The Effect of Differences in Buyer and Non-Buyer Characteristics on Equilibrium Price-Elasticities: an Empirical Study on the Italian Automobile Market

Franco Mariuzzo*
Department of Economics
Trinity College of Dublin
Ireland†

August 4, 2005

Abstract

This paper provides an empirical analysis of own- and cross-price elasticities of substitution for the 1989-2000 Italian automobile industry. We use product-level data consistent with a structural model of equilibrium in a differentiated products oligopolistic industry and follow a random coefficient model to get reliable price elasticities of substitution. We enrich our data with a special section having information on individuals buying and non-buying vehicles. Tracking the different characteristics of individuals buying and non-buying ends to be determinant for more precise estimations. Our data show that models that disregard this refined information tend to consistently overestimate price elasticities.

KEYWORDS: Differentiated Products, Discrete Choice, Automobile Industry, Structural Models, Simulation-Based Models.

JEL classifiers: JL11, JL62.

*Tel. +353 1 608 3227; fax +353 1 677 2503; e-mail: mariuzzf@tcd.ie.

†Special thanks to Editoriale Domus, Fiat S.p.A. and Bank of Italy for having provided me with the rich dataset. The responsibility for any errors is my own.

1 Introduction

Price elasticities are a fundamental key variable in the understanding of markets: computations of the percentage change of the market share of a product to its own (or its competitors) percentage price change depict a determinant picture of the underlying competitive environment. The relevance of this information is, as a matter of fact, stressed by the frequent use the Lerner Index has in the Industrial Economics literature.[1] However, if this relation is straightforward in case of homogenous products (and symmetric firms), its extension to the case of differentiated products asks for more sophisticated tools that are to be introduced in the next sections.[2] The main contribution of our paper is to shed light in the important implications of a refinement in these tools.

Our paper uses aggregate industry data and estimates the own- and cross-price elasticities of substitution for the 1989-2000 Italian automobile industry. Aware that Berry Levinsohn and Pakes (2004) and Petrin (2002) highlight a certain difficulty in getting good demand estimates from aggregate industry data, we add our aggregate data a special microdata section with rich information on the characteristics of households buying and non-buying vehicles. We use this information to recover proper income distributions that will be lymph of some of our simulations. We are not aware previous literature dealing with aggregate level data and exogenous individual information was able to separate

[1] The Lerner Index is defined as a weighted average of each firm's margin, with weights given by the firms' market shares. It is often used in the literature to represent the relation between market power (the level of competition) and price elasticities.

[2] Mariuzzo, Walsh and Whelan (2003) uses a nested logit approach to question the existence of a relation between firm size and market power in the Irish differentiated Carbonated Soft Drinks.

the characteristics of buyers from those of non-buyers. The fact that often distributions of these characteristics largely differ can raise a specification issue. Our data show that the income distributions of individuals buying new vehicles significantly differ from those of the entire sample. This difference is not only confined to the means of the distributions but concerns, also, their concentration, and it is this latter that makes our microdataset extremely important. In fact, controlling for the subset of individuals buying vehicles let us not only to reduce the simulation errors (increasing, subsequently, the efficiency of our estimates) but, also, to properly identifying the parameters that enter (directly) the computation of our price elasticities.

Two main approaches can be distinguished in the empirical automobile literature. A first one that uses disaggregate consumer data and a second one that uses aggregate industry data. The former, is mainly based on logit models that estimate demand at an individual level either directly Berkovec (1985) or, through nested versions assuming a priori ordering [Ben-Akiva (1973), McFadden (1978), Berkovec and Rust (1985)]. Data are, in this case, required to match product characteristics with consumer characteristics. In such a way, one allows both for a high degree of product differentiation and for consumer heterogeneity but, as drawbacks, pays both the price of neglecting the supply side, with all the subsequent equilibrium considerations, and of having a sample that seldom is fully representative of all marketed models. The former of these issues is promptly overcome in Goldberg (1995). Goldberg assumes the existence of a Nash equilibrium and provides an equilibrium analysis of demand and supply using a nested model. Another attempt to address an equilibrium analysis, and robust also to the later of the issues, is proposed by Berry Levinsohn and Pakes (2004). On the other side, the alternative approach of aggregate industry data (often the only source of data available) addresses demand and supply and, survives the critics of efficiency by adding exogenous information on individual characteristics [Berry Levinsohn and Pakes, (1995) - BLP onwards - and Petrin (2002)]. Among this empirical aggregate industry literature the main paper is undoubtedly BLP. It offers estimates of demand and supply in the U.S. differentiated automobile markets and suggests fine econometric tools to get more reliable own- and cross- price elasticities. The authors provide results using a GMM estimator and suggest simulations to recover market shares [Mac Fadden (1989), Pakes and Pollard (1989) for details]. In order to get more efficient estimates, they enrich their product level data with exogenous information on consumers' income characteristics but no distinction between buyers and non-buyers is made. Aware of possible poor demand estimates produced by market level data, and rather general exogenous individual characteristics, the authors extend BLP by adding, this time, microdata enriched with consumers' second choice information [Berry Levinsohn and Pakes (2004)]. They find that unobservable consumer attributes (our σ_k, *infra*) are both relevant to obtain reliable substitution patterns and to get better estimates. A similar aim of providing more precise parameters' estimates leads Petrin (2002) to improve the market level data with readily available data that relate the average characteristics of consumers to the characteristics of the products they purchase. These more precise estimation results are then used by the author to evaluate the welfare benefits of the minivans' introduction.

