# Empirical Industrial Organization

Notes for Summer School
Moscow State University, Faculty of Economics

Andrey Simonov*

June 2013

**Attributions**: This notes gained a lot from various courses in Tilburg University and University of Chicago, in particular courses by Bart Bronnenberg, Jaap Abbring, Tobias Klein, Derek Niel, Ali Hortacsu, Jean-Piere Dube, Guenter Hitsch and Pradeep Chintagunta. Section on production function function heavily relies on excellent review by Ackerberg et al. (2007). Of course, all errors and typos are of my own[1].


*University of Chicago, Booth School of Business
[1] Please send any comments/reports on typos to asimonov@chicagobooth.edu

# Contents

# 1 Static Demand Estimation

- Why do we care about demand system estimation? One of the key ingredients in a number of Industrial Organization (IO) questions. I.e.:
  - Entry/exit decisions;
  - Market size definition (merger analysis);
  - Production decisions.
- Lets focus on the differentiated product markets.
- Consider the market with $J$ heterogeneous goods.
- There are two ways to proceed with estimation: treat each product as a good itself, or treat a product as a bundle of characteristics.

## 1.1 Product Space

### 1.1.1 Linear Demand

- Think about a market with $j$ goods: $j = 1, \cdots, J$. Proposed demand equation:

$$q_i = a_i + \sum_j b_{ij} p_{ij} + \epsilon_{ij}$$

  where $q_i$ is demand for good $i$, $p_{ij}$ are price coefficients, $b_{ij}$ are marginal effects, $a_i$ are product-specific intercepts and $\epsilon_{ij}$ are the random shocks.
- Allow to compute own and cross price elasticities.
- The most obvious problem: we can get negative value for $q_i$

### 1.1.2 Log-linear/Constant Elasticity Demand

- Solution: use log-linear/constant elasticity demand:

$$log q_i = \alpha_i + \eta_i log m + \sum_{j=1}^{k} \epsilon_{ij} log p_j + \epsilon_i$$

  where $m$ is income and $\eta_i$ is income elasticity
- Looks like Marshallian demand curve.
- Well, another problem: adding up constraint does not hold unless $\eta_i = 1\ \forall i$: $\sum_{i=1}^{k} w_i \eta_i = 1$ where $w_i$ are the expenditure shares.

### 1.1.3 Almost Ideal Demand System

- Deaton and Muellbauer (1980) [14] address it by finding a specific form of the cost function and specifying the market share equation:

$$w_i = \alpha_i + \sum_{j=1}^{k} \gamma_{ij} logp_j + \beta_i log(\frac{X}{P}) + \epsilon_i$$

where $P$ is a price index

$$logP = \alpha_0 + \sum_j \alpha_j logp_j + \frac{1}{2} \sum_j \sum_l \gamma_{jl} logp_j logp_l$$

This is referred to as Almost Ideal Demand System (AIDS).

- We can back out elasticities from the parameters:

$$e_{ij}^M = -I\{i = j\} + \frac{\gamma_{ij}}{w_i} - \frac{\beta_i}{w_i} w_j$$

$$\epsilon_{ij}^H = e_{ij}^M + \eta_i w_j$$

$$\eta_i = 1 + \frac{\beta_i}{w_i}$$

- Can decrease the number of estimated parameters imposing a number of constraints: $\sum_{i=1}^{k} \alpha_i = 1$, $\sum_{i=1}^{k} \gamma_{ij} = 0$, $\sum_{i=1}^{k} \beta_i = 0$, $\gamma_{ij} = \gamma_{ji}$.
- Still have a number of problems:
  - There are $K^2 + 2K + 1$ parameters to estimate.
  - People often ignore the nonlinear specification and simply estimate OLS. We need an instrument for each price!
  - Can say nothing about new products entering the markets.

## 1.2 Characteristic Space: Models

- Treat each product as a bundle of characteristics $x_j$.
- Can define utility function consumer $i$ obtains from purchasing product $j$ is

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

where $p_j$ is the price of product $j$, and $\nu_i$ represents the vector of consumer tastes, distributed with density $f(\nu)$ over the population.

- Assume products are strong substitutes, that is consumer needs to buy only one product. In this case consumer would buy product $j$ if utility of product $j$ exceeds utility of all other products:

$$I(i \text{ choose } j) = I(u_{ij}(\cdot) > u_{ik}(\cdot)\ \forall k \neq j) \tag{1}$$

- We can define the region of $f(\nu)$ where consumers prefer product $j$ as

$$A_j(X,p) = \{\nu_i \in R_L | (u_{ij}(\cdot) > u_{ik}(\cdot)\ \forall k \neq j\}$$

- Now we can define the market share of the good $j$ as

$$s_j = \int_{A_j(X,p)} f(\nu) d\nu \tag{2}$$

  Knowing the market share and the market size, we can easily back out the demand by

$$q_j = M s_j$$

- How do we solve this integral? There are two way to proceed:
  - Assume a particular distribution which will make the solution analytic (i.e. logit or nested logit models)
  - Use simulations or quadratures to approximate the integral.
- Similarly, as econometricians we can use distribution $f(\nu)$ to define the probability of consumer $j$ choose $i$:

$$Pr(i \text{ choose } j) = Pr(u_{ij}(\cdot) > u_{ik}(\cdot)\ \forall k \neq j) \tag{3}$$

  In this case

$$Pr(i \text{ choose } j) = \int_{A_j(X,p)} f(\nu) d\nu$$

- In what follows we stick to the interpretation (2) as we discuss aggregate demand. Interpretation (3) would be useful later in the discussion of individual choice models.

### 1.2.1 Logit Model

- Assume additively separable utility functions:

$$u_{ij} = X_j \beta + \alpha p_j + \epsilon_{ij}$$

  where $\epsilon_{ij}$ denotes consumer idiosyncratic taste for product $j$ (which before was included into $\nu_i$) (was is the dimension of $\epsilon_i$ now?)

- Assume a particular functional form for $\epsilon_{ij}$:

$$\epsilon_{ij} \sim iid\ T1EV(\epsilon)$$

  which is Type 1 Extreme Value distribution[2] with location parameter 0 and scale parameter 1:

$$F(\epsilon_{ij}) = exp(-exp(-\epsilon_{ij}))$$

[2] Also known as Gumbel distribution

- Using the properties of T1EV distribution (difference of two independent T1EV is distributed logistically) we can write the market shares as

$$s_j(X,p) = \int \cdots \int_{A_j} f(d\epsilon_{i1}, \cdots, d\epsilon_{iJ}) = \frac{e^{X_j\beta+\alpha p_j}}{\sum_k e^{X_k\beta+\alpha p_k}} = \frac{e^{\delta_j}}{\sum_k e^{\delta_k}} \tag{4}$$

  where $\delta_j$ is the mean utility of product $j$.

- We do not observe the actual utility, which requires us to make some normalizations: on the location and on the scale:
  - Scale is implicitly normalized as logit error term has a fixed variance;
  - Location is usually normalized by $\delta_0 = 0$, which reflects the outside option.

- Examine price derivatives and elasticities in order to characterize demand function:

  Own-price marginal effect: $$\frac{\partial q_j}{\partial p_j} = \frac{\partial s_j}{\partial p_j} M = M\left[\alpha \frac{e^{\delta_j}}{\sum_k e^{\delta_k}} - \alpha\left(\frac{e^{\delta_j}}{\sum_k e^{\delta_k}}\right)^2\right] = M\alpha s_j(1-s_j)$$

  Cross-price marginal effect: $$\frac{\partial q_j}{\partial p_k} = \frac{\partial s_j}{\partial p_k} M = M\left[-\alpha \frac{e^{\delta_j} e^{\delta_k}}{(\sum_k e^{\delta_k})^2}\right] = -M\alpha s_j s_k$$

  (Check that own and cross-price elasticities exhibit similar problems)

- Own-price derivatives vary only with the market size of good $j$. This is restrictive as two products with the same market size (say BMW and Kia) should have the same own-price elasticity (completely unintuitive in the example of BMW and Kia).

