Empirical Industrial Organization

June 8, 2015

1 Introduction

These notes follow Ackerberg et al (2011). Empirical IO has many objectives that differ from those I talked about for partial equilibrium continuous demand analysis.

The model setting is a general equilibrium environment, so both the supply side and demand side are players, and both suppliers and demanders are relevant when we think about welfare analysis. Objects of interest are, among other things: a) firm markups; b) consumer price responses; c) new goods; d) market structure and branding; e) the cost of living; f) firm decisions; g) production and firm cost functions.

The data environment is one where we observe at least prices and quantities at the market level over many markets. It may be also that we observe many periods. It may be also that we observe individual firm supplies and/or individual consumer demands within each market (or market/time period). The most typical data set is a market panel for a particular type of good (e.g., market-level quantities of different breakfast cereals in US cities over 1980-2010).

To address the issues faced in empirical IO, we need to construct models that can

1. Deal with market level data, even though the choice model is for individual consumers;
2. Deal with choice among a huge number of products, and be able to speak to new products;
3. Deal with discrete aspects of product choice, like brand choice, and discretely consumed products, like automobiles.

2 Market-level Demand Systems

The data we have are market-level quantity demands q_i for $i = 1, ..., N$, but the model we have is of individual consumer demands. Here, q_i is a $J-$vector, where J could be a big number of possible goods. Pakes (1986) suggested a simulation estimator to bridge this gap. Such an estimator uses several steps:

1. draw incomes y_{it} and characteristics vectors z_{it} for consumer i in market t from the actual consumer characteristics in each market, e.g., from auxiliary data on consumers characteristics in each market.

2. let some integrable demand model $\hat{q}_{it} = q(y_{it}, z_{it}; \theta)$, where θ is a parameter vector, give the quantity vector chosen by each household. It is most likely a nonlinear model.

3. define aggregate demands $\widehat{q}_i(\theta) = \sum_{i \in t} q(y_{it}, z_{it}; \theta)$ given the parameter vector.

4. find the parameter vector θ that brings $\widehat{q}_i(\theta)$ as close as possible to observed market demands q_i.

3 Products and Product Characteristics

If there J products (aka: goods) in a standard continuous demand system, then you have at least $J(J-1)/2$ parameters governing price effects (Slutsky symmetry gets rid of half of them) and $J-1$ parameters governing budget effects (this would be for the simplest model, where e.g., hicks demands are linear in all prices and the budget). When J gets large, this numbers of parameters get large. Further, it is quite hard to deal with new products in a continuous demand approach.

When we say that J is a big number, we are referring to the fact that you can buy on the order of 50,000 distinct products at a supermarket. If we treat each good as completely distinct, we would need to estimate roughly 1,250,000,000 price parameters. But, do we really think bran flakes and corn flakes are soooo different that they needed to be treated entirely separately?

A *characteristics model* is one wherein each product $j = 1...J$ actually delivers us a $K-$vector of characteristics x, and our utility is a function of those K characteristics, and not of any other information about good j. Thus, in a characteristics model, there is a vector function $x(q)$ that tells us how much of each characteristic we get when we buy the quantity vector q, a $J-$vector of products. Here, K is taken to be much smaller than J. Thus, the characteristics model takes the J dimensional product space and reduces it to a K dimensional characteristics space. That is, a characteristics model offers dimension reduction, from J products to $K << J$ characteristics. This is a big win on parameter reduction if K is small.

A *linear characteristics model* is one wherein the utility is a function of a linear index of those characteristics, so that $x = Aq$. Here, A is a KxJ matrix giving the amount of each of the K characteristics contained in a unit of each product.

Suppose utility follows a linear characteristics model. Then,

$$U(q) = U(x) = U(Aq).$$

This is a very simple model, with a few important features (see, Blow, Browning and Crawford 2008):

1. In characteristics space, every product defines a ray out of the origin. If you buy a unit of the product, you get A units of the characteristics.

2. The budget constraint can be expressed in terms of the end points of these rays. You can buy any quantity up to where you spend all your money on a particular product. Connecting the end points of the rays gives you the budget constraint.