Our paper takes an intermediate position between the aggregate market level data literature and these later subsequent micro-refinements. It departs from the use of aggregate data and shows the relevance of adding separate information on the different characteristics of the group of individual purchasing and non purchasing the product. This kind of information is usually available at any national micro-level survey. The availability of this resource let the researcher to control for differences in the characteristics of the group of individuals: i) purchasing the good; ii) preferring the outside alternative. Especially in the case of a durable good, such as automobiles, these differences can be remarkably, and matter.

Eventually, a main pitfall in all the empirical automobile literature cited above is due to a non satisfactory treatment of dynamics. Depending on their expectations about future economic and family conditions, households may prefer to defer their purchase of a new car. The static nature of the considered models (mainly dictated by data availability) fails to take intertemporal substitution effects into account and fails, also, to consider parent houses strategic entry and exit.

The paper is organized as follows. The next section summarizes the evolution of the Italian automobile industry and section 3 describes our data. In section 4 we outline the underlying theoretical model. Section 5 highlights our estimation methods which, mainly, follow BLP. Section 6 addresses the computational mechanisms. In section 7 we provide our estimation results. Finally, the paper concludes in section 8.

2 The Italian Automobile Industry

A particular feature of the Italian automobile industry is the dominant role played by Fiat. Figure I shows the evolution of the Italian (new) automobile market for the 1989-2000 period split into four different geographical areas.[3] The graph on the top left highlights the substantial reduction (from a market share of 60% in 1989 to a market share of 36% in 2000) of the national, mainly Fiat, unit sales due to the progressive opening of the national market to foreign competition. This is partly explained by the progressive elimination of import duties as requested by the European Union and, mostly, by the accelerated competition produced by the European integration. A particular attention deserves the graph on the bottom right (Far East market shares), for the Far East sales were originally bounded to market quota. This quota restriction has steadily been removed up to disappear in 1999.[4]

A full picture of the state of health of the Italian automobile market is represented by the trend of total unit sales in Figure II. We observe a sharp collapse, about 20%, in the total unit sales in year 1992. This strong reduction goes along with the exit of the Italian Lira from the European Monetary System [(EMS), October 1992] due to an attempt of the Bank of Italy to maintain the exchange

[3]We ascribe each parent house an area by virtue of the country where its head office is located. We are in this way not addressing the labyrinth of all possible cross mergers, acquisitions or other market transactions.

[4]In our estimates we use cross dummies parent house-time to control for this macro effect.

rate anchored with the other EU currencies (an effect in line with the strong reduction in the Italian market shares - see Figure I -) and the subsequent economic destabilization which followed. The 1992 crisis motivated the heavy Amato's financial act which behaved an economic stagnation up to 1997, year when the policy makers intervened with the "*scrap-incentives*".[5] This turbulence seemingly affected the level of concentration in the Italian market. Table I offers three different measures of the level of concentration (the indexes C4, HHI and Gini coefficient). As expected concentration has decreased over time passing from a high concentrated industry (HHI above 1800) to a moderately concentrated industry (HHI between 1000 and 1800).[6] All indexes well support this trend.

Table II shows the 1989-2000 trend for some major physical automobile characteristics. We observe increases in cubic capacity and speed to go simultaneously along with a reduction in the fuel consumption. This trend is explained by a change in individual tastes towards faster but more fuel saving cars (dictated by the almost doubling of the gasoline real prices - Figure III -). Although Table II does not report information on airbags and ABS as standard, we approximate safety by the variable length and (somehow) trunk size. The increased variability in all characteristics, but trunk size, let us think to a greater differentiability in the marketed products.

Finally, Figure IV offers a picture of the real price-distribution trend. The time trend can be split into two subperiods: i) 1989-1992: a period where the

[5]Mariuzzo (2005) offers a story on the effects of the scrap incentives on the level of competition in the market.

[6]See the 1992 Guidelines. Reprinted in *Trade Regulation Reports*, June 5, 1992.

average price (of the marketed automobiles) goes, in his trend, along with the price variability; ii) 1993-2000: a period of price stabilization associated to a process of reduction in the price variability which effect might be due to an increased level of competition in the market.

