SIMPLE ESTIMATORS FOR CROSS PRICE ELASTICITY PARAMETERS WITH PRODUCT DIFFERENTIATION: PANEL DATA METHODS AND TESTING

JAVIER CORONADO SALEH[1]

Abstract

This article shows how to use simple Panel Data methods to consistently estimate demand parameters that come from the Random Utility model where products are differentiated. We followed Berry (1994) and obtain the Nested Logit specification and show that for a sample of information from the market of beer in Peru, the Random Effects (GLS) assumption in Panel Data is strongly rejected even when using Instrumental Variables. Yet, we test whether the alternative Fixed Effect (FE) model, estimated with the Within Groups estimator with and without Instrumental Variables, provides a correct specification. We show that preliminary evidence favors the FE model using the Angrist and Newey (1991) approach to test Chamberlain (1982) over-identifying restrictions implied in the FE specification.

Keywords: Panel Data, Fixed Effects, Chamberlain, Nested Logit, Random Utility.

I. INTRODUCTION

Relevant market analysis is the cornerstone for the economic study of market power both for antitrust procedures and regular market research and strategy. In antitrust international cases this analysis is typically conducted without scientific accuracy, mainly because a proper relevant market analysis, for antitrust cases, should not respond to the interest of academic precision but need to focus on a consistent study of the rationality of the firms that make profit. [2]

However, economic modeling and quantitative methods have gained prestige as valid tools for the systematic analysis of the definition of relevant markets. Modern econometrics have benefited enormously of the availability of richer datasets that can help to reveal in a rigorous way robust evidence that will help to objectively shape the boundaries of product and geographical markets for antitrust cases.

---

[1] Chief Economist at Indecopi, economist by the Pontificia Universidad Católica del Perú, Ph.D. Economics by the Universitat Pompeu Fabra (2010), MSc. Economics (2005), MSc. Economics by the London School of Economics (2003). *Affiliated Faculty* at the Barcelona Graduate School of Economics, Professor of Applied Microeconomics and Competition Policy at the Master in Economics at the PUCP.

[2] See for example Rubinfeld (1993).

In particular, economic modeling and econometrics can be fundamental to provide convincing evidence on the degree of demand substitutability among products, a key concept for the analysis of relevant markets.[3] Demand substitution is particularly controversial when marketed products are strategically differentiated in several dimensions. Indeed, in some circumstances products that are thought to belong to the same relevant market because of their common utility for the consumer are actually not interchangeable from the point of view of the same consumer. Those products might not, in the end, place any relevant competitive pressure on each other. Product differentiation, however, nowadays is the rule rather than the exception in massive consumption markets.

Sound econometric analysis starts with a reasonable modeling of the available observed market outcomes. In turn, modeling market outcomes requires an approach from alternative assumptions the way in which both consumers and firms make decisions as well as the way in which observed and unobserved information interact. In this way the researcher will have a diversity of options to reconcile observed data with modeling and assumptions and make an educated judgment on the model that better fits the data.

In this article, I present a simple and schematic way to estimate the parameters for cross price demand elasticity from firm level data that can be easily extended to product level data, based on the early developments of Berry (1994) and Berry, Levinson and Pakes (1995), among many other authors who have made key contributions to the empirical modeling of market outcomes with product differentiation.

The main contribution of this article is to show how Panel Data methods can help to identify the parameters of interest and test for consistent estimates, controlling many of the usual features of market data coming from markets with product differentiation. I provide an illustration applied to data from the Peruvian beer market. Part of the data has been constructed ad-hoc for this exercise so that the results should be taken only as a mean to show the methods and not necessarily as conclusive.

This approach is particularly useful when information on the characteristics of products or firms is not available and has to be treated as unobservable from the point of view of the researcher. Fixed Effects (FE) specifications with Panel Data can be useful to control correlated time invariant unobserved firm or product characteristics whereas further Instrumental Variables regressions can be used to test for the validity of more restrictive empirical specifications such as that of the Random Effects (GLS) assumption. Finally, we also show how to test for the FE specification to provide further evidence of the consistency of the estimations, resorting to Angrist and Newey (1991) version of the over-identifying restrictions test first suggested by Chamberlain (1982).

[3] A good guide for the use of own and cross-price demand substitution can be found in Motta (2004), Chapter 4.

The paper is divided in the following way: section 2 provides the basics of the Random Utility model for consumer choice in the context of differentiated products, section 3 shows how to integrate out consumers' preferences by imposing some simple but yet relatively flexible distributional assumptions for consumers' preferences and describes how to connect the theory with firm or product level data. Section 4 provides an illustration of the methods using firm level data for the Peruvian beer industry. Being an illustration, the results should be taken with care and cannot be regarded as conclusive because ideally, product level data should be used when available.

II. THE THEORY OF RANDOM UTILITY FOR DIFFERENTIATED PRODUCTS

When products are differentiated it is usually assumed that consumers select one variety of the product that is both available and feasible to them. Hence, consumers' choice is said to be discrete in that they will decide for one of the varieties and not for a combination of them.

Random utility models for consumer choice provides a flexible way to model consumer decision making when multiple varieties of a product are available and consumer decision making is fundamentally discrete. Possibly the best source for learning on random utility models is Ben-Akiva and Lerman (1985) which is a comprehensive, yet friendly reference for the early developments of Manski (1977). In short, the random utility model postulates that the benefits from individual consumption are essentially unobservable for the researcher so that they are better modeled through a random variable. [4]

Say there are N consumers; this number could be a sample or the entire population. Consider a market for a specific consumption product in which there is a set $\mathcal{J}$ of both available and feasible products.[5] That set has in total J products. To keep things simple, I will say that one active firm produces just one variety so that an element j included in $\mathcal{J}$ denotes both a firm and its corresponding differentiated product.

Now consider an individual i from the population and denote her utility from consuming product j as u_{ij}. Imagine we are able to observe her choice, revealing some important information on preferences. In particular, if consumer i chooses product j it must be the case that $u_{ij} \geq u_{is}$ for all $s \neq j$ included in $\mathcal{J}$. Given that the utility level for any product is a random variable it is possible to define the choice probability for product j in the following form:

[4] This might explain why classic text book consumer theory is usually not very useful to explain observed consumer behavior.

[5] I shall use from the very beginning the notation that is now standard.

$$P(j|\mathcal{J}) = Prob(u_{ij} \geq u_{is}|\mathcal{J})\ \forall\ s \neq j \quad (1)$$

So far, the only assumption is that consumers share the same probability distribution over preferences. The next step is to specify a parameterization for u_{ij}. In general Berry (1994), Berry et al. (1995) and Nevo (2001), and other related literature, propose a very flexible way to model the consumers' decision making by resorting to a random utility function that is linear in the product characteristics but can be non-linear in the consumer preferences.

