “modern” empirical literature that takes up where the older “monopolistic competition” literature(s) left off. Consumers demand products differentiated by product characteristics. Fixed costs and product differentiation imply a limited number of products and therefore imperfect competition, perhaps Nash price-setting. Multi-product firms are common.

A more recent literature models how firms choose what products to produce (in a multi-dimensional horizontal and/or vertical space.) This endogenizes the choice of products.
The recent empirical literature looks at market-specific data, at least detailed models of products and firms and perhaps additional data on consumers. The consumer data matches consumers (and their attributes or demographics) to their purchases.

The modern literature often treats unobservables as explicit latent variables, with interpretations as demand or supply shifters. This is in the “structural modeling” tradition of supply-and-demand analysis.

The advantages of the approach are [i] explicit assumptions and [ii] the ability to do detailed counter-factual policy analysis. In practice, there is often a dependence on functional form, although there is now a literature on non-parametric identification.
Instruments, Demand and Supply

To a large degree, the models here are the differentiated products extension of simple homogenous goods supply and demand. As in the S & D literature, much of our discussion will be about instruments. We now need to learn about substitution patterns within the market, often in cases where the exact same products do not occur in each market. We also need to model markups on the supply side, as prices do not depend on costs alone.
**Demand**

On the demand side, there are difficult questions of functional form, endogeneity and identification.

- There are often many products, leading to high-dimensional demand, necessitating product aggregation and/or other ways of reducing dimension.
- There is at least price endogeneity and, realistically, endogeneity of locations, quality and/or characteristics.
- We have to identify not just “price effects,” but also cross-products substitution patterns. What are good “instruments for “substitution patterns”?
The empirical differentiated products literature is often wrongly treated as being mostly about demand. But supply is necessary for many counter-factuals. Furthermore, under imperfect competition, supply decisions reflect demand elasticities and so are informative about demand.

In differentiated goods markets, there is often sufficient product variety and fixed costs are large relative to market size and perfect competition is not a good assumption. Positive markups are necessary to ensure at least zero profits.
Supply Side Issues

A classic issue is that costs (especially marginal costs) are often not observed. Therefore, they are inferred from firm behavior.

Other issues here include what is the proper model of supply (Nash in prices? quantities? dynamics?) How to handle multi-product firms? Is price chosen simultaneously with location / quality or after?
Dynamics

Unfortunately, we will not have time to discuss dynamics, on either the firm or consumer side. On the firm side, one could start with Ackerberg, Benkard, Berry, and Pakes (2007).
Review of Differentiated Products Theory

The differentiated products theory literature is largely concerned with firm behavior, but firms’ incentives are partly driven by demand.

Two broad traditions in deriving demand.

1. A representative consumer who has a taste for consuming a variety of products (much of classic “consumer theory” focuses on this)
2. Heterogeneous consumers facing products with differing characteristics

Representative consumer models almost always continuous demand, heterogenous consumers often discrete choice. Sometimes aggregate to discrete choice to multiple choices or continuous demand.
Representative Consumers and a Taste for Variety

In this tradition, the demand curve is typically derived from a well-specified utility function that features decreasing marginal utility from the consumption of each good in the market. This provides the incentive for the representative consumer to spread consumption across a variety of goods. The reason that goods are differentiated is typically buried in the parameters of the utility function, rather than being made explicit.

or ...
Characteristics Models

This tradition locates goods in a space of product characteristics. Consumers have heterogeneous tastes and place differing utility weights on the different product characteristics. Each consumer is often modeled as buying at most one unit of the good, yielding a “discrete choice” model of demand at the consumer level. Product variety is then a response to the variety of consumers preferences, rather than the “taste for variety” of a representative consumer. Aggregate (market level) demand is then found by summing up the demands of the individual consumers.

Sometimes consumers are located in the same product space as the products, sometimes just have random tastes for characteristics (and/or products).
Examples of Representative Consumer Models

One tradition assumes a CES representative consumer model, e.g. Dixit and Stiglitz (1977). Utility is a function of an aggregate of the differentiated goods and of money spent elsewhere. The aggregate is something like

\[ U(Z) = U\left(\sum_j \beta_j z_j^\rho\right), \]

with \( \rho < 1 \) giving a taste for variety.

CES style models are good for the stylized treatment of equilibrium (as in trade theory), but have not been used frequently in applied Industrial Organization (IO). They lead naturally to constant markups and can be difficult to reconcile with discrete purchasing patterns.
Heterogeneous CES?

As noted by Anderson, DePalma, and Thisse (1992), there is a close relationship between the CES and the logit that we consider later. Indeed, the logit social welfare function can be considered as the utility of a representative agent. At the market level, both CES and logit are too restrictive. We will add heterogeneity to the logit, but could have started with the CES and added heterogeneity.
Example: General Flexible Demand Systems

The problem of too many parameters

Could use general “flexible” demand systems: e.g. constant elasticity demand or translog demand, etc. But what are the underlying consumer preferences? Jorgenson, et al, Jorgensen, Lau, and Stoker (1982) suggest translog indirect utility, for example.

A basic problem in demand estimation in product space is that to we have to estimate many parameters from very little data. Thus for a hundred goods (which is not a lot), an unrestricted first order approximation would require a hundred cross price elasticities and one income elasticity for each of a hundred products. As a result a basic question is when can we aggregate goods, and still obtain a “structural” form which has all the appropriate interpretations.
Representative Consumer Example: Almost Ideal systems

How to justify demand in terms of consumer theory, but without “too many” own- and cross-elasticities? Look for a system in which consumers care about the “subutility” from various groups of goods and then look for a (justified) price index for the groups.

At the lowest level, model demand for a good as depending on the prices of within group goods, conditional on group expenditure. At the next level up, model group expenditure as depending on the group price index.

For example, Almost Ideal demand system functional form allows this (see Deaton and Muellbauer (1980)). (Note that the CES and logit can also “nest”.)
The Characteristics Approach

Products are defined by their characteristics. Consumers are defined by preferences for characteristics.

These models are particularly useful when the nature of products changes across markets. Less preferable when the characteristics are hard to define and measure, but the products are stable across markets. (Hausman: what about champagne – the number of bubbles?)

The choice between discrete and continuous models of demand will typically be decided by the nature of the market at hand.
Discrete Choice Models.

Many of the product differentiation models we consider are of the discrete-choice form. An advantage of these models is that they build demand from a well-specified utility for the characteristics of products. An unfortunate restriction is that they usually constrain each consumer to consider buying at most one unit of a good. This restriction can be relaxed: e.g. Hendel (1999). Examples of discrete choice models used in IO theory include the Hotelling and vertical models. Anderson, DePalma and Thisse (1992) provide a lengthy discussion of discrete choice models as used in the theory of product differentiation.
Discrete Choice Utility

A general discrete choice model starts with by specifying the utility of consumer $i$ for product $j$ as

$$ u_{ij} = U(x_j, p_j, \nu_i), $$

where $x_j$ is a vector of product characteristics, $p_j$ is the price of the product and $\nu_i$ is a vector of consumer characteristics. (Rather than model utility directly as a function of price, it might be preferable to model it as a function of expenditures on other products and then derive the “indirect” utility $U$ as a function of price.)
**Intro and Broad Ideas**

**Theory: Discrete Choice Characteristics Models**

**Hotelling**

As example, the utility function in the Hotelling model with quadratic transportation costs is

$$u_{ij} = \bar{u} - p_j - (x_j - \nu_i)^2,$$

where $x_j$ is the location of the product along the line and $\nu_i$ is the location of the consumer.
Vertical Model

In the vertical model Shaked and Sutton (1982), Bresnahan (1987))

\[ u_{ij} = v_i x_j - p_j \]

where \( x_j \) is “quality”, and \( v_i \) is the consumers “taste” for quality. Or, the model can be written as

\[ u_{ij} = \delta_j - \alpha_i p_j, \]

We normalize the utility of the outside alternative \((u_{i,0})\) to zero.
Multiple Fixed Dimensions of Preferences

Vertical and Hotelling utility functions can be extended to multiple characteristics. For example, an analog of the Hotelling model, with $K$ characteristics, is

$$u_{ij} = \bar{u} - p_j - \sum_{k=1}^{K} \alpha_k (x_{jk} - \nu_{ik})^2,$$

while an extension of the vertical model is the “pure random coefficients model”:

$$u_{ij} = \left[ \sum_{k=1}^{K} \nu_{ik} x_{jk} \right] - p_j$$
Finite Dimensional Models

In models where the dimension of tastes is potentially smaller than the number of products, a particular product at a particular price can be strictly dominated in preferences for all consumers. This product will have zero sales. This is maybe realistic in some markets, but the zero sales cause problems for empirical work Berry and Pakes (2007).
Traditional Econometric Discrete Choice

See, e.g., McFadden (1981)
A traditional econometric specification is:

\[ u_{ij} = x_j \beta_i - \alpha_i p_j + \epsilon_{ij} \]

where the consumer “tastes” are given by

\[ \nu_i = (\beta_i, \alpha_i, \epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{iJ}) \]

But what are the \( \epsilon_{ij} \)? Product and consumer specific random terms?
The $\epsilon$’s in the traditional discrete choice model have full support and are iid (or have an iid component). Therefore every good is purchased, no matter its price or characteristics.
Logit

In the traditional econometric specification, the $\epsilon_{ij}$ are assumed to be i.i.d. across products and consumers. In the simplest case, there are no random coefficients on the product characteristics, ($\beta_i = \beta$ and $\alpha_i = \alpha$) and the $\epsilon_{ij}$ have the “type 2 extreme value” distribution

$$F(\epsilon) = e^{-e^{-\epsilon}}.$$
Logit

This gives the traditional logit model, where the probability that good $j$ is purchased (i.e. the market share of product $j$) is

$$s_j = \frac{e^{\delta_j}}{\sum_{r=1}^{J} e^{\delta_r}},$$

with $\delta_j = x_j \beta - \alpha p_j$. If we add an “outside good” with $\delta_j = 0$, the logit market share becomes:

$$s_j = \frac{e^{\delta_j}}{1 + \sum_{r=1}^{J} e^{\delta_r}}.$$
Logit Substitution

While the logit market share is easy to calculate, the model has unintuitive properties. In particular, cross-price effects do not depend on the degree to which the products have similar $x_j$’s, but only on the values of the sum $\delta_j$. In the logit,

$$\frac{\partial s_j}{\partial p_j} = -\alpha s_j (1 - s_j)$$

and

$$\frac{\partial s_j}{\partial p_k} = \alpha s_j s_k.$$
Logit Substitution

The counter-intuitive substitution patterns do **not** come only from the specific distributional assumption in the logit model, but from the assumption that the only variance in consumer tastes comes through the i.i.d. product-specific terms $\epsilon_{ij}$. Since these terms are i.i.d., there is no source of correlation in consumer tastes across similar products. This is in strong contrast to the finite dimensional tastes models (Hotelling, vertical, etc.)
More Flexible Substitution

When variance is added to the terms $\beta_i$ and/or $\alpha_i$, then substitution patterns can become more reasonable. Now, a consumer who buys a good with a large value of some characteristic is more likely than the average consumer to have as a second choice another good with a large value of that characteristic.
Random Coefficients Logit

Assume that $\epsilon$ is extreme value, like the logit, but keep $(\beta_i, \alpha_i)$ random in utility

$$u_{ij} = x_j\beta_i - \alpha_i p_j + \epsilon_{ij}.$$ 

This is relatively easy to compute and the own- and cross-derivatives are more flexible.
$\epsilon$?

However, as long as the distribution of $\epsilon$ is i.i.d., there are effectively as many product characteristics as there are products. Indeed, we can think of $\epsilon_{ij}$ as the consumer taste for a product characteristic that is defined to be equal to one for product $j$ and zero otherwise. If the $\epsilon$’s have an unbounded distribution, then all goods are strict substitutes for one another (e.g. $\partial s_j / \partial p_r$ is positive for all $j$ and $r$.) Contrast this to the one-dimensional taste models (Hotelling, vertical) in which each product is a substitute with only 2 other “nearby” products.
The $\epsilon$’s are effectively product-specific tastes. When a new product is introduced, the dimension of tastes automatically increases. With logit errors, as $J \to \infty$, a single-product firm becomes a monopolist against the “fat tail” of consumers who care only about that good and markups go to a constant (not zero.) Also, welfare is automatically increased as the product spaced increases. For somewhat ad hoc ways to adjust for this, see Ackerberg and Rysman (2005).

The more important are random coefficients, the less important is this effect.
Equilibrium

Having discussed some methods of deriving demand, we can turn to a consideration of equilibrium. The simplest models of product differentiation would consider a set of single product firms each producing a differentiated product. We could begin by specifying a demand system for this set of related products, together with cost functions and an equilibrium notion.
The usual assumption is Nash-in-prices. To analyze the case of equilibrium with differentiable demand, note that the profits of firm \( j \) are given by

\[
\pi_j(p) = p_j q_j(p) - C_j(q_j(p)).
\]

The first-order condition is:

\[
q_j + (p_j - mc_j) \frac{\partial q_j}{\partial p_j} = 0.
\]

Note that we can rewrite this as

\[
p_j = mc_j + b_j(p),
\]

where the markup is

\[
b_j(p) = \frac{q_j}{\left| \frac{\partial q_j}{\partial p_j} \right|}.
\]
We can write the product Lerner index in terms of the usual “inverse elasticity” rule.

\[
\frac{(p_j - mc_j)}{p_j} = \frac{1}{\eta_j},
\]

where \( \eta_j \) is the absolute value of the product-specific elasticity.
Existence and Uniqueness

In a 1991 paper, Caplin and Nalebuff (1991) consider a broad class of discrete-choice models and provide a result on the existence of equilibrium in such models. They also provide a partial characterization of the equilibrium. They require two assumptions on the distribution of consumer utility. The first requires utility to be linear in consumer characteristics; note that all of our examples above satisfy this requires (although the quadratic transport cost example requires some re-writing to see this.) The second assumption places a restriction on the density of consumer tastes which restricts the density to a set that is broader than the set of log-concave densities.
Quantity vs. Price Setting

In differentiated products models, one could also consider quantity setting firms, as in the Cournot model. In this case, the first-order condition becomes

$$p_j + q_j \frac{\partial p_j}{\partial q_j} - mc_j = 0.$$ 

Note that $\frac{\partial p_j}{\partial q_j}$ is the $j^{th}$ diagonal element of the J by J matrix

$$\left[ \frac{\partial q}{\partial p^j} \right]^{-1},$$

which is not the same as the inverse of $\frac{\partial q_j}{\partial p_j}$ (except in the monopoly case). For many examples, the quantity-setting markup will be higher than the price setting markup.
Complete Information in Static Models

The world isn’t actually static, so the static model must be a “metaphor” for a relatively stable situation without strong dynamic linkages.

If the situation is stable, then it is likely that the economic actors have learned about many of the unobservables that are unobserved by us, leading to the frequent use of complete information models. In complete information static Nash, in equilibrium the firms only have to condition on the actions of their rivals, not necessarily the unobserved profit shifters.

Contrast to first-price auction models, where the game is often actually “one-shot” and bidders can really suffer from ex-post regret.
Empirical models follow the theory literature in many respects. To describe the data, we have to add sources of randomness. In demand, unobservables can be at the level of the consumer and the market. On supply, these are at the level of the firm, probably correlated across firms within markets.

Markets are typically separated in time and/or space. We will typically ignore time-series correlation, but this is not good. Spatial correlation is an issue in at least some contexts.
Again: types of Demand Models

As in the theory, empirical demand models usually take on one of two types, although one can mix and match.

1. Continuous Product Demand
   ▶ Utility defined directly over products (not characteristics)
   ▶ continuous quantity choice
   ▶ often representative consumer
   ▶ problems with “too many own and cross-price elasticities,” \( J^2 \) when number of products = \( J \).

2. Characteristic Models
   ▶ Utility over characteristics (helps with parsimony, counterfactual predictions)
   ▶ Usually discrete choice with heterogenous consumers

We will spend more time on characteristics models.
Continuous Demand, Representative Consumer Models

As an example, Hausman has a series of papers using Almost Ideal functional forms to ask questions about competition and/or price indices/benefit of new goods. In Hausman (1996), he measures the benefits of new goods by integrating under the demand curve as price decrease from the “choke-price” to the observed level.
Example: Hausman ’96

Hausman (1996) studies breakfast cereals with Almost Ideal system. The Almost Ideal model of product demand is [i] consistent with the utility of a representative consumer while [ii] expressing demand in a “nested” form, which the demand for products within subgroups depending on own and cross-price elasticities and cross-group demand depending on group price indices.

There is still a problem of “too many elasticity parameters” as the number of products grows large. On the other hand, the same cereals are sold in many markets and it is hard (but see Nevo (2001)) to list “characteristics” of cereal (“mushy”? ). Reduce the number of coefficients by ignoring small products and grouping the rest.
Price Endogeneity

The “error” in Almost Ideal models is usually “tacked on” to the end of the demand curve, but if we think of this as a “demand shock” there is plausibly an endogeneity problem with the coefficient on (log) price. This is just the usual “supply and demand” endogeneity problem—unobserved demand shocks drive demand.

Natural instruments are cost shifters and in the Almost Ideal system there are many prices and so we need many instruments. There are only a few observed cost shifters (e.g. Nevo uses price of sugar.)
“Hausman” instruments

Hausman uses the prices of goods in other markets as “cost” instruments. These are valid if price variation across market-time is driven by unobserved cost shocks, not unobserved demand shocks. Bresnahan criticized this assumption, but the Hausman instruments do provide rich cross-product variation if the underlying assumption is correct.
Empirical Characteristics Models

Leaving Hausman style models, we return to characteristics models where firms and consumers care about \((p_t, x_t)\). The simplest kind of empirical work in this context is “descriptive hedonics,” so let’s take a brief look at that.
Aside: Hedonics

The simplest sort of empirical work on differentiated products seeks to descriptively characterize the relationship between product characteristics and the prices and/or quantities of each firm. A simple regression of prices on product characteristics is called a “hedonic regression.” à la Griliches Griliches (1961). This type of regression is often used to show how the “reduced form” relationship between prices and quantities changes over time.

