Studying Differentiated Product Industries using Plant-Level Data

Sergio DeSouza
Universidade Federal do Ceara

Abstract

This paper develops a methodology to uncover consumer preferences from a discrete-choice demand model of product differentiation using plant-level data. When prices and quantities are observed, the appropriate strategy for estimating such model is well developed. However, most plant-level data sets only report revenue (sales) and total cost, hampering initial attempts to estimate the model according to standard approaches. This paper offers a way to circumvent this problem by bringing the extra information provided by usually observable aggregate data to determine the relevant parameters. The methodology consists of solving for the demand parameter that matches the total quantity implied by the model of demand and supply to its observable counterpart. Once this parameter is determined it is possible to define welfare measures and perform counterfactual simulations. This methodology is applied to measure the economic impact of a regulatory agency imposing a monopoly break up.

I would like to thank James Tybout and Mark Roberts for their useful comments and the Brazilian agencies (CAPES and FUNCAP) for financing this research.
Citation: DeSouza, Sergio, (2006) "Studying Differentiated Product Industries using Plant-Level Data." *Economics Bulletin*, Vol. 12, No. 3 pp. 1-11
Submitted: January 16, 2006. **Accepted:** July 7, 2006.
URL: http://economicsbulletin.vanderbilt.edu/2006/volume12/EB-06L10002A.pdf

1. Introduction

Many differentiated product industry studies rely on detailed information on markets shares and prices to determine consumer preferences and perform policy analysis (see Berry,1994, and the literature surveyed therein). However, such detailed data sets may be difficult to obtain in many instances. For example, due to confidentiality issues or strategic purposes, some firms are unwilling to release the relevant data. On the other hand, official statistical agencies usually find more welcoming firms when gathering information for plant-level surveys of the manufacturing sector. One of the reasons may rely on the fact that, in most cases, firms are asked to report only sales revenue and input expenditures, rather than (more revealing) information on prices and quantities. Such surveys are commonly available for many countries and cover most industries in the manufacturing sector, providing an easily accessible source for empirical studies. Surprisingly, only a few studies (to my knowledge) have used this source to study differentiate product industries (Klette and Griliches,1996, Melitz,2004 and DeSouza,2004). They all rely on Klette and Griliches framework which assumes a CES demand system, product differentiation and monopolistic competition. However, this model may not be reasonable for many industries. Monopolistic competition assumes that firms are not big enough to influence the aggregate market variables and therefore a price change by one firm has an irrelevant effect on the demand of any other firm. This assumption says that each product has no neighbor in the product space, which strongly restricts interaction between products (Tirole, 1988).

Building on Katayama, Lu and Tybout (2003), KLT henceforth, this paper develops an alternative methodology to study (using plant-level data) differentiated product markets in which consumer preferences are given by a discrete-choice demand function and firms play according to the Bertrand model. This setup relaxes some of the restrictions imposed by the Klette and Griliches framework, but poses some extra challenges.

When prices and quantities are observed, the standard strategy goes as follows. First, one sets up a parametric theoretical model of consumer and producer choice in a differentiated product market. Then, one matches the model's predictions about market shares to their empirical counterparts to obtain the relevant parameters. Once these parameters are pinned down it is possible to perform simulation by changing the market environment and recalculating the equilibrium prices and quantities. This strategy has been pursued in a number of works. A well known example can be found in Berry, Levinsohn, and Pakes (1999), where they measure the economic impact of a voluntary export restraint placed on exports of automobiles from Japan to the United States.

However, most plant-level data sets do not come in a directly usable form. Usually, they report only revenue (sales) and total cost instead of price and quantities, hampering initial attempts to estimate the model according to the standard approach. To overcome this restriction, KLT, using assumptions on a discrete-choice demand system, technology and firms' behavior, show that, given the demand parameters, there is a mapping from revenue and cost data to marginal costs, quality, prices and quantities. But, due to the lack of observables (e.g. prices, quantities and product characteristics), it is not possible to estimate the demand parameters using standard inference methods like GMM or ML.

One way to solve this problem is to introduce more information into the model. KLT do this by imposing prior distributions on the unknown parameters and performing Bayesian inference. However, as the economic model and the data set available are insufficient to perform standard inference, identification relies heavily on the priors. Thus, the parameters estimates are prone to be very sensitive to the priors.