Consider the following specification:

$$u_{ij} = x_j'\beta_i + \alpha_i p_j + \xi_j + \epsilon_{ij} \quad (2)$$

Where x_j is a vector of information containing k observable attributes for product j, β_i is a vector of random coefficients that depends on the preferences of the individual, p_j is the price of the product, α_i the corresponding random coefficient, ξ_j is an unobservable random variable that contains information on unobserved attributes of the product and ϵ_{ij} is a zero mean i.i.d.[6] random variable containing unobserved preferences of the individual for product j.

In general one would like to place the minimum set of restrictions on the way preferences condition the random coefficients so that the correlations between any two pair of utilities for any two products are only marginally restricted by the researcher. The usual way to do this is to assume that the random coefficients are specified in a way that conditions how a consumer is thought to perceive or weigh every generic observable *k* attribute.

Let ζ_{ik} be an i.i.d. random variable with a zero mean which represents a specific individual's attitude towards a certain attribute. Note that this random variable does not vary across products but models how a specific attribute of a product is weighted by a specific consumer. For instance, when modeling consumer choices for mobile devices, one might want to define for each consumer how she will weigh the screen size, battery duration, compatibility, total size, weight, and so on. Each ζ_{ik} will have a coefficient σ_k so that one might use a simple linear specification in the following form:

$$\beta_{ik} = \beta_k + \sigma_k\zeta_{ik} \quad (3)$$

By constructing the unconditional mean $E(\beta_{ik}) = \beta_k$ is the mean coefficient for attribute k across individuals. Something similar can be done for α_i, where α will be the

[6] Recall i.i.d. refers to the assumption that each realization of the random variable will come from the same probability distribution and is independent from any other realization.

mean coefficient for the price. This specification is convenient because it will naturally lead to a key object of interest known as the mean utility level for product j across individuals. In particular let us redefine (2) in the following way:

$$u_{ij} = \delta_j + v_{ij} \quad (4)$$

The object $\delta_j = x_j'\beta + \alpha p_j + \xi_j$ is the mean utility level of product j across individuals and v_{ij} is a mean zero random variable that is uncorrelated across individuals but will exhibit heteroskedasticity and is specified in the following way

$$v_{ij} = \sum_{k=1}^{K} x_k \sigma_k \zeta_{ik} + \epsilon_{ij} \quad (5)$$

The latter expression will condition the way in which utilities among products by a specific individual will correlate, so that it will determine the substitution patterns of each individual between any two products.

Now we can go back to (1) in order to define the probability of observing a specific choice for a product. The density distribution function of v in (5) will depend on the information contained in (x, σ), so that generically we can denote that function in the following way: $f(v, x, \sigma)$. Indeed, one could think of all the circumstances that will make a specific consumer choose one product, that is all the realizations of the attributes, the parameter σ, and therefore realizations of v that will predict the consumer to choose a specific product j from the set of feasible products.

Formally, an individual consumer will choose product j if the following condition holds:

$$u_{ij} = \delta_j + v_{ij} > u_{ih} = \delta_h + v_{ih} \; \forall \; h \neq j \quad (6)$$

Therefore, given the information contained in the mean utility levels, δ_j, we can define the set of realizations of v that will make a specific consumer to choose product j as follows:

$$\mathrm{A}_j(\delta) = \{\mathrm{v_i} \colon \delta_j + \mathrm{v_{ij}} > \delta_h + \mathrm{v_{ih}}, \forall \; \mathrm{h} \neq \mathrm{j}\} \quad (7)$$

Finally, as v_i is i.i.d. it is possible to integrate in one dimension over the density of this random variable and obtain an expression for the probability of choosing j across individuals which represents a theoretical prediction for the market share of product j.

$$s_j(\delta(x, p, \xi), x, \theta) = \int_{A_j(\delta)} f(v, x, \sigma)\, dv \quad (8)$$

It is very important to notice that in principle, adding the predicted probabilities across products should give one construction however this system cannot be complete, at least theoretically speaking, without considering a circumstance in which consumers would rather drop out from the market. Consider for example a situation in which there is a generalized increase in prices for the whole set of products available for consumers. If consumers cannot drop out from the market, everyone will be forced to choose one of the options, when it might be the case that some individuals would prefer to stop consuming the product all together.

Berry (1994) discussed this situation and suggested including an outside option as a device to include a realistic circumstance in which some marginal consumers will decide to change to a different set of products instead of selecting one of the products in the set $\mathcal{J}$. In some cases there is a natural way to define an outside option; however in some others this could amount to adopting an arbitrary assumption that could condition the estimation in unexpected ways. Think for instance of the decision that consumers face when choosing one mode of transportation over another to commute from home to work, university, and so on. Several modes of transport might be available such as private car, bus, train, and motorbike. Some consumers could, notwithstanding, prefer not to use those modes of transportation and for example use a bicycle or walk. These outside options seem natural in this context.

Therefore, a more realistic approach requires defining an outside option which, in general, will not be marketed in the market under analysis but corresponds to some sort of unique alternative that consumers can resort to, whenever the benefits from participating in the market are lower than their reservation value. In practice, the researcher will have to consider a set of $J+1$ products where the attributes of the outside option cannot be observed and are usually normalized (restricted to a certain value).

In fact, the usual normalization implies restricting the mean utility level of the outside option to zero. That is setting $\delta_0 = 0$ where the sub-index "0" is common to denote the mean reservation value.

III. FROM INDIVIDUAL DECISIONS TO PRODUCT LEVEL DATA: IT IS ALL ABOUT PREFERENCES

3.1 Observables meet predictions

In the last section, we discussed the formal approach to consumer decision making where individual preferences are key to predict theoretical decisions. In practice it is common to use revealed preferences of consumers so that individual decisions are observed and can be confronted with the theory.

From the point of view of the researcher $\mathscr{s}_j(\delta(x, p, \xi), x, \theta)$ in (8) is a theoretical prediction that cannot be perfectly observed. Notice, however, that if a priori value for the size of the market in terms of value or units is available, say M, and the corresponding quantities produced per product are observable, say q_j, preliminary market shares can then be calculated $s_j = \frac{q_j}{M}$. If this is the case, once individual preferences are integrated out of the analysis, the researcher can try to equate observables and theoretical predictions in the following way:

$$s_j = \mathscr{s}_j(\delta(x, p, \xi), x, \theta) \quad (9)$$

In fact, the researcher will be able to define a system of equations in the following form:

$$\begin{aligned} s_0 &= \mathscr{s}_0(\delta(x, p, \xi), x, \theta) \\ s_1 &= \mathscr{s}_1(\delta(x, p, \xi), x, \theta) \\ &\dots \\ s_J &= \mathscr{s}_J(\delta(x, p, \xi), x, \theta) \end{aligned}$$

with $J + 1$ equations and the same number of unknowns. The key issue to recognize up to this point is that expressions for $\mathscr{s}_j(\delta(x, p, \xi), x, \theta)$ are essentially unknown and integrating out preferences might not be feasible analytically although numerically, for example if a full random coefficient specification is contemplated. However, once we solve this essential problem then a natural protocol for parameter identification will emerge: choose θ that make theoretical predictions as close as possible to the observable market shares.