In the interpretation of Griliches and Pakes 2001, the hedonic coefficients combine the equilibrium effects of costs and demand (via markups) in a characteristics model.
Hedonics, cont

For example, hedonic regressions are used to correct the producer price index for computers. The data is a panel of prices, $p_{jt}$ and characteristics, $x_{jt}$, of products over time. Simplifying, say that in period $t$, price is regressed on $x$ to obtain the parameters of

$$p_{jt} = x_{jt} \beta_t + \epsilon_t.$$ 

In period $t + 1$, a new set of products is on the market. It would be a mistake to look at the unadjusted price of, say, a mainframe computer because most likely this year’s model is far better than last year’s model (i.e. the $x$’s have improved.). The hedonic helps us ask how much today’s models would have cost yesterday.
Hedonics, cont

We can get a predicted price by using last year’s coefficients on this year’s characteristics

\[ \hat{p}_{jt+1} = x_{jt+1} \beta_t. \]

Alternatively, we could put year dummy variables in a pooled regression of log price on \( x \)’s across years and treat the changes in the dummy variable as percentage changes in price, adjusted for quality.
Hedonics, cont

This hedonic technique does not generate an ideal utility-based price index, but Pakes (2003) discusses bounds on the difference between the hedonic and an ideal index. Consider a non-parametric regression of \( p_t \) on \( x_t \) generating the function \( \hat{p}_t(x) \). We can then construct an index comparing \( p_{i,t+1} \) to \( \hat{p}_t(x_{i,t+1}) \) – thus asking what this product would have cost last year. By the usual price-index logic (Paache/Lespeyres) this bounds the true index. Note that this does not deal with the welfare affects of truly new \( x' \) – in that case \( \hat{p}_t(x_{i,t+1}) \) will not be defined.
Many of the product differentiation models we consider are of the discrete-choice form. Advantages include

- a clearly-specified utility function for the characteristics of products
- relatively easy to handle heterogeneous consumers
- a relatively parsimonious treatment of demand for many products, since parameters are often modeled as growing in the number of products, not characteristics
- handles different levels of aggregation from the consumer to the market (although aggregation not always trivial, it is well-defined)
- easy to incorporate full or partial information on characteristics of purchasing consumers.
Multiple Discrete Choice

An unfortunate restriction of traditional discrete choice models is that they often constrain each consumer to buy at most one unit of a good. This restriction can be relaxed, see

- Hendel (1999) (firm demand for PCs)
- Björnstedt and Verboven (2013), a combination of logit and CES, allows for continuous choice).
Example of Discrete Choice Applications on Market Data

First, look at a classic paper that combines supply and demand, not using classic econometric discrete choice methods. Then, turn to demand and supply separately, and together.
Bresnahan’s Paper on the 1955 Auto Price War.

Bresnahan (1987) starts with a puzzle, which is the apparent decrease in automobile prices that occurred in 1955, which was an economic boom year, as compared to the surrounding years of 1954 and 1956, when automobile prices seemed to be higher. Bresnahan wanted to know whether differences in competition could account for the price change. One hypothesis is that collusive behavior collapsed in the face of the economics boom. The idea, then, as in the homogeneous goods literature on “testing the competition”, is to look for evidence on what sort of competitive regime best explains the data in each of years 1954-1956.
Bresnahan, cont

Bresnahan assumes that we observe, for each product (= car), a set of product characteristics $x_j$ as well as the market outcomes of price and quantities sold, $p_j, q_j$. We observe the data only at the overall market (aggregate) level, without any information on the actions of individual households. Such data is readily available from industry-oriented publications (such as *Automotive News* and Ward’s.)
For simplicity, Bresnahan uses the demand vertical model, so there is only one attribute per product. In the data, there are a number of different characteristics, such as size, horsepower and luxury, so he creates an index say $\delta_j = x_j \beta$. He assumes a uniform distribution of tastes for quality and derives the demand function.

$$q_j = q_j(\delta_j, \delta_{-j}, p_j, p_{-j}).$$ (1)

that we derived before.
The prices are endogenously determined by a price-setting Nash equilibrium. Note that in the automobile industry, the first-order condition has to be modified to account for multi-product firms. GM, for example, offers dozens of products in the market today. To solve for equilibrium prices, one must also specify a cost side of the model. Costs, particularly marginal costs, are not typically observed so, as in the Cournot models above, we often are forced to infer marginal costs with the help of the oligopoly first order condition. Bresnahan assumes that marginal costs are a convex function of quality. (This helps to ensure an equilibrium with positive shares.)
Bresnahan, cont

Bresnahan can then:

- Solve for reduced-form $p$ and $q$
- Tack on errors (say they’re normal.)
- Estimate by MLE under two equilibrium assumptions: Nash in price and collusion.

Given the estimates under the two competing equilibrium assumptions, Bresnahan does a “non-nested” hypothesis test and finds, as conjectured, that collusive pricing fits better in 1954 and 1956, but Nash pricing fits better in 1955!
Bresnahan on Markups and Characteristics

Bresnahan emphasizes that under non-collusion oligopoly, markups are heavily affected by whether or not the firm faces close substitutes. He makes this point in the context of multi-product firms and multi-product first order conditions. He uses this insight to try to distinguish collusive and non-collusive behavior, finding that collusion really did collapse in the boom.

The insight is also useful when we think about “instruments for substitution” patterns.
Empirical Demand Models

Example: Bresnahan on the 1955 Auto Price War

(b) competitive

Figure 2(b)
Bresnahan, cont

Critiques:

▶ The vertical model is way too restrictive. (Maybe better for computers.)
▶ What identifies demand elasticities when there is no within-year price variation? ("The model").
▶ The error structure does not allow for unobservables.

So, how about traditional econometrics discrete choice models?
Review of Econometric Discrete Choice Models

A classic probit or logit

\[ u_{ij} = x_j \beta - \alpha p_j + \epsilon_{ij}, \]

with \( \epsilon \) either standard normal or else “extreme value” (double exponential, \( \epsilon \sim exp(-exp(-\epsilon)) \)).

- Traditional Probit
- Traditional i.i.d. Multinomial Probit
- Multinomial Probit with correlated errors and the high dimensionality of integration.
Logit again

The i.i.d. logit solves the problem of high dimensional integration with a function form of:

\[ s_j = \frac{e^{x_j \beta - \alpha p_j}}{\sum_k e^{x_k \beta - \alpha p_k}} \]  \hspace{1cm} (2)

Again, note the derivatives are:

\[ \frac{\partial s_j}{\partial p_j} = -\alpha s_j (1 - s_j) \]

\[ \frac{\partial s_j}{\partial p_k} = \alpha s_j s_k \]

These depend only on shares, not on \( x's \).
Traditional Logit with Consumer Data

Estimating a traditional logit at the individual level within a single market:

\[ u_{ij} = x_j \beta - \alpha p_j + \left( \sum_r z_{ir} x_{jr} \gamma_r \right) + \epsilon_{ij} \]  (3)

\[ = \delta_j + z_i' \Gamma x_j + \epsilon_{ij} \]  (4)

where \( \delta_j = x_j \beta - \alpha p_j \) is McFadden’s “alternative specific constant.”

The probability that individual \( i \) chooses good \( j \) is

\[ Pr(i \mid z_i, \delta, \Gamma) = \frac{e^{\delta_j + z_i' \Gamma x_j}}{\sum_k e^{\delta_k + z_i' \Gamma x_j}} \]  (5)
We can then estimate $(\delta_j, \Gamma)$ by MLE, where the contribution to the log-likelihood of an individual is

$$\ln \left( Pr(i \mid z_i, \delta, \Gamma) \right) = \delta_j + z_i' \Gamma x_j + \ln \left( \sum_k e^{\delta_j + z_i' \Gamma x_j} \right)$$

There are very standard algorithms for maximizing the log-likelihood. Note: because of the $\delta_j$ term, at the market level, should get an exact fit to product-market shares (integrating out across a mass of consumers.)
Traditional Discrete Choice on Micro Data

Just as we gave Bresnahan as an example of traditional discrete choice on micro data, so we can give a classic example of “econometric” discrete choice on consumer-level data.
Empirical Example: McFadden on BART

One classic example of discrete choice modeling with micro data is the McFadden, Talvitie, and Associates (1977) analysis of transportation mode choice, intended to provide policy guidance for the development of mass transit systems such as the Bay Area Rapid Transit system in California. Consider a slight variation on the model of Chapter 2, Table 7 of that study. In the model, each work commuter in metro area $t$ chooses a commuter transportation mode: automobile or bus. Bus transportation is further divided into the choice of walking or driving to the bus, for a total of 3 modes: auto alone, bus with walk access and bus with auto access.
The discrete choice logit utility function for person $i$ considering mode $j$ in market $t$ is

$$u_{jt} = \delta_{jt} + z_{ijt}\gamma + \epsilon_{ijt}, \quad (6)$$

where $\delta_{jt}$ is the “alternative specific constant,” $z_{ijt}$ is a vector of consumer attributes that are potentially specific to mode $j$ and $\epsilon_{ijt}$ is an i.i.d. extreme value idiosyncratic shock to tastes. The functional form assumption on $\epsilon_{ijt}$ creates the logit functional form for choice probabilities. The parameter $\gamma$ is to be estimated.
McFadden Attributes

The $z_{ijt}$ attributes are created from a survey of household attributes together with the exact geographic location of the household and workplace and the detailed way in which the transportation network connects those two household-specific locations. The list of mode-specific attributes includes, for example, transportation time in minutes, which varies across modes in different ways for different households. Consumers also care about the dollar cost of the mode, which can be broken into two parts. One part varies across consumers (for example gasoline costs vary with distance to work) and one part does not vary (the standard bus fare.) We denote the dollar cost that does not vary across consumers as $p_{jt}$. 
McFadden “Alternative Specific Constants”

The alternative specific $\delta_{jt}$ are estimated as parameters on dummy variables for each mode. Note, though, that metro-area bus fare, $p_{jt}$ for the bus-walk mode, is perfectly correlated with the alternative-specific dummy variable for the that mode. The authors therefore do not estimate a coefficient on bus fare but instead create a variable defined as fare divided by per-minute wage. This gives a cost, in minutes of work, that varies across consumers.
<table>
<thead>
<tr>
<th>Mode 1:</th>
<th>Auto</th>
<th>Data: Work Travel Survey, East Bay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode 2:</td>
<td>Bus, Walk Access</td>
<td>Model: Multinomial Logit, Fitted by the Maximum Likelihood Method</td>
</tr>
<tr>
<td>Mode 3:</td>
<td>Bus, Auto Access</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 2 Work-Trip Mode-Choice Basic Model**
All cost and time variables are calculated round-trip. Dependent variable is alternative choice (one for chosen alternative, zero otherwise). Sample size: 161.

a/ The variable is zero for the auto alternative, and takes the described value for the other alternatives.

b/ The variable is zero for the bus alternatives, and takes the described value for the auto alternative.

c/ The variable is zero for the bus-with-walk-access alternative, and takes the described value for the remaining alternatives.

d/ The variable is one for the bus-with-auto-access alternative and zero otherwise.

e/ Sum of home-to-work and work-to-home.
### Explanatory Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimated Coefficients</th>
<th>t-Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost divided by post-tax wage, in cents/ (cents per minute)</td>
<td>-.0358</td>
<td>2.78</td>
</tr>
<tr>
<td>On-vehicle time, in minutes</td>
<td>-.0185</td>
<td>1.38</td>
</tr>
<tr>
<td>Walk time, in minutes</td>
<td>-.0190</td>
<td>.972</td>
</tr>
<tr>
<td>Transfer-wait time, in minutes</td>
<td>-.0534</td>
<td>1.54</td>
</tr>
<tr>
<td>Number of transfers</td>
<td>-.0723</td>
<td>.249</td>
</tr>
<tr>
<td>Headway of first bus, with a ceiling of 8 minutes, in minutes</td>
<td>-.218</td>
<td>2.32</td>
</tr>
<tr>
<td>An index of distance to parking at home</td>
<td>-.318</td>
<td>.933</td>
</tr>
<tr>
<td>Family income with ceiling of $7000, in $ per year</td>
<td>.000434</td>
<td>1.51</td>
</tr>
<tr>
<td>Family income minus $7000 with floor of $0 and ceiling of $3000, in $ per year</td>
<td>.000785</td>
<td>1.66</td>
</tr>
<tr>
<td>Family income minus $10,000 with floor of $0 and ceiling of $5000, in $ per year</td>
<td>-.000617</td>
<td>2.47</td>
</tr>
<tr>
<td>Length of residence in community, in years</td>
<td>.143</td>
<td>2.90</td>
</tr>
</tbody>
</table>
Table 2, continued

<table>
<thead>
<tr>
<th>Explanatory Variables</th>
<th>Estimated Coefficients</th>
<th>t-Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>An index of population density in neighborhood b/</td>
<td>- .741</td>
<td>2.58</td>
</tr>
<tr>
<td>Dummy if respondent is over 44 years of age</td>
<td>- .781</td>
<td>1.22</td>
</tr>
<tr>
<td>Dummy if there is child in household b/</td>
<td>-1.63</td>
<td>2.54</td>
</tr>
<tr>
<td>Number of persons in household who can drive c/</td>
<td>1.10</td>
<td>2.68</td>
</tr>
<tr>
<td>Auto alternative dummy d/</td>
<td>-5.49</td>
<td>2.27</td>
</tr>
<tr>
<td>Bus-with-auto-access dummy d/</td>
<td>-2.76</td>
<td>3.45</td>
</tr>
</tbody>
</table>

Likelihood ratio index: .5983
Log likelihood at zero: -176.9
Log likelihood at convergence: .05
Alternative Specific Constants, again?

In the McFadden BART example, the characteristics are really just dummies for the alternatives, which interact with $z_i$. Further, there is just one market, so all identification is from variation in the consumer attributes within market.

The coefficients are used to predict demand for subway service, which is like the bus, but quicker and at a different price, with a different set of routes.

Question for later: would the alternative-specific constant be the same in some other market?
Micro Level Nested Logit

In the micro-level logit, consumers of “type” $z_{it}$ will substitute to other products that are predicted to be popular conditional on $z_{it}$, but the substitution patterns do not depend directly on the characteristics of the goods.

The nested logit model allows for tighter substitution among products who are in a similar group (or “nest”) of products.
Nested Logit Utility

The nested logit utility function include as common “taste” for within group products, see Ben-Akiva (1973), McFadden (1978) and Cardell (1997). The nested logit utility function for a product in group $g$ is

$$u_{ijt} = \delta_{jt} + z'_{it} \Gamma x_{jt} + \nu_{ig}(\sigma_g) + (1 - \sigma_g) \epsilon_{ijt}. \quad (7)$$

We can let the parameters $\beta_g$ and $\sigma_g$ vary by group, or not.
The within group shares still follow the logit function form,

$$s_{j/g} = \frac{e^{(\delta_j + z'_i \Gamma x_{jt})/(1-\sigma_g)}}{D_g}$$  \hspace{1cm} (8)

where the denominator is

$$D_g \equiv \sum_{k \in \mathcal{J}_g} e^{(\delta_k + z'_i \Gamma x_{kt})/(1-\sigma_g)}.$$  \hspace{1cm} (9)

The group shares are

$$\bar{s}_g = \frac{D_g^{(1-\sigma_g)}}{\sum_h D_h^{(1-\sigma_h)}}$$  \hspace{1cm} (10)

This gives the market share (purchase probability) as the product of the within group share and the group share.
Following McFadden, we have two ways of estimating the nested logit model via maximum likelihood. The first method estimates the models in two steps. At the “lower level,” we can estimate each nest separately using the within group shares and recover the nest-specific parameters. Then, at the upper level we can use the group share data to recover the $\sigma_g^\ell$ parameters. The second method would be to simply estimate the model via MLE all at once. The first method is intuitively and easier to program using existing logit code. However, the first method will not carry over to estimation with random coefficients.
For a nested logit application see the related papers by Goldberg (1995) or Goldberg (1998).
Random Coefficient Probit or Logit

\[ u_{ijt} = x_{jt} \beta_{it} - \alpha_{it} p_{jt} + z_{ijt} \gamma + \xi_{jt} + \epsilon_{ijt}, \]  

(11)

where vector \((\beta_{it}, \alpha_{it})\) is assumed follow a parametric distribution (normal for the \(\beta_{it}\)'s, log-normal for the \(\alpha_{it}\)'s) with parameters to be estimated and \(\epsilon_{ijt}\) is still extreme value (RCLogit) or else normal (RCProbit).
Simulation

With random coefficients, we lose the nice logit / nested logit functional form and are left with a high-dimensional integral. Solution: either numeric integration (getting a little easier) or else, traditionally, simulation methods.

Quickest idea:

- For given parameters, draw random terms from the assumed distribution for each individual in a sample.
- Use these draws and parms to construct simulated choices.
- Average these draws across simulated individual to construct a sample-average choice probability.
- Compare these simulated probabilities to the true individual choices in the data (or to the true probabilities in the data) and/or use in an MLE method (but be careful, see below.)
MLE takes the log of the predicted probability. This is a big problem when shares are small – a small error in $\hat{s}_{ij}$ is a large error in $ln(\hat{s}_{ij})$ when shares are small, because the $ln(\cdot)$ function is very steep when evaluated near zero.

With a relatively low dimension (one? three? five?) it is probably better to use modern numeric integration.
Market Level Data

In contrast to the “McFadden” tradition, there is an “IO tradition” that focuses on market-level data and oligopoly, as in the Bresnahan paper. After a bit we will try to put these back together.
The idea is to extend supply and demand empirical models to markets with product differentiation (Berry '94) Berry (1994) and Berry, Levinsohn and Pakes Berry, Levinsohn, and Pakes (1995).) As in the supply and demand literature, want to account for the implications of consumer and firm optimization together with equilibrium pricing.

First application: autos, again, with a focus on “realistic” own- and cross-price elasticities.
Data

Data are at least market-level observations on prices, quantities and characteristics of products. Might also have data on characteristics of purchasing consumers and on firm production (but this is rare.) Consumers modeled in a discrete choice framework. The utility of consumer $i$ for product $j$ depends on observed and unobserved (by us) characteristics, $(y, \nu)$, of the consumer and observed and unobserved (by us) characteristics of the product, $(x, \xi)$. Each firm produces a given set of products whose production cost depends on a vector of cost shifters $w$; presumably, $w$ includes the product characteristics, $x$. The equilibrium is Nash in prices.
Product Choices

To derive demand, we assume that choices are all products in the market and an outside good. The utility of a choice is determined by a parametric form for the interaction between consumer characteristics and product attributes (Lancaster (1971), McFadden, etc.) The demand function is then derived by explicitly aggregating over the choices of consumers with different characteristics.
Endogenous Prices

As in traditional homogeneous goods models, the econometric endogeneity of prices follows from the presence of unobserved characteristics.

In autos, unobserved characteristics include style, dealer quality, tradition. How to explain: prices and sales of similarly sized and powered Ford and BMW; very different sales of very similar Japanese compact cars.

Assuming no unobserved product characteristic also leads to over fitting problem.
Unobserved Characteristics

Unobserved characteristics were usually ignored in the traditional discrete choice literature. Micro studies sometimes add product-specific intercept, but no attempt to say how this would change in a counterfactual. Product intercept is co-linear with $p_j, x_j$.

Solution (from Berry (1994)): Allow for unobserved characteristic, $\xi_j$

\[
\begin{align*}
  u_{ij} & = x_j \beta - \alpha p_j + \xi_j + \epsilon_{ij}, \\
           & = \delta_j + \epsilon_{ij},
\end{align*}
\]

where $\delta_j = x_j \beta - \alpha p_j + \xi_j$. 
Unobserved Product Characteristics – cont

Problem: $\xi$ enters $s_j$ in a highly non-linear fashion and is correlated with (at least) prices. Can solve for $\xi$ given parameters of model. However, one $\xi$ for every observation, need more restrictions. Following traditional demand literature, we assume that $\xi$ is uncorrelated with some “demand shifters”; here these are product characteristics. Could use other restrictions (endogenous characteristics?)
Inverting the Market Share Function

For the true model: \( s_j = s_j(\delta, \theta), j = 1, \ldots, J \). This is \( J \) by 1 equations in \( J \) unknown \( \delta_j \). Note the necessity of normalizing the “outside good” (or some good) to zero.