3. Some end points of rays may lie inside the convex hull of other end points. These products are never bought, because other products provide the same characteristics at a better price. Such products are said to *suck*.

4. The convex hull of the end points of rays thus defines the budget constraint. It may not reach the axes (e.g., if there are no products that deliver just a single characteristic).

5. Indifference curves in x space just look like regular indifference curves, since the argument of utility is characteristics x.

6. You will never want to buy more products than there are characteristics. You might want to buy less products. If your indifference curve lands on an endpoint of 1 of the products' rays, then you will buy just 1 product (that one), but if your indifference curve lies on a flat bit of the convex hull, you will buy more than 1 product.

7. Characteristics models also have something to say about new goods. A new good is just a new (or changed) column of A. If we know indifference curves over x, then we can predict demand for something that has a different combination of characteristics x from what we've observed before. That is, we can draw a new ray out, and see what it does to the budget constraint.

4 Discrete Demands

The continuous demand model allows the consumer to choose any positive quantity of any product. But, in real life, you cannot buy 2 grams of rice—you have to buy a bag of rice. So, there is discreteness in real-life consumer demand decisions.

Suppose that for each product $j = 1, ..., J$ we could buy quantities only if they appeared on the list $q_1^j, .q_t^j.., q_{T_j}^j$. Here, T_j gives the number of choices of quantities for good j. Suppose, e.g., that we can buy tuna only in 100g tins. Then, q_t^j is multiples of 100g. The set $\{q_t^j\}_{j,t}$ defines all the possible combinations of quantities we could buy. This set defines the budget constraint, which will generally be a step function in quantity space.

Indifference curves are still over continuous quantities. All we have changed is the budget constraint. Nonetheless, this model has a few important features:

1. Consumers will always choose a quantity vector on a kink of the budget constraint. The flat parts are all dominated by the kinks.

2. Consumers will not in general spend all their money. This means that there must exist another good (sometimes called the "outside good", sometimes called "money") where that extra money can be spent. In a static model, this outside good may or may not generate utility. In a dynamic model, the outside good may be savings, which will generate utility in another period.

3. Because consumers land on kinks and there may be extra money lying around in the outside good, small price changes may not generate any behaviour response. This means, we may have to be careful about how we think about quantity derivatives with respect to price.

The simplest discrete demand model is one wherein the consumer can only choose zero or 1 of each product. In a 2-product world, this budget constraint looks like a square, with a point at 1,1. If the consumer cannot able afford 1,1, then their demand choice is a discrete choice between 1 unit of product 1 or 1 unit of product 2. This would be a binary choice problem.

5 Discrete Demands in Linear Characteristics Model

Putting together discreteness and characteristics means that the rays we positted in Section 3 above are dotted. That is, you cannot buy any quantity of a product yield its chararacteristics; rather, you can only buy certain fixed values of quantities of it.

If we further restrict the model so that the discreteness is binary for each product, we have a budget constraint that is a hypercube in characteristics space. If we further restrict that consumers never have enough money to buy more than 1 product, we have a multinomial choice problem: consumers can buy 1 of J products which deliver different amounts of the characteristics c.

For a multinomial choice problem, we can use standard limited dependent variable econometric methods, but we want to ensure that:

1. We allow for error term distributions that are not crazy;
2. We allow for both observed characteristics and for unobserved characteristics.

The latter point means that we have to allow for unobserved characteristcs in a nonlinear model (because discrete choice models are inherently nonlinear). Berry (1994) showed that for a wide class of discrete choice models, we can formulate unobserved characteristics in such a way that there is a one-to-one correpondence between choice probabilities and unobserved characteristics parameters under the model. Berry, Levinsohn, and Pakes (1995; BLP) show that for a narrower class of models, there is a contraction mapping that is linear in the unobserved characteristics so that these models can be estimated relatively easily.

6 BLP

You know you've made the big leagues when the acronym for a paper you wrote becomes a verb.