3 The Data

Our dataset consists of three dimensions: A) Individuals; B) Products; C) Time.

A) Individuals are households drawn from the Bank of Italy Surveys on Households' Income and Wealth (SHIW - see Table III -). Apart from data on households characteristics such as disposable income, family components, area, age etc., the dataset holds a special section on vehicles' purchase (Table IV). Households are asked whether or not they bought/sold a vehicle in the year and, in case of transaction, the price they, respectively, paid/received. We use this information both to obtain better demand estimates and to get a measure of our outside good market share s_0 (share of households not buying a new car - Figure V -).[7] Unfortunately, our data don't let us to distinguish between used and new vehicles. We propose in Appendix A a method of minimum distances to recover the subset of households buying a new vehicle.

B) Products include information on sales, list prices and physical characteristics such as, engine attribute (kilowatt, cubic capacity), dimension (length), comfortability (number of doors, trunk size) and performance variables (fuel consumption, acceleration time, maximum speed). All this information is avail-

[7]In our estimates we compute the outside good market share from the total number of households in the economy.

able in three different datasets (two furnished by *Editoriale Domus-Quattroruote* and one by *Fiat).*[8] To be more specific:

i) a former *Quattroruote* database offers information on prices. We have for the 1989-2001 period 65715 quarterly prices. We, then, reduce the dataset to yearly prices by averaging the quarterly prices.

ii) a latter *Quattroruote* database furnishes information on all auto characteristics introduced above. The original dataset contains 16111 observations, of which only 11125 are not the same model repeated.[9]

From 1996 on, the variable fuel consumption (liters*100km) is marked by three different EU standards: 1) urban, 2) sub-urban, 3) mixed, while before that period the distinction was among: a) urban, b) 90 km/h (sub-urban), c) 120 km/h. Averaging a), b) and c) one gets a good approximation of the mixed fuel consumption in 3).

iii) Finally, the *Fiat* database (11246 different models) offers information on market segments (28 exogenous different segments), quantities sold each year, body, type of engine and few characteristic variables such as kilowatt, cubic capacity and number of doors.

By merging ii) and iii) we obtain, for the 1989-2000 period, a database of 11055 models and, once we consider the models reported in different years to be different observations we have a pooled dataset of 46533 observations.[10]

[8] A special thank to Andrea Battiston for having patiently added the *Fiat* dataset the *Editoriale Domus-Quattroruote* id codes (*Infocar-anno-mese*) necessary to merge the two different datasets.

[9] It could be the case that some of the observations that we consider repeated (the difference between 16111 and 11125) differ each other for some characteristics unobservable in our data.

[10] Models with unit sales below 500 a year have been excluded.

C) <u>Time</u> includes the years 1989, 1991, 1993, 1995, 1998, 2000. This particular spell is constrained to the availability of our Bank of Italy data.

In next section we define our *model/year* and *different model/year* which numerosity is reported in Table V.

4 The Model

The number of competing parent houses (see Table V) is small enough to let us think the Italian automobile industry as an oligopolistic market structure with highly differentiated automobile models. The underlying game is assumed to be a differentiated product static game with prices as strategic variables. We model parent houses as price-setting oligopolists and consumers as price takers and assume the existence of Nash equilibrium in prices.[11] We model demand as a discrete choice setting where each consumer (household in our data) decides to buy the car that gives him the highest utility (considering also the utility of the outside good: not buying a new car) and we model the other side of the market as multiproduct firms (the parent houses in Table V) that choose the prices that maximize their profits.

Let F_t be the set of parent houses in our market and J_t the set of all different models produced at time t and let each parent house $f_t \in F_t$ to produce at time t a $J_{ft} \subset J_t$ subset of models.

In order to save some notation we drop onwards, when not especially necessary, subscripts f and t.

[11] Singh and Vives (1984) show the duality of price and quantity assumptions in a differentiated product setting.

We assume consumers' utility to depend on product characteristics, prices and individual taste parameters. The aggregation of our discrete-choice model of consumer behavior produces the market demands; while the cost side is based on the assumption of a functional form and obtained from the first order conditions of the profit maximization.

Before proceeding further we need some definitions of what is a model in our framework. The previous section presented our original dataset as containing 11055 models (p*arent house-name plate-types*).[12] If one considers the models reported in different years to be different observations, one finds himself with a pooled dataset of 46533 observations. Both numbers are far too large to run our implemented algorithms. We have to restrict the number of models. Table V shows our process of aggregation. We distinguish between *model-year* and *different-model-year*. A *model-year* is a string *parent house-name plate* which vector of observed characteristics is the weighted (by unit sales) average of characteristics of *types* with the same string *parent house-name plate*. Furthermore, as our dataset has a panel component, we define a *model-year* to be the same over time if none of its characteristics has changed more than ±20% over a period. It follows what is a *different-model-year*. In the rest of the paper, if not explicitly stated, we mean by model our <u>*model-year*</u> j_t and interpret j_t as an integer.[13]

In the following three subsections we follow closely the BLP notation and

[12]Examples are Fiat-Panda-Young, Fiat-Panda-1100, etc.