If possible, invert to find mean levels of utility:

\[
\delta = s^{-1}(s, \theta)
\]  

(12)

Now could estimate demand parameters by IV from, e.g., :

\[
\delta_j = x_j \beta - \alpha p_j + \xi_j,
\]

(13)

Alternatively, can think of solving for \( \xi_j \) directly. Note that the parameters are \( \alpha, \sigma \)!

We will greatly generalize this later.
Instruments, again

- Cost shifters, maybe as Hausman instruments
- Changes in Rival’s $x$’s over time/markets (sub-market?) This is the logic of Bresnahan and the oligopoly supply-side. Where you are in the characteristics space helps to determine your markups.
- Panel Data structures: (e.g. same $\xi$ for different product in different markets, or an AR(1) process with innovations that are uncorrelated with past observables.)

*Intuition*: Elasticities are “identified” by changes in the choice set over time (including changes in price), holding preferences constant (or controlling for changes in preferences from observed data.)
Logit Example

Shares:

\[ s_j(\delta) = \frac{e^{\delta_j}}{1 + \sum_r e^{\delta_r}}, \]  \hspace{1cm} (14)

Share of outside good:

\[ s_0(\delta) = \frac{1}{1 + \sum_r e^{\delta_r}}. \]  \hspace{1cm} (15)

\[ \Rightarrow \ln(s_j) - \ln(s_o) = \delta_j = x_j \beta - \alpha p_j + \xi_j. \]  \hspace{1cm} (16)

Berry (1994) also gives analytic inverse for vertical model, nested logit. More complicated models require a numerical inverse.
Example of Inversion

As in Berry (1994), in the logit case, we can estimate

\[ y_j \equiv \ln(s_j) - \ln(s_o) = \delta_j = x_j \beta - \alpha p_j + \xi_j. \quad (17) \]

by 2SLS or another IV method. Note that \( p_j \) is exogenous – it is likely correlated with the unobserved term \( \xi_j \). Potential instruments include other-firms \( x' \)'s, cost shifters – or else panel data assumptions.

Recall that the mean utility level, \( \delta_j \), determines all behavior including market share and cross-price effects. Markups for price-setting firms vary only with market share.
Table 3 from BLP
Results with Logit Demand and Marginal Cost Pricing
2217 observations)

<table>
<thead>
<tr>
<th>Variable</th>
<th>OLS Demand</th>
<th>IV Logit Demand</th>
<th>OLS ln(price) on w</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-10.068 (0.253)</td>
<td>-9.273 (0.493)</td>
<td>1.882 (0.119)</td>
</tr>
<tr>
<td>HP/Weight *</td>
<td>-0.121 (0.277)</td>
<td>1.965 (0.909)</td>
<td>0.520 (0.035)</td>
</tr>
<tr>
<td>Air</td>
<td>-0.035 (0.073)</td>
<td>1.289 (0.248)</td>
<td>0.680 (0.019)</td>
</tr>
<tr>
<td>MP$</td>
<td>0.263 (0.043)</td>
<td>0.052 (0.086)</td>
<td>-</td>
</tr>
<tr>
<td>MPG*</td>
<td>–</td>
<td>–</td>
<td>-0.471 (0.049)</td>
</tr>
<tr>
<td>Size *</td>
<td>2.341 (0.125)</td>
<td>2.355 (0.247)</td>
<td>0.125 (0.063)</td>
</tr>
<tr>
<td>trend</td>
<td>–</td>
<td>–</td>
<td>0.013 (0.002)</td>
</tr>
<tr>
<td>Price</td>
<td>-0.089 (0.004)</td>
<td>-0.216 (0.123)</td>
<td>-</td>
</tr>
<tr>
<td>No. Inelastic Demands (+/- 2 s.e.'s)</td>
<td>1494 (1429-1617)</td>
<td>22 (7-101)</td>
<td>n.a.</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.387</td>
<td>n.a.</td>
<td>.656</td>
</tr>
</tbody>
</table>
A well-known solution to problems with logit: Interact Product and Consumer Characteristics. Random coefficients logit or probit Hausman and Wise (1978) example:

\[ u_{ij} = x_j \beta_i - \alpha p_j + \epsilon_{ij}. \]  

(18)

More generally,

\[ u_{ij} = \delta_j + \mu(x_j, p_j, \nu_i, \theta) + \epsilon_{ij}. \]

(19)

To get aggregate demand have to compute a complicated integral. Aggregation problem solved by simulation, as in Pakes (1986). On simulation see also Pakes and Pollard (1989) and McFadden (1989).
Automobile “Supply and Demand” in order to do policy counterfactuals: trade policy, environmental policy, merger policy, etc.

Data only at the market level: prices, quantities and market shares. Treat time dimension as different markets (not great: dynamics? correlation?)

Extend Berry ’94 “invert for the unobservable” to random coefficients logit. On the supply side, solve for marginal cost via multi-product Nash pricing assumption.
Details on the Berry-Levinsohn-Pakes (BLP) Utility Functions

“Cobb-Douglas”

\[
U(\nu_i, p_j, x_j, \zeta_j; \theta) = (y_i - p_j)^\alpha G(x_j, \xi_j, \nu_i) e^{\epsilon(i,j)}, \tag{20}
\]

Taking logs

\[
u_{ij} = \alpha \log(y_i - p_j) + x_j \beta_i + \xi_j + \epsilon_{ij}. \tag{21}\]

Let

\[
\beta_{ik} = \bar{\beta}_k + \sigma_k \nu_{ik}, \tag{22}
\]
“Income spent elsewhere” is probably too literal for this market. Berry, Levinsohn, and Pakes (1999) put a log-normal coefficient on price and make other improvements (while looking at trade policy.)
where

$$\delta_j = x_j \bar{\beta} + \xi_j$$

and

$$\mu_{ij} = \alpha \log (y_i - p_j) + \sum_k \sigma_k x_{jk} \nu_{ik} + \epsilon_{ij}$$
“Income” draws are assumed to be log-normal

\[ y_{it} = \exp(\bar{y}_t + \sigma \nu_{iy}), \]  

with parameters calculated from CPS distribution of income. This gives some macro effects in demand (again, not great.)
Calculating the Market Share via Simulation

Condition on \( \nu_i, y_i \) – this is a logit and get closed form. Take \( ns \) draws on \( \nu_i, y_i \) and average over the implied logit shares:

\[
\frac{1}{ns} \sum_{i=1}^{ns} \frac{e^{\bar{\mu}_{ij} + \delta_j}}{\sum_k e^{\bar{\mu}_{ik} + \delta_k}},
\]

where \( \bar{\mu}_{ij} = \alpha \log(y_i - p_j) + \sum_k \sigma_k x_{jk} \nu_{ik} \).

The error in simulated shares enters linearly – not such a problem if shares enter the estimation “linearly” so that the simulation error will average out. (See McFadden and Pakes, 1986 on simulation.) But it is not linear here, simulation error is a big problem (see Berry, Linton, and Pakes (2004)). BLP use “importance sampling” to reduce simulation variance.
BLP Contraction Mapping

BLP then provide an algorithm (a contraction mapping) that solves for $\delta$ given the parameters and a set of simulation draws. Have to hold the simulation draws fixed as the parameters chance (or incorrectly risk that change in objective function is due to change in simulation draws.) Have to account for the simulation variance – the inversion means that the simulation error does not enter linearly. The simulation error gets worse as the shares get small – hard to accurately simulate small shares. See Berry, Linton, and Pakes (2004).
GMM Estimation of Demand-side BLP

Review of the GMM estimation algorithm for the demand-side alone. See also Nevo (2000b).

- Guess a parameter
- Solve for $\delta$ and therefore $\xi$.
- Interact $\xi$ and instruments $z$ – these are the moment conditions $G(\theta)$.
- Calculate an objective function – how far is $G(\theta)$ from zero? $f(\theta) = G'AG$ for some positive definite $A$.
- Guess a new parameter and try to minimize $f$.
- Variance of $\hat{\theta}$ includes variance in data across products and simulation error as well as any sampling variance in the observed market shares.
  (Can simplify the algorithm since $\delta$ in linear in some parameters.)
Oligopoly Supply Side

It is not at all clear that there is enough variation in the choice sets to get precise estimates of the substitution patterns from demand-side data along. In an oligopoly, pricing behavior also contains information on the substitution patterns, via the markups. (This is also Bresnahan’s point about pricing in oligopoly.)
Adding the Supply Side

Adding a supply side allow us to both

- Gain additional information on demand-side substitution and
- Learn marginal costs, which are necessary for many counterfactuals.

The cost is committing to a particular model of oligopoly competition (e.g. Nash in prices). One can do robustness checks on this.
Rosse (1970) suggests an estimation equation from an imperfectly competitive first order condition.

This is a direct generalization of the perfectly competitive model. In perfect competition, the first-order condition for optimal output says to set

\[ p = mc \]

If marginal costs vary with characteristics \( x_t \) and an error \( \omega_t \), then we would have something like a hedonic regression

\[ p_t = x_t \gamma + \omega_t \]

But perfect competition doesn’t make much sense given differentiated products. That is why the hedonic coefficients are involve a projection on both markups and marginal cost.
Rosse 1970

Consider the single-firm monopoly case, with linear demand

\[ q_j = x_j \beta - \alpha p_j + \xi_j. \]

and linear marginal cost

\[ mc_j = w_j \gamma + \lambda q_j + \omega_j. \]

The f.o.c. for optimal price gives us

\[ q_j + p_j \frac{\partial q_j}{\partial p_j} = mc_j \frac{\partial q_j}{\partial p_j} \]

\[ q_j - \alpha p_j = -\alpha mc_j \]

\[ p_j = \frac{q_j}{\alpha} + mc_j \]

\[ p_j = \frac{q_j}{\alpha} + w_j \gamma + \lambda q_j + \omega_j. \]

\[ p_j = q_j \left( \lambda + \frac{1}{\alpha} + w_j \gamma + \omega_j. \right) \]
Firm Behavior with Multi-Product Firms

Marginal Cost:

\[ mc_j = w_j \gamma + \omega_j. \]  \hfill (28)

Might also model \( \ln(mc_j) \), make \( mc \) a function of \( q \), etc.

Profits:

\[ \pi_f = \sum_{j \in J_f} (p_j - mc_j) M s_j(p, x, \xi; \theta), \]  \hfill (29)
First-order Condition:

\[ s_j(p, x, \xi; \theta) + \sum_{r \in J_f} (p_r - mc_r) \frac{\partial s_r(p, x, \xi; \theta)}{\partial p_j} = 0. \]  
\( (30) \)

Given the demand function, it is possible to solve this for the vector of \( mc \)'s and so for markup, \( b_j(p, x, \xi, \theta) \) and for \( \omega_j \).
In particular, write the foc as

\[ s + \Delta(p - mc) = 0, \]
\( (31) \)

where \( \Delta \) has \( \partial s_j / \partial p_j \) on the diagonal, \( \partial s_r / \partial p_j \) on the off-diagonals of jointly owned products and zeros elsewhere. MC is then:

\[ mc = p + \Delta^{-1}s. \]
\( (32) \)
Estimating from the FOC

Given the demand parameters, can think of estimating the equation,

\[ mc_j = p_j - b_j(p, x, \xi, \theta) = w_j \gamma + \omega_j. \]  \hspace{1cm} (33)

Just as in estimating demand, estimates of the parameters \( \gamma \) can be obtained from orthogonality conditions between \( \omega \) and appropriate instruments. Recall that \( b_j \) can be calculated from the demand parameters alone (once again there is a problem of simulation.) Can also estimate supply and demand together. Here, we change the prior algorithm to solve for both \( \omega \) and \( \xi \) and interact these with the instruments to form the moment conditions.
Notes:

- We do not require a unique equilibrium.
- The markup depends on $\xi$, $\omega$ and so is econometrically endogenous.
- Other static equilibria are easy (e.g., qty-setting, collusion) by changing the $\Delta$ matrix. See Berry, Levinsohn, and Pakes (1999) for quantity setting and mixed-price and quantity setting foc's.
Caveats:

- Nash Pricing
- no dynamics
- no production data
- no direct data on consumers
Instruments

Instruments? Again: costs (for prices) and changes in choice set. BLP use supply-side restrictions, some cost instruments and also “BLP Instruments” (I don’t like that name), which are “characteristics of other goods”. This is in the spirit of Bresnahan and maybe should be Bresnahan-BLP instruments.

Berry, Levinsohn, and Pakes (1999) introduce better cost shifters. Intuitively, changes in the choice set identify $\sigma$ and changes in cost identify the coefficient on price. However, changes in the choice set also in principle change markups and can help with the price coefficient, whereas exogenous changes in price (due to costs) can help to trace out substitution.
Optimal Combinations of Instruments

How to approximate the optimal combination of instruments? BLP ’95 tries to use a “flexible” combination of characteristics—basis functions of a symmetric polynomial in own-product, own-firm and rival characteristics.

Linear: own $x$, sum (or mean of) of own-firm $x$, sum (or mean) of rival firm $x$.

These are not great. The sum of “other product” characteristics hardly varies in a market with a large number of firms. The own- and rival-product distinction helps here. We could move to a higher dimensional polynomial, but this creates a problem of too-many instruments and possible over-fitting.
Optimal Instruments

Berry, Levinsohn, and Pakes (1999) use Chamberlain’s result on “optimal” GMM instruments to get a better combination of \( x \)’s. Reynaert and Verboven (2014) show in Monte Carlos and an auto application that this is \textit{much, much} better.
Optimal Instruments

Thinking just of demand. Under some sampling assumptions, the optimal instruments (Chamberlain or Newey (1990)) are

\[ E \left[ \frac{\partial \xi(\theta)}{\partial \theta} \mid z \right] \]

Note that in the classic linear case

\[ \epsilon = y - x \beta \]

the derivative of the error wrt \( \beta \) is just \( x \) and so the optimal instrument would be \( E[x \mid z] \), as expected.

The question is how to implement this in the BLP example.
Implement Optimal Instruments

The Reynaert Verboven method for optimal instruments is to predict price from $z$ (regression on an exchangeable polynomial in $z$). For any guess at $\xi$, say $\xi^k$, we can then predict the “true” market shares for the predicted $p$ and the given $\xi$. This creates a “counterfactual” data set which defines $\xi(\theta)$ and $\frac{\partial \xi(\theta)}{\partial \theta}$ for that data.

To approximate the optimal instruments, we then average over the above calculation for $K$ draws on $\xi^k$ from the estimated distribution of $\xi$ (given our demand estimates.) As in Berry, Levinsohn, and Pakes (1999), one could alternatively use the supply side to get the prediction of $p$, but R & V say this doesn’t work as well.

Could use optimal instruments for demand and supply together, using

$$E \left[ \frac{\partial \xi(\theta)}{\partial \theta} \mid z \right], E \left[ \frac{\partial \omega(\theta)}{\partial \theta} \mid z \right],$$
next table: some semi-price elasticities (related to single-firm markups)
<table>
<thead>
<tr>
<th>Mazda 323</th>
<th>Nissan Sentra</th>
<th>Ford Escort</th>
<th>Chevy Cavalier</th>
<th>Honda Accord</th>
<th>Ford Taurus</th>
<th>Buick Century</th>
<th>Nissan Maxima</th>
<th>Acura Legend</th>
<th>Lincoln LS400</th>
<th>Town Car</th>
<th>Seville</th>
<th>LS400</th>
<th>735i</th>
</tr>
</thead>
<tbody>
<tr>
<td>323</td>
<td>-125.933</td>
<td>1.518</td>
<td>8.954</td>
<td>9.680</td>
<td>2.185</td>
<td>0.852</td>
<td>0.485</td>
<td>0.056</td>
<td>0.009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentra</td>
<td>0.705</td>
<td>-115.319</td>
<td>8.024</td>
<td>8.435</td>
<td>2.473</td>
<td>0.909</td>
<td>0.516</td>
<td>0.093</td>
<td>0.015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Escort</td>
<td>0.713</td>
<td>1.375</td>
<td>-106.497</td>
<td>7.570</td>
<td>2.298</td>
<td>0.708</td>
<td>0.445</td>
<td>0.082</td>
<td>0.015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cavalier</td>
<td>0.754</td>
<td>1.414</td>
<td>7.406</td>
<td>-110.972</td>
<td>2.291</td>
<td>1.083</td>
<td>0.646</td>
<td>0.087</td>
<td>0.015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accord</td>
<td>0.120</td>
<td>0.293</td>
<td>1.590</td>
<td>1.621</td>
<td>-51.637</td>
<td>1.532</td>
<td>0.463</td>
<td>0.310</td>
<td>0.095</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taurus</td>
<td>0.063</td>
<td>0.144</td>
<td>0.653</td>
<td>1.020</td>
<td>2.041</td>
<td>-43.634</td>
<td>0.335</td>
<td>0.245</td>
<td>0.091</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Century</td>
<td>0.099</td>
<td>0.228</td>
<td>1.146</td>
<td>1.700</td>
<td>1.722</td>
<td>0.937</td>
<td>-66.635</td>
<td>0.773</td>
<td>0.152</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maxima</td>
<td>0.013</td>
<td>0.046</td>
<td>0.236</td>
<td>0.256</td>
<td>1.293</td>
<td>0.768</td>
<td>0.866</td>
<td>-35.378</td>
<td>0.271</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legend</td>
<td>0.004</td>
<td>0.014</td>
<td>0.083</td>
<td>0.084</td>
<td>0.736</td>
<td>0.532</td>
<td>0.318</td>
<td>0.506</td>
<td>-21.820</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Town Car</td>
<td>0.002</td>
<td>0.006</td>
<td>0.029</td>
<td>0.046</td>
<td>0.475</td>
<td>0.614</td>
<td>0.210</td>
<td>0.389</td>
<td>0.280</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seville</td>
<td>0.001</td>
<td>0.005</td>
<td>0.026</td>
<td>0.035</td>
<td>0.425</td>
<td>0.420</td>
<td>0.131</td>
<td>0.351</td>
<td>0.296</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS400</td>
<td>0.001</td>
<td>0.003</td>
<td>0.018</td>
<td>0.019</td>
<td>0.302</td>
<td>0.185</td>
<td>0.079</td>
<td>0.280</td>
<td>0.274</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>735i</td>
<td>0.000</td>
<td>0.002</td>
<td>0.009</td>
<td>0.012</td>
<td>0.203</td>
<td>0.176</td>
<td>0.050</td>
<td>0.190</td>
<td>0.223</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Direct application of BLP auto model to:

BLP ’99

Berry, Levinsohn, and Pakes (1999) add cost-side shifters in the form of exchange rates and wages interacted with country of production. They also use the optimal instruments and continue to use both demand and supply to pin down the substitution patterns. (In retrospect, maybe still too many random coefficients in the specification, despite the improvements.)