To that aim, first the researcher needs to invert the system of equations to solve a unique solution. If s is the $J + 1$ vector of observed market shares and $\mathscr{s}$ the vector of theoretical predictions as a function of the mean utility levels (remember that at this point preferences have been integrated out), then the system will be solved for the mean utility levels in terms of observables:

$$\boldsymbol{\delta}(\mathbf{x}, \mathbf{p}, \boldsymbol{\xi}) = \mathscr{s}^{-1}(\mathbf{s}) \quad (10)$$

Then, conditional on observed information (x, p) and assuming only little regarding the behavior of $\boldsymbol{\xi}$ the variations across functions of market shares can be used to estimate the parameters of the system of equations.

3.2 Distributional assumptions and possible specification

One way to conduct a practical approach to parameter estimation is to specify restrictions in the way individual preferences are distributed in terms of the probability of their possible realizations, or more precisely, in terms of restrictions on the density of realizations. Restrictions, however, will impose not only a practical way to identify the parameters of interest, but condition, sometimes crucially, the way in which one expects the individuals to behave in practice.

A somehow flexible approach to restrict preferences requires defining a structure of correlations of individual preferences for certain categories of products. Cardell (1997) suggested that individual preferences might be restricted in terms of "variance components" that resembles a sequential decision making of consumers taking a series of independent probabilities of steps along the consumer choice.

Suppose, for simplicity, that products can be grouped in $G + 1$ mutually exclusive subsets where the outside option corresponds to a specific one-element subset. Products can be grouped based on some observed characteristic that make products in a specific sub-set closer from the point of view of the consumer.[7] This, for example, is the approach conducted in Rubinfeld (1993) for the Ready-to-Eat Cereal case presented in Kraft Foods vs Court of New York.

For each group, further clusters can be defined, giving birth to a multilevel nested model; however to keep things simple, let us consider a case in which consumers are supposed to first decide for a group g, where $\mathcal{J}_g$ products are considered and then choose one $j \in \mathcal{J}_g$.

Following Cardel (1997), the distribution of ζ can be assumed to be uniquely determined by a parameter σ $(with\ 0 \leq \sigma < 1)$ and that ϵ follows a Type I Extreme Value distribution function, that is the typical distributional assumption that follows, for example, a standard Conditional Logit model (See Wooldridge (2000)). Then, it can be shown that the combination $\zeta + (1 - \sigma)\epsilon$ will also follow the Type I Extreme Value distribution function. Cardel (1997) considers ζ as a variance component of the error structure that is taken as a group-specific parameter. This parameter allows for a specific correlation of preferences within groups.

To better understand why this approach may be suitable for a more realistic modeling of consumer preferences, think about the role of the parameter σ. If $\sigma = 1$, the "randomness" of the utility function will be determined by ζ, therefore, any two utilities from products belonging to the same group will exhibit the largest possible correlation. In this case, substitution patterns will be driven basically by comparisons of products from

[7] Coronado (2010) used this approach to analyze consumer decisions in pharmaceutical markets. Pharmaceutical products where naturally grouped in terms of active ingredients and where further distinguished inside a sub-set depending on whether the product was a generic or a branded one.

the same group, whereas products belonging to different groups could be considered not very closed substitutes.

On the other hand a value of $\sigma = 0$ will place a disproportionate weight to ϵ as a source of correlations between any two products' utilities. Given that ϵ is assumed to be i.i.d. only the comparison of the characteristics of the two products will be relevant for decision making, regardless of other feasible products to be closer substitutes of one or another. Therefore $\sigma = 0$ is nothing but the particular case in which preferences satisfy the Independence of Irrelevant Alternatives (IIA) axiom which is at the core of the standard MacFadden (1978) Conditional Logit model (Wooldrige (2000)).

In the construction, by introducing the variance component approach will be equivalent to including it as an extra observable attribute for each product, giving information about the group to which it belongs and assume each consumer will have a specific taste for that particular group. Hence, we will present such information by means of $G + 1$ dummy variables, so that, for example, d_{jg} takes the value of one if product j belongs to group g $(d_{jg} = \mathbb{I}(j \in \mathcal{J}_g)$). In that case the variance component will be a random coefficient ζ_{ig}. Therefore, the corresponding utility level will be as follows:

$$u_{ij} = x_j'\beta + \alpha p_j + \xi_j + (\zeta_{i0}d_{i0} + \zeta_{i1}d_{i1} + \cdots + \zeta_{iG}d_{iG}) + (1-\sigma)\epsilon_{ij} \quad (11)$$

Where the ζ_{i0} corresponds to the variance component of the outside option. Then it follows that the utility levels of two products belonging, for example, to group $g = 1$ will exhibit correlation through ζ_{i1}. Note that parameter σ is assumed to be constant across products, implying automatically that the within group correlation of utility levels are the same across groups, a restriction that can be further tested empirically. In the latter, there will be one different σ_g for each group, in which case the substitution patterns among products of the same group may be different across groups.

The distributional assumption will prompt to a theoretical probability for choosing product j of the form:

$$s_j(\delta,\sigma) = s_{j/g}(\delta,\sigma)s_g(\delta,\sigma) \quad (12)$$

Where:

$$s_{j/g}(\delta,\sigma) = \frac{e^{\frac{\delta_j}{1-\sigma}}}{\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}}} \quad (12a)$$

$$\mathscr{s}_g(\delta,\sigma) = \frac{\left[\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}}\right]^{1-\sigma}}{\sum_g\left[\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}}\right]^{1-\sigma}} \qquad (12b)$$

The last expression (12b) indicates the theoretical probability of an individual choosing one product of group g, whereas the previous one, (12a) is the predicted probability of choosing product j given that the consumer has chosen group g. The joint probability of choosing group g and product j from it is the unconditional probability of choosing product j from the set of feasible products shown in (12).