[13]In our estimates we control for *different-model-years* introducing the variable number of years the model is marketed.

describe the demand, the supply (the cost side) and the market equilibrium.[14]

4.1 The Demand Side

We derive our demand by aggregating a discrete choice model of individual consumer behavior. We are aware that, when choosing among different models of cars, individuals do not restrict their decisions only on prices but, they also consider the different characteristics. This approach, suggested by Lancaster (1971, 1991), offers the possibility of moving from the product space to the characteristic space, which is quite useful when one has (as in our case) to deal with many products and few characteristics. With this approach we explain better why products, although physically similar, may differ in consumers' perception about quality, durability, status, or services.[15] Unfortunately, some characteristics such as style, reputation and past experience are unobservable to us but, meanwhile, they are rather frequent determinants of consumers' demand and we don't want to neglect their effects in our model.[16]

We represent the utility derived by consumer $i \in I$ from consuming product j to be $U(\boldsymbol{\zeta}_i, p_j, \mathbf{x}_j, \boldsymbol{\xi}_j; \boldsymbol{\theta})$. Where I is the number of individuals in the economy (households in our case), $\boldsymbol{\zeta}_i$ is a vector of individual i's characteristics whereas $(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi})$ are vectors and matrices of product characteristics. In our notation $\mathbf{p}$ represents the price vector of our products and $\mathbf{X}$ and $\boldsymbol{\xi}$ are our matrix

[14]In our paper the simultaneous estimate of demand and supply is confined to an efficiency reason. However, any normative analysis or policy experiment is conditioned to an underlying equilibrium and a simultaneous estimate of demand and supply.

[15]See Anderson, De Palma and Thisse (1992) other than for a good revision of discrete choice models, for the conditions of a one to one correspondence between the discrete choice and the address (characteristics) models.

[16]Manski (1977) argues that the randomness in observed consumer behavior is mainly due to unobservable characteristics that influence consumers' choice.

of observed and vector of unobserved products' characteristics, respectively. Finally, $\boldsymbol{\theta}$ includes any parameters that determinate the distribution of consumer characteristics $\{\alpha, \boldsymbol{\sigma}\}$, as well as, conditional on these characteristics, the utility parameters that describe the utility surface $\boldsymbol{\beta}$ and the marginal costs $\boldsymbol{\gamma}$. We partition $\boldsymbol{\theta}$ in $\boldsymbol{\theta} = \{\boldsymbol{\theta_1}, \boldsymbol{\theta_2}\}$ where $\boldsymbol{\theta_1} = \{\boldsymbol{\beta}, \boldsymbol{\gamma}\}$ is the subset of parameters that are to be concentrated out of our objective function and $\boldsymbol{\theta_2} = \{\alpha, \boldsymbol{\sigma}\}$ the parameters that enter our objective function.

We avoid, onwards, to represent vectors in bold letters and matrices in capital and bold letters.

From the discrete choice literature McFadden (1981), consumer i chooses model $j \in J$ if and only if it maximizes his utility

$$U(\zeta_i, p_j, x_j, \xi_j; \theta) \geq U(\zeta_i, p_r, x_r, \xi_r; \theta) \text{ for } j, r = 0, 1, ..., J \tag{1}$$

where $r = 1, ..., J$ alternatives represent purchases of the competing differentiated products whereas, alternative zero $r = 0$, or the outside alternative, represents, in our case, both the option of not purchasing any of those products (allocating therefore all expenditures to other commodities) and the option of purchasing a used car.

We expect consumers with different individual characteristics to make different choices. For this reason, we define

$$A_j = \left\{\zeta : U(\zeta, p_j, x_j, \xi_j; \theta) \geq U(\zeta, p_r, x_r, \xi_r; \theta), \text{ for } r = 0, 1, ..., J \right\}$$

to be the set of values of individual characteristics ζ that induce the choice of good $j \in J$. Assuming ties occur with zero probability (which means the distribution function P of ζ is absolute continuous with respect to the Lebesgue measure), we obtain, by aggregation, the market share of good j

$$s_j\left(p, x, \xi; \theta\right) = \int_{\zeta \in A_j} P_0\left(d\zeta\right) \qquad j = 0, 1, ..., J \tag{2}$$

where $P_0\left(d\zeta\right)$ is nothing but the density of ζ in the population and $0 < s_j < 1$ [with $\sum_{j=0}^{J} s_j\left(\cdot\right) = 1$] the market shares. We describe in section 6 how to compute $s_j\left(\cdot\right)$.