How to modify foc for quota

\[
\max_{p_{jt}} \sum_{k \in J_t} \left[ p_{kt} M_t s_{kt}(p_t) - C_{kt}(M_t s_{kt}(p_t)) - F_{kt} \right]
\]

s.t.

\[
\sum_{k \in J_t} M_t s_{kt}(p_{kt}) = \bar{Q}_{ft}
\]

FOC:

\[
s_{jt} + \sum_{k \in J_t} \left[ (p_{kt} - m c_{kt} - \lambda_{ft}) \frac{\partial s_{kt}}{\partial p_{jt}} \right] = 0
\]

where \(\lambda_{ft}\) is the shadow value of the quota for firm \(t\). In practice, constrain \(\lambda_{ft}\) to be equal across some firms and estimate as a coeff. on a

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demand-side parameters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Means ((\bar{\beta})'s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-5.901</td>
<td>0.712</td>
</tr>
<tr>
<td>HP/Weight</td>
<td>2.946</td>
<td>0.486</td>
</tr>
<tr>
<td>Size</td>
<td>3.430</td>
<td>0.342</td>
</tr>
<tr>
<td>Air</td>
<td>0.934</td>
<td>0.199</td>
</tr>
<tr>
<td>MP$</td>
<td>0.202</td>
<td>0.084</td>
</tr>
</tbody>
</table>
# Table 5b

<table>
<thead>
<tr>
<th>Standard deviations $(\sigma_\beta \text{'}s)$</th>
<th>Constant</th>
<th>HP/Weight</th>
<th>Size</th>
<th>Air</th>
<th>MP$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.112</td>
<td>0.167</td>
<td>1.392</td>
<td>0.377</td>
<td>0.416</td>
</tr>
<tr>
<td>Term on price $(\alpha)$</td>
<td></td>
<td>$(−p/y)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>44.794</td>
<td>4.541</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Cost-side parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate 1</th>
<th>Estimate 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.035</td>
<td>0.310</td>
</tr>
<tr>
<td>ln(HP/Weight)</td>
<td>0.604</td>
<td>0.063</td>
</tr>
<tr>
<td>ln(Size)</td>
<td>1.291</td>
<td>0.106</td>
</tr>
<tr>
<td>Air</td>
<td>0.484</td>
<td>0.043</td>
</tr>
<tr>
<td>Trend</td>
<td>0.018</td>
<td>0.004</td>
</tr>
<tr>
<td>Japan</td>
<td>3.255</td>
<td>0.667</td>
</tr>
<tr>
<td>Japan*trend</td>
<td>−0.036</td>
<td>0.008</td>
</tr>
<tr>
<td>Euro</td>
<td>3.205</td>
<td>0.525</td>
</tr>
<tr>
<td>Euro*trend</td>
<td>−0.032</td>
<td>0.006</td>
</tr>
<tr>
<td>lag[ln(e-rate)]</td>
<td>0.026</td>
<td>0.024</td>
</tr>
<tr>
<td>ln(wage)</td>
<td>0.356</td>
<td>0.079</td>
</tr>
</tbody>
</table>
## VER dummies

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>VER81</td>
<td>-0.085</td>
<td>0.187</td>
</tr>
<tr>
<td>VER82</td>
<td>-0.022</td>
<td>0.228</td>
</tr>
<tr>
<td>VER83</td>
<td>0.001</td>
<td>0.248</td>
</tr>
<tr>
<td>VER84</td>
<td>0.403</td>
<td>0.245</td>
</tr>
<tr>
<td>VER85</td>
<td>0.361</td>
<td>0.303</td>
</tr>
<tr>
<td>VER86</td>
<td>0.675</td>
<td>0.307</td>
</tr>
<tr>
<td>VER87</td>
<td>1.558</td>
<td>0.353</td>
</tr>
<tr>
<td>VER88</td>
<td>1.490</td>
<td>0.379</td>
</tr>
<tr>
<td>VER89</td>
<td>1.277</td>
<td>0.458</td>
</tr>
<tr>
<td>VER90</td>
<td>1.063</td>
<td>0.469</td>
</tr>
</tbody>
</table>
Computation in BLP

For a introductory description of the BLP algorithm, see Nevo (2000b) and the code on Nevo’s website.

However, most importantly see the caveats and the alternative (in many cases better) computational approach Dub, Fox, and Su (2012). One key point: the contraction that solves for $\delta$ has to have a very tight tolerance (unlike the code on Nevo’s website) or else the “outer loop” search for parameters will behave very badly.

Dub, Fox, and Su (2012) suggest an alternative approach that uses modern optimizers to solve for $\delta$ and $\theta$ simultaneously rather than in a nested loop. In many cases this is faster and in some cases much faster.

Also, probably better to use pure modern numeric integration (or at least deterministic sequences for the simulation), see Skrainka and Judd (2011).
Characteristics Based Models with Micro Data.

“Micro Data” matches consumers to choices, as in the McFadden transportation example, but we continue to worry about price endogeneity and the supply side.

Two data scenarios

1. Good market level data but limited micro data either [a] small sample of purchases or [b] limited cross-tabs of purchases by consumer attributes (or other summaries)

2. A large micro sample that matches consumers’ attributes to their purchases.

In case [1], we can add a few “micro moments” to aggregate model. In case [2] we can estimate the model at the
Petrin on the Minivan

Petrin (2002) is an example that, like the original BLP, makes use of some aggregated data on the consumers who buy a car. In Petrin’s case, this lets him better estimate the demand for a new characteristic.

Idea of the minivan: build a “car” with many seats that isn’t a “truck” by expanding the vertical height of a passenger car. This innovation was immediately popular with families. What is the social value of the innovation in the product space?

Since this is a “new characteristic,” we can only analyze ex post (pre-introduction, there can be no data on demand for the characteristic.)
Petrin’s exercise:

Petrin (2002): estimate a demand and cost side from the observed data, and then recompute equilibrium prices and quantities from a choice set that does not include the minivan. Of course the problem is that the other cars produced might not have been the same had the Minivan not been introduced ... but you have to start somewhere.

Petrin uses pretty much the same specification as BLP, but brings in also micro first-choice moments obtained from the CEX, by simply adding moment restrictions corresponding to purchase probabilities and car characteristics interacted with individual attributes.
Welfare Gain

There are two sources of welfare gain. One is the added consumer surplus of the non-marginal Minivan producers. The other is the fall in prices for most other family cars. Petrin results indicate that the added compensating variation for those who purchased a Minivan; it is about 1/7th the cost of the vehicle. Or roughly a 3% gain in income for about 1% of US Households (roughly one in ten households buy a car every year, and one in ten of those cars are Minivans).
Petrin discuss the role of the logit errors in possibly over-stated the return to new product introduction, particularly for the second, third minivan and so forth.

The consumer data reduce the importance of the $\epsilon$’s and reduce the welfare gain, but it is still a big number.

It is these kind of numbers that get the profession excited about new goods, R&D, and growth; but remember this was a highly successful innovation. Maybe one or two innovations like this have occured in the car industry over the last twenty years, and they do alot of R&D.
Example: Berry, Levinsohn, and Pakes (2004). Even given the methodological improvements in Berry, Levinsohn, and Pakes (1999), it was still not clear that we could estimate substitution parameters from demand alone. Therefore, we looked for information on

- consumer data
- any direct data on substitution?
MicroBLP Data

The micro data we add here is the CAMIP data. General Motors Corporation’s data used for marketing and product quality programs. Data include

- vehicle characteristics, prices, and sales (similar to product level data already in use except of higher quality)
- household characteristics by vehicle purchased (age, income, family size, ... broken down by vehicle purchased)
- second choice vehicles (generated as the reply to the question: “If you did not purchase this vehicle, what vehicle would you purchase?”)

Note however that unlike previous studies, this has only one cross section.
“Micro BLP” utility

Utility includes both random coefficients and interactions of consumer attributes with $x$’s.

$$U_{ij} = \delta_j + \sum_{kr} x_{jk} z_{ir} \beta^o_{rk} + \sum_{kl} x_{jk} \nu_{il} \beta^u_{kl} + \epsilon_{ij},$$  \hspace{1cm} (34)

where

$$\delta_j = \sum_k x_{jk} \lambda_k + \xi_j.$$  \hspace{1cm} (35)
As noted, interactions of consumer characteristics and product characteristics are needed for reasonable cross price elasticities. Now have two such interaction terms.

- Observed consumer characteristics (the $z_i$) and product characteristics (Term is $\Sigma_{kr} x_{jk} z_{ir} \beta^o_{rk}$.)
- Unobserved consumer characteristics (the $\nu_i$) and product characteristics (Term is $\Sigma_{kl} x_{jk} \nu_{il} \beta^u_{kl}$.)
Added information in first choice micro data: matches individuals characteristics to:

- probabilities of purchasing a car.
- match between car characteristics and consumer attributes
- sometimes also second choice vehicles.

(1) and (2) should provide alot of information on the value of the outside alternative, and on the importance of interactions of observed consumer and observed product characteristics. (3) is a little like having a panel, as it changes the choice set; and in a particularly relevant way. This lets us pick up the effects of unobserved individual attributes. (1) and (2) should be able to determine the impact of observed attributes on second choices; if second choices are even closer to first choices than predicted by (1) and (2), then it is because there are unobserved attributes.
Sources of Information.

- random sample of consumers
- choice-based sample; samples purchasers and finds out their characteristics.

Often, data from marketing firms is choice-based, requires correction for “choice-based” sample.
Here are some general points to keep in mind.

- The benefit of having the “first choice” micro data is that it matches observed individual attributes to the products those individuals choose. Consequently, the contribution of that data will be determined by the importance of the observed attribute data in determining individual demands (by the $\beta^o_{rk}$, and the variance of the $z_i$). Formally if $\beta^o_{rk} \sim 0$ the aggregate purchase proportions (i.e. aggregated data on sales and characteristics) are sufficient statistics for the micro first choice data (i.e. we revert to BLP’s problem).
The unobserved individual attributes (the $\nu_i$) differentiates this model from the standard microdata based logit model, and focuses attention on two issues.

1. A substantive issue: what is the “quality” of the observed individual attribute data (are most of the attributes that cause different preference intensities for different product characteristics observed?)

2. A computational issue: if the $\nu_i$ are important the individual choice probabilities have to be simulated, and this complicates the estimation algorithm. Leads to desire to “Test” the null $\beta_{lk}^u = 0$. 
There is little intuition leading us to expect that the first choice data will help us identify the impact of unobserved individual characteristics on preferences when the choice set doesn’t change across observations (as is the case with a single cross-section of data; recall that then if $\beta_{rk}^o = 0$, then the product level data are sufficient statistics for the problem). However the unobserved characteristics do affect the relationship between first and second choices (regardless of the importance of observed attributes).
The micro data enables us to estimate a separate constant term for each choice, our

\[ \delta_j = \sum_k x_{jk} \lambda_k + \xi_j \]

and the vector

\[ \beta \equiv (\beta^o, \beta^u) \]

I.e. \( \beta \) can be estimated without assuming anything about the distribution of the \( \xi_j \).
However, since

$$\delta_j = \sum_k x_{jk} \lambda_k + \xi_j$$

(36)

we need $\lambda$ to compute elasticities w.r.t. the product attributes (including price). Different assumptions on the joint distribution of the observable characteristics and the $\{\xi_j\}$ would allow us to estimate the $\lambda_k$, but the assumptions that would identify the model when we add the micro data are no different then the assumptions that would identify it when only macro data are available.
Once again, *It follows that the basics of the simultaneity problem are the same whether we have micro, or only macro, data.* On the other hand we can get estimates of $\beta$ without $\lambda$, so we can do things like analyze the impact of new goods without solving the simultaneity problem.
Estimation Issues

Assumptions on Unobserved Product Attributes.

- Estimate \((\beta, \delta)\) pointwise.
- Make assumption on the distribution of \(\xi\) (e.g. \(E[\xi|x] = 0\), as in BLP), and estimate \((\beta, \lambda)\) instead of \((\beta, \delta)\).

Tradeoff: We gain efficiency if the assumption is right but lose consistency if it is wrong. Choose to estimate \((\beta, \delta)\) (and then perhaps investigate assumption) because there is so much data.
Computational Choices; Dimension of Search.

- Search over \((\beta, \delta)\). [\(\delta\) alone has over 200 parameters.]
- Two step estimator. Use the contraction mapping in BLP and product level data to solve for \(\delta\) as a function of \((\beta, s^N, P^{ns})\), i.e. for the \(\delta^{N, ns}(\beta)\) that solves

\[
s(\beta, \delta, P^{ns}) = s^N
\]

then search over \(\beta\) to match the model’s predictions to the CAMIP data.

Choose to use \(\delta^{N, ns}(\beta)\).
Summary: Nested Method of Moments Algorithm.

- First step. As in BLP we first find, for any given $\beta$, the value of $\delta$ which makes the aggregate shares derived from the model just fit the product level demands.

- Second step. Substitute $\delta^{N,ns}(\beta)$ for $\delta$, into the model, and compute the model’s predictions for the micro data moments as a function only of $\beta$. Select $\beta$ that minimizes a distance between the moments predicted by the model and the moments in the data.

- Calculation of Moments. The CAMIP sample is “choice based” – it contains $n_j$ random draws on the characteristics of households who purchased car $j$. So we match the data on consumer attributes conditional on a purchase of car $j$ to the model’s predictions. (Use Bayes’ Law).
Estimating the $\lambda$

Recall that

$$\delta_j = \Sigma_k x_{jk} \lambda_k + p_j \lambda_p + \xi_j.$$  

We expect $p_j$ to be correlated with $\xi_j$, so even if the $\delta_j$ were known, we would need to use instruments in order to obtain consistent estimates of the parameters of this equation. Unlike BLP, who had twenty cross-sections with which to estimate this equation, we only have the data for 1993. This suggests a precision problem similar to BLP’s; but this time only for a subset of the parameters of interest (the $\lambda$). We are particularly interested in $\lambda_p$, since it plus the other parameters already estimated determines all own and cross price elasticities.
To add information we can add an assumption on marginal cost and a pricing equation as in BLP or we can find prior information on $\lambda_p$. We used three estimates. One of zero (implicit in usual stuff), one from our own two equation model as in BLP, and one from GM’s estimate of an aggregate elasticity of one. Levels vary markedly with the estimate of $\lambda_p$ but cross sectional pattern do not (see tables 14 and 15). Note that as might have been expected from looking at the second choice data elasticities are particularly low for light trucks. This means markups are higher here, and we might expect the addition of new vechicles in these “niches”.

Steven T. Berry  Yale University, Cowles Four Empirical Models of Differentiated Products  June 18-19, 2015 162 / 341
Prediction Exercise:

Introducing a Mercedes SUV.*

<table>
<thead>
<tr>
<th>Model</th>
<th>Price</th>
<th>Old Share</th>
<th>New Share</th>
<th>New - Old Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEWCAR</td>
<td>33.659</td>
<td>0.0000</td>
<td>0.0762</td>
<td>0.0762</td>
</tr>
</tbody>
</table>

Biggest Declines in Sales.

<table>
<thead>
<tr>
<th>Model</th>
<th>Price</th>
<th>Old Share</th>
<th>New Share</th>
<th>New - Old Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>FORDEXPL</td>
<td>24.2740</td>
<td>0.2518</td>
<td>0.2373</td>
<td>-0.0144</td>
</tr>
<tr>
<td>JPGRNDCH</td>
<td>25.8490</td>
<td>0.1475</td>
<td>0.1376</td>
<td>-0.010</td>
</tr>
<tr>
<td>CHS10BLZ</td>
<td>22.6510</td>
<td>0.1106</td>
<td>0.1071</td>
<td>-0.0036</td>
</tr>
<tr>
<td>TO4RUNNR</td>
<td>25.5480</td>
<td>0.0380</td>
<td>0.0347</td>
<td>-0.0033</td>
</tr>
<tr>
<td>NISSPATH</td>
<td>24.943</td>
<td>0.0397</td>
<td>0.0375</td>
<td>-0.0022</td>
</tr>
<tr>
<td>Luxury cars**</td>
<td>top 4</td>
<td>.1610</td>
<td>.1565</td>
<td>-.0045</td>
</tr>
<tr>
<td>All Vehicles</td>
<td>n.r.</td>
<td>9.711</td>
<td>9.711</td>
<td>.000</td>
</tr>
</tbody>
</table>

* See the text for the characteristics of the new car.
** Cars priced above $30,000.
Discontinue Oldsmobile Division.

<table>
<thead>
<tr>
<th></th>
<th>Old Share</th>
<th>New Share</th>
<th>New-Old Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Oldsmobiles</td>
<td>.237</td>
<td>0</td>
<td>-.237</td>
</tr>
<tr>
<td>All GM</td>
<td>3.126</td>
<td>3.016</td>
<td>-.110</td>
</tr>
<tr>
<td>All Cars</td>
<td>9.711</td>
<td>9.695</td>
<td>-.016</td>
</tr>
</tbody>
</table>

Non-Olds Share Changes.

<table>
<thead>
<tr>
<th></th>
<th>Old Share</th>
<th>New Share</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHELUM</td>
<td>0.1354</td>
<td>0.1548</td>
<td>0.0194</td>
</tr>
<tr>
<td>BUILES</td>
<td>0.1216</td>
<td>0.1336</td>
<td>0.0120</td>
</tr>
<tr>
<td>PONGAM</td>
<td>0.1322</td>
<td>0.1441</td>
<td>0.0119</td>
</tr>
<tr>
<td>HONACC</td>
<td>0.2955</td>
<td>0.3039</td>
<td>0.0084</td>
</tr>
</tbody>
</table>

Olds Share Changes

<table>
<thead>
<tr>
<th></th>
<th>Old Share</th>
<th>New Share</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLDCIE</td>
<td>0.0676</td>
<td>0.00</td>
<td>-0.0676</td>
</tr>
<tr>
<td>OLDSUP</td>
<td>0.0594</td>
<td>0.00</td>
<td>-0.0594</td>
</tr>
<tr>
<td>OLDn88</td>
<td>0.0498</td>
<td>0.00</td>
<td>-0.0498</td>
</tr>
<tr>
<td>OLDACH</td>
<td>0.0333</td>
<td>0.00</td>
<td>-0.0333</td>
</tr>
<tr>
<td>OLDn98</td>
<td>0.0187</td>
<td>0.00</td>
<td>-0.0187</td>
</tr>
<tr>
<td>OLDSBRAV</td>
<td>0.0083</td>
<td>0.00</td>
<td>-0.0083</td>
</tr>
</tbody>
</table>
“MicroBLP” uses only *one market* and so relies on *second choice data* to “vary the choice set” and identify substitution patterns. Still no interesting variation in price.

Cross-market micro data is more useful.