- Things are even worth in cross-price derivatives: if $s_{BMW} = s_{Kia}$ cross-price marginal effect of price of Mercedes on both BMW and Kia should be the same. Obviously makes no sense.

- In a way market shares are "sufficient statistics" for substitution patterns. This is often said to be the result of the logit model; however, it rather comes from iid tastes of consumers over products:

$$u_{ij} = \delta_j + \epsilon_{ij} \Rightarrow s_j = f(\delta_j, \delta_1, \cdots, \delta_J)$$
$$s_j = s_k \Rightarrow \delta_j = \delta_k$$

- Logit models have another problem: Independence of Irrelevant Alternatives (IIA) property. It says that $\frac{s_j}{s_k}$ does not depend on other products.

- Classical example is when consumer needs to choose 1) between red bus and a car; 2) between red bus, blue bus and a car. If $\frac{s_{RED\ BUS}}{s_{CAR}} = 1/2$ in the first model (33.3% of consumers choose a bus), then $\frac{s_{RED\ BUS}}{s_{CAR}} = 1/2$ in the second model (but now 50% of consumers choose a bus: 25% a red bus and 25% a blue one). Once again makes no sense.

- These problems can be solved by allowing for correlation in consumer tastes, i.e. people with high $\epsilon_{RED\ BUS}$ would also have high $\epsilon_{BLUE\ BUS}$. However, if try to estimate all elements in the variance-covariance matrix $\Sigma_{\epsilon_i}$ we need to estimate $(J^2 + J)/2$ elements, which is the same problem as we had in case of product space demand estimation.

- We can try to put some restrictions on $\Sigma_{\epsilon_i}$ such that the space of parameters is not too big.

### 1.2.2 Nested Logit Model

- We can use some prior knowledge about the structure of the market to divide products into $G$ nests.
- Assume that nests are exclusive (product can be only in one nest).
- Redefine consumer utility as

$$u_{ij} = X_j\beta + \alpha p_j + \zeta_{ig_j} + (1-\sigma)\epsilon_{ij} = \delta_j + \nu_{ij}$$

  where

  - $\delta_j = X_j\beta + \alpha p_j$
  - $\nu_{ij} = \zeta_{ig_j} + (1-\sigma)\epsilon_{ij}$
  - $\zeta_{jg_j}$ is a specific taste of group $g_j$ and is iid over groups.

- What is the structure of $\Sigma_\nu$?

$$E(\nu_{ij}\nu_{ik}) = \begin{cases} \sigma_\zeta^2 & \text{if } (j,k) \in g \\ 0 & \text{otherwise} \end{cases}$$

- Market share is defined as in (4):

$$s_j = \int \cdots \int_{A_j} f(d\zeta_i, d\epsilon_i) \tag{5}$$

- Cardell (1996) [10] shows that $\forall\sigma$ there exists a unique distribution of $\zeta_i$ such that if $\epsilon_i \sim iid\ EV$ then $\nu_{ij} \sim iid\ EV$
- Nested Logit model makes this particular assumptions, which makes an integral in (5) analytical:

$$s_j(X,p) = s_{j|g}s_g = \frac{exp(\frac{\delta_j}{1-\sigma})}{\sum_{k\in g_j} exp(\frac{\delta_k}{1-\sigma})} \left( \frac{\left(\sum_{k\in g_j} exp(\frac{\delta_k}{1-\sigma})\right)^{1-\sigma}}{\sum_g \left(\sum_{k\in g_j} exp(\frac{\delta_k}{1-\sigma})\right)^{1-\sigma}} \right) \tag{6}$$

  where $s_{j|g}$ stands for the market share of good $j$ in the nest $g$, and $s_g$ stands for the market share of nest $g$.

- Price elasticities in this case would be stronger within group than between groups (derivation is the same as for logit).
- What happens as we change $\sigma$?
  - as $\sigma \to 1$, substitution would be only within group;

  - as $\sigma \to 0$, we get back to logit model.
- Extensions to this include multiple level of nests (Cardell 1997 [10]), overlapping nests (Bresnahan, Stern, Trajtenberg 1997 [7]), ordered nests.
- Nested models are quite powerful. However, often criticized for setting the groups a-priori.

### 1.2.3 Random Coefficients Models

- Redefine consumer utility as

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

  Notice that now we allow both $\beta$ and $\alpha$ to vary across consumers: intuitively we would expect that now price elasticities should be very flexible.
- However, we need to impose some structure of $\beta_i$ and $\alpha_i$, otherwise we have $(M+1)N$ parameters to estimate with $N$ observations ($M$ is the number of characteristics).
- One way to proceed is to assume a particular distribution for $\beta_i$ and $\alpha_i$, i.e.:
  - $\beta_i$ consist of $M$ independent marginals, where

$$\beta_{im} \sim N(\beta_m, \sigma_m^2)$$

  - $\alpha_i$ can be interpreted as consumer distaste for price:

$$ln\alpha_i \sim N(\alpha, \sigma_\alpha^2)$$

    (why do we use log-normal instead of normal in this case?)
- As before, we assume that $\epsilon_{ij} \sim iid\ T1EV$.
- Obviously, it is a generalized version of the logit model. It also appears to be a generalized version of the nested logit model.
- Define

$$\beta_{im} = \beta_m + \sigma_m\tilde{\beta}_{im} \qquad \tilde{\beta}_{im} \sim N(0,1)$$
$$\alpha_i = exp(\alpha + \sigma_\alpha\tilde{\alpha}_i) \qquad \tilde{\alpha}_i \sim N(0,1)$$

  Now we can rewrite $u_{ij}$ as

$$u_{ij} = X_j\beta - \alpha p_j + \left[\sum_m \sigma_m x_{jm}\tilde{\beta}_{im} - \sigma_\alpha p_j\tilde{\alpha}_i + \epsilon_{ij}\right] = \delta_j + \nu_{ij}$$

- Notice that $\nu_{ij}$ would be correlated across products as it contains $x_j$ and $p_j$. That is, cross-price elasticities with products that have more similar characteristics would be higher.
- It seems like the result we want. What are the caveats?

- Well, one thing is that once again we make very specific assumptions on parameters. This can be slightly improved by using empirical distribution of demographics.
- Another problem is more severe. Rewrite the market share equation:

$$s_j = \int \cdots \int_{A_j} f(d\tilde{\beta}_i, \tilde{\alpha}_i, d\epsilon_i) \tag{7}$$

Unfortunately there is no analytical solution for (7). One way to go is to use quadratures, but these becomes infeasible as the number of integrals gets $> 5$ (what number of integrals would you expect say for cola market? automobile market?)

- This is not a problem for bayesian econometrics (discussed today at the Data Analysis section).
- If we want to stay in the frequentists domain we need to simulate this integral. For an arbitrary function $f(\cdot)$ and arbitrary distribution with pdf $g(\cdot)$:

$$Y = \int f(z)g(z)dz \approx \frac{1}{NS} \sum_{NS} f(z_{ns})$$

where $z_1, \cdots, z_{NS}$ are $NS$ random draws from $g(\cdot)$.

- This is sometimes referred to as pure frequency simulator. Easy to show that it is unbiased.
- Suits for solving (7), but requires setting $NS$ to a very big number. Can we do better than that?
- Recall that for $\epsilon_{ij} \sim T1EV$ we could solve the integral analytically. Use this fact to rewrite (7) as

$$s_j = \int \cdots \int \frac{exp\left(X_j\beta - \alpha p_j + \left[\sum_m \sigma_m x_{jm}\tilde{\beta}_{im} - \sigma_\alpha p_j \tilde{\alpha}_i\right]\right)}{1 + \sum_k exp(\cdot)} f(d\tilde{\beta}_i, \tilde{\alpha}_i) \tag{8}$$

(why did we get one in the denominator?)