Finally, given the number of consumers I in the economy one can derive the aggregated demand functions

$$q_j\left(\cdot\right) = I s_j\left(p, x, \xi; \theta\right), \text{ for } j \in J \tag{3}$$

which are nonlinear functions of the observable and unobservable product characteristics.

The next step is to assume some shape to the utility function. As well known from the discrete choice literature [well documented in Anderson, De Palma and Thisse (1992)] a utility additively separable in its product characteristics [our δ_j defined in (6)] and consumers' characteristics (the error component ϵ_{ij}) provides poor substitution effects. That is, conditional on market shares, elasticities of substitutions do not depend, in that case, on the observable characteristics of the product. This can easily be reformulated with an example. Suppose Fiat 500 (a cheap Italian car) and a Ferrari (which prestige everybody knows) have close market shares (our s_j) then, a change in the price of a Porsche (another luxury car which prestige is well known) should affect both models in the same way. Which is hard to believe. We need a specification that captures the idea that goods with closer characteristics are expected to have higher cross-price elasticities. What we have in mind is, whenever individuals have preferences for some specific characteristics, we expect them to have a potential second choice in the subset of cars with similar characteristics. Furthermore, whenever a new car enters the market, we expect it to have a higher effect on the demand of cars with similar characteristics. In order to get more reasonable substitution patterns we suggest a functional form that allows for interaction between individual and product characteristics (known in the literature as *random utility models*). Following BLP we nest a random coefficient model into the following Cobb-Douglas utility function

$$U(\zeta_i, p_j, x_j, \xi_j; \theta) = (f(y_i) - p_j)^{\sigma_y} D\left(p_j, x_j, \xi_j, \nu_i; \cdot\right) \exp(\epsilon_{ij}) \tag{4}$$

where $f(y)$ is some function of the individual income.[17] Finally, ϵ_{ij} (assumed to be i.i.d. across products and consumers). We assume $D(\cdot)$ to be linear in logs and $K \geq K_1$, then

$$u_{ij} = -\alpha p_j + \sigma_y \ln(f(y_i) - p_j) + \sum_{k=1}^{K} \beta_k x_{jk} + \xi_j + \sum_{k=1}^{K_1} \sigma_k x_{jk} \nu_{ik} + \epsilon_{ij} \tag{5}$$

for $j = 1, ..., J$, while

$$u_{i0} = \alpha_0 \ln(f(y_i)) + \xi_0 + \sigma_0 \nu_{i0} + \epsilon_{i0}$$

where $(\nu_{i0}, \nu_{i1}, ..., \nu_{iK_1})$ is a vector of idiosyncratic consumer tastes that interact with product characteristics (*alias* marginal utility of characteristics); x_{j1}

[17] We assume the function of income to have the following functional form:
$$\begin{array}{c} p_j + 1 \text{ if } y_i < p_j + 1, \forall i \in I, j \in J \\ y_i \text{ otherwise} \end{array}.$$

is a dummy vector and σ_k the standard deviation of the marginal utility distributions. This representation assumes individuals to have different preferences for each different observable characteristic. The effect of x_k units of characteristic k on the marginal utility is $(\beta_k + \sigma_k \nu_{ik})$. One may observe, from the distribution of tastes for characteristic k, how, higher values of β_k (the mean) or σ_k (the standard deviation) explain an increase in the share of consumers buying cars with higher k characteristic values. Moreover, the value of σ_k is relevant in explaining the substitution effects. Let's suppose an increase in the price of a car with high k characteristic. In this case, consumers who substitute away from that car will: i) in case of a low variance of the marginal utility associated with characteristic k (low σ_k), not tend to substitute towards other high k cars; ii) whereas, in case of high σ_k, the opposite is true (similar products become better substitutes). This effect is simply explained by the marginal utility for the k characteristic.[18] A similar argument is valid also for the price income effect.

It is important to notice how the utility in (5) can be decomposed into:

i) a common (to all consumers) mean component

$$\delta_j \equiv -\alpha p_j + \sum_{k=1}^{K} \beta_k x_{jk} + \xi_j \tag{6}$$

ii) a deviation from that mean

$$\mu_{ij} = \sigma_y \ln \left(f\left(y_i \right) - p_j \right) + \sum_{k=1}^{K_1} \sigma_k x_{jk} \nu_{ik} \tag{7}$$

[18] Once we scale $E(\nu_{ik}^2) = 1$, we get that, the mean and variance of the marginal utility associated to the k characteristic are, respectively, β_k and σ_k^2. In our estimates we use the Delta method to control for the relation between the variance parameters σ_k^2 and the estimated standard deviations $|\sigma_k|$.

where μ_{ij} (the heterogeneity in consumer tastes) depends on the interaction between consumer preferences (ν_i) and product characteristics (x_j) and the relation between the simulated incomes $[f(y_i)]$ and prices (p_j)

iii) an ϵ_{ij} error term i.i.d. across products and consumers.