Example just accepted to *Science*: work by Levinsohn and Mobarak on the choice of latrines (or not) in Bangladesh. Price variation and price-variation is exogenous and person-specific from experimental design. But: want to know the role of peer effects: *endogenous market share* is in the micro utility function. Variation in experimental intensity (across villages) is a potential instrument for the endogenous village market share.
Goeree (2008): firms choose both advertising and price. Advertising alters the consumer’s “consideration set” while price influences demand conditional on the consumer being aware of the good. Could have a logit probability of being aware of the good times a BLP style demand conditional on consideration.
3.2. Information Technology

In industries where new product introductions are frequent, the full information assumption is not innocuous. This paper considers a model of random choice sets, where the probability that consumer $i$ purchases product $j$ depends upon the probability she is aware of $j$, the probability she is aware of the other products competing with $j$, and the probability she would buy $j$ given her choice set. Assuming consumers are aware of the outside option with probability 1, the (conditional) probability that consumer $i$ purchases $j$ is

$$s_{ijt} = \sum_{S \in C_j} \prod_{l \in S} \phi_{ilt} \prod_{k \notin S} (1 - \phi_{ikt}) \frac{\exp{\delta_{jt} + \mu_{ijt}}}{y_{it}^\alpha + \sum_{r \in S} \exp{\delta_{rt} + \mu_{ir}}},$$

where $C_j$ is the set of all choice sets that include product $j$. The $\phi_{ijt}$ term is the probability $i$ is informed about $j$. The $y_{it}^\alpha$ term is from the presence of the outside good. The outside sum is over all the choice sets that include product $j$. 
The information technology, $\phi_{ijt}$, describes the effectiveness of advertising at informing consumers about products. Suppressing time notation, it is given by

\begin{equation}
\phi_{ij}(\theta_\phi) = \frac{\exp(\gamma_j + \lambda_{ij})}{1 + \exp(\gamma_j + \lambda_{ij})},
\end{equation}

which is a function of medium advertising, where the $m = 1, \ldots, M$ media are magazines, newspapers, television, and radio. The $m$th element of the $M \times 1$ vector $a_j$ is the number of ads for $j$ in $m$. The components of $\phi_{ij}$ that are the same for all consumers are given by

$$\gamma_j = a'_j(\varphi + \rho a_j + i_m \Psi_j) + \vartheta x_{j}^{\text{age}},$$

where the vectors, $\varphi$ and $\rho$, measure the effectiveness of advertising media at informing consumers. I include fixed effects for those firms that offered a product every quarter (the $\Psi_j$), but do not estimate a fixed effect for each medium, so $i_m$ is a column vector of 1’s. Finally, consumers may be more likely to know a product the longer it has been on the market: this is captured by $\vartheta$, where $x_{j}^{\text{age}}$ is the PC age measured in quarters.
One more “micro” example: Capps, Dranove, and Satterthwaite (2003) (see also Ho (2009).)

Competition between hospitals in different markets. Great consumer data: diagnosis of illness, location, etc. These interact with hospital characteristics in different markets: location of hospital with location of patient, severity of illness with “university hospital”, etc. The hospital may in “in network” or not (much more expensive.)
Supply side is harder: bargaining between hospitals and insurance companies.

The demand estimates intuitively set the value of the hospital to the network (how unhappy are consumers if the hospital is left out of the network) and this should set the “threat point” of the hospital in negotiations.

See also new work by Gautam Gowrisankaran and Town (2015).
10-15 years after the original BLP paper, there were still questions about identification. Is the model “really identified” on market-level data? Given consumer-level data? What is the role of the BLP instruments? Of cost shifters? Other possible instruments?
Nonparametric Identification: Why?

Identification: assuming away problems of sample size, what could we learn in principle if we saw the population distribution of observed data?

1. How should we think about parametric estimates?

   *Are functional form/distributional assumptions approximations for estimation or essential maintained hypotheses?*

2. What can, in principle, be learned from typical observables?

   ▶ what restrictions on the model are important?
   ▶ what types of variation in the data are important?

3. Step toward new nonparametric/semiparametric estimators?
Berry and Haile (2014) on Identification using Market Level Data

Berry and Haile (2014)
Formal identification is interesting even if we are going to use parametric models in the end.

Focus on what instruments are needed / useful
Issues in Identification

Goal: counter-factual policy analysis in an extended “supply and demand” differentiated products framework.

Two classic issues in the Identification of Demand

1. Endogeneity of prices
2. Role of Functional Form Assumptions

In the end, we will need exogenous variation in the data that reveals [1] the effect of prices and [2] the nature of substitution patterns.
Endogeneity of price (and possibly of product characteristics), comes from the presence of market-level unobserved demand factors. In now common empirical IO practice, we model an unobserved product/market level demand unobservable that is possibly correlated with price and that also explains why model doesn’t fit perfectly.
Functional Form Assumptions can often impose answers about cross-price elasticities and markets. CES and pure market-level Logit restrictions are classic – these can “estimate” elasticities from almost no data, but the answer comes almost exclusively from functional form.

We want to allow for more flexible functional forms. In principle, prove results for “non-parametric” case, although in practice probably use “flexible” functional forms.
Index, Inversion and Instruments: Logit Example

Begin with the simplest example: logit demand of consumer $i$ for product $j$ in market $t$

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

Note that we have a logit index in this model:

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

The logit market shares (purchase probabilities) are non-linear in the product index, $\delta$:

$$s_{jt} = \frac{e^{\delta_{jt}}}{1 + \sum_{k} e^{\delta_{kt}}}$$
The index, shares and prices

This is trivially solved – inverted – for the index that includes the unobservable, here using the share of the “outside good” 0:

$$\delta = \ln(s_{jt}) - \ln(s_{0t}).$$

Here, the index is a function of the share vector. We will greatly generalize this. Remembering the definition of the index,

$$\ln(s_{jt}) - \ln(s_{0t}) = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}.$$

This looks like a 2SLS regression, need instruments for price (recall that $x_{jt}$ instruments “for itself.”) To look even more like what we do later, re-write again as:

$$x_{jt} + \xi_{jt} = \frac{1}{\beta} (\ln(s_{jt}) - \ln(s_{0t})) + \frac{\alpha}{\beta} p_{jt}.$$

LHS is a tightly parameterized function of shares and price.
Non-logit substitution

The logit model has ridiculous substitution patterns that depend only on market shares, not $x$. If put log-price in the index, looks like CES with many goods, also ridiculous for actual empirical work.
Nested Logit

Logit “within” and “across” groups of products (“nests”) but at least richer than the pure logit.

At the market level, has an extra parameter on the log “within group share.”

\[
\ln(s_{jt}) - \ln(s_{0t}) = x_{jt}\beta - \alpha p_{jt} + (1 - \lambda) \ln(s_{j/g,t}) + \xi_j.
\]

Useful, but still restrictive. What we the extra IVs for with-group share? Numbers / quality of other goods? To look even more like what we do later, re-write again as:

\[
x_{jt} + \xi_{jt} = \frac{1}{\beta} \left( \ln(s_{jt}) - \ln(s_{0t}) - (1 - \lambda) \ln(s_{j/g,t}) \right) + \frac{\alpha}{\beta} p_{jt}
\]

LHS is a more complicated function of shares and price.
Random Coefficients Logit

McFadden, Hausman and Wise (1978) and Berry, Levinsohn and Pakes (BLP, 1995) add random coefficients to the logit, e.g.

\[ u_{ijt} = \sum_k x_{jkt} \beta_{ik} - \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt} \]

\[ \beta_{ik} = \bar{\beta}_k + \sigma_k \nu_{ik} \]

With \( \nu_{ik} \) standard normal. But what identifies the new parameters \( \rho \)? Instruments that shift substitution patterns? BLP instruments? Are these sufficient? Necessary?
The “BLP Inversion”

It is now nonlinear in the parameters.

Not the usual way to write it, but if one $x$ does not have a random coefficient can write the “BLP inverse” as

$$x_{jt}^{(1)} + \xi_{jt} = \frac{1}{\beta_1} \tilde{\delta}_j \left( s_t, p_t, x^{(2)}, \theta \right)$$

where “mean utility” $\tilde{\delta}$ has to be computed numerically and depends on a small number of parameters.

Again, our method generalizes this to an arbitrary unknown function of shares and prices, which is like a generalized discrete choice model with an index restriction on one of the $x$’s and on the demand unobservable $\xi$. Need instruments for shares and prices. The way the shares enter the function govern the “substitution” patterns, as in the nested logit.
Classic Instruments

- Cost shifters or proxies for cost shifters.
- “Hausman” instruments: prices of goods in other markets (assuming common cost, but not demand, shocks.)
- “BLP” instruments: exogenous characteristics of other goods (variation in rival’s $x$’s changes substitution patterns.)
- “Waldfogel” instruments: conditioning on the demand and demographics of an individual consumer, use as instruments the distribution of attributes of other consumers (which shift available products.) Or, like Ying Fan, use consumer attributes (or market size) in cost-related markets.
In the existing empirical literature, no formal argument about when, for example, the “BLP instruments” are sufficient, alone, to identify demand? With the functional form assumption, “easy” to get “enough” instruments, but is this all functional form?
Demand Notation

- consumer $i$, market $t$, products $j \in \mathcal{J}_t$
- “market” $\iff$ choice set
- “product” = “good” = “choice”
- $x_{jt}$, exogenous observables (product dummies ok)
- $p_{jt} \in \mathbb{R}$, endogenous observables (price)
- $\xi_{jt} \in \mathbb{R}$, market/choice-specific unobservable
- $0 \in \mathcal{J}_t$, “outside good”, $J_t = |\mathcal{J}_t| - 1$
- choice set (market) $\iff \{ \mathcal{J}_t, \{x_{jt}, p_{jt}, \xi_{jt}\}_{j \in \mathcal{J}_t} \}$
Preferences: Random Utility

The random utilities, conditional on prices and on observed and unobserved characteristics, are i.i.d. across consumers and markets, with the joint distribution

\[ F_v(v_{i1t}, \ldots, v_{iJt} | x_t, p_t, \xi_t) \]

where we condition on the “choice set.”

Unless we want cardinal utility or to consider Pareto improvements (where we need to map utility into the “addresses” of consumers) we don’t need to consider the “utility function” and the “distribution of utility parameters (tastes).”

This more general treatment of utility makes the identification results follow much more easily (as opposed to starting with semi-parametric functional forms.)
Restrictions

- All market-specific unobserved heterogeneity is in $\xi_{jt}$. General until we assume it is a scalar.
- Next: add an index assumption.
Restrictions on Preferences

Two real restrictions:

- Assume a scalar unobservable $\xi_{jt}$.
- Quality Index:
  - Partition $x_{jt} = (x_{jt}^{(1)}, x_{jt}^{(2)})$, $x_{jt}^{(1)} \in \mathbb{R}$ and let

$$
\delta_{jt} \equiv x_{jt}^{(1)} \beta + \xi_{jt} \\
= x_{jt}^{(1)} + \xi_{jt}
$$

Henceforth condition on $x^{(2)}$ and drop the superscripts. The model is then completely general in $x^{(2)}$. 
In the index $\delta$, $x^{(1)}$ and $\xi$ are perfect substitutes, a strong assumption. In some cases, can relax to

$$\delta_{jt} = g(x_{jt}) + \xi_{jt},$$

$g(\cdot)$ an unknown strictly monotonic function.

Intuition is that we are going to use the index to capture the effect of an exclusion restriction that $x_{jt}$ and $\xi_{jt}$ only enter the utility of good $j$. 
Market Demand

Integrating over utility in the usual way gives us the market share functions:

\[ s_{jt} = \sigma_j(\delta_t, p_t), \quad j = 0, \ldots, J \]

\( s_{jt} \) is observed data

Note that if we identify the unobservable, we see demand directly. This is all we want in many cases. To recover the distribution of cardinal utility, we can further assume a utility function that is linear in price (otherwise, not necessary.)
Another Example: Continuous Demand

Don’t need discrete choice. With continuous demand, can define shares as

\[ s_{jt} = \frac{q_j(\delta_t, p_t)}{q_0 + \sum_k q_k(\delta_t, p_t)} \]

with fixed \( q_0 > 0 \)

With continuous demand, the index \( \delta_{jt} \) might shift the marginal rate of substitution between good \( j \) and “money spent outside the market.” See Beckert and Blundell (2008).

We will need the goods to be weak substitutes in terms of the share \( s_{jt} \), but it turns out this is consistent with a degree of complementarity in \( q_{jt} \).
Monotonicity

Our assumptions will imply that $s_j(\delta_t, p_t)$ is strictly increasing in $\delta_j$. 
The Problem of Multidimensional Demand

Even outside of discrete choice, in a differentiated products demand system, an unobserved increase in the desirability of one product will effect the demand for all products.

We deal with this by “inverting” the demand system, solving for the unobserved shock to each product and then identify demand via nonparametric “instrumental variables” style restrictions on the unobservables.
Inversion

Berry, Gandhi, and Haile (2013) extend prior proofs to show that the non-parametric inverse

\[ \delta_{jt} = \sigma_j^{-1}(s_t, p_t) \]

is unique.

This result extends / complements older demand / math-econ literature on the uniqueness of demand inverses (can also apply to inverse in \( p \).)
Conditions for Unique Inverse

- **Weak** substitutes across all goods (in shares). \( \delta_{jt} \) up \( (p_{jt} \) down) implies \( s_{kt}, k \neq j, \) down.

- Some strong substitution: in particular **Connected Strict Substitutes**. There must be a path of strict substitution from each good to an “outside good.”

The outside good can be “natural” (purchase no automobile) or else artificially defined to make the condition work. With discrete choice weak substitutes is not a restriction. With continuous goods, can have a high (but restricted) degree of complements and still have substitution in the shares as defined above.
Substitution Patterns for Inversion

We can represent the pattern of substitution with the directed graph of a matrix $\Sigma(\delta)$ whose elements are

$$\Sigma_{j+1,k+1} = \begin{cases} 
1 & \{\text{good } j \text{ substitutes to good } k \text{ at } \delta\} \quad j > 0 \\
0 & j = 0.
\end{cases}$$

The directed graph of $\Sigma(\delta)$ has nodes (vertices) representing each good and a directed edge from node $k$ to node $\ell$ whenever good $k$ substitutes to good $\ell$ at $\delta$. 
Connected Substitutes

Assume:
For all $\delta \in \mathcal{X}^*$, the directed graph of $\Sigma(\delta)$ has, from every node $k \neq 0$, a directed path to node 0.

The condition requires that for any distinct products $j$ and $j'$, there be a “path of substitution”, possibly indirect, from $j$ to $j'$

An implication (given substitution and monotonicity) is that if all the $\delta$'s in set $K$ strictly increase, there is some good $k \notin K$ (maybe good 0) whose share declines.

“All the goods are in the same market.”
Figure 1: Directed graphs of $\Sigma(x)$ for $x \in \tilde{X}^*$ ($x$ equals minus price) in some standard models of differentiated products. Panel a: multinomial logit, multinomial probit, mixed logit, etc.; Panel b: models of pure vertical differentiation, (e.g., Mussa and Rosen (1978), Bresnahan (1981b), etc.); Panel c: Salop (1979) with random utility for the outside good; Panel d: Rochet and Stole (2002); Panel e: independent goods with either an outside good or an artificial good 0.
Identification via Instruments

Once we know that the inversion is unique, in the market data case we can write:

\[ x_{jt} + \xi_{jt} = \sigma_j^{-1}(s_t, p_t) \]

Estimating the inverse is as good as estimating the market share (demand) function. And now its linear in the errors, so can use existing and extended nonparametric IV literature.

To estimate, put restrictions of the form: unobservables are [a] mean independent or [b] fully independent
Identification via Instruments
cont

Our equation

$$x_{jt} + \xi_{jt} = \sigma_{jt}^{-1}(s_t, p_t).$$

“almost” has the form of a classic non-parametric IV regression with a linear error, as in Newey & Powell (2003)

$$y_i = \Gamma(x_i) + \epsilon_i$$

Although the endogenous and exogenous variables enter differently, we can easily repeat Newey-Powell argument using our equation. Instruments have to be “rich enough” in a precise way. But: what are the instruments?
Instruments in the Market Case

In our model

\[ x_{jt} + \xi_{jt} = \sigma_j^{-1}(s_t, p_t). \]

There are 2\(J\) endogenous variables on the RHS: the prices and the shares. As in the nested logit, the shares are capturing the “substitution patterns.”

The \(J\) \(x_{jt}\)’s (“BLP” instruments) are available and necessary; assume these are mean independent of \(\xi_{jt}\). But still need \(J\) more instruments – cost shifters or Hausman instruments?

One possibility: variation in the “BLP instruments” reveals the role of shares (substitution) while product-specific cost-side instruments reveal the role of price.

In this fully non-parametric context, need a lot of instruments! Maybe quite reasonable if dimension of choice set is small (voting?)
How to Reduce the Number of Needed IVs?

1. **More functional form restrictions.**

2. **More Data** Consumer-level data (matching choices to consumer attributes, holding the unobserved product characteristics fixed) can identify substitution patterns without instruments, using some restrictions on utility.
Can use the fully parametric models from the beginning, with a finite-dimensional parameter:

\[ x_{jt} + \xi_{jt} = \sigma^{-1}_j(s_t, p_t, \theta). \]

Or impose semi-parametric restrictions. Putting \( p \) into the index gives

\[ x_{jt} \beta - p_{jt} + \xi_{jt} = \sigma^{-1}_j(s_t), \]

which can be re-written as a problem of identifying the price-inverse of demand

\[ p_{jt} = \sigma^{-1}_j(s_t) + x_{jt} \beta + \xi_{jt}. \]

With no interactive “brand \( j \) dummies” can impose symmetry and exchangeability

\[ p_{jt} = \sigma^{-1}(s_j, s_{-j}) + x_{jt} \beta + \xi_{jt} \]

giving the one equation across all products (can use within-market variation.)
Example: Analog of Nested Logit
Inverse is exchangeable within nest.

Example: Local Competition
The assumption that competition is “local” can reduce the number of product shares and prices that enter the inverse. Extreme case is a “generalized vertical model” with products ordered by price to give

\[ x_{jt} + \xi_{jt} = \sigma^{-1}(s_{j-1}, t, s_j, t, s_{j+1}, t, p_{j-1}, t, p_j, t, p_{j+1}, t) \]

regardless of the total number of products.
Different Kinds of Functional Form Restrictions

1. Parametric on the utility function + distribution of “tastes” (McFadden, BLP – traditional.)
2. Semi-parametric on the index or the properties of $\sigma^{-1}$ (exchangeability)
3. directly parameter on the inverse function itself. Have to obey monotonicity and invertiability assumptions
Example of (2) + (3): generalized non-nested groups

\[ x_{jt} \beta - \alpha p_{jt} + \xi_{jt} = \]

\[ \ln(s_{jt}) - \theta_0 \ln(s_0) + \theta_1 \ln(\bar{s}^1) + \theta_2 \ln(\bar{s}^2) + \theta_3 \ln(\bar{s}^3) \]

where \( \bar{s}^1 \) is the sum of shares of the products in group 1, etc. If the \( \theta \)'s are all positive, the monotonicity assumptions are satisfied. Don’t know in advance what random utility model generates this, but with support conditions could solve for the random utilities afterwards.
Micro Data

See Berry and Haile (2010) (revision in process).

Key data:
- $y_{it}$: indicator for the choice of consumer $i$ in market $t$.
- $z_{ijt}$, involves an individual-product interaction, e.g.
  - distance to hospital, school, retailer
  - family size $\times$ car size
  - household exposure to product advertising
  - consumer / plan specific predicted Rx drug plan cost
Utility Restrictions with Micro Data

We now assume an index in the consumer / product attribute:

$$\lambda_{ijt} = g(z_{ijt}^1) + \xi_{jt}$$

for $g(\cdot)$ unknown and increasing. Similar to before, condition on $z^{(2)}$ (perfectly general) and drop the superscripts on $z$.

Gives product choice probabilities of

$$s_j = \sigma_j(\lambda_{ijt}, x_t, p_t)$$

Again, don’t have to specify the functional form of utility and the distribution of consumer tastes. Again, quasi-linearity will allow us to uncover the distribution of utility.
Observed Consumer Purchase Probabilities

In each market \( t \) we observe the purchase probabilities associated with different values of \( z \). Let the vector \( \bar{z}^t(s) \) be the \( z \) that generates a purchase prob. vector \( s \) in market \( t \). Our inversion result makes sure this is unique.