Now there is "only" $M+1$ integrals to simulate and $2(M+1)$ parameters to estimate.

## 1.3 Characteristic Space: Estimation

### 1.3.1 Berry's inverse

- Usually we observe prices, quantities (market shares) and product characteristics. Theoretical model predicts that

$$s_j = s_j(x, p; \beta, \alpha)$$

We need to find such $\beta$ and $\alpha$ that this equality holds (at least approximately)

- Intuitive way to go is to minimize the distance between the two. Usual way to think about it is to add an unobservable and find $\beta$ and $\alpha$ that minimize it:

$$s_j = s_j(x, p; \beta, \alpha) + \eta_j$$

- Assuming that $\eta_j$ is independent of $X$ and $p$ (that is, $\eta_j$ is a corresponding measurement error) we can simply use the corresponding moments for estimation.
- As econometricians we might not observe all $X$, so we use only a subset of $X$ for estimation. In this case independence of $\eta_j$ and $p$ is a very strong assumption (why?)
- Berry (1994) [5] criticized this specification by pointing that unobservables should be inside of the market share. He suggested using the following specification:

$$u_{ij} = X_j\beta + \alpha p_j + \xi_j + \epsilon_{ij}$$

where $\xi_j$ correspond to product characteristics that are unobserved to econometrician (but observed to the firms and consumers). Under the assumption of T1EV on $\epsilon_{ij}$ we are back in the case of logit specification.

- Same way as in (4), we can write

$$s_j(X, p) = \int \cdots \int_{A_j} f(d\epsilon_{i1}, \cdots, d\epsilon_{iJ}) = \frac{e^{X_j\beta + \alpha p_j + \xi_j}}{\sum_k e^{X_k\beta + \alpha p_k + \xi_k}} \tag{9}$$

- Berry (1994) shows that there exist a very simple inversion of this non-linear function. Specify one of the goods as an outside good with $u_{i0} = \epsilon_{i0}$. In this case

$$s_j = \frac{e^{X_j\beta + \alpha p_j + \xi_j}}{1 + \sum_{k \neq 0} e^{X_k\beta + \alpha p_k + \xi_k}}$$

$$s_0 = \frac{1}{1 + \sum_{k \neq 0} e^{X_k\beta + \alpha p_k + \xi_k}}$$

Denominator of any market share is the same. Moreover, the ratio of market share of product $j$ and market share of the outside good would not depend on characteristics and price of other products than $j$:

$$\frac{s_j}{s_0} = e^{X_j\beta + \alpha p_j + \xi_j}$$

Taking the logs:

$$ln\frac{s_j}{s_0} = X_j\beta + \alpha p_j + \xi_j$$

This looks like a simple linear regression, which can be estimated via OLS or IV.

- The question of endogeneity of $\xi_j$ still remains, especially the assumption of independence of $p_j$ and $\xi_j$: would we expect prices to be independent of some product characteristics?
- There is a number of instruments that were proposed:
  - Cost shifters - variables that shift costs but do not shift demand;

  - Berry, Levinsohn, Pakes (1995) [6]: In oligopolistic markets - characteristics of competitors (as a proxy for the closeness of competition);
  - Nevo (2001) [31]: In case of a cross-section of markets: prices of products in another markets.

- But we have just discussed that logit model is very restrictive. Can we derive anything like that for nested model and for random coefficients?
- There is an analytical expression for the Nested Logit:

$$ln\frac{s_j}{s_0} = X_j\beta + \alpha p_j + \sigma ln(s_{j|g}) + \xi_j$$

  where $s_{j|g}$ is the share of good $j$ in the nest $g$. Once again, if $\sigma = 0$ this simplifies to a simple logit model.

- What about Random Coefficients? Recall that we need to solve

$$s = s(X, p, \xi, \theta)$$

  where $\theta = \{\beta, \alpha, \sigma, \sigma_\alpha\}$. By taking logs, adding $\xi$ to both side and rearranging:

$$\xi = \xi + ln(s) - s(X, p, \xi, \theta)$$

- BLP prove that operator

$$T(\xi) = \xi + ln(s) - s(X, p, \xi, \theta)$$

  is a contraction mapping. By finding a fixed point we get the correspondence between $\xi$ and $\theta$.

- In case of Logit and Nested Logit, we could use IV to estimate parameters $\theta$. Analogously we could use GMM procedure. In case of Random Coefficients, GMM is the only option.
- Notice that GMM requires solving for the fixed point on every step of the estimation. For people who attended micro session yesterday - does it ring a bell? This class of algorithms got the name Nested Fixed Point algorithms in the literature (same procedure as used in estimation of dynamic games).

### 1.3.2 BLP 1995

- In their paper BLP use a slightly different version of utility function

$$u_{ij} = X_j\beta_i + \alpha ln(y_i - p_j) + \xi_j + \epsilon_{ij}$$

  where $y_i$ stands for household $i$ income. Does it make sense? Well, first is that it coincides with Cobb-Douglas utility function. Second is that this is an analogue of random coefficient distribution - we can use empirical distribution of income. Arguably good a-priori expectation of the shape of distribution.

- BLP also add supply side to the model. Assume firm $n$ owns $K$ cars. Then its profits are
$$P_n = \sum_k (p_k - mc_k) M s_k$$
where $mc_k$ stands for marginal costs. Assuming constant marginal costs the first order condition ofr price is
$$\frac{\partial P_n}{\partial p_j} = s_j + \sum_k (p_k - mc_k) \frac{\partial s_k}{\partial p_j} = 0 \tag{10}$$

- Specify marginal costs as
$$mc_j = exp(X_j \gamma + \omega_j)$$
and plug it back to (10):
$$s_j + \sum_k (p_k - exp(X_j \gamma + \omega_j)) \frac{\partial s_k}{\partial p_j} = 0 \tag{11}$$

- This gives us $J$ equations (number of products) and $J$ unobservables $\omega_j$. Equations can be inverted in terms of $\omega_j$. Notice that we know all components of (11) except parameters $\gamma$ and unobservables $\omega_j$ (how do we know $\frac{\partial s_k}{\partial p_j}$?).
- BLP use the following moment conditions to estimate parameters:
$$E\left[\begin{pmatrix} \xi_j \\ \omega_j \end{pmatrix} \mid X\right] = 0$$

- Writing it is sample analogues:
$$G_J(\theta) = \frac{1}{J} \sum_j \left[\begin{pmatrix} \xi_j(\theta) \\ \omega_j(\theta) \end{pmatrix} \otimes f_j(X)\right]$$
which is then multiplied by some (optimal) weighting matrix $B$ ($dim(\theta) \times J$) and set to zero[3] :
$$BG_J(\theta) = 0$$

- Notice that Berry's inverse works not only for an unobservable $\xi_j$, but also for the mean utility $\delta_j$. For Random Coefficients:
$$u_{ij} = X_j \beta + \alpha p_j + \xi_j + \left[\sum_m \sigma_m x_{jm} \tilde{\beta}_{im} + \sigma_\alpha p_j \tilde{\alpha}_i + \epsilon_{ij}\right] = \delta_j + \nu_{ij}$$
$$T(\delta) = \delta + ln(s) - s(X, p, \delta, \theta)$$

[3] We can also think of this as forming the objective
$$G_J(\theta)' A G_J(\theta)$$
and minimizing it with some optimal matrix $A$. Two methods are equivalent, and optimal $A$ is
$$A = Var(G_J(\theta))^{-1} = B'B$$

- If we are interested only in mean parameters $\beta$ and $\alpha$ we can simply use IV given the fixed points of $\delta$:
$$\delta_j = X_j\beta + \alpha p_j + \xi_j$$
This was proposed by Nevo (2001) in the context of market for cereals (Nevo uses cross-section of markets as opposed to BLP).
- In their paper BLP use GMM. However, nothing restricts us from specifying a particular error structure and use MLE.