For what remains it is useful to show the explicit solution for the prediction of the market share of our outside good, applying the normalization $\delta_0 = 0$:

$$\mathscr{s}_0(\delta,\sigma) = \mathscr{s}_{0/1}(\delta,\sigma)\mathscr{s}_1(\delta,\sigma) = \frac{1}{\sum_g\left[\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}}\right]^{1-\sigma}} \qquad (13)$$

In order to perform the parameter estimation and a subsequent calculation of the empirical cross elasticities of substitution, the researcher needs to solve the equation $s_j = \mathscr{s}_j(\delta,\sigma)$. That is solve for the mean utility levels, that are obviously constant across individuals, in terms of the observed a priori market shares, s_j. Formally, the next step is to obtain an expression for $\delta_j = \mathscr{s}_j^{-1}(s_j)$ which in this case, under the distributional assumptions, will provide a closed form solution albeit cumbersome to obtain (See the appendix for the mathematical derivation):

$$\ln(s_j) = \frac{\delta_j}{1-\sigma} + \ln(s_0) + \sigma\ln(s_{j/g}) - \sigma\frac{\delta_j}{1-\sigma} \qquad (14)$$

Or rearranging:

$$\ln(s_j) - \ln(s_0) = \delta_j + \sigma\ln(s_{j/g}) \qquad (15)$$

where:

$$\delta_j = x_j'\beta + \alpha p_j + \xi_j \qquad (15a)$$

Therefore, under the distributional assumptions it is possible to obtain a linearized expression for parameter estimation:

$$\ln(s_j) - \ln(s_0) = x_j'\beta + \alpha p_j + \sigma\ln(s_{j/g}) + \xi_j \qquad (16)$$

where $\{\alpha, \beta, \sigma\}$ are the population parameters to identify with some econometric method, $s_{j/g}$ is the observed market share of product j within its designated cluster and ξ_j is taken as the unknown variation of the dependent variable across products.

Usually in demand estimation, p_j and $s_{j/g}$ are variables that are determined endogenously in the market, therefore they will be structurally correlated with ξ_j given that equilibrium values of observables are dependent on given attributes, both observable and unknown. The researcher will need to carefully select valid instruments for these two endogenous variables.

3.3 Cross Price elasticites

Analytical expressions for cross price elasticities are straightforward and will depend upon the potential rival product in the same cluster. Let us for instance consider product $j \in \mathcal{J}_g$, then for any particular product *h*, the formal cross price elasticity of demand, denoted ϵ_{jh}, will be given as follows:

$$\epsilon_{jh} = \begin{cases} -\alpha s_h p_h & , \text{if } h \notin \mathcal{J}_g \\ -\alpha \left[\frac{\sigma}{1-\sigma} + s_g\right] s_{h/g} p_h, & \text{if } h \in \mathcal{J}_g \end{cases} \qquad (17)$$

Estimation of cross-price elasticities then requires estimating consistently α and σ.

Note that for the particular case in which the rival product does not belong to the same cluster of the product of interest in (17), the cross price elasticity of demand resembles that of a simple Conditional Logit model, with the particular (and not always sensible) feature that it will depend only on information of the relative market power (measured in terms of its market share) and price of the rival product.[8]

The cross price elasticity with respect to a rival product that belongs to the same cluster of the product of interest will also depend on information about the rival product, but it is conditioned by the parameter σ which weights the correlation of the utilities between two products supposed to be close substitutes with respect to the correlation of two products assumed to be only poor substitutes.

As it is easy to see in (1) when the population parameter σ happens to be equal to zero, the substitution patterns will in general reduce to the Conditional Logit. A specification

[8] In Berry et al. (1995), it is also noted that in the Conditional Logit case each firm´s or product´s own price elasticity depends only on its market share. Therefore, two firms having the same market share will also have the same price elasticity and the same margin, which might not reflect other factors explaining price to cost margins.

test for the Conditional Logit (restricted) versus de one-nest Nested Logit (unrestricted) then naturally arises as a simple single significance test on σ.

IV. EMPIRICAL ILLUSTRATION: DATA FROM THE BEER MARKET

4.1 The Data

Identification requires a sufficiently large sample of information that provides enough variation that is exogenous to the error term. In many cases the number of relevant products, or firms, is not as numerous as one would like, limiting the empirical implementation with cross sectional data at the product or firm level.

Rubinfeld (1993) presented a similar approach that he developed in this article to estimate cross-price elasticities of substitution in the well-known case *Kraft Foods/ Nabisco vs Court of New York*, a case of a merger between Post Cereals and Nabisco Cereals in the Ready-to-Eat cereal market in the United States. Professor Rubinfeld was able to gather information of about 200 different varieties of cereal that amounted to a significant source of degree of freedom.[9]

This is certainly a very unusual case since most cases, at best, only have tenths of their products available in the market. However, nowadays market outcomes can be observed either in certain geographical areas and/or through time at the firm or product (brand) level. This for example is the case of Rojas and Peterson (2008), where they use a set of information of 64 brands of beers observed in 58 cities (metropolitan areas) for 20 quarters.

Nevo (2001) used a Panel Data set for the Ready-to-Eat cereal market of the U.S. using the methods described in section 3, however, he doubted that utilities could actually be properly explained by observable attributes and introduced for the first time a way to circumvent the problem of having to control the endogenous variables that signal some important unobserved attributes. In particular, the author was interested in controlling quality perceptions that he modeled through brand fixed effects (time invariant) that, once considered in the estimation, eliminates the need to control for instance,, advertising as sunk costs (see 4.2.2), that is often difficult to obtain and might in general be endogenous. The literature then suggests that Panel Data structures are helpful to pursue the difficult task of properly identify the parameters required to calculate reasonable and credible cross-price elasticities of substitution.

[9] In fact, the Nested Logit approach was very useful in that case also because it helped to place restrictions to the enormous cross-price elasticities that should be identified in principle.

Along that line of reasoning we have collected information from public sources of the markets for white beer in Peru. In particular we will use a sub-set of information taken from monthly data on price index and quantities sold by 5 firms in the country, not distinguishing between brands because the most prominent ones belong to the same firm as we will describe later on, in addition data from public sources does not include such level of detail. We observed information between January 2012 to April 2014, however one of the brands, which is focused in a very specific regional market, is taken as the "national" outside good. Table 1 summarizes some descriptive statistics for prices and shares for each firm.

TABLE N°1

SUMMARY OF BASIC DESCRIPTIVE STATISTICS: JANUARY 2012 - APRIL 2014[/a]

Firm	Statistic	Sales Share (Liters sold)	Beer Average Prince Index (2012=100)	Price of Barley (S/. / Metric Tons)	
				National Whole sale	International
Grupo Aje	mean s.d. obs.	0,017 0,007 28	105,631 15,262 28	-.-	-.-
Amazónica	mean s.d. obs.	0,054 0,015 28	-.-	-.-	-.-
Backus	mean s.d. obs.	0,093 0,017 28	109,902 8,809 28	-.-	-.-
San Juan	mean s.d. obs.	0,093 0,017 28	102,301 2,388 28	-.-	-.-
Total	**mean s.d. obs.**	**0,200 0,319 140**	**106,979 12,013 112**	**1,624 0,046 112**	**0,557 0,107 112**

Notes:
a/ Shares are predicted from production indexes and volume of sales as of 2012.
Source: Ministry of Production, National Institute of Statistics.

Having data only at the firm level, naturally, is an important limitation for this exercise, given that the model derives from mean utility levels that are normally defined at the product level. Therefore, at the most, substitution patterns have to be interpreted only as averages with respect to the aggregate productions and prices of a specific firm.