Common Choice Probability We assume that the support of \( z \) is sufficiently large (relative to the support of \( \xi \)) that (at least conditional on a price \( p \)) in every market there is a consumer type \( z \) who purchases the good with probability \( s^0 \).
Nonparametric Identification

Identification: Micro Data

Inversion in the Micro Case

For a given purchase probability, $s$, that is observed in a market we can once again \textit{invert} to find the index that explains it:

$$\bar{\lambda}_j^t(s^0) = \sigma_j^{-1}(s^0, x_t, p_t)$$

Substituting in for the index:

$$g(\bar{\lambda}_j^t(s^0)) + \xi_{jt} = \sigma_j^{-1}(s^0, x_t, p_t)$$

Have to choose an $s^0$ such that the support of $z$ is big enough to find $\bar{z}^{rt}$ in every market.

Turns out, we can learn the function $g(\cdot)$ from local variation in $s$ \textit{within markets} (the unobservable drops out!) and then use (e.g.) Newey-Powell across markets to uncover the rest.
Given some assumptions, the derivative of $\tilde{z}_{jt}(s^0)$ exists via the implicit function theorem and is given by

$$\frac{\partial z_{jt}(s^0)}{\partial s_j} = \frac{\partial \sigma_j^{-1}(s^0, p_t)}{\partial s_j} \frac{\partial g_j(z_{jt}(s))}{\partial z_j}.$$  \hspace{1cm} (37)

Further, we can observe this derivative by varying $s$ about $s^0$.

These derivatives are “observables” that we use to identify the function $g(\cdot)$.
\[ g(\bar{z}_j^t(s^0)) + \xi_{jt} = \sigma_j^{-1}(s^0, p_t) \]

Given differentiability,

\[ \frac{\partial g(\bar{z}_j(s^0))}{\partial z} \frac{\partial \bar{z}_j(s^0)}{\partial s_j} = \frac{\partial \sigma_j^{-1}(s^0, p_t)}{\partial x_j} \]

Note that \( \xi \) drops out.
For some market $t$ (say $t = 1$), with price $p$ we can WLOG choose $\tilde{z}_{t=1}(s^0)$ as a value at which we normalize $\partial g_j(\hat{z}_j)/\partial z_j = 1$. For each $j$, we can then identify the derivative

$$\frac{\partial \sigma^{-1}_j(s^0, p_{t=1})}{\partial q_j} = \frac{\partial g_j(\hat{z}_j)}{\partial z_j} \frac{\partial \tilde{z}_{j,t=1}(s^0)}{\partial q_j} = \frac{\partial \tilde{z}_{j,t=1}(s^0)}{\partial q_j}. \quad (38)$$

This gives us the derivatives of $\sigma^{-1}$ at a given $(s_0, p)$. 
Now define the support of $\tilde{z}_t(s^0)$ conditional on a price vector $p = p_t$ as $\tilde{Z}(p_t)$. It follows from (38) that for each value of $z \in \tilde{Z}(p_{t=1})$ we can now identify

$$\frac{\partial g_j(z)}{\partial z_j} = \frac{\partial \sigma_j^{-1}(q, p_{t=1})/\partial q_j}{\partial z_j(q)/\partial q_j}$$

(39)

This gives us the derivatives of $g$ over some region.
What we need next is a series of *cumulatively overlapping* sets

\[ \tilde{Z}(p_1), \tilde{Z}(p_2), \tilde{Z}(p_3), \ldots \]

whose union is \( \bigcup_{p \in \mathcal{P}} \tilde{Z}(p) \).
Knowing $g(\cdot)$, we now have a non-parametric IV estimating equation

$$g \left( \bar{z}^t_j(s^0) \right) + \xi_{jt} = \sigma_j^{-1}(s^0, x_t, p_t)$$

running across markets. It’s important that $s^0$ is fixed across markets, so we don’t need instruments for the substitution patterns, we learn these from the micro data!
Instruments for $p$ include

- Cost shifters, including Hausman IVs.
- “Waldfogel” Instruments (features of the distribution of $z$.) Works because we have ruled out spill-overs or sorting, so that conditioning on a given consumer’s $z_{ijt}$ is all that matters for consumer-level demand. But aggregate demographics (income, etc.) matter for price (and later, for entry, location and variety).
Even Stronger Micro Results

Stronger conditions give even stronger results on micro data, see again Berry and Haile (2010).

Assume (roughly)

- Utility is linear in $z_{ijt}$
- and $z_{ijt}$ has full support.

Then we can uncover the *marginal* distribution of utility for one good, which is shifted by only one scalar endogenous variable, $p_{jt}$. One cost shifter, or else one “BLP instrument” alone, are is enough!
In summary, then, micro data can greatly restrict the number of required instruments by using variation in consumer attributes to “trace out” substitution patterns, leaving us “only” to instrument for the effect of prices.

If price enters in a simple way, or else under the strongest of conditions on micro utility and the support of attributes, we can identify the model with as few as one instrument.
Supply Side

Can estimate classic oligopoly supply side either

1. by imposing a static equilibrium assumptions, backing out marginal costs and then estimating a marginal cost function, or

2. with stronger instruments, uncovering the marginal cost unobservable before imposing the oligopoly supply side and then

3. can test a variety of equilibrium assumptions

For a full treatment of supply side identification, see Berry and Haile (2014).
Oligopoly MC

The f.o.c.’s of “all” multiproduct static oligopoly models can be inverted to solve for marginal costs

\[ mc_{jt} = \psi_j (s_t, M_t, D_t (s_t, p_t), p_t) . \]

where the \( \psi_j \) function varies by model. Everything on the RHS is known once demand is identified and the oligopoly assumption is imposed. Putting a classic error structure on \( mc \), we can identify it from

\[ mc_{jt} = \bar{c}_j (Q_{jt}, w_{jt}, ) + \omega_{jt} \]

using “classic” nonparametric IV methods. There is only one endogenous variable (quantity), but an abundance of instruments so the oligopoly assumption is testable.
Alternative Approach to $MC$

- Don’t specify the oligopoly model, but assume that an unknown $\psi_j$ solves for $MC$
- Put an index assumption on $MC$: $\kappa_{jt} \equiv w_{jt}^{(1)} \gamma_j + \omega_{jt}$.
- Can identify the cost index without knowing the oligopoly model. This puts stronger requirements on the IVs (have to be truly excluded from demand) but permits a particularly strong (pointwise) test of any oligopoly assumption.
This figure shows the “Bresnahan” intuition, but there are many possible changes (in the true model) that should keep qty and the cost index fixed while changing the appropriate oligopoly analog of marginal revenue. In a false null, these changes will incorrectly predict changes in quantity.
Simultaneous Identification of S & D

The series of papers, building on earlier work by Matzkin, also discusses how to constructively prove identification by change of variable methods that use the entire system of demand and supply (oligopoly first-order conditions) together.

This approaches strengths the IV assumption to independence between errors and instruments.

The results show a trade-off between [i] support conditions on the exogenous data and [ii] quality shape restrictions on the density of the unobservables.
Practical Advice

▶ Intuition for identification is largely correct: need exogenous changes in “choice set” (characteristics of rival’s etc.) to identify substitution patterns and cost/markup shifters to deal with price endogeneity.

▶ The most general nonparametric approach would be possible only if the number of choices is very small and the number of markets is large. Otherwise, functional form assumptions will play their usual roles: [i] to smooth the data and [ii] to reduce dimensionality.

▶ Functional form assumptions might be made as usual on the utility function, or else on the inverse market share function directly.

▶ Alternatively, micro consumer data can identify substitution patterns without instruments, leaving only the problem of instrumenting for price.
Conclusion

Our interpretation: mostly positive overall message

- with limited structure, identification holds under standard IV conditions
- key requirement: adequate instruments. Nature and number of required of instruments depends on assumptions and on availability of micro data.
*Before*: some doubts in profession about identification, even for semi-parametric BLP model

*After*: even less restrictive models are NP identified, under same conditions required for elementary models (e.g., IV regression). More clarity as to what sources of variation (what instruments) are required under different conditions.
Product Variety

In differentiated products markets, competition in the space of horizontal and vertical quality may be as important as competition in price. Understanding competition in quality and in product characteristics is important to

- the welfare analysis of markets,
- anti-trust policy,
- marketing,
- etc.

For example, in anti-trust analysis the U.S. Dept. of Justice focuses on changes in price and not much on changes in product characteristics and quality. In many markets, this is quite odd.
There are not typically good “welfare theorems” about endogenous variety and location in an oligopoly model. We can have too much or too little enter, firms can locate too close to each other or else too far. Firms may “waste” resources in entry or in building better products, to the degree that they are trying to shift market share rather generate new consumer welfare.
Endogenous Product Characteristics

Models often involve either

1. Entry into a discrete space of product characteristics: “Entry models” (later), or

2. Conditional on Entry, continuous product choice via a first-order condition.
Continuous Characteristics

With continuous characteristics, we can think of modeling the choice of observed characteristics via a Nash equilibrium first-order condition for each observed characteristic. Need sufficient instruments to do this!

Rosse (1970) modeled monopoly newspapers as choosing number of pages (size) in addition to ad quantity and subscription price. All his functional forms were linear, making estimation into a classic linear simultaneous equations problem.
Example: Fan (2013) Newspapers Mergers

Fan (2013) newspaper competition and mergers.

Newspapers are a straightforward multi-sided market. Charge price to both subscribers and advertisers. Newspapers also choose “quality” as well as amount and type of content (local vs. national news.)

In a merger, changes to characteristics may be as important as changes in price.
Endogenous Product Characteristics

Example: Fan (2013) Newspapers Mergers

Fan, cont

Fan (2013) considers overlapping county-level markets for newspapers. She also emphasizes the endogeneity of both price and (non-political) newspaper characteristics. Depending on the county, $c$, readers face different choice sets that include various suburban papers within larger metropolitan regions. The utility function of reader $i$ in county $c$ for newspaper $j$ is

$$u_{ijc} = x_j \beta + y_{jc} \phi + z_c \phi - \alpha p_j + \xi_{jc} + \epsilon_{ijc}$$

where $p_j$ is the price of the newspaper and $x_j$ is a vector of endogenous characteristics (news quality, local news ratio and news content variety). The vector $y_{jc}$ contains within county newspaper characteristics assumed to be exogenous (e.g. whether the headquarters city is in the county) and $z_c$ is a vector of county demographics. The term $x_{ijc}$ again captures unobserved tastes for the newspaper in a given county. Once again, the vector of county-level market shares can be inverted to obtain the mean utility terms.
Fan, cont

There are now four endogenous variables, price and the vector of endogenous characteristics. It is likely that these are correlated with the unobserved taste. For example, the price is likely to be higher when the newspaper is unobservably more popular. This makes the IV problem more difficult. Broadly speaking, Fan also makes use of modified Walfogel style instruments, interacted with exogenous rival characteristics (BLP instruments). In her case, she uses consumer demographics in the other counties served by a rival newspaper. For example, the headquarters county is taken as exogenous, so that the demographic levels in that county are available as instruments.
Fan, cont

In Fan (2013), the Waldfogel style instruments include other counties education level, median income, median age and urbanization. Fan (2013) considers extending the model of to include random coefficients on some of the characteristics. As noted, random coefficients allow for richer substitution patterns than simple logit models. For example, if some consumers have a larger than average taste for local news, they will likely substitute from one locally focused paper to another one, whereas the pure logit generates substitution patterns that vary only with demographic-specific market shares. It not clear whether Fan has adequate instruments for this purpose, and so she models only one random coefficient (on local news).
Fan, advertising

Fan also models advertising demand, with similar instruments, giving us an advertising quantity equation that depends on the newspaper subscription quantity and on the price of advertising, $r_j$. 
Fan-like Profits and Costs

Simplifying, profits for a single-country paper are then

\[ p_t q(p_t, x_t, z_t, y_t, \xi_t) + r_t a(r_t, q_t, \eta_t) - C(q_t, x_t, y_t, \omega_t) \]

with

\[ C = \bar{C}(q_t, a_t, y_t, x_t, \theta) + \omega^q q_t + \omega^a a_t + \omega^y y_t \]

It is very helpful that the errors \( \omega \) are linear shocks to incremental costs.
Endogenous quality via a first-order condition

We can easily write focs for the prices, similar to the ones we have discussed before. If quality is chosen simultaneously to price, the foc for quality is

$$\left[ p_t - \left( \frac{\partial \bar{C}}{\partial q_t} + \omega^q \right) \right] \frac{\partial D^q}{\partial y_t} + \left[ r_t - \left( \frac{\partial \bar{C}}{\partial a_t} + \omega^a \right) \right] \frac{\partial D^q}{\partial y_t} \frac{\partial D^a}{\partial q_t} - \left( \frac{\partial \bar{C}}{\partial y_t} + \omega^y \right)$$

This is linear in the $\omega$’s, as are the focs for $q$ and $r$. If these three focs are invertible to solve for the three cost-shocks $\omega$’s, then we can implement a supply-side GMM procedure, with the unobservable cost shocks mean independent of instruments. (Fan actually has quality chosen before price).
Discrete Entry into Differentiated Locations

We can apply techniques from the IO entry literature to generate endogenous variety via entry into a discrete product space. In an oligopoly equilibrium, we may still need exogenous variation that shifts my rival’s location, which affects my entry decision.
Some Models

- Bresnahan and Reiss (1991b) looked at symmetric entry, ex-post differentiation.
- Berry (1992) and Ciliberto and Tamer (2009) consider models where the differentiation is ex ante, prior to entry.
- Reiss and Spiller (1989) and Berry and Waldfogel (1999) estimated variable profits outside the entry model,
- Mazzeo (2002) considered discrete product segments (“quality”) and ex-post differentiation (ordered models), needs strong assumptions on order to get unique equil.
- Seim (2006) uses private info
- Jia (2008) adds network effects in geographic entry
- Manski (many papers) – use incomplete models, maybe get bounds.
Intro to Bresnahan and Reiss

Bresnahan and Reiss (1991b), Bresnahan and Reiss (1988), Bresnahan and Reiss (1991a)

Look at retail and professional firms in small isolated markets.

Profits in market $i$ are

$$\pi(N_i) = M_i v(N_i, x_i, \theta) - F_i$$

with $M_i$ being market size, $N_i$, the number of firms, $x_i$ are firm characteristics and $F_i$ are fixed costs.

Nash Equilibrium implies that

$$\pi(N_i) > 0 > \pi(N_i + 1).$$

or

$$M_i v(N_i, x_i, \theta) > F_C > M_i V(N_i + 1, x_i, \theta).$$
Estimation by Ordered Probit

Given a distributional assumption on $FC$, the Nash Equilibrium condition

$$M_i \nu(N_i, x_i, \theta) > FC_i > M_i \nu(N_i + 1, x_i, \theta).$$

generates an ordered probit. If $FC \sim \Phi(\cdot)$ then the likelihood of $N$ firms is

$$\Phi(M_i \nu(N_i, x_i, \theta)) - \Phi(M_i \nu(N_i + 1, x_i, \theta)).$$

and one can estimate $\theta$ by MLE.
Review: The Monopoly Entry problem

Here we review the results of Matzkin (1992) and others, using the potential monopoly entry example. Profits of an entering firm are:

\[ \pi(x_i, F_i) = v(x_i) - F_i, \]

where \( v \) is the deterministic variable profit and \( F \) the random fixed cost. In a cross-section of markets, entry occurs when \( v(x_i) > F_i \). We observe the entry probabilities \( p(x_i) \).

An immediate problem is that any monotonic transformation of both \( v \) and \( F \) results in the same entry probabilities. How bad a problem is this?
Non-Robustness to Monotonic Transformations

How bad is the problem? For many issues, not bad at all. Assume $F$ is i.i.d. Then $p(x_i) = \Phi(v(x_i))$ is one such monotonic transformation of $v$, and it reveals

- $\partial p / \partial x$
- The sign of the effect of an $x$ on $v$
- The relative effects:

$$\frac{\partial v / \partial x^1}{\partial v / \partial x^2}$$

This is the kind of problem in, e.g. Berry ’92. (What is the sign and relative magnitude of “airport presence” in entry and profits?)
Qualitative Shape Restrictions

But what if we want to know the full shape of $V$? Matzkin ’92 suggests qualitative shape restrictions, preferably derived from theory, together with an i.i.d. assumption on $F$.

E.g. assume constant marginal costs, then for many models variable profit is proportional to population, $v(x) = z_i \bar{v}(x_i)$ where $z$ is population.

Sketch of Matzkin’s proof: for some $x'$ normalize units so that $\bar{v}(x') = 1$. Then $p(z, x') = \Phi(z)$, which reveals the distribution of $F$, $\Phi$, and from this get the other values of $v$. Done!

(Matzkin considers broader class of $v$’s that are h.d.1 in some subset of $x$.)

Steven T. Berry  Yale University, Cowles Four Empirical Models of Differentiated Products  June 18-19, 2015  245 / 341
Weaker Assumptions?

Following Manski and others, conditional quantile (median) restrictions on the distribution of $F$ can also reveal $v(x)$ up to units, although not the full distribution of $F$.
But in the Monopoly case, not sure why we even care about the full shape of $v$.
One issue: “nature of competition”.
Bresnahan and Reiss, Ordered Entry

The B & R model is a simple extension of the Monopoly model with identical potential entrants. Variable profits decline in the number of firms, $y$: $v_y(x)$.

With an i.i.d. $F$, we have

\[
\begin{align*}
\Pr(y = 0|x) &= 1 - \Phi(v_1(x)) \\
\Pr(y = 1|x) &= \Phi(v_1(x)) - \Phi(v_2(x)) \\
\Pr(y = 2|x) &= \Phi(v_2(x)) - \Phi(v_3(x)) \\
&\vdots
\end{align*}
\]

B & R estimate by fully parametric MLE and ask: how fast does $v$ decline in $y$?
The Economic Question in B & R

What is the value of $\frac{v_2(x)}{v_1(x)}$?

Think of benchmark models:

1. Fixed Prices: $\frac{v_2(x)}{v_1(x)} = 1/2$.

2. Cournot Competition: $\frac{v_2(x)}{v_1(x)} \in (0, 1/2)$?

3. Homogeneous Goods Bertrand: $\frac{v_2(x)}{v_1(x)} = 0$?

Can think of similar ratios for other $y$.

**But these ratios are not robust to monotonic transformations, and so the economic parameter of interest is not non-parametrically identified in B & R without Matzkin-style shape restrictions.**

In fact, there is a monotonic transformation that sets the ratio to anything between 0 and 1.
B & R use the Shape Restriction in Population

We can write the $B&R$ model as series of threshold-crossing models:

$$
\Pr(y \geq 1| x) = \Phi(v_1(x))
$$

$$
\Pr(y \geq n| x) = \Phi(v_n(x))
$$

and so restricting variable profits to be proportional to population will identify $v_y$ (and indeed over-identify $\Phi$, so that we could allow $\Phi_y(F)$ as with perhaps upward-sloping supply of the fixed asset.)
Identification Lessons from the B & R Model

Without qualitative shape restrictions, the object of interest (the “nature of competition”) cannot be identified, but with one natural (though restrictive) shape assumption, the nature of competition is fully identified. Here, the binary threshold crossing literature is enough, but it will not enough be in more complicated models.
Additional Data

Add $p$ and $q$ to the data set, estimate variable profit from this. All that is left to the entry model is fixed cost.
Using price and qty data in entry model

Example: Berry and Waldfogel (1999) on radio entry. We will consider a (2014) version of this model, focusing on entry into horizontal and vertical entry locations.
Discrete Types of Firms

Mazzeo (2002) Mazzeo considers the model without heterogeneity in fixed costs, but with different types of firms: 
Profits for any firm choosing quality level \( t = (1, \ldots, T) \) in market \( m \) are assumed to be

\[
\pi_{tm} = x_m \beta_t + g(N_1, \ldots, N_T, \theta_t) + \epsilon_{tm}.
\]

where \( \bar{N} = N_1, \ldots, N_T \) is the vector of the number of firms of each type. The parameters (specific to each type) are \( \beta_t \) on the market level variables and \( \theta_t \), which parameterizes the effect of own-type and other-type competition.