### 1.3.3 MPEC method

- In some cases NFXP can be very slow - remember that it requires solving a contraction algorithm on each iteration. Judd and Su (2012) [11] proposed an alternative method (in the context of Rust's (1987) [33] engine replacement problem discussed yesterday). Can also be applied to BLP-type problems.
- Basic idea: instead of solving $\xi_j$ as a function of $\theta$ treat $\xi_j$ as a parameter subject to constraints. Define
$$G_J(\xi_j,\omega_j) = \frac{1}{J}\sum_j \left[\begin{pmatrix} \xi_j \\ \omega_j \end{pmatrix} \otimes f_j(X)\right]$$
and solve
$$min_{\omega_j,\xi_j,\theta}\ G_J(\xi_j,\omega_j)' A G_J(\xi_j,\omega_j)$$
subject to
$$s_j = s_j(X,p,\xi,\theta) \quad \forall j$$
$$s_j + \sum_k (p_k - exp(X_j\gamma + \omega_j))\frac{\partial s_k}{\partial p_j} = 0 \quad \forall j$$
- Judd and Su claim that in some cases (i.e. Rust's engine replacement) this can be approximately 50 times faster than solving NFXP. Currently in the literature there is an ongoing debate if they are correct or not.

## 1.4 Alternatives and extensions

### 1.4.1 Probit

- Alternative way of dealing with correlation of the error terms is to use probit.
- As before, consumer utility is
$$u_{ij} = X_j\beta - \alpha p_j + \epsilon_{ij}$$
- Assume
$$\epsilon_i \sim N(0,\Sigma)$$
where $\Sigma$ is a $(J \times J)$ dimension matrix.

- Once again, we are back in the case of $J^2/2$ parameters to estimate.
- Another problem is that analytic form of the integral that we used in (4) is no longer available. In fact, there is no analytic solution to this $J$ dimensional integral.
- However, there is a number of nice properties of normal distribution, i.e. conjugacy of two normals. This made probit estimation popular among Bayesians.

### 1.4.2 Extensions

- So far we have looked at cases were agents:
  - Have one to one mapping from purchase to consumption;
  - Are not forward looking;
  - Choose today independently from yesterday;
  - Buy only one type of good;
  - Choose the type of good, not the quantity;
  - Consider all products that are available for purchase;
  - Have stable and known (to them) preferences.
- We have to relax the first assumption when we think about storable goods (i.e. soft drinks): people might behave strategically and respond to promotions, therefore maintaining a stock of a product. Most of the papers that relax the purchase-consumption link involve dynamics.
- Another reason to include dynamics is to assume that past purchases can drive current purchases, i.e. due to addiction (Becker/Murphy 1988 [4]), cost of thinking (Hoyer 1984 [24]), habit formation (Bronnenberg et al. 2012 [8])
- Discrete choice does not allow for complementarity between the products; a number of papers aimed to relax it: Hendel (1999) [23], Gentzkow (2007) [17], Rossi et al. (2002) [25], etc.
- Some papers when even further and tried to consider several markets (Pradeep 2007 [12])
- To estimate most of these models we need individual level data.

## 1.5 Individual level data

### 1.5.1 Estimation

- There is an increase in availability of individual level data sets (i.e. on consumer packaged goods, such as IRI Data Set (Bronnenberg et al. 2007 [9]) or Nielson Scanner Data);
- Assume we observe consumer $i$ purchasing good $j$ at time period $t$.

- Consider a simple utility form:

$$u_{ijt} = X_{jt}\beta + \alpha p_{jt} + \epsilon_{ijt}$$

Before we assumed logit error to get the expression for the market shares. Now for every particular choice we can compute the probability of this choice:

$$Pr(i \text{ choose } j)(X, p) = \int \cdots \int_{A_j} f(d\epsilon_{i1}, \cdots, d\epsilon_{iJ}) \tag{12}$$

- Making logit assumption on $\epsilon_{ijt}$ we get

$$Pr(i \text{ choose } j)(X, p) = \frac{e^{X_{jt}\beta + \alpha p_{jt}}}{\sum_k e^{X_{kt}\beta + \alpha p_{kt}}} \tag{13}$$

- In case of aggregate level data we observed the market shares. In case of individual data we observe choices.
- Intuitive way to approach this estimation problem is to maximize probabilities of observed choices. Can be done via maximum likelihood given a particular assumption on the error term. Likelihood specification is

$$L = \prod_i \prod_t \prod_j Pr(i \text{ choose } j)(X, p)^{I(i \text{ choose } j)}$$

Can easily rewrite it as a log-likelihood:

$$logL = \sum_i \sum_t \sum_j I(i \text{ choose } j) log(Pr(i \text{ choose } j)(X, p))$$

- In the same way we can approach nested logit or random coefficients.
- Marketing literature specifies another type of models: latent class models. $\alpha$ and $\beta$ are assumed to be different for different segments of buyers (say $S$ segments):

$$u_{ijt} = X_{jt}\beta_s + \alpha_s p_{jt} + \epsilon_{ijt}$$

- Assign some probability $p_s$ of consumer being in each segment. Then we can write down the likelihood as

$$L = \prod_i \sum_s Pr(i \text{ is in } s) \prod_t \prod_j Pr(i \text{ choose } j)(X, p)^{I(i \text{ choose } j)}$$

- However, latent class models are just a particular case of random coefficients (with $\alpha$ and $\beta$ having some discrete distribution)
- So how can we adjust this simple model to account for various extensions?

### 1.5.2 State Dependence

- Simplest way to incorporate dynamics is to include a variable $l_{ijt}$ that captures past choices into the model:

$$u_{ijt} = X_{jt}\beta + \alpha p_{jt} + \lambda l_{ijt} + \epsilon_{ijt}$$

- Guardini and Little (1983) [20] specify $l_{iht}$ as a weighted average of past time period $l_{ijt-1}$ and choice at period $t-1$:

$$l_{ijt} = \omega l_{ijt-1} + (1-\omega)I(i \text{ choose } j \text{ at } t-1)$$

  with $\omega$ being between 0 and 1. If we choose omega big enough $l_{ijt}$ is highly autocorrelated, which implies that choice in period $t_1$ would affect choices in periods $t_2 = t_1 + T$ with $T$ being quite big (by "quite big" here I mean around 10). What are the problems with this specification?

- In the more recent marketing papers $\omega$ is simply set to 0: utility from choosing a good today affects only your choice in the next period.

- Heckman[4] (1991) [22] discuss the difference in structural and spurious state dependence, and points out the identification difficulty in case of both heterogeneity of agents and state dependence.

- In Dube et al. (2012) [15] authors use

$$u_{ijt} = \alpha_{ij} + \eta_i p_{jt} + \gamma_i I\{s_{it} = j\} + \epsilon_{ijt}$$

  where $p_{jt}$ is price, and $s_{it}$ is state (previous purchase). They allow for a very flexible form of parameters $\{\alpha_1, \cdots, \alpha_J, \eta, \gamma\}$ - mixture of 5 normal distributions. To identify the parameters authors use MCMC methods (discussed today in data analysis section).

- Dube et al. test the source of state dependence (structural versus spurious) via including past prices instead of state variable: past prices serve as a proxy for previous choice. This approach has some limitations (which ones?)

- Marketing literature studying CPG (consumer packaged goods) usually do not instrument for prices - prices are assume independent of consumer shock in particular period $t$ (usually purchase occasion) as changing prices in each period can be very costly. For obvious reasons a bad strategy in car market analysis.