This paper uses the same ideas behind the KLT framework, i.e. it assumes a discrete-choice demand and price-setting firms to establish a mapping between observed revenue and cost data and firms' decisions. But this paper markedly differs from KLT's contribution with respect to the identification strategy. Rather than imposing prior distributions, this paper proposes bringing usually observable data on aggregate physical quantities to pin down the model parameters through a calibration technique.

Although it may be difficult to obtain detailed data on quantities at the plant level, the same is not true for aggregate variables. For instance, in the beverage sector, information on aggregate consumption of beer is widely available for many countries. Shortly, the methodology consists of matching the total quantity implied by the mapping to its observable counterpart. The intuition is that aggregate quantity also carries information on consumer preferences. Therefore, observing this aggregate measure help uncover the demand parameters.

After pinning down the preference parameters, qualities and marginal costs it is possible to construct consumer and producer welfare measures and perform counterfactual experiments. This paper last objective is to illustrate the methodology using plant-level data from the Colombian beer industry.

This paper is organized as follows. The following section introduces the demand and supply models. The ensuing section presents an empirical strategy to uncover the unobservables that takes into account the limitations imposed by commonly available plant-level data. The fourth section shows how to evaluate the welfare effects of different market environments (policy simulation) once the parameters of interest are determined. The fifth section shows an application of the methodology to the Colombian beer market. And finally, the last section presents some concluding remarks.

2. Model

In this section I lay out the model that governs consumers and producers decisions. The demand system has the familiar logit form while the supply (and equilibrium) equation is given by a price-setting game.

Demand Assume that consumers rank products according to their characteristics and prices. There are *N+1* choices in the market, *N* inside goods and one outside good. Consumer *i* chooses one of these goods, given prices p_j, quality ξ_j and idiosyncratic preferences ε_{ij}. Consumer *i* derives utility from consuming firm *j*'s product according to $u_{ij} = \delta_j + \varepsilon_{ij}$ (*j=0,1,..,N*), where $\delta_j = -\alpha p_j + \xi_j$.

McFadden (1981) shows that one can integrate out ε_{ij}, which is assumed to be an identically and independently distributed extreme value random variable, to obtain a closed form solution for the market shares of an inside good *j*

$$s_j = \frac{\exp[\delta_j - \delta_0]}{\sum_k \exp[\delta_k - \delta_0]} \; ; j=1,2,\ldots,N \tag{1}$$

Here, the subscript zero identifies the outside good. Further, taking the log-difference between s_j and s_0 the demand equation takes the simple linear relation

$$\ln s_j - \ln s_0 = \alpha(p_j - p_0) + \xi_j - \xi_0 \;; \quad j=1,2,\ldots,N \tag{2}$$

Supply When setting prices, firms take into account demand and cost determinants. Therefore, the pricing decision also contains valuable information on consumer preferences and market equilibrium. First, assume that each firm *f* produces a subset F_f of the goods sold in this market and maximizes the sum of profits given by

$$\Pi_f = \sum_{j \in F_f} (p_j - mc_j).q_j \tag{3}$$

where mc_j is the marginal cost of producing brand *j*. If firms behave according to Bertrand, then, it can be shown that the price p_j of any product *j* produced by firm *f* must satisfy the following F.O.C

$$s_j + \sum_{r \in F_f} (p_r - mc_r) \frac{\partial s_r}{\partial p_j} = 0 \;\; ; j=1,2,\ldots,N \tag{4}$$

The outside good pricing decision is assumed to be exogenous and therefore does not interact strategically with the pricing decision of the inside goods. Note that (4) is flexible enough to accommodate different market structures. The first structure is the single firm product, in which the firm can only control the price of its unique brand. The second is the multi-product firm, in which the firm internalizes the price decision of all of its brands. A third example is a monopoly, where one firm produces all the varieties offered in the market.

3. Empirical Strategy

Obviously, in the absence of disaggregated data on prices and quantities the demand and supply equations, (2) and (4) respectively, cannot be directly used for estimation. However, from (2) and (4) it is possible to establish a mapping from observed revenue and cost data to variables that describe firms' behavior (e.g quantities). Then, this paper shows how to calibrate the demand parameter by asking the model to match the observed aggregate quantity.