Notwithstanding, it is worth recalling that the aim of this exercise is to provide an empirical illustration of the methods that can be carried out in order to bring to the empirical field, the modeling of consumer choice and observable market level data in section 3. Our aim is to show how a Panel Data structure may help to use many different sources of identification along with the classical instrumental variables approach.

As it is common in demand parameter estimation we will require some exogenous variation to identify the coefficients of interest, especially for covariates such as price or shares that are market equilibrium outcomes. To that aim, we consider information that is supposed to be correlated with market outcomes but is determined outside the market. It is usual to make use of information related to costs, which is why in this case we will use information from international and national wholesale prices of barley, a key ingredient for the beer industry (see Table 1).

Finally, another source of information that will be used for identification is in relation to recent changes in the application of the specific consumption tax in this market that will be briefly discussed in the following section.

4.2 Specification and Sources for Identification

4.2.1 General Panel Data Specification

The basic specification follows what has been developed so far in section 3, however, given the Panel Data structure of the data, we include a sub-index t to denote that shares, prices and attributes (if any), are observed through time.

$$\ln\left(s_{j,t}\right) - \ln(s_{0t}) = x'_{j,t}\beta + \alpha p_{j,t} + \sigma \ln\left(s_{j/g,t}\right) + \xi_{j,t} \qquad (18)$$

$$j = 1, \ldots, J \; and \, t = 1, \ldots, T$$

Note that we do not expect specific attributes that are observable to vary in the short term for a specific firm or product, but they will vary across firms or products.

The error term $\xi_{j,t}$ could be specified in several ways to account for different structures that allow us to control for unobserved information. In this case we will adopt a two-way error component structure[10], which includes a fixed effect for each firm, η_j, times the dummy variables, λ_t and a remaining error term that may vary through time υ_{it}.

[10] See Baltagi (2008)

$$\xi_{j,t} = \eta_j + \lambda_t + \upsilon_{it} \qquad (19)$$

As discussed in previous sections, we should expect $p_{j,t}$ and $s_{j/g,t}$ to be correlated with the error term as they are endogenous equilibrium values. We allow for multiple sources of correlation. For example, prices and shares could be correlated with fixed attributes that are not observable (or not measurable in a concrete and objective way) for both η_j and with υ_{it}.

Whenever the researcher believes that right hand side variables are endogenous to the fixed effects, the natural reaction would be to treat η_j as coefficients, to be estimated and, for example, perform a Within Groups (WG) transformation to deal with that potential source of bias. Therefore, the two-way error component helps to control a firm specific time invariant unobserved characteristics that are correlated with equilibrium outcomes.

Note however that the FE specification may not provide a solution for existing endogeneity of prices, shares, and possibly other observable covariates, with respect to the remaining time variable error component, υ_{it}. For instance it is often contemplated that prices and shares are endogenous to the error term because they come from the simultaneous nature of price and shares determination in demand and supply interactions.

Thus, for that last reason, obtaining consistent estimates may require using some instrumental variables approach. Notably instruments are sometimes difficult to find especially for such aggregated type of data as the one we have at hand.[11] In this case, we propose to use the following sources of identification:

i. The recent change in the specific tax applied to beer consumption that might affect average prices in different ways across firms;

ii. Information from market prices of some costs such as international and local barley prices;

iii. We assume that some firms are focused on some local markets (e.g. San Juan in the east and Ambev in Lima), whereas others have a national scope (e.g. Backus). Under that assumption, we may use extra sources of variation at the individual firm level, correlated with factor markets that are exogenous to the Beer market. In

[11] In Nevo (2001), the author used the fact that regional markets are independent to each other so that information from regional markets is exogenous to error term to a specific market, but information across regional markets shares common shocks, so that they appear to be relevant instruments.

this case, we collected information on the level of labor market occupation of the manufacturing sector as a proxy for information related to labor costs.[12]

4.2.2 Advertising

As previously noticed, advertising is nowadays a prominent source of differentiation and it is often considered as a means to persuade consumers much in the same way as analyzed by Sutton (1991). Indeed, previous works on the beer industry have considered that advertisement is used as a means to produce changes in the quality of perceptions rather than a means to disseminate critical information; this for example is the case of Rojas and Peterson (2008).

Firms devote efforts to advertisement in order to persuade consumers. A standard theoretical way to model this assumption is to consider that consumers have a taste for a "perceived quality" so that advertisement outlays are supposed to increase the mean valuation of a specific product.

Including data from advertisement introduces at least two challenging questions. First, whether advertisement should be considered in the classical way as a sunk cost that is not, generally considered from the supply side for pricing but conditions consumer perceptions and; second, whether it has short run effects. In the first case, the researcher might want to consider the effects of advertisement as time invariant at the firm level but variable across firms, whereas in the second case advertisement outlays will enter the specification as a covariate that varies across time. This last case for example, was used by Nevo (2001) in his analysis of market power in the ready-to-eat cereals industry in the U.S.

Advertisement is done at the product level, such that if we consider a firm *j* having products, and let A_(l,j) denote product specific advertisement expenditures, and following Rojas and Peterson (2008), it is common to assume that advertisement and marketing efforts will exhibit decreasing returns to scale so that these values would enter the mean utility function with $A_{l,j}^{\gamma}$, where $\gamma = 0{,}5$ is a commonly used technological parameter in the literature. We follow Nevo (2001) by considering that advertising might have short run effects so that it will enter our specification in the following manner[13]:

$$\mathcal{A}_{jt} = \sum_{l=1}^{n_j} A_{l,j,t}^{0,5} \qquad (20)$$

[12] The assumption requires that short run changes in the level of employment in the regional manufacturing sectors be positively correlated with the labor opportunity costs for workers.

[13] As an alternative $\mathcal{A}_{jt}$ could be the simple sum of advertisement expenditures.

We obtained estimated advertisement outlays from television, radio and newspapers for the main products in the market, based on information from a Media Monitor Service. The following Table shows the estimated average share of advertisements expenditures for the period between 2012-2013:

TABLE N°2

ESTIMATED AVERAGE SHARES OF ADVERTISEMENT EXPENDITURES 2012-2013

Firm	%
Grupo Backus /1	76,8%
Ambev	19,0%
AJE Group	3,0%
Otros	1,1%

Notes:
1/ Does not include San Juan.
Source: Media monitoring service.

4.3 Estimation and hypothesis testing

4.3.1 Alternative hypothesis and estimators

The empirical specification that will be considered is:

$$\ln(s_{j,t}) - \ln(s_{0t}) = \alpha p_{j,t} + \sigma \ln(s_{j/g,t}) + \sum_{l=1}^{n_j} \delta_{lj} A_{l,j,t}^{0,5} + \xi_{j,t} \qquad (21)$$

$$\xi_{j,t} = \eta_j + \lambda_t + \upsilon_{it}$$

In order to empirically implement this specification we assume, a priori, that the sales from Backus can be grouped with the sales of Ambev, considering that a sizeable proportion of their respective sales are targeted for premium products. On the other hand, San Juan and Aje Group sales are considered as a group in which firms are much more focused on regular beers. The "outside option" in this exercise is given by Amazonica. Hence, σ would be capturing the relative mean utility level correlation inside each of the groups defined.