### 1.5.3 Extensions

- Among other way the basic model can be extended are:
  - Relax the assumption that people know their preferences over the product - incorporate learning (Ackerberg 2003 [1], Erdem and Keane 1996 [16], Gentzkow and Shapiro 2008 [18], Sridhar et al. 2005 [30]) and forgetting (Mehta 2004 [28]).

[4] In the cited paper and a number of other papers

- Relax the assumption that people have perfect information about prices - search models: non-sequential and sequential search (Stigler 1961 [34], De los Santos et al. 2011 [13], Kim et al. 2009 [26], Moraga-Gozalez 2006 [29])
- Relax the assumption that people consider all products in the markets - considerations sets

# 2 Production function

- Production function is a fundamental component of all economies, relate inputs to outputs
$$Y_j = A_j K_j^{\beta_k} L_j^{\beta_l}$$
where $Y_j$ - output of firm $j$, $K_j$ - capital inputs, $L_j$ - labor inputs, and $A_j$ - some technology level. In logs:
$$y_j = \beta_0 + \beta_k k_j + \beta_l l_j + \epsilon_j$$
where $log(A_j) = \beta_0 + \epsilon_j$. $\beta_0$ can be interpreted as mean efficiency across firms, and $\epsilon_j$ can contain innate technology, management differences, measurement errors or any unobserved sources of variation of output.

- OLS estimation? Problematic, as inputs are correlated with $\epsilon_j$.

- I.e. assume competitive input and output markets, capital being fixed, $\epsilon_j$ fully observed by a firm, and labor choice impacts only current profits. Then
$$L_j = \left[\frac{p_j}{w_j}\beta_l A_j K_j^{\beta_k}\right]^{1/(1-\beta_l)} = \left[\frac{p_j}{w_j}\beta_l e^{\beta_0+\epsilon_j} K_j^{\beta_k}\right]^{1/(1-\beta_l)} \tag{14}$$
where $w_j, r_j, p_j$ denote prices of inputs and output. That is, $L_j$ depends on $\epsilon_j$.

- More generally, $K_j$ also depends on $\epsilon_j$, although researchers consider that bias on $\beta_l$ is usually more severe than bias on $\beta_k$ (Why? Will it change if we think about neoclassical growth model?)

- Another source of endogeneity: selection due to firm attrition. Firms observe $\epsilon_j$ and decide whether to stay in the market. We observe *selected* sample (conditional on staying in the market).

- I.e. assume monopolistic firms exogenously endowed with fixed level of capital. Assume away dynamics. Firm would exit if
$$P(\epsilon_j, K_j; p_j, w_j, \beta) < \Phi$$
where $\beta \equiv (\beta_0, \beta_l, \beta_k)$ and $\Phi$ is the selloff value. The higher the $K_j$, the lower the value of $\epsilon_j$ the firm can bear. Hence, selection will generate *negative* correlation between $\epsilon_j$ and $K_j$.

## 2.1 Traditional solutions

- Lets generalize the production function:
$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \omega_{jt} + \eta_{jt} \tag{15}$$
where index $t$ stand for time period, and $\epsilon_{jt}$ is decomposed into $\omega_{jt}$ (observed to firm) and $\eta_{jt}$ (unobserved to firm). Now endogeneity comes only from $\omega_{jt}$ (call it "unobserved productivity").

### 2.1.1 IV estimation

- One way to deal with it is *instrumental variables* approach: find variables which are correlated with inputs but uncorrelated with $\omega_{jt}$. Examining (14) we can propose using $r_{jt}$ and $w_{jt}$: they affect input and under some conditions are independent of $\omega_{jt}$ (what are the conditions?)
- Even if input price are valid instruments (uncorrelated with $\omega_{jt}$), there are several problems:
  - Firms do not report input prices, and when they do, they report adjusted input prices (i.e. average wage per worker) (why is it a problem? think about quality of labor);
  - We require variation in the instruments *across firms*, while input markets are often national in scope;
  - $\omega_{jt}$ might depend on some unobserved input; this often makes $r_{jt}$ and $w_{jt}$ invalid instruments (why?);
  - IV approach does not address selection problem: input prices would be correlated with $\omega_{jt}$ (which way? might be bad question)
- One possibility to correct for selection is by using selection models (i.e. Heckman 1979 [21])

### 2.1.2 Fixed effects estimation

- A second traditional approach is *fixed effects* estimation.
- Assume that $\omega_{jt}$ is constant over time. Then equation (15) is

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \omega_j + \eta_{jt} \tag{16}$$

- Depending on the amount of assumptions on $\eta_{jt}$ we want to make we can use various ways to deal with fixed effects. I.e. assuming strict exogeneity ($\eta_{jt}$ are uncorrelated with input choices $\forall t$) we can estimate

$$(y_{jt} - \bar{y}_j) = \beta_k(k_{jt} - \bar{k}_j) + \beta_l(l_{jt} - \bar{l}_j) + (\eta_{jt} - \bar{\eta}_j) \tag{17}$$

  via LS procedure.
- This approach solves selection problem, but might also have some problems:
  - Assumption $\omega_{jt} = \omega_j$ is quite strong (we may want to examine economic environmental changes which affect $\omega_{jt}$);
  - If there is measurement error in inputs, fixed effect can produce even worse results then OLS (Griliches and Hausman 1986 [19]);
  - In practice fixed effect produce unreasonably low estimates of capital coefficients (i.e. returns to scale of 0.6)

## 2.2 Olley and Pakes (1996)

- Olley and Pakes (1996) [32] (OP) propose a different approach to solve the endogeneity problem;
- They make a number of specific assumptions:
  - Time of choosing inputs and dynamic nature of inputs;
  - Econometric unobservable is scalar;
  - Investment level is strictly monotonic in the scalar unobservable.
- We start with a basic OP model without selection, and add selection issue later.

### 2.2.1 OP basic model

- Assume production function similar to (15):

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_a a_{jt} + \omega_{jt} + \eta_{jt} \tag{18}$$

  where additional input $a_{jt}$ is the natural log of the age of a plant.
- Assume that $\omega_{jt}$ follows exogenous first order Markov process:

$$p(\omega_{jt+1}|I_{jt}) = p(\omega_{jt+1}|\omega_{jt})$$

  where $I_{jt}$ is the information set of the firm at time $t$. $p(\omega_{jt+1}|\omega_{jt})$ is stochastically increasing in $\omega_{jt}$.
- This is both econometric assumption on unobservables and economic assumption on firms' expectations about $\omega_{jt+1}$.
- Labor is assume to be a non-dynamic input;
- Capital is determined by a deterministic dynamic process:

$$k_{jt} = \kappa(k_{jt-1}, i_{jt-1}) = (1-\delta)k_{jt-1} + i_{jt-1}$$

  This implies that capital of the firm in $t$ is determined by capital of the firm in $t-1$ and investments in $t-1$.
- This implies that unexpected change in $\omega_{jt}$, $\omega_{jt} - E(\omega_{jt}|\omega_{jt-1})$, is independent of $k_{jt}$.
- Firm maximizes expected profits; single period profit is

$$\pi(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t) - c(i_{jt}, \Delta_t)$$

  This profit is conditional on optimal level of $l_{jt}$ as it is a function of $\omega_{jt}$, and $\Delta_t$ represents economic environment.
- Given this, firm solves

$$\begin{aligned} V(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t) &= \max\Big\{\Phi(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t), \max_{i_{jt}\geq 0}\{\pi(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t) - \qquad (19)\\ &- c(i_{jt}, \Delta_t) + \beta E[V(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t)|k_{jt}, a_{jt}, \omega_{jt}, \Delta_t, i_{jt}]\}\Big\} \qquad (20)\end{aligned}$$

  where $\Phi(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t)$ represents the selloff value.