Mapping More formally, this mapping is derived (and computed) as follows.

Equation (4) in the text can be rewritten as

$$1+\sum_{r\in F_f}(p_r-mc_r)\frac{\partial s_r}{\partial p_j}\Big/ s_j=0 \qquad (5)$$

Note that firm *j*'s revenue (R_j) and variable cost (TC_j) can be written as $R_j=p_j.q_j$, $TC_j=mc_j.q_j$, where q_j represents firm *j*'s production. Thus, one can write the market share for firm *j* as $s_j=q_j/(Q+Q_0)$, where Q and Q_0 represent respectively total production of inside goods and total output production of the outside good (both represent physical quantities). Hence, substituting these equations into the pricing rule (5) and solving for quantity of plant *j* belonging to firm *f* ($j\in F_f$) one obtains

$$q_j=\left(\frac{1}{\alpha.(R_j-TC_j)}+\left(\left(\frac{1}{Q+Q_0}\right).\sum_{r\in F_f}\frac{(R_r-TC_r)}{(R_j-TC_j)}\right)\right)^{-1} \qquad (6)$$

Aggregating over the q_j's results in

$$Q=\sum_{f=1,2,..NF}\ \sum_{j\in F_f}\left(\frac{1}{\alpha.(R_j-TC_j)}+\left(\left(\frac{1}{Q+Q_0}\right).\sum_{r\in F_f}\frac{(R_r-TC_r)}{(R_j-TC_j)}\right)\right)^{-1} \qquad (7)$$

where *NF* is the total number of firms. This non-linear equation can be solved numerically for *Q* given (α, ***R*** , ***TC***,Q_0), where ***R*** $=\{R_j ; j=1,...NF\}$ and ***TC*** $=\{TC_j ; j=1,...NF\}$. Then, given the same parameters and data, q_j is determined from (6), whereas p_j , mc_j and s_j follow trivially from $p_j=R_j/q_j$, $mc_j=TC_j/q_j$ and $s_j=q_j/(Q+Q_0)$ respectively. Finally, the log-linearized version of the demand system (2) can be solved for relative quality a_j

$$a_j\equiv\xi_j-\xi_0=\alpha.(p_j-p_0)+\ln s_j-\ln s_0$$

In sum, for a given set (α, ***R*** , ***TC***,Q_0, p_0), this mapping shows how to obtain total quantity (Q) and plant-level information on price, marginal cost, relative quality and quantity.

If good instruments were available, it would be possible to combine them to a_j and impose moment restrictions to identify the demand parameter using the familiar GMM (see Berry, 1994). The data set one has in mind however does not report product characteristics (the usual instruments). The model is therefore not identified. One way to solve this problem is to introduce more information into the model. KLT use a similar mapping to the one developed above[1] and introduces the required additional information into the model by imposing prior distributions on the unknown parameters and performing Bayesian inference. However, as the economic model and

[1] On the one hand, the mapping developed in this paper generalizes the one found in KLT in order to accommodate multi-plant ownership as required by this paper's application. In the Colombian beer industry one firm owns more than one plant. On the other hand it restricts the discrete-choice model to have the simple logit form and not its more general version (the nested the logit) as assumed in the KLT mapping. This restriction is necessary to implement the calibration methodology. This will be clarified in the calibration sub-section below.

the data set available are insufficient to perform standard inference, identification relies heavily on the priors. Thus, the parameters estimates are prone to be very sensitive to the priors. This paper offers an alternative route by bringing commonly available numbers on aggregate quantity to uncover the demand parameter and other empirically relevant unobservables.