Now existence and equilibrium are even worse, but ad-hoc assumptions on the order of entry help.
Application is to motels at highway intersections (1-star, 2-star, etc.)
Seim

Seim (2006) introduces asymmetric information into an econometric models of potential entrants’ location decisions. Specifically, Seim models a set of $N$ potential entrants deciding in which one, if any, of $L$ locations they will locate. In Seim’s application, the potential entrants are video rental stores and the locations are Census tracts within a town.

If potential entrant $i$ enters location $l$, it earns

$$\pi_{il}(n_l, x_l) = x_l \beta + \theta_{il} \sum_{j \neq i} D_{jl} + \sum_{h \neq l} \theta_{lh} \sum_{k \neq i} D_{kh} + \nu_{il}$$

(40)

Each store treats the other stores’ $D_{jh}$ as random variables when computing the expected number of rival stores in each location $h$. 
By symmetry, expected prob of rival entry as $p_h = E(D_{kh})$ and the N of expected rivals is $(N - 1)p_h$.

Note convenient linearity of $\pi$ in $D_{kh}$

$$E_D \pi_{il}(\bar{n}, x_i) = x_i \beta + \theta_{il} (N - 1)p_l +$$

$$+ \sum_{j \neq l} \theta_{lj} (N - 1)p_j + \nu_{il}$$

$$= \bar{\pi}_l + \nu_{il}.$$
To calculate the Nash equilibrium entry probabilities $p_1, \ldots, p_L$ we must solve the nonlinear system of equations

\begin{align}
    p_0 &= \frac{1}{1 + \sum_{l=1}^{L} \exp(\bar{\pi}_l)} \quad (42) \\
    p_1 &= \frac{\exp(\bar{\pi}_1)}{1 + \sum_{l=1}^{L} \exp(\bar{\pi}_l)} \quad (43) \\
    \vdots & \vdots \vdots \quad (44) \\
    p_L &= \frac{\exp(\bar{\pi}_L)}{1 + \sum_{l=1}^{L} \exp(\bar{\pi}_l)} \quad (45)
\end{align}
Seim argues that for fixed $N$, an solution to this equilibrium system exists and is typically unique, although as the relative variance of the $\nu$’s declines, the problem approaches the discrete problem and it seems that the non-uniqueness problem faced in perfect information simultaneous-move could reoccur. The single location model discussion above illustrates how and why this could happen.
Using price and qty data in entry model

Example: Berry and Waldfogel (1999) and Berry, Eizenberg and Waldfogel (2015). Here, we do not try to learn about variable profits, but rather use price and quantity data to learn variable profits and then all that is left are the values of fixed cost.

Berry, Eizenberg and Waldfogel (2015) consider entry into a two dimensional product location space of horizontal and vertical differentiation.
Berry, Eizenberg and Waldfogel

In Berry, Eizenberg and Waldfogel (2014), we present a very simple model of entry into a discrete product space that allows for point estimates of the parameters of variable profits and bounds on fixed costs.

Applying this model to the Radio Industry, we consider optimal product variety in terms of the number of stations in different radio formats ("rock", "country", etc.)

Extensions include: vertical quality, joint ownership, merger analysis.
Background on Radio

- There is a long theoretical literature on the inefficiency of free entry into oligopolistic markets. New firms “steal business” from existing firms: a negative externality. Lower prices for existing consumers and the intro of new varieties create an offsetting positive externality.
- Excessive entry into radio industry has often been suggested.
Berry and Waldfogel, 1999

They use new data and simple methods to estimate the extent of and welfare loss from excess entry in radio broadcasting.

**Results from BW ’99**

- First, look only at market participants: broadcasters advertisers. Welfare loss from free entry, as opposed to the socially optimum $N$, is 40% of industry revenue. A big number?

- There is still the positive externality to listeners. If listeners value an hour of listening at about 15 cents an hour, then welfare loss to market participants would be just offset by external benefit to listeners.

But they had to assume symmetric stations, no differentiation by format, etc.
Variety and Multiple Equilibria

- Can easily introduce variety into the post-entry variable profits model (e.g. BLP, nested logit, etc.), although “product characteristics” can now be endogenous.
- BUT: often lose unique equilibrium
- Example: for 2 varieties ($N_1, N_2$), both (2,1) and (1,2) might be equilibria.
- This is why Berry & Waldfogel assumed symmetry: otherwise can’t estimate via MLE.
Estimation with Multiple Equilibria

- Here, we use a simple extension of the “semi-parametric” Bresnahan and Reiss bounds, avoid estimating the distribution of $F$ altogether.
- Much simpler than current, general econometric method, but very specific to the model.
- Harder part is extending to unobserved vertical quality.
Outline of Model

1. Stations produce listeners, who make a free choice as to listening. Listeners care about format and within format stations are more “similar”. Formally, use nested logit. Most general model includes discrete unobserved quality levels.

2. Stations sell listeners to advertisers. Advertisers’ demand is downward sloping in the share of the population who listen. Simple constant elasticity functional form.

3. There is free entry into a discrete product space (formats) and a static Nash equilibrium. No unique equilibrium: entry problem is no longer a Bresnahn-Reiss style ordered probit.

(1) and (2) give variable profit function, (3) adds fixed costs.
Observed Data and Variable Profits

No Variable Cost (but add endogenous fixed cost of “quality” later).

In market $t$, format $k$, We observe:

- ad price $p_t$,
- format share $s_{kt}$,
- stations numbers $N_{kt}$,
- market demographics $x_t$,
- population $M_t$.

At observed vector $N_{kt}$, observed variable profits are

$$V_{kt} = p_t(s_t)M_t s_{kt}$$

At market outcome, variable profit $V_{kt}$ is just observed revenue, $R_{kt}$. 
Counter-Factual Variable Profits

To create bounds on fixed cost, also need variable profits at $N_{kt} + 1$.

To get this counter-factual, need to

1. Estimate model of listening demand $s_{kt}(x_t, N_{kt}, N_{-k,t}, \theta_d, \xi_{kt})$,
2. Estimate model of Advertising Price $p_t(x_t, s_t, \omega_t, \theta)$. 
How to Model the Product Space

- **“Ex-Ante” vs. “Ex-Post”** differentiation: with ex-ante, have to specify number and characteristics of potential competitors. For airlines (Berry ’91) and Chain Stores (Jia, ’06) this might make sense, but other times is quite arbitrary.

- **Continuous v.s Discrete** product space. Easier to specify “counterfactual profits” (profits of the “next entrant”) with discrete space.

Here

we use ex-ante identical entrants into a discrete space of product “segments”, both horizontal and vertical.
Listening Model: Horizontal case

- Discrete-choice model: listen to one of the “inside” stations, or choose outside option (not listening to commercial radio)
- Nested-logit, 11 nests (ten format categories + “not listening”)
- Listener $i$’s utility from listening to station $j$, which belongs to format category $g$, in market $t$, is given by:

$$u_{ijgt} = \underbrace{x_{gt}\beta + \xi_{gt}}_{\delta_{gt}} + \nu_{igt}(\sigma) + (1 - \sigma)\epsilon_{ijgt}$$

- $x$ includes: market average income, share college educated, share Black & Hispanic, regional dummies, format dummies, interactions (“country $\times$ South”); $\xi_{gt}$ taste shock
Variety via Discrete Entry

$u_{ijgt} = x_{gt}\beta + \xi_{gt} + \nu_{igt}(\sigma) + (1 - \sigma)\epsilon_{ijgt}$

- Comments:
  1. Implies within-format symmetric mean-utility & market share
  2. Complication: account for in-metro vs. out-metro ("home dummy")
  3. $0 \leq \sigma < 1$ a business-stealing parameter (highest as $\sigma \to 1$)
     - $\nu_{igt}$ has the unique distribution derived by Cardel (1997) which depends on the parameter $\sigma$
     - $\nu_{igt} \to 0$ as $\sigma \to 0$
Estimation of Horizontal Differentiation Model

- For a station in format $g$, market $t$ (follow Berry 1994):

$$\ln(s_{jt}) - \ln(s_{0t}) = x_{gt}\beta + \sigma\ln(s_{j/g,t}) + \xi_{gt}$$

1. One observation for each format-market pair; Within-format symmetry imposed: $s_{jt} = S_{gt}/N_{gt}$, $s_{j/g,t} = 1/N_{gt}$
2. The above adjusted to allow for home vs. nonhome stations (so really two observations for each format-market pair)
3. Estimation using 2SLS accounting for the endogeneity of $s_{j/g,t}$ with instruments (i) market population (ii) number of out-metro stations (taken to be exogenous) (iii) number of out-metro stations in same format
4. Selection challenge for “Urban,” “Spanish,” “Religious”
### Table 1: Description of the Dataset

<table>
<thead>
<tr>
<th>Variable</th>
<th>Units</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share in-metro</td>
<td>%</td>
<td>0.111</td>
<td>0.026</td>
<td>0.030</td>
<td>0.151</td>
</tr>
<tr>
<td>Share Out-metro</td>
<td>%</td>
<td>0.015</td>
<td>0.023</td>
<td>0.000</td>
<td>0.104</td>
</tr>
<tr>
<td>N1 (in-metro)</td>
<td>integer</td>
<td>19.644</td>
<td>7.565</td>
<td>4.000</td>
<td>45.000</td>
</tr>
<tr>
<td>N2 (out-metro)</td>
<td>integer</td>
<td>7.209</td>
<td>8.320</td>
<td>0.000</td>
<td>37.000</td>
</tr>
<tr>
<td>Population</td>
<td>millions</td>
<td>1.016</td>
<td>1.687</td>
<td>0.075</td>
<td>14.481</td>
</tr>
<tr>
<td>Ad Price</td>
<td>$</td>
<td>570.480</td>
<td>237.653</td>
<td>258.222</td>
<td>2691.177</td>
</tr>
<tr>
<td>Income</td>
<td>10,000$</td>
<td>4.584</td>
<td>0.860</td>
<td>2.482</td>
<td>8.010</td>
</tr>
<tr>
<td>College</td>
<td>%</td>
<td>21.200</td>
<td>5.370</td>
<td>10.200</td>
<td>37.100</td>
</tr>
</tbody>
</table>

Computed using the 163 markets with full data, see text.
## Format Classification

Table 2: Classification of Formats into Ten Categories

<table>
<thead>
<tr>
<th>Format Group</th>
<th>Formats Included</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adult Cont.</td>
<td>Classic Hits 80s Hits</td>
</tr>
<tr>
<td>CHR</td>
<td>CHR</td>
</tr>
<tr>
<td>Rock</td>
<td>Rock Active Rock Modern Rock Classic Rock</td>
</tr>
<tr>
<td>Oldies</td>
<td>Oldies</td>
</tr>
<tr>
<td>Urban</td>
<td>Urban Urban AC Urban Oldies Rhythmic Old.</td>
</tr>
<tr>
<td>Spanish</td>
<td>Spanish-Span.-Oldies-Span.-EZ-Span.-Cl. Hits-Span.-Talk-Span.-Ranchero</td>
</tr>
<tr>
<td></td>
<td>Spanish.-Oldies-Span.-EZ-Span.-Cl. Hits-Span.-Talk-Span.-Ranchero</td>
</tr>
<tr>
<td></td>
<td>Spanish.-Oldies-Span.-EZ-Span.-Cl. Hits-Span.-Talk-Span.-Ranchero</td>
</tr>
<tr>
<td>Sports</td>
<td>Sports Farm</td>
</tr>
<tr>
<td>Other</td>
<td>Other Variety Pre-teen A30 Classical</td>
</tr>
<tr>
<td></td>
<td>Other Bluegrass Ethnic N/A Adult Stand.</td>
</tr>
<tr>
<td></td>
<td>Other Blues Silent Jazz Easy List.</td>
</tr>
<tr>
<td></td>
<td>Other cp-new A22 Smooth Jazz Dance</td>
</tr>
<tr>
<td></td>
<td>Other Americana A26</td>
</tr>
</tbody>
</table>

Steven T. Berry  Yale University, Cowles Four Empirical Models of Differentiated Products  June 18-19, 2015 272 / 341
## Table 4: The listening equation - base case (horizontal differentiation)

<table>
<thead>
<tr>
<th>Region Dummies</th>
<th>northeast</th>
<th>0.122***</th>
<th>(0.042)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>midwest</td>
<td>0.0974**</td>
<td>(0.041)</td>
</tr>
<tr>
<td></td>
<td>south</td>
<td>-0.0506</td>
<td>(0.041)</td>
</tr>
<tr>
<td>Demographics</td>
<td>black</td>
<td>-0.0681***</td>
<td>(0.014)</td>
</tr>
<tr>
<td></td>
<td>hisp</td>
<td>-0.0233**</td>
<td>(0.0097)</td>
</tr>
<tr>
<td></td>
<td>income</td>
<td>-0.00258</td>
<td>(0.017)</td>
</tr>
<tr>
<td></td>
<td>college</td>
<td>-0.0630**</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Format Dummies</td>
<td>included</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>1919</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Interactions      | hispXspan | 0.352*** | (0.036) |
|                   | blackXurban        | 0.506*** | (0.050) |
|                   | southXreligious    | 0.809*** | (0.095) |
|                   | southXcountry      | 0.316*** | (0.072) |
|                      | Corr. Parm.       | σ        |         |
|                      |                   | 0.519*** | (0.063) |
|                      |                   | In-metro dummy | 0.639*** | (0.082) |

| Constant          | -5.325*** | (0.15) |

σ = 0.519***
R-squared = 0.72

* * * indicates level of statistical significance (*** = p < 0.01, ** = p < 0.05, * = p < 0.1)
Allowing for Both Horizontal & Vertical Differentiation

- Two important new challenges:
  1. How to define / measure quality?
     - Allow quality to be an *unobserved* station characteristic
  2. Deal with endogeneity of quality (more likely to enter as “high quality Country” if market’s unobserved taste for Country is high?)
     - Incorporate market-format fixed effects

- Caveats: allowing quality differentiation in Mainstream and News/Talk only (explain below); out-metro automatically defined as low quality
Intuition for Identifying Unobserved Quality

High-quality stations have high demand—a high “mean utility” $\delta$. But across markets, $\delta$ can be because of an unobserved taste for ratio formats. Stations get high preference for free, but have to pay for quality, so want to distinguish these.

Therefore, look at within market demand difference within format. Simply, within market-format, more popular stations are higher quality. Only problem is this requires multiple levels of quality per market-format.
Utility with Station Quality

The utility for listener $i$ from listening to station $j$ in format $g$, in market $t$ is assumed to have the usual nested logit structure we defined before,

$$u_{i,j,t} = \delta_{jt} + \nu_{igt}(\sigma) + (1 - \sigma)\epsilon_{ijt},$$

with “mean utility” for station $j$ now given by

$$\delta_{jt} = \gamma^q \cdot q_{jt} + \gamma^h \cdot h_{jt} + \psi_{gt}.$$

In the mean utility, $q_{jt}$ is the quality level of a station, $h_{jt}$ is a “home” dummy variable for in-metro stations and $(\gamma^q, \gamma^h)$ are parameters to be estimated. For simplicity quality takes on two values: 0 (“low”) and 1 (“high”).
## Discrete Quality Levels

<table>
<thead>
<tr>
<th>Quality</th>
<th>$q_{jt}$</th>
<th>$h_{jt}$</th>
<th>utility term</th>
</tr>
</thead>
<tbody>
<tr>
<td>out-metro</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>in-metro low</td>
<td>0</td>
<td>1</td>
<td>$\gamma^h$</td>
</tr>
<tr>
<td>in-metro high</td>
<td>1</td>
<td>1</td>
<td>$\gamma^h + \gamma^q$</td>
</tr>
</tbody>
</table>
The term $\psi_{gt}$ is a format-market fixed effect, capturing the mean taste for format $g$ in market $t$. This depends in turn on both observed and unobserved components,

$$\psi_{gt} = d_{gt}\lambda + \xi_{gt},$$

where $d_{gt}$ is a vector of observed variables, $\lambda$ is parameter to be estimated and $\xi_{gt}$ is still the unobserved listener taste for format $g$ in market $t$. 
Identifying and Estimating Station Specific Quality

The number of listeners in each market is large and quality levels should be reflected in listening shares. We thus try to identify and estimate a discrete quality parameter for every station in every market.

Note that the relevant asymptotics for the estimation of quality involves the number of sampled listeners per market, not the number of stations or markets. Within a market, stations shares have a multinomial distribution. Since quality levels are discrete, conditional on identification, estimation will be super-efficient.
Identification of Station Level Quality

- As usual, for identification arguments, we assume an infinite sample of listeners.
- If there are two levels of within format/market share, then the higher level is higher quality.
- If there are not, then the problem is harder. We can also use the comparison to any out-metro stations shares. If the in-metro shares are “much higher,” then they are high quality. This is shown formally below.
- If, for example, there is only one in-metro station (or all in-metro shares are the same) and no out-metro, then we have *set-identification*: either the stations are all high quality or all low quality.
Consider a market where all the in-metro stations have the same market share within some format. For any guess at the quality level of stations in the market, there is a value of $\psi_{gt}$ that explains the observed common level of shares.
Using Within Market Shares

Because the market-format taste does not effect the within format shares, we can avoid this potential problem of non-identification if we focus on the within format shares. This is similar to the idea of “differencing out” a fixed effect to deal with endogeneity. To proceed, let

\[ \kappa_1 \equiv \gamma^q/(1 - \sigma), \]

\[ \kappa_2 \equiv \gamma^h/(1 - \sigma), \]

and let the vector \( \kappa \equiv (\kappa_1, \kappa_2) \).
Within Market Nested Logit

The nested logit then implies that conditional on choosing format $g$ the expected probability of choosing station $j$ in market $t$ (the “within format share”) is given by

$$p_{j/gt}(\kappa, q) = \frac{\exp(\kappa_1 \cdot q_{jt} + \kappa_2 \cdot h_{jt})}{\sum_{\ell \in g} \exp(\kappa_1 \cdot q_{\ell t} + \kappa_2 \cdot h_{\ell t})},$$

where $q$ is notation for the long vector of quality levels for all markets’ stations. The expression in depends only on the quality levels in the given market-format.
Identification Steps for Quality

1. identify $\kappa$ using data on markets where differences in shares identify quality levels.
2. use $\kappa$, together with out-metro shares, to identify quality levels in additional markets.
Remaining Partial Identification

This leaves us with one remaining case of partial identification: market/formats with no out-metro station and identical shares for in-metro stations. In these market/formats, we know that either [i] all stations are high quality or [ii] all stations are low quality.