Consider a logit model for utility for a consumer i who can choose to consumer exactly 1 unit of 1 product among J products (this is very close to McFadden 1974 "random utility model"):

$$U_{ij} = x_j\beta - \gamma p_j + \xi_j + \varepsilon_{ij},$$

where x_j is the observed row-vector of characteristics of a unit of product j, p_j is its price, ξ_j is the value of an unobserved characteristic of product j, β is a parameter vector of valuations for each characteristic, and ε_{ij} is unobserved taste variation.

The optimal choice is the one where utility is the highest. Since there are unobserved variables, all consumers do not make the same choice. We solve the model for *choice probabilities*, which are driven by the distribution of ε_{ij} and the distribution of ξ_j. If we let these be extreme value distributions, then the choice probabilities follow the familiar logit form.

In particular, suppose that $\xi_j = 0$ and ε_{ij} are iid Type-1- Extreme-Value (T1EV) random variates (or suppose that $\xi_j + \varepsilon_{ij}$ are iid T1EV) satisfying $f(\varepsilon_{ij}) = \exp(-\exp(-\varepsilon_{ij}))\exp(-\varepsilon_{ij})$.

Then,

$$P[x_j = 1] = \frac{\exp(x_j\beta - \gamma p_j)}{\sum_k \exp(x_k\beta - \gamma p_k)}.$$

If ε_{ij} is iid (both across products j for a consumer i, and across consumers i for a product j), then the distribution of a consumer's preferences over products other than the product it bought does not depend on the product they bought. The iid assumption comes at some cost.

Suppose the price of the product you buy rises. Then, you may want to switch. The iid assumption means that every switcher sees every new alternative equally, and so would be equally likely to switch to it.

Suppose 2 products have the same choice probabilities. Then, they have the same price derivatives. This means they'd have the same markups (as these are determined in oligopoly by demand responses).

These problems with the iid errors do not disappear if we include the unobserved characteristic ξ_j. Rather, these just act like fixed effects in the logit probability expression above.

BLP address these problems by allowing for valuations β that vary across consumers. Let x_j include both observed product characteristics and p_j, the price of good j. (That is, let the price be one of the K characteristics.) Let z_i and ν_i be a T^o-vector observed consumer characteristics and a T^u-vector of unobserved consumer characteristics. Then, let utility be given by

$$U_{ij} = x_j\theta_i + \xi_j + \varepsilon_{ij},$$

where

$$\theta_i = \overline{\theta} + \Theta^o z_i + \Theta^u \nu_i.$$

This expression for utility is identical to the simple form given above in that it depends on observed characteristics, price (one of the characteristics now), an unobserved characteristic and an error term. Here, θ_i is a $K+1-$vector of parameters (K characteristics and 1 price) that varies across individuals. For a consumer with $z_i = \nu_i = 0$, $\theta_i = \overline{\theta}$. But, for other consumers, the vector θ_i adds the T^o-vector of observed characteristics z_i multiplied by the $T^o xK+1$-matrix of parameters Θ^o, and the T^u-vector of unobserved characteristics ν_i multiplied by the $T^u xK+1$-matrix of parameters Θ^u. The unobserved consumer characteristics break the restriction that everyone sees every alternative to their good identical.

All utilities are relative to that of an outside good (U_{i0}), so we think of these as "utility above the outside good". Now the assumption of iid ε_{ij} does not bite so terribly, because we have Θ^u to allow for correlation across equations. So, let ε_{ij} be iid T1EV (so that we get a multinomial logit at the end).

Now, we want to express this utility function in terms of the bit that is easy to deal with, and the hard bits. Substituting the valuations above into the utility function, we get

$$U_{ij} = \delta_j + x_j \Theta^o z_i + x_j \Theta^u \nu_i + \varepsilon_{ij},$$

where

$$\delta_j = x_j \overline{\theta} + \xi_j.$$

At the market level, we are going to have δ_j which are not troublesome because the unobserved characteristic does not vary across i and because $\overline{\theta}$ does not vary across i. However, we do have 2 potentially troublesome terms: the interaction between observed product characteristics x and observed consumer characteristics z; and the interaction between observed product characteristics x and unobserved consumer characteristics ν.