If the above decomposition has the advantage of providing more reliable substitution patterns, it has also the drawback of requiring more complex econometric procedures.

4.2 The Cost Side

Our partial equilibrium analysis is also partial in the sense we are only considering the Italian market. The limit of this approach is that, we are assuming the prices fixed by different parent houses in the Italian market, to be independent of the set of prices decided by the same companies (and competitors) on the foreign markets. We face such restriction by assuming a cost function additive in the Italian and foreign production which raises, unfortunately, the drawback of confining our analysis to a linear cost function. On the supply side we assume the following additive total cost function

$$\widetilde{C}\left(q_j;\cdot\right) = \left[\left(q_j^{Ita} + q_j^{\overline{Ita}}\right)\exp\left(w_j\gamma + \omega_j\right)\right] + F_j \tag{8}$$

where subscripts Ita and $\overline{Ita}$ stand, respectively, for the Italian and foreign market and $q_j = q_j^{Ita} + q_j^{\overline{Ita}}$. We denote with w and ω respectively, the observed and unobserved subset of cost characteristics, and with γ the coefficients to be estimated.[19] F is a fixed cost. We need to distinguish between Italian and

[19] It could be the case that a *parent house-name plate* has characteristics that vary among the different countries they are marketed. In that case, the following functional form would

foreign markets, for we only have data on the Italian production (we know only q_j^{Ita}). This drawback forces us to address variable costs in the following conditionally linear (in quantity) functional form

$$C\left(q_j^{Ita};\cdot\right) = q_j^{Ita}\exp\left(w_j\gamma + \omega_j\right). \tag{9}$$

We omit onwards in our notation superscript Ita.

Since $C\left(q_j;\cdot\right) > 0$ we get our marginal cost to be loglinear in the following vector of cost characteristics

$$\ln\frac{dC\left(q_j;\cdot\right)}{dq_j} \equiv \ln(mc_j) = w_j\gamma + \omega_j. \tag{10}$$

We expect w to be inclusive of the relevant characteristics observed by all consumers. For example larger cars or, cars with higher unobserved (to us) characteristic values, are expected to be more costly to produce.

4.3 Market Equilibrium

Given the demand system in (3), the profits of multiproduct parent house f (relative to the sales on the Italian market) are

$$\prod\nolimits_f = \sum_{j\in J_f}\left(p_j - mc_j\right)q_j \tag{11}$$

Maximizing (11) we get, for every $f \in F$ parent house, the common first order conditions

$$s_j\left(\cdot\right) + \sum_{r\in J_f}\left(p_r - mc_r\right)\frac{\partial s_r\left(\cdot\right)}{\partial p_j} = 0, j \in J_f \tag{12}$$

fit better

$$\widetilde{C}\left(q_j;\cdot\right) = \left[q_j^{Ita}\exp\left(w_j^{Ita}\gamma^{Ita} + \omega_j^{Ita}\right) + q_j^{\overline{Ita}}\exp\left(w_j^{\overline{Ita}}\gamma^{\overline{Ita}} + \omega_j^{\overline{Ita}}\right)\right] + F_j.$$

from which we get our price equilibria.

In order to write our markup relation in compact form we define

$$\Delta_{jr} = \begin{cases} \frac{-\partial s_r(\cdot)}{\partial p_j}, & \text{if models } r, j \in J_f \text{ are produced by the same parent house;} \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

where

$$\begin{aligned} \frac{\partial s_j(\cdot)}{\partial p_j} &= \int \phi_j(\zeta, \cdot)\left(1 - \phi_j(\zeta, \cdot)\right)\left[\frac{\partial \mu_j(\zeta, \cdot)}{\partial p_j}\right] P_0(d\zeta) \\ \frac{\partial s_r(\cdot)}{\partial p_j} &= \int -\phi_j(\zeta, \cdot)\,\phi_r(\zeta, \cdot)\left[\frac{\partial \mu_r(\zeta, \cdot)}{\partial p_j}\right] P_0(d\zeta) \end{aligned} \tag{14}$$

which components of (14) are going to be explained in our computational section 6.

It is possible now to rewrite our first order condition in vector notation

$$s - \Delta(p - mc) = 0 \tag{15}$$

and solve it for the price-cost markup

$$p = mc + \Delta^{-1}s. \tag{16}$$

We define the markup vector to be

$$b \equiv \Delta^{-1}s$$

such that the problem spreads in the following pricing equation

$$\ln(p_j - b_j) = w_j\gamma + \omega_j. \tag{17}$$

The next section introduces our adopted econometric procedures.