<u>Calibration</u> Equilibrium aggregate quantity also carries information on consumer preferences. Therefore, observing this aggregate measure may help uncover the demand parameters. Indeed, from the discussion above, the model implies that the total production of the inside goods (Q) is given by a function of (α, $\boldsymbol{R}$, $\boldsymbol{TC}$, Q_0). Assuming that one can observe the total production of the inside goods (Q^{obs}), one can ask the model to match this observed quantity according to

$$Q(\alpha, R, TC, Q_0) = Q^{obs} \tag{8}$$

The vector (R, TC, Q_0, Q^{obs}) is observed. Therefore, α can be determined[2] from (8). Then, p_j, mc_j and a_j are obtained by reapplying the mapping. Note however that (8) is based on a simple logit model instead of more sophisticated setups like nested logit model and the full-random coefficient model (Berry, Levinhson and Pakes, 1995; Nevo, 2000). The assumptions underlying the simple logit model place restrictive assumptions on consumer preferences such that cross-price effects have notoriously undesirable properties[3]. The model could be appended to accommodate more sophisticated setups, and consequently, more plausible cross-price effects. However, observing more variables would be necessary to devise other calibrating equations and uncover the extra parameters introduced by these new setups. Thus, due to the limitation commonly found in plant-level data sets the logit model restrictive assumptions cannot be relaxed in this calibration framework. For instance, notice that a nest-logit model would introduce another parameter on the left side of (8). Then, one would have to solve for two parameters with only one equation (8).

It is worth noticing that the assumptions underlying the simple logit are nevertheless less restrictive than those imposed by previous plant-level studies of differentiated product markets where firms are assumed to behave according to the monopolistic competition setup. In this market set cross-price effects are even more restricted, since the price change by one firm has an irrelevant effect on the demand of any other firm.

4. Policy Simulation

An advantage of structural estimation is that, once the parameters of interest are determined, one can simulate the effect of different market environments using the usual welfare metrics. The framework for counterfactual simulations laid out in this section is standard in discrete-choice demand models. The distinctive difference is that the entries on the welfare metrics are obtained by the calibration equation and the

---

[2] This model could potentially be estimated through some non-linear econometric method as NLLS. However, the series for aggregate quantity is usually small, hampering quality inference.

[3] The cross-price derivative between varieties r and j ($\partial s_r / \partial p_j$) depend only on the product of their market shares, and not, as one should intuitively expect, on their own characteristics and the characteristics (and market shares) of the other varieties.

transformation algorithm. Indeed, α is pinned down by (8) and then prices, marginal costs and qualities are obtained from the transformation algorithm.

The counterfactual experiment goes as follows. Determine the demand parameters and uncover prices, quantities, marginal costs and qualities according to the calibration methodology described above. Next, impose an exogenous market environment change (e.g. one can reduce the price of the outside good, split up a monopoly into smaller firms or merge the largest firms) and recalculate the new equilibrium prices (p_j^*) and quantities (q_j^*) from (4) holding relative quality and marginal cost fixed. The remaining task is to calculate the welfare variation for consumers and producers. McFadden (1981) shows that, in the logit model, consumer surplus variation (ΔCS) is given by

$$\Delta CS = \frac{1}{\alpha}\left\{\ln\left(\sum_{j=1}^{N}\exp[a_j - \alpha(p_j^* - p_0)]\right) - \ln\left(\sum_{j=1}^{N}\exp[a_j - \alpha(p_j - p_0)]\right)\right\} \tag{9}$$

In turn, producer surplus (PS) is given by a simpler equation. It is just the sum of profits of all active firms. Thus, the variation in producer surplus can be calculated from the following equation

$$\Delta PS = \left(\sum_{j=1}^{N}(p_j^* - mc_j)q_j^*\right) - \left(\sum_{j=1}^{N}(p_j - mc_j)q_j\right) \tag{10}$$

The section below shows an application of the methodology developed so far to study the impact of a regulatory action against an established monopoly in the Colombian beer industry. I emphasize though that the purpose of this application is to show how to manipulate the calibration strategy and a counterfactual experiment, rather than providing a detailed study of the beer industry.

5. An Application to the Colombian Beer Industry

After an aggressive horizontal merger strategy in the early seventies, *Cerveceria Bavaria* became a monopoly in the domestic production of beer by acquiring rivals like *Cerveceria Aguila, Cerveceria Union, Cerveceria Andina* and other smaller producers. Imports never accounted for more that 7% of the Colombian domestic beer market during the 1970's. Monopolist firms are expected to set prices higher than competitive levels and therefore to enjoy high profits at the expense of consumers' welfare. In this way, an interesting experiment consists in calculating the welfare effects of a more competitive environment.