Assuming that all the covariates are mean independent of the error term, $\xi_{j,t}$, implies that the Ordinary Least Squares (OLS) estimator is consistent but far from being efficient, as the error term that will follow the typical Panel Data autocorrelation structure.[14] Under the hypothesis that the right hand-side variables are mean independent of the error term, every fixed effect in the error term can be taken as a random shock so that an efficient estimator will require a Generalized Least Squares (GLS) approach which will use an estimate of the variance-covariance matrix of the error term. In that case the usual choice is the Nerlove and Balestra (1996) estimator for which it is always a good idea to estimate the variance components of the error term using the small sample adjustment suggested by Swamy and Arora (1972).[15]

A more realistic approach would be to consider that the right hand side variables- especially prices and shares- are correlated with η_j because market equilibrium values are supposed to be conditioned by firm specific characteristics that are not observable (for example, reputation, facilities, etc.) that are thought to be constant in the short run. In this case, the typical Within Groups (WG) estimator which transforms every observation as a deviation from the within group time mean will be a consistent but an inefficient estimator. This transformation wipes out the η_j (as they are constant through time and its time mean) clearing the potential endogeneity of the covariates with the fixed effect.

The WG estimator will not be consistent though, if some right hand side variables are not strictly exogenous from the error term, which requires controlling the remaining endogeneity of prices and shares that are typical from demand estimations. The estimation procedure should then combine a WG estimator with an Instrumental Variables' approach (WG-IV). To that aim, at least two valid instruments should be considered. In this case, we will use the value of the application of the specific consumption tax (ISC), a price index considering variations of the international and national wholesale prices of Barley (Barley) and information of the level of employment occupation in the manufacturing industry by region (Labor).

Note that under the strong assumption that the right hand side variables are mean independent of the error term $\xi_{j,t}$, the GLS estimator is efficient but under alternative assumptions, only the WG or the WG-IV are consistent. Therefore, a series of Hausman specification tests[16] arise naturally to test different sets of hypotheses regarding the consistency of alternative estimators. In particular the following:

---

[14] See Baltagi (2008)

[15] In this last case, the procedure is to run a regression considering η_j as particular coefficients to be included in the deterministic part of the model and obtain an estimate of the variance of υ using the corresponding residuals. Then to obtain an estimator of the second variance component with typical formula $\sigma_1^2 = T\sigma_\eta^2 + \sigma_\upsilon^2$ the procedure uses the Between Groups (BE) estimator's residuals. The feasible variance-covariance matrix will be constructed using these estimators. The Swamy-Arora approach can now be obtained at no computational effort using any standard econometric package.

[16] The Hausman (1978) specification test is based upon the difference between an efficient estimator under the

Table 3 shows the corresponding results in which all the estimators briefly described above are considered under their corresponding labels. All the estimations consider time dummy variables, meaning that we include 27 dummies in the specification to account for common trends[17].

Let us focus first in the estimated coefficient for price (p in the table). The OLS estimator provides a rather unconvincing result, as the value of the estimator is positive. This is already signaling specification problems possibly due to simultaneous equation biases as well as correlation of covariates with respect to the fixed effects. In short, the OLS estimator is deemed biased under more general assumptions. The GLS estimator does not make significant progress with respect to the previous estimator, again because the strong assumption required for it to be consistent does not seems reasonable in this context.

Yet, when we perform an extra set of estimations by computing the Baltagi (1981) Random Effects Two Stage Least Squares estimator (GLS-IV) still results seem poor.

The set of WG estimators seem to perform better which was expected from the very beginning. Indeed the WG estimator is consistent under the weaker assumption of correlation of the covariates and υ Test (I) shows the classical test for the joint significance of the fixed effects coefficients. In this case the *F* statistic is large enough so as to reject the null of not significance whereas the relative values of the estimated σ_η^2 and σ_υ^2 suggests that most of the variance in the error term comes from those fixed effects. Test (II) which compares the WG and the GLS estimator following Hausman's specification test, distributed as $\chi^2(3)$ rejects the assumption for GLS to be consistent (its p-value, not shown is close to zero).

null hypothesis and a consistent one. Under the alternative, only the latter remains consistent. By construction, the Hausman Statistic, which is based on a quadratic form, requires the variance covariance matrix of the difference in coefficients to be positive definite, something that is known to be difficult to observe with Panel Data estimators. For instance, the scale of the estimated coefficients may cause that variance covariance matrix not to satisfy the asymptotic properties required to conduct the Hausman test. We did not observe this kind of problem with our basic estimations, however the Hausman statistic seems to be very sensible to the way in which advertisement is included in the empirical specification.

[17] It is always useful to recall that although 28 periods of time are available we need to drop one dummy to avoid the so called dummy trap.

TABLE N°3

ESTIMATION RESULTS FOR VARIOUS SPECIFICATIONS FOR $ln(s_j) - ln(s_0)$ /1 /2 /3

Variable	OLS	GLS	GLS-IV	WG	WG-IV
p	0,013* (2,008)	0,013 (1876)	0,018* (2,305)	-0,004*** (-4,169)	-0,004*** (-3,297)
$\mathcal{A}$	0,001*** (12,814)	0,001*** (12,281)	0,001*** (12,285)	0,000 (0,738)	0,000 (0,911)
$s_{j/g}$	0,517*** (8,553)	0,517*** (8,34)	0,479*** (7,35)	1,026*** (39,334)	1,108*** (14,428)
Constant	2,082** (3,146)	2,082** (2,596)	1,506 (1,712)	5,055*** (42,377)	5,149*** (27,734)
σ_n σ_v		- 0,091	0,096	1,221 0,091	1,243 0,096
obs R^2	112 0,852	112 0,852	112 0,850	112 0,546	112 0,547
Tests (I) F:n_1=...n_j=0 (II)Hausman				1 588,92 79,21	1 264,11 136,59

Notes:
1/ *t*-stats in parenthesis; *: coefficient significant at the 5% level, ** : significant at the 1% level, ***: significant at the 0,5% level.
2/ VI regressions are performed assuming deviations to the mean of prices and shares are endogenous.
3/ Test (I) is the standard test of joint significance of the *J-1* fixed effects whereas Test (II) shows the standard Hausman specification tests.
Source: Own estimations.

Finally, the WG-IV estimator controls further sources of endogeneity, although it reduces the precision of the estimations. As with the WG estimator, the price coefficient appears with the correct sign but, in contrast with that it suffers from a large loss of precision which is reflected in the smaller t-statistic. Again, the Wald test of joint significance of the fixed effects (Test I in the Table 3) rejects the null that all of them are zero.