5 GMM Estimator

The fact that producers know the value of the unobserved (to us) product characteristics generates correlation between prices and unobserved product characteristics (cars with higher unmeasured quality should be sold at higher prices) and one has to face a simultaneity issue.[20] Concerned with this issue, we assume, as common in the literature, our unobservables to satisfy a conditional mean independency property

$$E\left[\xi_j\left(\cdot;\theta_0\right)|z\right] = E\left[\omega_j\left(\cdot;\theta_0\right)|z\right] = 0 \tag{18}$$

with $z = [x, w]$ our demand and supply observed characteristics and θ_0 the true parameters value. Although relatively strong, this assumption does not require prices to be uncorrelated with unobservables but only the observed product characteristics to be exogenous in our model. We make use of this conditional mean independency property together with the condition

$$E\left[\left(\xi_j, \omega_j\right)'\left(\xi_j, \omega_j\right)|z\right] = \Omega\left(z\right)$$

and use a General Method of Moments [GMM, Hansen (1982)] to simultaneously estimate our demand and supply parameters

$$G(\cdot;\theta) = E\left\{H_j(z)T\left(z\right)\left(\begin{array}{c}\xi_j\left(P_0;\theta\right)\\ \omega_j\left(P_0;\theta\right)\end{array}\right)|z\right\} \tag{19}$$

where P_0 is the population distribution, $H_j(z)$ is the matrix of functions of our exogenous observable characteristics (instruments, *infra*) and $T\left(z\right)$ is a 2x2 matrix that adjusts for correlation between demand and supply unobservables.[21]

[20] Moreover, aggregate demand (3) is a non linear function of product characteristics.

[21] See Hayashi (2000) for a good reference on GMM and pp. 856-857 of BLP for details on the correction matrix $T(z)$.

Given that P_0 is unknown we replace it with its simulated distribution P_{ns} and our GMM becomes

$$\underset{L\text{x}1}{G(P_{ns};\theta)} = \left\{ \frac{1}{J} \sum_{j=1}^{J} H_j(z) T\left(z\right) \left(\begin{array}{c} \xi_j\left(P_{ns};\theta\right) \\ \omega_j\left(P_{ns};\theta\right) \end{array} \right) \right\} \tag{20}$$

where $L = L_\xi + L_\omega$ is the total number of instruments used, respectively, on our demand and cost side. The fact that $L > |x| + |w|$ (overidentification) confines our problem to find the minimum distance

$$\underset{\theta_2}{\text{Min}} G'\left(\cdot\right) W G\left(\cdot\right) \tag{21}$$

where $\underset{L\text{x}L}{W}$ is the weigthing matrix (defined *infra* in the computational section). Notice that the minimization of (21) is only with respect to the parameter set $\theta_2 \subset \theta$ which means we use a two step non linear GMM estimation procedure. This procedure produces the following asymptotic Variance-Covariance

$$\begin{aligned} \sqrt{J}\left(\widehat{\theta} - \theta_0\right) &= \left(\Gamma' W \Gamma\right)^{-1} \Gamma' W V W \Gamma \left(\Gamma' W \Gamma\right)^{-1} \\ V &= V_1 + V_2 \end{aligned} \tag{22}$$

where Γ is the gradient of (20) with respect to the θ parameters.[22] V_1 arises from the process generating the product characteristics and V_2 from the simulation process.[23]

The following subsection lists and explains the choice of our instruments.

[22]The gradient of the non linear parameters on the cost side is computed numerically.

[23]In our estimates we compute the estimated simulation variance $\widehat{V}_2$ by employing a Monte Carlo procedure that replicates the simulations 50 times.

5.1 Instruments

We define $H_j(z)$ to be the matrix of functions of our exogenous observable characteristics. We use as instruments, apart from the variables x and w themselves, the following functional forms:

i) Each period, the average product characteristics and standard deviations of other products ($\neq j$) produced by parent house f. These instruments are good cost shifters, for they capture economies of scale: the more similar are the different products produced by the same parent house, the higher are the derived economies of scale. Cost shifters are common across products of the same parent house and short run shocks are (once allowed for cross dummies parent house-time) uncorrelated with these factors.

ii) Each period, the average product characteristics produced by other parent houses ($\neq f$). As cost shifter they capture the cost efficiency of parent house f with respect to its competitors. Short run shocks are (once allowed for cross dummies parent house-time) uncorrelated with these instruments.

iii) Each period, the sum of the products produced by parent house f. These are both cost and demand shifters and capture scale economies and demand spillovers.

In our estimates we run an overidentification test to verify the validity of the above instruments.

The next subsection describes our computation procedure.