Variables Definitions and Data Description The data set consists of an unbalanced panel of plants in the Colombian beer industry, with more than 10 employees, in 1977. These data were originally collected by the Colombia's *Departamento Nacional de Estadistica* (DANE) and have been cleaned as described in Roberts (1996). The revenue series are constructed as the total sales revenue and the total variable costs are defined as the sum of payments to labor, intermediate input purchases and energy purchases. Since plant-level prices and quantities are not directly observed the methodology described in the previous sections applies. From an additional source (UN database) I obtain the total quantity of beer (in hectoliters) produced in the country for the same sample period. Ideally, one would want to have data on the quantity of beer consumed in the country. However, the data in hand is not so restrictive since there is very little export activity in this sector.

Then, I define an inside good as a domestically produced beer. I also use auxiliary data to uncover the price of the outside good, defined here as imported beer (p_0)[4], as well as its imported quantity in hectoliters (Q_0). In a separate publication DANE also reports the net weight (in kilos) and the monetary value of beer imports (in pesos). Assuming that beer has the same density as water[5] (1kg per liter), it is easy to transform the net weight in kilos to volume of imported beer in hectoliters (Q_0). Then, p_0 follows from the ratio of the peso value of imports to Q_0.

Simulation Experiment and Results Below, I develop a counterfactual experiment that simulates the welfare effects of a regulatory action against the Bavaria monopoly. First, assume that the government was considering action against the Bavaria monopoly in 1977. Calibrating the model for this year yields an "estimate" for the demand parameter (α is found to be equal to 2.77). Then, suppose that a regulatory agency plans to split up the Bavaria conglomerate into smaller independent companies. Next, calculate the new equilibrium prices and market shares given the new ownership structure from (4). Finally, compute welfare variation for consumers and producers from equations (9) and (10).
The second column of Table 1 presents the new plant owners after the break-up. For instance, firm A owns plants 1 to 8, Firm B plants 9 to 14, and so on[6]. As the market becomes more competitive with more rival brands, prices are expected to drop. Indeed, prices experience a sharp decrease, the average price falls by 47,43%. Market shares do not change considerably; they all increase by a small percentage, mostly at the expense of the imported variety, which goes from a 6% market participation to virtually zero (0.01%). This is simply the standard finding that more competitive environments generate higher output[7]. It should be noted, however, that this result is driven by the assumption that the price of the outside good is exogenously determined in the model. If this were not the case, one would probably see a decrease in the imported good price, dampening the decline in its market share. Driven by lower prices, consumer welfare goes up by 8,79 millions (in 1977 pesos), while total profits go down by 8,24 millions. The consumer surplus effect dominates the producer surplus effect such that the increase in social welfare is 0.55 millions[8].

A legitimate question is to ask whether the calibration methodology gives a reasonable value for α. The answer is affirmative. The welfare analysis, calculated after calibrating α ($\alpha = 2.77$), shows that the social gains (0.55 millions) from

---

[4] This is a composite good that bundles together all the different imported varieties.
[5] More than 96% of beer content is water.
[6] I have also simulated different ownership arrangements. For example, instead of five firms with similar shares (the case analyzed in this paper), I tried the scenario where the split up would result in only 3 firms. Results were similar.
[7] Domestic output since total output is fixed.
[8] These figures could change if one factors in non-price competition effects, e.g, advertising and brand introduction. For instance, the market structure may influence the decision to introduce new brands and therefore change welfare calculation. In order to determine if these effects exist one needs a dynamic model of brand introduction. However, an empirical model of dynamic decisions, such as advertising and brand introduction, has yet to be developed. Pakes and Macguire (1994) framework gives a lead on how to model such dynamic decisions.

breaking up the monopoly are not very high, implying that monopoly power is limited. This finding is consistent with previous studies of the beer industry. One example is Hausman et al. (1994).

6. Conclusion

This paper develops a methodology to study differentiated product industries using plant-level data sets that report only revenue and input expenditures, not prices or quantities. By bringing the extra information on aggregate quantity, a calibration technique is derived to uncover the empirically relevant unobservables. Once these unobservables are uncovered it is possible to perform counterfactual experiments. I demonstrate the methodology by conducting a simulation of a monopoly break-up that gives monetary figures to consumer gains and producers losses.