- Optimal exit decision rule can be written as

$$\chi_{jt} = \begin{cases} 1 & \text{if } \omega_{jt} \geq \bar{\omega}_t(k_{jt}, a_{jt}) \\ 0 & \text{otherwise} \end{cases}$$

  under assumption that $\Phi(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t)$ does not increase in $\omega_{jt}$ too fast (why?) and if equilibria exists.

- Investment demand function is

$$i_{jt} = i(k_{jt}, a_{jt}, \omega_{jt}, \Delta_t) = i_t(k_{jt}, a_{jt}, \omega_{jt}) \tag{21}$$

  Pakes (1994) through in some conditions under which this $i_t(\cdot)$ is strictly increasing in $\omega_{jt}$ (is it intuitive?)

- Given that (21) is strictly increasing in $\omega_{jt}$, we can invert it:

$$\omega_{jt} = i_t^{-1}(k_{jt}, a_{jt}, i_{jt}) \tag{22}$$

  Exploits the assumption that $\omega_{jt}$ is the only unobservable in the investment equation (other assumptions?).

- Use this inversion in (18):

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_a a_{jt} + i_t^{-1}(k_{jt}, a_{jt}, i_{jt}) + \eta_{jt} \tag{23}$$

- Big question is how to estimate (23). One way is to try to solve for the parametric form on $i_t^{-1}(\cdot)$. Note that this requires complete specification of the underlying model (demand functions, sunk costs, etc.), and then solving a complicated dynamic game (finding solution in (19)).

- Another way is to use non-parametric estimation (i.e. fit a polynomial in $k_{jt}, a_{jt}, i_{jt}$ to approximate $i_t^{-1}(k_{jt}, a_{jt}, i_{jt})$, or kernel methods). The downside is that this polynomial will be collinear with a constant, $k_{jt}$ and $a_{jt}$, so we would not be able to identify $\beta_0, \beta_k, \beta_a$.

- Rewrite (23) as

$$y_{jt} = \beta_l l_{jt} + \phi_t(k_{jt}, a_{jt}, i_{jt}) + \eta_{jt} \tag{24}$$

  where $\phi_t(k_{jt}, a_{jt}, i_{jt}) = \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + i_t^{-1}(k_{jt}, a_{jt}, i_{jt})$. Notice that we need to allow $\phi_t$ to be different for different $t$.

- From (24) we can estimate $\hat{\beta}_l$ and $\hat{\phi(\cdot)}$ via LS. This is the first stage of the estimation.

- In the second stage we would like to identify $\beta_k$ and $\beta_a$. In the first stage we got an estimate of

$$\phi(k_{jt}, a_{jt}, i_{jt}) = \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + \omega_{jt}$$

- Conditional on a set of parameters $\beta_0$, $\beta_k$ and $\beta_a$ we can compute

$$\hat{\omega}_{jt} = \hat{\phi}(k_{jt}, a_{jt}, i_{jt}) - (\beta_0 + \beta_k k_{jt} + \beta_a a_{jt})$$

- Decompose $\omega_{jt}$ into expected and unexpected parts:

$$\omega_{jt} = E(\omega_{jt}|I_{it-1}) + \xi_{jt} = E(\omega_{jt}|\omega_{jt-1}) + \xi_{jt} = g(\omega_{jt-1}) + \xi_{jt}$$

  Notice that $k_{jt}$ and $a_{jt}$ are functions of the information set in time $t-1$. That is, by construction $\xi_{jt}$ is uncorrelated with $k_{jt}$ and $a_{jt}$. (reason that $g$ does not depend on time - assumption on transition matrix)

- Now we can rewrite (18) as

$$\begin{aligned} y_{jt} - \beta_l l_{jt} &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + \omega_{jt} + \eta_{jt} = \\ &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g(\omega_{jt-1}) + \xi_{jt} + \eta_{jt} = \\ &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g(\phi_{jt-1} - \beta_0 - \beta_k k_{jt-1} + \beta_a a_{jt-1}) + \xi_{jt} + \eta_{jt} = \\ &= \beta_k k_{jt} + \beta_a a_{jt} + \tilde{g}(\phi_{jt-1} - \beta_k k_{jt-1} + \beta_a a_{jt-1}) + \xi_{jt} + \eta_{jt} \end{aligned}$$

- Using $\hat{\beta}_l$ and $\hat{\phi}_{jt-1}$, and treating $\tilde{g}$ non-parametrically, we can estimate this equation (i.e. with NLLS in case of polynomial approximation) using $\xi_{jt} + \eta_{jt}$ as error term.

- Ackerberg et al. (2005) [3] propose to use GMM with a set of moments $E(k_{jt}\xi_{jt}) = 0$ and $E(a_{jt}\xi_{jt}) = 0$. They state that this produces lower variance compared with OP approach.

- Wooldridge (2004) proposes using one-stage procedure with the following moments:

$$E\left[\begin{array}{c} \eta_{jt} \otimes f_1(k_{jt}, a_{jt}, i_{jt}, l_{jt}) \\ (xi_{jt} + \eta_{jt}) \otimes f_2(k_{jt}, a_{jt}, k_{jt-1}, a_{jt-1}, i_{jt-1}) \end{array}\right] = 0$$

  with appropriate choice of $f_1$ and $f_2$.

### 2.2.2 Endogenous exit

- Now we relax the assumption that exit is exogenous.

- Remember that exit decision depends on

$$\chi_{jt} = \begin{cases} 1 & \text{if } \omega_{jt} \geq \bar{\omega}_t(k_{jt}, a_{jt}) \\ 0 & \text{otherwise} \end{cases}$$

- Nothing change in the first stage (24): we found a proxy for $\omega_{jt}$, which takes care of endogeneity.

- The second stage regression contains both $\eta_{jt}$ and $\xi_{jt}$. Firms decision to exit depends on $\omega_{jt}$, and $\xi_{jt}$ is a component of $\omega_{jt}$.

- Rewrite the second stage regression as

$$E[y_{jt} - \beta_l l_{jt}|I_{jt-1}, \chi_{jt} = 1] = \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + E[\omega_{jt}|I_{jt-1}, \chi_{jt} = 1] \quad (25)$$

  where $\eta_{jt}$ canceled out as it is orthogonal to $I_{jt-1}$ by construction. Also notice that $k_{jt}$ and $a_{jt}$ are known at $t-1$.

- Given that $\omega_{jt}$ is a Markovian process we can rewrite $E[\omega_{jt}|I_{jt-1}, \chi_{jt} = 1]$ as

$$E[\omega_{jt}|I_{jt-1}, \chi_{jt} = 1] = g(\omega_{jt-1}, \bar{\omega}_t(k_{jt}, a_{jt})) \tag{26}$$

- Once again, we do not know $\bar{\omega}_t(\cdot)$, but we can model it non-parametrically. However, this would not allow us to make any inference on $\beta_0, \beta_k, \beta_a$.
- We can go another way and define the propensity score

$$Pr(\chi_{jt} = 1|I_{jt-1}) = Pr(\chi_{jt} = 1|\omega_{jt-1}, \bar{\omega}_t(k_{jt}, a_{jt})) = \phi(i_{jt-1}, k_{jt-1}, a_{jt-1}) = P_{jt} \tag{27}$$

- (27) can be estimated non-parametrically, giving us $\hat{P}_{jt}$.
- Under some conditions we can invert (27) in terms of $\bar{\omega}_t(k_{jt}, a_{jt})$:

$$\bar{\omega}_t(k_{jt}, a_{jt}) = f(w_{jt-1}, P_{jt}) \tag{28}$$

- Rewriting (25) using the results in (26), (27) and (28) we get

$$\begin{aligned} E[y_{jt} - \beta_l l_{jt}|I_{jt-1}, \chi_{jt} = 1] &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g(\omega_{jt-1}, f(w_{jt-1}, P_{jt})) = \\ &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g'(\omega_{jt-1}, P_{jt}) = \\ &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g'(\phi_{jt-1} - \beta_0 - \beta_k k_{jt-1} - \beta_a a_{jt-1}, P_{jt}) \end{aligned}$$