On the other hand, the Hausman specification test reflects the test suggested in Baltagi (2004) based on the comparison of the WG-IV and the GLS-IV. This test, distributed as $\chi^2(29)$, rejects that the information at hand meets the assumptions for the GLS-IV estimator to be consistent.

This last result must be taken with care. Although we presume that the WG-IV is superior, even with respect to the WG estimator, we can only say that the GLS is not ideal because either the unobserved fixed characteristics of the firms are conditioning the market outcomes (prices and shares) or because there is a bias due to simultaneous equations or both, but we are not able to distinguish which one is the source of the problem.

This leads us to a different set of questions relating the validity of the fixed effects specification that will be covered in the following sub-section.

Focusing on our apparently superior specifications, WG and WG-IV, the estimator for σ, the coefficient of the "own group share" variable, $s_{j/g}$, is clearly not statistically different from the unity. This means that imposing our a priori assumption that Backus and Ambev might belong to a specific market segment whereas San Juan and AJE Group might belong to a separated was not unreasonable. However, the cross-price elasticities within each group are implausible as they will be extremely high (just plug $\sigma \cong 1$ in the expression (17).

Note however that surprisingly, in all of the five specifications the estimator for σ appear consistently different from zero within a range between ~0,5 and ~1,0 so that it seems to be the case that the Conditional Logit assumption depicted in section 3 is at least difficult to verify in practice and more flexible specifications are, in principle, more appealing such as that of the Nested Logit.

4.3.2 Alternative hypothesis and estimators

As noted in the previous sub-section, the data at hand and the methods for controlling various sources of bias indicates that the GLS estimators are not consistent even with a complementary IV approach. The alternative comparison is the WG estimators but the result should not be taken as evidence that the true specification for the data is necessarily that of a FE model.

It is however a common mistake to take as given that whenever the Hausman test rejects the random effects (GLS) models the FE model, estimated by the WG estimator that it is free from specification problems. To understand why this conclusion is flawed in general, consider the assumption behind the WG estimators. The WG estimator considers all of the right hand side variables potentially correlated with the fixed effects, in which case, wiping them out for good makes sense. But this is not however the general case.

Chamberlain (1982) proposed a way to test whether the FE model, with those effects correlated with the right hand side variables, is appropriate. In particular, Chamberlain modeled the relation behind the WG estimator by resorting to an auxiliary equation that imposes that the fixed effects are determined by a linear combination of all leads and lags of the explanatory variables. Consider a vector $X'_{jt} = \left[p_{jt}\, , \mathcal{A}_{jt}\, , ln(s_{j/g,t})\right]$ then the fixed effects are modeled in the following way:

$$\eta_j = X'_{j1}\gamma_1 + \cdots + X'_{jT}\gamma_T + \omega_j \qquad (22)$$

where every γ_t is a (3×1) vector of coefficients in our beer market illustration. Introducing this specification into the original model will, in practice, produce a reduced form equation where testable hypothesis on the restrictions imposed over the coefficients are available to the researcher. By construction ω_j is independent of the explanatory variables and as in the WG estimator, the explanatory variables have to be independent of the remaining time variable error term to provide for consistent estimates.

As noted before when imposing the auxiliary equation the fixed effect model is actually imposing restrictions on the reduced form parameters with respect to the structural parameters. For example, for period $t = 1$ the reduced form equation will be

$$\ln\left(s_{j,1}\right) - \ln(s_{01}) =$$

$$(\alpha + \gamma_{11})p_{j,1} + (\sigma + \gamma_{21})ln\left(s_{j/g,1}\right) + (\delta + \gamma_{31})\mathcal{A}_{j1} + X'_{j2}\gamma_2 + \cdots + X'_{jT}\gamma_T + \lambda_t + \upsilon_{jt} \qquad (23)$$

So that in effect one would be estimating for period $t = 1$:

$$\ln\left(s_{j,1}\right) - \ln(s_{01}) = \pi_{11}p_{j,1} + \pi_{21}\ln\left(s_{j/g,1}\right) + \pi_{31}\mathcal{A}_{j1} + X'_{j2}\gamma_2 + \cdots + X'_{jT}\gamma_T + \lambda_t + \upsilon_{it}$$

Therefore if N is sufficiently large, the reduced form parameters, for each period of time, could be estimated consistently using the specific cross-firm variation.

Therefore, if one is able to estimate the structural parameters for example, by means of a simultaneous equations estimation (the typical choice would be the efficient Three-Stage Least Squares estimator) it would be straightforward to test the Fixed Effects specification by testing simultaneously all the restrictions implied to the structural reduced form parameters.

Angrist and Newey (1991) show that the procedure is equivalent to try to infer the $Tk + k$ parameters implied in the FE model using the T^2k parameters of the reduced form model. Notice that from equation (23) we will be able to estimate $k + (T-1)k$ parameters, but then this will have to be done for each T so that in effect we will have $(k + (T-1)k) \times T = T^2k$ reduced form parameters from which "deduce" the structural coefficients of the FE specification.

The procedure loosely described in the previous paragraphs could be cumbersome. However, Angrist and Newey (1991) showed that Charmberlain's approach to test the FE specification is equivalent to construct a test statistic by estimating T independent linear regressions by OLS of the WG residuals, $\widehat{\tilde{v}_{Jt}} = \widehat{v_{Jt} - \bar{v}_J}$, one for each period of time, against all leads and lags of the covariates of the original model. The statistic is simply the sum of T elements each of which is the degrees of freedom times the R^2 of each of the linear regressions.[18]

In our case, such procedure is not feasible because for each period of time we would have only four observations. For illustration purposes we reduced the dimensionality problem dividing our sample into four periods or waves so that $T' = \frac{T}{4} = 7$ and take each observation within one period of time as independent of the rest, so that we will have $N' = 7 \times 4 = 28$ observations in each of four "artificial" waves. Each regression will have $4 \times 3 + 1 = 13$ parameters to estimate (three for the covariates plus the constant times the number of artificial waves).

Finally, we can construct what Angrist and Newey (1991) call the $h(\alpha, \sigma, \delta, \gamma_1, \dots, \gamma_T)$ statistic that will have a Chi-square distribution function with $T'^2k - (T'k + k\ ) = 33$ degrees of freedom. We calculate $h(\alpha, \sigma, \delta, \gamma_1, \dots, \gamma_T) = 26{,}05$ whereas the $\chi^2(33)$ critical value at the 5% level of significance is $47{,}4$. This means that in our exercise we cannot reject the FE specification.

V. CONCLUDING REMARKS

We have shown how Panel Data estimators can be used to identify consistent estimates of discrete choice specifications that come from the random utility model with product differentiation. In particular we showed a relatively flexible discrete choice specification for consumer preferences that is close to the Nested Logit model and applied it to a sample of firm level data from the Beer industry in Peru.