Due to limitations commonly found in plant-level data sets the restrictive assumptions underlying the logit model can not be relaxed in the framework developed here. These assumptions are nevertheless less restrictive than the ones imposed in previous plant-level studies of differentiated product industries, where firms are assumed to behave according to the monopolistic competition setup. In this way, this paper can be seen as a step forward in the effort to devise (reasonable) empirical models to analyze differentiated product industries using plant-level data.

References

Berry, S. (1994). "Estimating Discrete-Choice Models of Product Differentiation," *Rand Journal*, 25(2), 242-262.

Berry, S., Levinsohn, J., and Pakes, A. (1995) "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841-890.

_________ (1999) "Voluntary Export Restraints in Automobiles," *American Economic Review*, 89(3), 400-430.

Caplin, A. and Nalebuff, B. (1991) "Aggregation and Imperfect Competition: On the Existence of Equilibrium," *Econometrica*, 59(1), 25-59.

Hausman, J., Leonard, G., Zona J. (1994) "Competitive Analysis with Differentiated Products" *Annales d'economie et de Statistique*, 34, 159-180

Katayama, H., Lu, S. and Tybout, J. (2003) "Why Plant-Level Productivity Studies Are Often misleading, and an Alternative Approach to Inference". NBER Working Paper No.9617.

Klette, T. and Griliches, Z. (1996) "The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous," *Journal of Applied Econometrics*, 11(4), 343-61.

McFadden, D. (1981) "Econometric Models of Probabilistic Choice'', in C. Manski and D.McFadden (Eds), *Structural Analysis of Discrete Data.*

Melitz, M. J. (2000) "Estimating Firm-Level Productivity in Differentiated Product Industries". Mimeo. Harvard University.

Nevo, A. (2000) "A Practitioner's Guide to Random Coefficients Models of Demand". Journal of Economics and Management Strategy, 9, 266-275.

Pakes, A. and McGuire, P. (1994) "Computing Markov-Perfect Nash Equilibria: Numerical Implications of a Dynamic Differentiated Product Model" *RAND Journal of Economics*, 25(4), 555-589.

Roberts, M. (1996) "Colombia, 1977-85: Producer Turnover, Margins, and Trade Exposure," in *Industrial Evolution in Developing Countries*, eds. Roberts, M. and Tybout, J., Oxford University Press, Oxford and New York.

Schmalensee, R. (1978) "Entry Deterrence in the Ready-to Eat Breakfast Cereal Industry". *Bell Journal of Economics* (9), 305-327.

Scherer, F. (1980) *Industrial Market Structure and Economic Performance*., Chicago, IL: Rand McNally, College Publishing.

Table 1
Market shares and price variation as a result of the monopoly break-up

Plant ID	Ownership after break-up	Ownership before break-up	Market shares before break-up in %	Market shares gain in %	Price variation in %
1	A	BAV	1.99	0.08	-55.16
2	A	BAV	2.06	0.08	-56.14
3	A	BAV	4.65	0.19	-58.65
4	A	BAV	0.48	0.02	-19.06
5	A	BAV	2.51	0.10	-55.69
6	A	BAV	3.85	0.15	-46.40
7	A	BAV	1.97	0.08	-53.85
8	A	BAV	3.78	0.15	-57.09
9	B	BAV	2.22	0.38	-57.61
10	B	BAV	2.06	0.35	-56.99
11	B	BAV	4.50	0.77	-30.85
12	B	BAV	1.68	0.29	-22.81
13	B	BAV	1.12	0.19	-54.05
14	B	BAV	0.60	0.10	-66.35
15	C	BAV	22.11	0.64	-79.20
16	D	BAV	1.38	0.10	-54.00
17	D	BAV	3.21	0.24	-68.12
18	D	BAV	9.15	0.67	-59.73
19	E	BAV	19.46	1.25	-55.31
20	D	BAV	5.04	0.37	-66.26

Other important Measures

Consumer surplus variation	8798004
Producer surplus variation	-8243465
Social Surplus variation	554538.8
Average Price variation	47.43%

Column 2 presents the fictitious denominations (A, B, C, D and E) for the firms that are created after the split up. Before the counterfactual scenario all plants (identified in column1) belong to Bavaria (BAV). The fifth column, in turn, presents the slice of the market they add to their previous shares. The welfare measures are given in 1977 Colombian pesos.