- That is, as before we can write

$$\begin{aligned} y_{jt} - \beta_l l_{jt} &= \beta_0 + \beta_k k_{jt} + \beta_a a_{jt} + g'(\phi_{jt-1} - \beta_0 - \beta_k k_{jt-1} - \beta_a a_{jt-1}, P_{jt}) + \zeta_{jt} = \\ &= \beta_k k_{jt} + \beta_a a_{jt} + \tilde{g}(\phi_{jt-1} - \beta_k k_{jt-1} - \beta_a a_{jt-1}, P_{jt}) + \zeta_{jt} \end{aligned}$$

  where we compress $\beta_0$ into the non-parametric function $\tilde{g}(\cdot)$ as before.
- Substitute $\hat{P}_{jt}$, $\hat{\phi}_{jt}$ and $\hat{\beta}_l$ and estimate it with NLLS, approximating $\tilde{g}(\cdot)$ either with polynomial or with kernel function.

## 2.3 Levinsohn and Petrin

- What if we observe some periods with zero investment? Levinsohn and Petrin (2003) [27] (LP) report more than 50% percent of firms with zero investment in their data set.
- Model of OP allows to adjust for that: OP require investments to be strictly monotonic only on the subset of data. That is, we can use the data with $i_{jt} > 0$.
- This produce consistent results: conditioning on $i_{jt} > 0$ does not have anything do say about both $\eta_{jt}$ and $\zeta_{jt}$.
- However, for obvious reasons dropping 50% of the data is inefficient.
- LP propose an alternative estimation method for the case of zero investments: they propose to use some other variables (not investments) to proxy for $\omega_{jt}$. They use intermediate outputs (i.e. electricity, fuels, materials) which are rarely zero.

- Rewrite (18) with additional input $m_{jt}$ (materials):

$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + \omega_{jt} + \eta_{jt} \tag{29}$$

  LP assume that $m_{jt}$ is a non-dynamic input which is determined by

$$m_{jt} = m_t(k_{jt}, \omega_{jt}) \tag{30}$$

  Assume that $m_t$ is monotonic in $\omega_{jt}$ (LP state conditions that are sufficient for this).

- Materials is a static choice, so parametric assumption can be imposed here
- Given this, we can proceed as in OP model. Start with inverting $m_t$ function:

$$\omega_{jt} = m_t^{-1}(k_{jt}, m_{jt}) \tag{31}$$

  Substitute (31) into (29):

$$\begin{aligned} y_{jt} &= \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + \beta_m m_{jt} + m_t^{-1}(k_{jt}, m_{jt}) + \eta_{jt} = \\ &= \beta_l l_{jt} + \phi_t(k_{jt}, m_{jt}) + \eta_{jt} \end{aligned}$$

  Estimate $\beta_l$ and $\phi_{jt}$, and proceed to the second stage where

$$\tilde{y}_{jt} = \beta_k k_{jt} + \beta_m m_{jt} + \tilde{g}(\phi_{jt-1} - \beta_k k_{jt-1} - \beta_m m_{jt-1}) + \xi_{jt} + \zeta_{jt} \tag{32}$$

  where we use a nonparametric estimate for $\phi_{jt}$ from the first stage and non-parametrically estimate $\tilde{g}(\cdot)$.

## 2.4 Collinearity issues

- Assumptions required by OP and LP models are quite restrictive.
- It turns out that even these assumptions hold, there might be some identification issues.
- Ackerberg, Caves and Frazer (2006) (ACF) point to possible collinearity of $l_{jt}$ and non-parametric function of $k_{jt}, a_{jt}, i_{jt}$ in the first stage ($k_{jt}, m_{jt}$ in case of LP).
- Examine the case of LP : $l_{jt}$ and $m_{jt}$ are chosen simultaneously, perfectly variable, non-dynamic inputs. It is intuitive to expect that they are decided in a similar way:

$$m_{jt} = m_t(\omega_{jt}, k_{jt}) \tag{33}$$

$$l_{jt} = l_t(\omega_{jt}, k_{jt}) \tag{34}$$

  From the results in (30) we know that we can invert $m_j$ function in $\omega_{jt}$. Plug it into $l_t$ instead of $\omega_{jt}$:

$$l_{jt} = l_t(m_t^{-1}(k_{jt}, m_{jt}), k_{jt}) = g(m_{jt}, k_{jt}) \tag{35}$$

  We get that $l_{jt}$ is a function of $m_{jt}, k_{jt}$. But if we want to use non-parametric identification on function of $\omega_{jt}, k_{jt}$ we would not be able to identify $\beta_l$ given that $l_{jt}$ is a function of $m_{jt}, k_{jt}$.

- The same problem holds for OP model.
- ACF discuss various additional assumption that make DGP such that $\beta_l$ can be identified. They find two specifications for LP model: one relies on optimization error in $l_{jt}$ but require no optimization error in $m_{jt}$, other sets additional timing assumptions and assumes another independent shock between time periods which affects $l_{jt}$ is independent of everything else (see ACF, section 3.1). Both seem to be quite unrealistic.
- For OP model there is a more plausible assumption:
  - Assume that $l_{jt}$ is chosen between $t-1$ and $t$ (it is not perfect variable now). Denote this point $t-b$, $0 < b < 1$;
  - Assume that $\omega_{jt}$ is Markovian between subperiods $t-1$, $t-b$, etc;
  - Then
  $$l_{it} = l_t(\omega_{it-b}, k_{jt})$$
  which implies that $l_{jt}$ is not a function of $\omega_{jt}$;
  - Now there is not necessarily collinearity as $l_{jt}$ is not a function of $\omega_{jt}$.
- This method would not work in case of LP as it requires $l_{jt}$ to be set after $m_{jt}$ (otherwise $l_{jt}$ would enter a non-parametric equation and $\beta_l$ would not be identified in the first stage.

## 2.5 ACF Procedure

- But is it necessary to identify $\beta_l$ in the first stage?
- ACF propose a procedure that does not estimate any coefficients in the first stage, but only gets rid of $\eta_{jt}$.
- Assume the same structure as for OP correction procedure before:
  - $l_{jt}$ is not perfectly variable and is chosen in $t-b$, $0 < b < 1$, that is before materials;
  - $\omega_{jt}$ is Markovian between sub-periods $t-1$, $t-b$, etc.
- Now firm's material inputs would be determined by
$$m_{jt} = m_t(\omega_{jt}, k_{jt}, l_{jt}) \tag{36}$$
- as before, invert (36) and plug into the output function to get the first stage equation:
$$y_{jt} = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + m_t^{-1}(m_{jt}, k_{jt}, l_{jt}) + \eta_{jt} \tag{37}$$
  $\beta_l$ is clearly not identified, but we get an estimate $\hat{\Phi}_{jt}$ of the term
  $$\Phi_{jt}(m_{jt}, k_{jt}, l_{jt}) = \beta_0 + \beta_k k_{jt} + \beta_l l_{jt} + m_t^{-1}(m_{jt}, k_{jt}, l_{jt})$$
- ACF proceed the same way as LP, but two moment conditions are required as both $\beta_l$ and $\beta_k$ should be estimated.