We showed that the FE specification, and its WG estimator that accounts for correlated covariates with unobservable fixed effects at the firm level performs better than the RE specification estimated by a GLS estimator. This could be an appropriate approach as most likely unobserved product or firm characteristics are fixed through time and will always be potentially correlated with observed prices and market shares.

We also considered that given the potential simultaneity bias in demand regressions an IV approach would further complement the WG estimator. We therefore performed the

[18] This requires also that the time variant error term of the model to be homoskedastic.

WG-IV estimator using as instruments the beer price index and shares estimated amounts of the ISC tax, employment information and a combination of international and local prices of Barley. The estimates does not change notably with respect to the WG estimator, however we test a GLS-VI estimator against the WG-VI as suggested in Baltagi (2004), rejecting the assumptions for the validity of the GLS-VI. The latter result suggests that in this case the GLS-VI is not consistent (even the price effect appeared with the wrong sign).

We took a step forward and text the Fixed Effects empirical specification following Chamberlain (1981) and Angrist and Newey (1991). For that, however, we required to reduce the dimension of the parameters to be estimated in the test. With this preliminary approach we could not reject the Fixed Effects specification (i.e. fixed effects correlated with the covariates), which suggests that the WG estimator is the recommended choice.

Estimates of the mean utility correlation for firms pertaining to the same group in the Nested Logit appear to be in a range between 0,5 and 1,0, which means that simple consumers' preferences specifications, such as the Conditional Logit, would not be appropriate. However the implied cross-price elasticities within a group of similar firms is excessive as the parameters approached the unity.

Nonetheless, the Random Utility model combined with Panel Data techniques at the firm or product level can be used in a simple way to provide evidence of different levels of substitution among products with more detailed information that can help to better support decisions regarding relevant market analysis in antitrust cases.

VI. REFERENCES

ANGRIST, J. and NEWEY, W. (1991). *Over-identification tests in earnings functions with fixed effects*. Journal of Business & Economic Statistics 9, pp. 317-323.

BALTAGI, B. (1981b). *Simultaneous equations with error components*. Journal of Econometrics 17, pp. 189–200.

BALTAGI, B. (2004). *Panel Data: Theory and Applications*. Physica-Verlag, Heidelberg.

BALTAGI, B. (2008). *Econometric Analysis of Panel Data*. 4th Edition, Wiley

BEN-AKIVA, M, and LERMAN, S. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, Cambridge, Ma.

BERRY, S. (1994). *Estimating Discrete Choice Models of Product Differentiation*. Rand Journal of Economics, 25, pp. 242-262.

BERRY, S., LEVINSOHN, J. and PAKES, A. (1995). *Automobiles Prices in Market Equilibrium*. Econometrica, 63, pp. 841-890.

CARDELL, N. (1997). *Variance Components Structures for the Extreme-Value and Logistic Distributions with and Application to Models of Heterogeneity*. Econometric Theory, 13, pp. 185-213.

CHAMBERLAIN, G. (1982). *Multivariate Regression Models for Panel Data*. Journal of Econometrics, 18, pp. 5-46.

CORONADO, J. (2008). *Can Market Structure Explain High Price Cost Margins in the Pharmaceutical Industry: The Case of Antihypertensives*, un-published Ph.D. Thesis Chapter presented in the "Swiss IO Day", University of Bern.

MANSKI, C. (1977). *The Structure of Random Utility Models*. Theory and Decision, 8, pp. 229-254.

MCFADDEN, D. (1978). *Conditional Logit Analysis of Qualitative Choice Behavior*. Frontiers in Econometrics, ed. by P. Zarembka. Academic Press.

MOTTA, M. (2004). *Competition Policy*. Cambridge University Press.

NERLOVE, M. and BALESTRA, P. (1996). *Formulation and Estimation of Econometric Models for Panel Data*. Chapter 1 in L. Mátyás and P. Sevestre, eds., The Econometrics of Panel Data: A Handbook of the Theory With Applications (Kluwer Academic Publishers, Dordrecht), pp. 3–22.

NEVO, A. (2001). *Measuring Market Power in the Ready-to-Eat Cereal Industry*, Econometrica, 69(2), pp. 307-342.

ROJAS C. and EVERETT P. (2008). *Demand for Differentiated Products: Price and Advertising Evidence from the U.S. Beer Market.* International Journal of Industrial Organization 26, pp. 288–307.

RUBINFELD, D. (2000). *Market Definition with Differentiated Products*. Antitrust Law Journal, Vol. 68 No. 1.

SUTTON, J. (1998). *Technology and Market Structure, Theory and History*. MIT Press.

SWAMY, P.A.V.B. and ARORA, S.S. (1972). *The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models*. Econometrica 40, pp. 261–275.

WOOLDRIDGE, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Appendix: Derivation of the linearized nested logit specification

Consider the following definitions:

$$\mathcal{s}_g(\delta,\sigma) = \frac{\left[\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}}\right]^{1-\sigma}}{\sum_g \left[\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}}\right]^{1-\sigma}} \qquad \text{(a1)}$$

$$s_{j/g}(\delta,\sigma)=\frac{e^{\frac{\delta_j}{1-\sigma}}}{\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}}} \qquad \text{(a2)}$$

As noticed in section 3.2, the unconditional probability of choosing product *j* is given by:

$$s_j(\delta,\sigma)=\frac{e^{\frac{\delta_j}{1-\sigma}}}{\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}}}\times\frac{\left[\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}}\right]^{1-\sigma}}{\sum_g\left[\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}}\right]^{1-\sigma}} \qquad \text{(a3)}$$

Where:

$$s_0(\boldsymbol{\delta},\sigma)=s_{0/1}(\boldsymbol{\delta},\sigma)s_1(\boldsymbol{\delta},\sigma)=\frac{1}{\sum_g\left[\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}}\right]^{1-\sigma}}$$

This last expression can be plugged into (a3) to obtain:

$$s_j(\delta,\sigma)=\frac{e^{\frac{\delta_j}{1-\sigma}}}{\left[\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}}\right]^{\sigma}}\times s_0 \qquad \text{(a3')}$$

Replace the theoretical probabilities with the corresponding market shares to obtain:

$$\ln(s_j)=\frac{\delta_j}{1-\sigma}+\ln(s_0)-\sigma\ln\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}} \qquad \text{(a4)}$$

Now, recall:

$$s_{j/g}(\delta,\sigma)=\frac{e^{\frac{\delta_j}{1-\sigma}}}{\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}}} \qquad \text{(a5)}$$

Take logs:

$$\ln(s_{j/g})=\frac{\delta_j}{1-\sigma}-\ln\left(\sum_{j\in\mathcal{J}_g} e^{\frac{\delta_j}{1-\sigma}}\right) \qquad \text{(a6)}$$

And replace in (a4)

$$\ln(s_j)=\frac{\delta_j}{1-\sigma}+\ln(s_0)+\sigma\ln(s_{j/g})-\sigma\frac{\delta_j}{1-\sigma}$$