7 Estimation of BLP

Consider the case where we have product x market level data that gives the market share of every product in every market (it doesn't need to be a balanced panel though). Also, assume there exist some instruments w satisfying $E[\xi|w] = 0$. Since we only observed these data, we obviously do not observe any consumer characteristics z. All are loaded onto the unobserved consumer characteristics ν. We assume that we do know the distribution(s) of these consumer characteristics.

The parameters of the model are δ, θ. Note that δ includes ξ.

1. Conditional on the unobserved consumer characteristics ν, and given a particular parameter vector δ, θ, market shares have the logit form $\exp(\tau_j)/1 + \sum_s \exp(\tau_s)$. So, we integrate over the distribution of ν (there are no observed z's) to get market shares σ_j:

$$\sigma_j(\delta, \theta) = \int \frac{\exp(\delta_j + x_j \Theta^u \nu_i)}{1 + \sum_s \exp(\delta_s + x_s \Theta^u \nu_i)} f(\nu) \partial \nu.$$

   Nobody likes integrating. Pakes (1986) suggests sampling some ν from a distribution, and summing:

$$\sigma_j(\delta, \theta; P^{ns}) = \sum_r \frac{\exp(\delta_j + x_j \Theta^u \nu_{ir})}{1 + \sum_s \exp(\delta_s + x_s \Theta^u \nu_{ir})}.$$

   Here, P^{ns} is a distribution from which you sample s times for each of the n observations in a particular market. This gives a vector of σ_j given $\delta, \theta; P^{ns}$. This step requires a distribution P^{ns} to sample from. You might choose a normal, e.g..

2. Then, solve for δ by iterating. BLP (1995) show that this reduces to iterating the following linear equation

$$\delta_j^{new}(\theta) = \delta_j^{old}(\theta) + \ln s_j - \ln \sigma_j(\theta, \delta^{old}; P^{ns})$$

until $\sigma_j(\delta, \theta; P^{ns}) = s_j$ for all j. This means we can identify δ given θ, including identification of ξ given θ. In particular, with θ in our pocket and δ_j identified, we can compute $\xi_j = \delta_j - \sum_k x_{jk}\theta_k$. This iteration step would typically use a *while* loop in Stata.

3. Now we use our instruments for ξ. Find θ that bring ξ as close as possible to orthogonal to w. In particular, one may use GMM for the moment conditions $E[\xi|w] = 0$. This is not the friendliest GMM, though, because for each value of θ, we have to do step 2 above. Luckily, it is canned in the Stata module *blp*.

4. The GMM requires instruments w that are uncorrelated with unobserved characteristics ξ. One might well suspect that observed product characteristics x (and consumer characteristics z, if used) are thus elements of w. But are prices p uncorrelated with unobserved characteristics? This seems a far cry. One might use supply-side instruments for prices. Or, one might model the supply side of the market, e.g., by modeling the pricing equations for oligopolistic supply (this would still require instruments).

8 References

1. Berry, Steve, 1994, ”Estimating Discrete Choice Models of Product Differentiation,” RAND, vol. 25, no. 2, pp. 242-262.

2. Berry, S., J. Levinsohn and A. Pakes, 1995, ”Automobile Prices in Market Equilibrium,” Econometrica, vol. 63, no. 4, pp. 841-890.

3. Blow, Laura, Browning, Martin and Crawford, Ian. 2008. "Revealed Preference Methods for the Consumer Characteristics Model" Review of Economic Studies, vol 75, pp. 371-389.

4. Pakes, Ariel, 1986, ”Patents as Options: Some Estimates of the Value of Holding European Patent Stocks,” Econometrica, vol. 54, pp. 755-784.

5. Daniel Ackerberg, C. Lanier Benkard, Steven Berry, and Ariel Pakes, 2011, “Econometric Tools for Analyzing Market Outcomes”, Handbook Chapter.

6. McFadden, Daniel, 1974, “The measurement of urban travel demand." Journal of public economics 3.4: 303-328.

7. McFadden, Daniel, and Paul Zarembka, 1974, "Conditional logit analysis of qualitative choice behavior” in Frontiers in Econometrics, 105-142.