- Decompose $\omega_{jt}$ as before:

$$\omega_{jt} = E(\omega_{jt}|I_{it-1}) + \xi_{jt} = E(\omega_{jt}|\omega_{jt-1}) + \xi_{jt} = g(\omega_{jt-1}) + \xi_{jt}$$

- One set of moment conditions could be also used before (proposed by ACF for OP):

$$E(\xi_{jt}|k_{jt}) = 0$$

- This would not hold for $l_{jt}$ as it is chosen after $t-1$, so it might be correlated with $\xi_{jt}$. However, $l_{jt-1}$ was chosen in $t-b-1$, and is in the information set $I_{jt-1}$. That is, second set of moments can be

$$E(\xi_{jt}|l_{jt-1}) = 0$$

- To recover $\xi_{jt}$ we compute the implied $\omega_{jt}(\beta_k, \beta_l)$:

$$\omega_{jt}(\beta_k, \beta_l) = \hat{\Phi}_{jt} - \beta_k k_{jt} - \beta_l l_{jt}$$

  and non-parametrically regress $\omega_{jt}(\beta_k, \beta_l)$ on $\omega_{jt-1}(\beta_k, \beta_l)$. Residuals of the regression are the implied $\xi_{jt}(\beta_k, \beta_l)$

- Use sample analogue of the moments

$$\frac{1}{T}\frac{1}{N}\sum_t \sum_i \xi_{jt}(\beta_k, \beta_l) \begin{pmatrix} k_{jt} \\ l_{jt-1} \end{pmatrix}$$

- LP and OP use moments

$$E(\xi_{jt}|l_{jt-1}) = 0$$

  to test for the over-identifying restriction. This is different from what ACF do as OP and LP do inference on $\beta_l$ in the first stage.

- Assumption is ACF can also be adjusted so that $b = 1$, so the second set of moment conditions becomes

$$E(\xi_{jt}|l_{jt}) = 0$$

- ACF also compare their procedure with dynamic panel models: ACF appears to be more flexible regarding serially correlated transmitted error $\omega_{jt}$ but less flexible regarding non-transmitted error $\eta_{jt}$ and fixed effect $\alpha_j$. Also, ACF requires additional assumptions.

- Ackerberg et al. (2007) [2] discuss how one can relax scalar unobservable assumption.

# References

[1] Ackerberg, Daniel (2003), Advertising, Learning, and Consumer Choice in Experience Goods Markets: An Empirical Examination, *International Economic Review*, Vol. 44 (3), 10071040.

[2] Ackerberg, D., Benkard, L., Berry, S., and Pakes, A. Econometric Tools for Analyzing Market Outcomes, Chapter in *Handbook of Econometrics: Volume 6A*, Edited by J. Heckman and E. Leamer, North Holland 2007.

[3] Ackerberg, D., K. Caves and G. Frazer (2005). "Structural Identification of Production Function", *mimeo*, 2006, R&R Econometrica.

[4] Becker, Gary and Kevin M. Murphy (1988) "A Theory of Rational Addiction", *Journal of Political Economy* 96: 675-700, August 1988.

[5] Berry, Steven (1994), Estimating Discrete Choice Models of Product Differentiation, *RAND Journal of Economics*, 25, 242262

[6] Berry, Steven, James Levinsohn and Ariel Pakes (1995), Automobile Prices in Market Equilibrium, *Econometrica*, 60(4), 889917.

[7] Bresnahan, Timothy F, Scott Stern and Manuel Trajtenberg Market Segmentation, Transitory Market Power, and Rents from Innovation: Personal Computers in the late 1980's, *RAND Journal of Economics*, 1997.

[8] Bronnenberg, Bart J., Jean-Pierre Dube and Matthew Gentzkow (2012) The Evolution of Brand Preferences: Evidence from Consumer Migration, *American Economic Review*, 102 (6), 2472-2508

[9] Bronnenberg, Bart J., Michael W. Kruger, and Carl F. Mela (2008), The IRI Marketing Data Set, *Marketing Science*, 4 (July-August), 745-748.

[10] Cardell, N. S. (1997). "Variance components structures for the extreme value and logistic distributions with applications to models of heterogeneity". *Econometric Theory*, 13, 185-213.

[11] CheLin Su and Kenneth L. Judd, 2012. "Constrained Optimization Approaches to Estimation of Structural Models," *Econometrica*, Econometric Society, vol. 80(5), pages 2213-2230, 09.

[12] Chintagunta, Pradeep and Inseong Song (2007) "A Discrete/Continuous Model for MultiCategory Purchase Behavior of Households", *Journal of Marketing Research*, Volume XLIV, November, 595612

[13] De Los Santos, Babur, Ali Hortasu, and Matthijs R. Wildenbeest. 2012. "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior." *American Economic Review*, 102(6): 2955-80.

[14] Deaton, Angus and John Muellbauer (1980) "Economics and Consumer Behavior", Cambridge University Press

[15] Dube, Jean-Pierre, Gunter Hitsch and Peter Rossi (2012), "State dependence and alternative explanations for consumer inertia", *RAND Journal of Economics* Vol. 41, No. 3, Autumn 2010 pp. 417445

[16] Erdem, Tulin and Michael Keane (1996), DecisionMaking Under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets, *Marketing Science*, Vol. 15(1), 120.

[17] Gentzkow, Matthew (2007) "Valuing New Goods in a Model with Complementarities: Online Newspapers", *American Economic Review*, 97 (3). June 2007.

[18] Gentzkow, Matthew and Jesse Shapiro (2010), What Drives Media Slant? Evidence from U.S. Daily Newspapers, *Econometrica*, Vol. 78, No. 1 (January, 2010), 3571.

[19] Griliches, Z., Hausman, J. (1986). Errors in variables in panel data. *Journal of Econometrics* 31 (1), 93118.

[20] Guadagni, Peter M. and John D.C. Little (1983) "A Logit Model of Brand Choice Calibrated on Scanner Data", *Marketing Science*, Vol. 2, No. 3, pp. 203-238.

[21] Heckman, James (1979). Sample selection bias as a specification error, *Econometrica* 47 (1), 153161.

[22] Heckman, James (1991) Identifying the Hand Of The Past: Distinguishing State Dependence from Heterogeneity, *American Economic Review*, (May, 1991), 81(2), 75-79.

[23] Hendel (1999) Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns, *Review of Economic Studies*, April 1999, pp. 423-446.

[24] Hoyer, W.D. (1984) "An Examination of Consumer Decision Making for a Common Repeat Purchase Product". *The Journal of Consumer Research*, Vol. 11, No. 3

[25] Kim, Jaehwan, Peter Rossi and Greg Allenby (2002) "Modeling Consumer Demand for Variety", *Marketing Science*, 21, Summer, 229-250.

[26] Kim, Jun B., Paulo Albuquerque and Bart J. Bronnenberg (2010), "Online Demand Under Limited Consumer Search", *Marketing Science*, 29:6, 1001-1023

[27] Levinsohn, J., Petrin, A. (2003). Estimating production functions using inputs to control for unobservables, *Review of Economic Studies* 70 (2), 317341.

[28] Mehta, Nitin, Surendra Rajiv, and Kannan Srinivasan (2004), Role of Forgetting in MemoryBased Choice Decisions, *Quantitative Marketing and Economics*, Vol. 2, 107140.

[29] Moraga-Gonzalez, Jose Luis (2006) "Estimation of Search Costs", *Working Paper*, University of Groningen

[30] Narayanan, Sridhar, Puneet Manchanda, and Pradeep Chintagunta (2005), Temporal Differences in the Role of Marketing Communication in New Product Categories, *Journal of Marketing Research*, Vol. 42(August), 278290.

[31] Nevo (2001) Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry, *The RAND Journal of Economics*, 31(3), 395-421, 2000

[32] Olley, G.S., Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry, *Econometrica* 64 (6), 12631298.

[33] Rust, John, 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica*, Econometric Society, vol. 55(5), pages 999-1033, September.

[34] Stigler, George J. (1961) "The Economics of Information", *The Journal of Political Economy*, Vol. 69, No. 3 (Jun., 1961), pp. 213-225