Our base case approach is to estimate demand from the market/formats with out-metro stations and then to look for bounds on counterfactuals for the set identified cases.
Back to IV for $\sigma$

Note that having identified qualities, we move back to the usual case of Berry (1994) with a nested logit where all characteristics are observed (although recall that the unobservable taste variable $\xi_{gt}$ is at the level of the format, not station.)

Thus, the remainder of the demand identification problem is standard and we will continue to need an instrument variables approach to identify the format-market level taste parameters ($\gamma^q, \gamma^h, \sigma$) (as opposed to the composite station-quality parameters ($\kappa_1, \kappa_2$) that are identified from the within format choice problem.)
Instruments Again

As in the pure horizontal model, we assume that the unobserved format-level taste shifter is mean-independent of a set of instruments,

$$E[\xi_{gt} | Z_{gt}] = 0.$$ 

In the empirical application, we let $Z_{gt}$ contain the market’s population, the number of the market’s out-metro stations, and the number of out-metro stations in the same format, as well as the $d$ covariates.
Table 5: Demand Parameters with Vertical Quality

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma$ – business stealing</td>
<td>0.589</td>
<td>0.017</td>
</tr>
<tr>
<td>$\gamma^q$ – high quality</td>
<td>0.604</td>
<td></td>
</tr>
<tr>
<td>$\gamma^h$ – lower quality</td>
<td>0.466</td>
<td></td>
</tr>
<tr>
<td>blackXurban/10</td>
<td>5.555</td>
<td>0.001</td>
</tr>
<tr>
<td>hispXspan/10</td>
<td>3.962</td>
<td>0.002</td>
</tr>
<tr>
<td>region dummies</td>
<td>included</td>
<td></td>
</tr>
<tr>
<td>format dummies</td>
<td>included</td>
<td></td>
</tr>
<tr>
<td>demographics</td>
<td>included</td>
<td></td>
</tr>
</tbody>
</table>
Demand from Advertisers

We treat stations as "producing" listeners and then selling them to advertisers. For now, a very simple inverse ad-demand function. The demand from advertisers for listeners in market $t$ is modeled by a downward-sloping, constant-elasticity specification:

$$\ln(p_t) = x_t \alpha - \eta \ln(s_t) + \omega_t$$

Popl. and out-metro stations are instruments for endogenous share. Might be able to have this vary by format / demographic, but data is pretty bad for this.
Notes on Advertiser Demand

It would be nice to let ad demand vary with demographics of listeners, but we have limited price variation in our data. Some recent work uses industry “estimated” prices for individual stations. Sometimes these appear to just be informed guesses. There is an obvious trade-off here in using less reliable data to estimate a more realistic model.

So far we have not used the less reliable data and so we keep the ad demand model (unrealistically) simple.
Table 6: Advertiser Demand

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>northeast</td>
<td>-0.0746</td>
<td>-0.0739</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.063)</td>
</tr>
<tr>
<td>midwest</td>
<td>0.0835</td>
<td>0.0799</td>
</tr>
<tr>
<td></td>
<td>(0.061)</td>
<td>(0.059)</td>
</tr>
<tr>
<td>south</td>
<td>0.0148</td>
<td>0.0132</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.059)</td>
</tr>
<tr>
<td>income</td>
<td>0.0567*</td>
<td>0.0606**</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>college</td>
<td>0.167***</td>
<td>0.164***</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>black</td>
<td>-0.0231</td>
<td>-0.0242</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>hisp</td>
<td>-0.0120</td>
<td>-0.0124</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>$-\eta$</td>
<td>-0.541***</td>
<td>-0.510***</td>
</tr>
<tr>
<td></td>
<td>(0.062)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>Constant</td>
<td>4.492***</td>
<td>4.554***</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.18)</td>
</tr>
<tr>
<td>Observations</td>
<td>163</td>
<td>163</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.52</td>
<td>0.52</td>
</tr>
</tbody>
</table>
Equilibrium in Product Segments

Once we have listening demand and the (inverse) advertising demand equation, we have estimated variable profits.

Segment Fixed Costs

To recover fixed-costs (constant across products within segments) need to have a model of equilibrium market structure.
A good assumption for work that relies on the cross-sectional nature distribution of market structure. With no explicit dynamics, we would like firms to choose the best-response to rival’s actions – otherwise why don’t they move? Justification for cross-sectional study is [i] population and demographics are strong instruments and [ii] firms are in “long-run” equilibrium.

In a dynamic model, some private info makes more sense – firms might be surprised to find themselves in a bad location and then move away.
Bounding the Distribution of Fixed Costs

Complete Info Static Nash Equilibrium

- No variable costs. $F$ has to be less than observed revenue.
- Also, $F$ has to be greater than counterfactual revenue at $(N_{kt} + 1)$.
- Construct counterfactual revenue from listening demand and ad-price equation (including values of unobservables.)
- Can’t do this for markets with $N_t = 0$; selection discussed below.
Bounds on $F$

We know that

$$R_{kt} > F_{kt}$$

This provides an upper bound for $F$, making only the assumption that $R$ and $F$ are constant within segment.

Lower Bound on $F$: in equilibrium,

$$V_k(N_{kt} + 1, y_t, x_t, \theta_0) < F_{kt}.$$  

We get bounds in each market without making any assumption on the distribution of fixed costs.
Table 9: Welfare analysis in the base case (horizontal only)

<table>
<thead>
<tr>
<th></th>
<th>Observed</th>
<th>Optimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Welfare ($ millions)</td>
<td>11,977</td>
<td>13,779</td>
</tr>
<tr>
<td>Mean In-Metro Listening Share (%)</td>
<td>11.10%</td>
<td>8.15%</td>
</tr>
<tr>
<td>Mean Ad Price ($)</td>
<td>570.48</td>
<td>662.54</td>
</tr>
</tbody>
</table>

**BUT!** This is for market participants only.
**Table 10: Optimal Structure with Quality**

Table 10: Optimal and observed market structures, horizontal and vertical differentiation

**A. Formats with a single quality level:**

<table>
<thead>
<tr>
<th>Format</th>
<th>Mean # observed</th>
<th>Mean # optimal</th>
<th>Mean optimal reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHR</td>
<td>1.06</td>
<td>0.85</td>
<td>20.2%</td>
</tr>
<tr>
<td>Country</td>
<td>2.10</td>
<td>1.02</td>
<td>51.3%</td>
</tr>
<tr>
<td>Rock</td>
<td>2.33</td>
<td>1.04</td>
<td>55.3%</td>
</tr>
<tr>
<td>Oldies</td>
<td>1.02</td>
<td>0.86</td>
<td>16.2%</td>
</tr>
<tr>
<td>Religious</td>
<td>1.66</td>
<td>0.77</td>
<td>53.5%</td>
</tr>
<tr>
<td>Urban</td>
<td>1.50</td>
<td>0.71</td>
<td>52.6%</td>
</tr>
<tr>
<td>Spanish</td>
<td>1.34</td>
<td>0.50</td>
<td>63.0%</td>
</tr>
<tr>
<td>Other</td>
<td>2.12</td>
<td>1.01</td>
<td>52.3%</td>
</tr>
</tbody>
</table>

**B. Formats with quality differentiation:**

<table>
<thead>
<tr>
<th>Format:</th>
<th>Mainstream</th>
<th>News/Talk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low $q$</td>
<td>High $q$</td>
</tr>
<tr>
<td>Mean # observed</td>
<td>1.18-1.95</td>
<td>1.40-2.17</td>
</tr>
<tr>
<td>Mean # optimal</td>
<td>0.37-0.66</td>
<td>0.56-0.86</td>
</tr>
</tbody>
</table>
Overprovision of Quality

Out of 90 markets with observed high-quality Mainstream stations, in 72 cases welfare can be unambiguously improved by converting one of those stations to low quality operation. An even higher rate, 94.9%, applies to the News/Talk format. On the other hand, there are no cases where a market has observed low-quality stations—in either format—and converting one of them to high quality would unambiguously improve welfare.

Our analysis of local changes to quality offerings, therefore, reveals a pattern of over-provision of quality at the margin.
BEW Conclusion

Methodologically, we develop a model of entry into an unobservable discrete quality space, and combine this with a straightforward set-identified model of entry and fixed costs.

Substantively,

▶ Introducing richer tastes for variety reduces, but hardly eliminates, the problem of excess entry due to profit shifting.

▶ Similarly, heterogeneous quality does not much effect the overall excess entry result, and there is some evidence of the (local) over-provision of quality.

▶ In this particular context, the excess entry may or may not be offset by unpriced gains to listeners.
Media Markets / Preference Externalities

Media markets have several interesting features:

▶ Often two-sided markets (revenue from consumers and advertisers). “Platforms”
▶ High fixed costs relative to marginal costs
▶ Public Policy issues: political/social diversity of viewpoints, distribution monopolies (cable, fiber to home).
Media Bias and Reputation

Self-confirming bias – can media competition reduce this? (to some degree)
“Preference Externalities”

Where do product choice sets come from?

Waldfogel story:

- Politics are thought to result in the *tyranny of the majority*.
- Classic competitive markets offer the “freedom to choose.”

*As Friedman (1962) famously put it, [e]ach man can vote, as it were, for the color of tie he wants and get it; he does not have to see what color the majority wants and then, if he is in the minority, submit.*
But when fixed costs are large relative to the market, only a subset of products will be offered.

Which products are offered will depend on your *neighbors* preferences. Implications for urban economics, trade, health care, media markets, etc.
Media Slant

Gentzkow and Shapiro (2010)

Newspaper bias in news stories: does it come from owner preferences or reader preferences?

Policy question: does owner concentration effect politics?
GS – Slant

- Develop a text-based measure of political “slant”
- Use this is a simple Hotelling-like differentiated products (monopoly) model.
- Calculate the optimal (profit max) political location of the firm – is there any evidence that owners systematically deviate from this?
GS Slant: text analysis

2005 Congressional Record:

Most Popular 2 word D phrases: private accounts, trade agreement, American People

Most Popular 2 word R phrases: stem cell, natural gas, death tax.

Weight them by the biggest difference in use.
WHAT DRIVES MEDIA SLANT?
GS Reader Problem

In zip code $z$, readers have their own political slant, $r_z$, measured by political contributions from the zip code (pretty noisy and weighted to the rich.)

Ideal ideological point:

$$ideal_z = \alpha + \beta r_z$$

Hotelling utility:

$$u_{izn} = \bar{u}_{zn} - \gamma(y_n - ideal_z)^2 + \epsilon_{ijn}$$
GS Demand

\[ \bar{u}_{zn} = X_z \phi^0 + W_{zn} \phi^1 + \xi_{mn} + \nu_{zn} \]

Expand the square, substitute in terms, use logit, get

\[ \ln \left( \frac{S_{zn}}{1 - S_{zn}} \right) = \delta_{mn} + \lambda^d_0 y_n r_z + \lambda^d_1 r_z + \lambda^d_2 r_z^2 + X_z \phi^0 + W_{zn} \phi^1 + \nu_{zn} \]
Measurement error in $y_n$ (but not $r_z$) handled by instruments – $R_n$ is the overall share of republicans and $r_z R_n$ is excluded. This is a Waldfofel like instrument – politics of others in the metro determines $y_n$. Fixed effect absorbs the market-level taste $\xi_{mn}$. 
### TABLE II

**Evidence on the Demand for Slant**

<table>
<thead>
<tr>
<th>Description</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OLS</td>
</tr>
<tr>
<td>(Zip share donating to Republicans) × Slant</td>
<td>10.66</td>
</tr>
<tr>
<td></td>
<td>(3.155)</td>
</tr>
<tr>
<td>Zip share donating to Republicans</td>
<td>−4.376</td>
</tr>
<tr>
<td></td>
<td>(1.529)</td>
</tr>
<tr>
<td>(Zip share donating to Republicans)^2</td>
<td>−0.4927</td>
</tr>
<tr>
<td></td>
<td>(0.2574)</td>
</tr>
<tr>
<td>Market–newspaper FE?</td>
<td>X</td>
</tr>
<tr>
<td>Zip code demographics?</td>
<td>X</td>
</tr>
<tr>
<td>Zip code X market characteristics?</td>
<td>X</td>
</tr>
<tr>
<td>Zip code FE?</td>
<td>X</td>
</tr>
<tr>
<td>Number of observations</td>
<td>16,043</td>
</tr>
<tr>
<td>Number of newspapers</td>
<td>290</td>
</tr>
</tbody>
</table>
### TABLE IV

**ECONOMIC INTERPRETATION OF MODEL PARAMETERS**

<table>
<thead>
<tr>
<th>Description</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual slant of average newspaper</td>
<td>0.4734</td>
</tr>
<tr>
<td></td>
<td>(0.0020)</td>
</tr>
<tr>
<td>Profit-maximizing slant of average newspaper</td>
<td>0.4600</td>
</tr>
<tr>
<td></td>
<td>(0.0047)</td>
</tr>
<tr>
<td>Percent loss in variable profit to average newspaper from moving 1 SD away</td>
<td>0.1809</td>
</tr>
<tr>
<td>from profit-maximizing slant</td>
<td>(0.1025)</td>
</tr>
<tr>
<td>Share of within-state variance in slant from consumer ideology</td>
<td>0.2226</td>
</tr>
<tr>
<td></td>
<td>(0.0406)</td>
</tr>
<tr>
<td>Share of within-state variance in slant from owner ideology</td>
<td>0.0380</td>
</tr>
<tr>
<td></td>
<td>(0.0458)</td>
</tr>
</tbody>
</table>
WHAT DRIVES MEDIA SLANT?

![Scatter plot showing the relationship between Market Percent Republican and Slant.](image)
Hospital Mergers in the US

Traditional analysis of hospital mergers was not based on any strong economic model. Hospitals are strongly differentiated in location, but also in speciality. “Community” hospitals are good for a broken arm, big-city university hospitals are much better for major illness (cancer, etc.)

How to analyze a merger? US Federal Trade Commission has adopted an explicit demand-based differentiated products framework. Recent research has focused on how to develop a “supply side” based on the fact that, in the US, hospital and private health insurers bargain contracts (how much the insurer will pay the hospital for non-elderly care.)
Capps, Dranove, and Satterthwaite (2003) and Ho (2009)) model competition between hospitals in different markets. Great consumer data: diagnosis of illness, location, etc. These interact with hospital characteristics in different markets: location of hospital with location of patient, severity of illness with “university hospital”, etc. The hospital may in “in network” or not (much more expensive.)

These authors estimate a logit model with rich consumer (diagnosis and location) data that interacts with rich hospital data. Substitution between hospitals varies a lot by diagnosis, not just hospital characteristics.
Supply side is harder: bargaining between hospitals and insurance companies. See Ho (2009) and more recent work by Gautam Gowrisankaran and Town (2015).

As intuition for supply, the demand estimates intuitively set the value of the hospital to the network (how unhappy are consumers if the hospital is left out of the insurer network) and this should set the “threat point” of the hospital in negotiations.
Merger analysis could be based on

- Change in Concentration
- Some Demand Based Measure (WTP for adding a hospital to a network, from a logit, Capps, Dranove, and Satterthwaite (2003))? 
- Full Merger Simulation?

Christopher Garmon (US Federal Trade Commission), “The Accuracy of Hospital Merger Screening Methods,” shows that logit-based WTP is reasonably well-correlated with post-merger price changes. Change in concentration only picks out the most extreme worst merger.

Full merger simulation based on the logit and a very simple supply model does not do well. Better models or data?
Example: Gowrisankaran Nevo and Town

Gautam Gowrisankaran and Town (2015)
At hospital $j$, patient $i$ with diagnosis $d$ covered by insurance $m$ gets utility

$$u_{ijd} = x_{ijd} \beta - c_{id} w_d p_{md} + \epsilon_{ijd}.$$  

where $c$ is a co-insurance term, $w_d$ is a diagnostic weight and $p$ is specific to the insurer/hospital. The choice set of hospitals varies by insurer (a good instrument or not?)

The price / insurance interaction is interesting, but this is a standard micro-logit, otherwise.
Define the ex-ante expected cost to the MCO and the employer that it represents to be $TC_m(N_m, p_m)$. The MCO pays the part of the bill that is not paid by the patient, hence

$$
(5) \quad TC_m(N_m, p_m) = \sum_{i=1}^{I} \sum_{d=1}^{D} 1\{m(i) = m\}(1 - c_{id})f_{id}w_d \sum_{j \in 0, N_m} p_{mjs_{ijd}(N_m, p_m)}.
$$

Define the value in dollars for the MCO and the employer it represents to be:
GNT Price Bargaining

Define the value in dollars for the MCO and the employer it represents to be:

\[
V_m(N_m, p_m) = \frac{\tau}{\alpha} \sum_{i=1}^{I} 1\{m(i) = m\} W_i(N_m, p_m) - TC_m(N_m, p_m),
\]
GNT Price Bargaining

To understand more about the equilibrium properties of our model, we solve the first order conditions of the Nash bargaining problems, \( \partial \log NB_{m,s} / \partial p_{mj} = 0 \). For brevity, we omit the \("*"\) from now on, even though all prices are evaluated at the optimum. We obtain:

\[
(12) \quad b_{s(m)} \frac{q_{mj} + \sum_{k \in J_s} \frac{\partial q_{mk}}{\partial p_{mj}} [p_{mk} - m c_{mk}]}{\sum_{k \in J_s} q_{mk} [p_{mk} - m c_{mk}]} = \frac{A}{\frac{\partial V_m}{\partial p_{mj}} - b_{m(s)} \frac{V_m(N_m, p_m) - V_m(N_m \setminus J_s, p_m)}{B}}.
\]
GNT Price Bargaining

We rearrange the joint system of \( \#(\mathcal{I}_s) \) first order conditions from (12) to write

\[
q + \Omega(p - mc) = -\Lambda(p - mc)
\]  

where \( \Omega \) and \( \Lambda \) are both \( \#(\mathcal{I}_s) \times \#(\mathcal{I}_s) \) size matrices, with elements \( \Omega(j, k) = \frac{\partial q_{mk}}{\partial p_{mj}} \)

and \( \Lambda(j, k) = \frac{b_{m(s)}}{b_{s(m)}} \frac{A}{B} q_{mk} \). Solving for the equilibrium prices yields

\[
p = mc - (\Omega + \Lambda)^{-1} q,
\]
A key point then, is that GNT can still solve for mc as a function of all the demand and bargaining parameters. They think of the sum of the traditional and bargaining terms in the foc as a “bargaining adjusted” demand effect.

After estimating the model, they find that a proposed hospital merger in Virginia, which was turned down by the FTC, would have substantially raised prices.
Conclusion

There is now a great variety of methods for analyzing product differentiation, including static demand and oligopoly pricing. Further extensions allow us to analyze endogenous variety and quality, either via first-order conditions for continuous variables or entry models of discrete differentiation.

We leave fully dynamic models to another day.
References I


References II


References III


References IV


References V


References VI


References VII


References VIII


References IX


References XI


References XIII


McFadden, D., A. Talvitie, and Associates (1977): *Demand Model Estimation and Validation*. Institute of Transportation Studies, Berkeley CA.


References XIV


References XV


References XVI