6 Computation

We write our individual i's utility for product $j = 1, 2, ...J$ in period $t \in T$ as

$$u_{ijt} = -\alpha p_{jt} + \sigma_y \ln \left[\begin{array}{c} y_{it} - p_{jt}, \text{ if } p_{jt} + 1 < y_{it} \\ 1, \text{ otherwise} \end{array} \right] + \sum_{k=1}^{K} \beta_k x_{jk} + \xi_{jt} + \\ + \sum_{k=1}^{K_1} \sigma_k x_{jkt} \nu_{ikt} + \epsilon_{ijt} \tag{23}$$

where y_{it} are the simulated individual incomes with $i \in ns$ and $t \in T$. Whereas, our utility for the outside good is

$$u_{i'0t} = \underbrace{\xi_{0_t}}_{\delta_{0t}} + \underbrace{\alpha_0 \ln (y_{i't}(\cdot)) + \sigma_0 \nu_{i'0}}_{\mu_{i'ot}} + \epsilon_{i'0_t}$$

Since each individual's choices are invariant to i) multiplication of utility by each person specific positive constant; ii) addition to utility of any person specific number (i.e. affine transformations), we can normalize $\alpha_0 = \sigma_0 = \xi_{0_t} = 0$.

6.1 Simulations

We draw our vector $\zeta = (f_{y,1}, \nu_{1,1}, ..., \nu_{K_1,1}; ...; f_{y,T}, \nu_{1,T}, ..., \nu_{K_1,T})$ from a multivariate normal distribution with zero mean and identity variance-covariance matrix and from consumer's income distribution (assuming independence between the different distributions).[24] We draw each period $t \in T$, nsx$(K_1 + 1)$ individual observations (nsxK_1 from the multinormal distribution and nsx1 from our Bank of Italy special data section).[25] A good feature of our Bank

[24] The underlying assumption of simulating observations from a multivariate normal distribution is that individuals like some characteristics and dislike others. For example individual i might like speed but not length etc.

[25] In our empirical procedure we take 100 draws for each of the 6 periods, for a total 600 draws and, in order to avoid problems in our minimization procedure, we censor our multivariate normal distribution to 99%.

of Italy microdataset is that it let us to distinguish between the income distributions of those who purchased a vehicle and those who did not. Table VI highlights the significant differences in the income distributions of those individuals who purchased a vehicle versus the entire sample. Individuals purchasing a new vehicle have not only higher incomes but, also, more concentrated income distributions. We observe that the differences in the two distributions are statistically significant both in their means and variance ratio. We name the income distributions of the entire sample as "*coarse*" income distributions; the income distributions of those purchasing a vehicle as "*fine*" income distributions. In our results we evaluate the differences in the estimated elasticities of substitution produced by the "*coarse*" (as in BLP) versus the "*fine*" income distributions. Of course using, as in BLP, exogenous information on the "*coarse*" income distributions of the entire population raises an issue of specification.

The next subsection outlines the full estimation procedure.

6.2 The Estimation Procedure

As in BLP we obtain our market shares in two stages. In the first stage we integrate out (each period/market) over the distribution of ϵ (assumed to be a Type 2 Extreme Value) and obtain the following logistic [conditional on individual characteristics (ζ)] market share functions

$$\phi_{ijt} = \frac{e^{\delta_{jt}(\cdot)+\mu_{ijt}(\cdot)}}{1+\sum_{jt=1}^{Jt} e^{\delta_{jt}(\cdot)+\mu_{ijt}(\cdot)}} \tag{24}$$

which notation have already met with in (14). ϕ_{ijt} is the probability individual i purchases product j in period t. In a second stage, we integrate out (each period/market) over $\zeta_t = (f_{y,t}, \nu_{1,t}, ..., \nu_{K_1,t})$ and obtain our market shares

$$s_{jt} = \int \phi_{jt}(\zeta_t, \cdot) P_0(d\zeta_t). \tag{25}$$

The non closed solution of (25) asks for a simulation procedure. An immediate simulation procedure consists of replacing the population density with its simulated distribution

$$s_{jt} = \int \phi_{jt}(\zeta_t, \cdot) P_{ns,t}(d\zeta_t) \cong \frac{1}{ns} \sum_{i=1}^{ns} \phi_{ijt}(\cdot) \tag{26}$$

Product j's market share is therefore the result of an average individual probability. At this point, we are only left with the determination of our δ_{jt} (the mean component of our utility function). As in BLP we avoid its lack of analytical solution by using a contraction mapping operator

$$T_{(P_{ns,t};\theta)}[\delta_{jt}] \simeq \delta_{jt} + \ln(s_{jt}) - \ln[s_{jt}(\cdot)] \tag{27}$$

which is nothing but a recursive method to determine $\delta(\cdot)$. A recursive method that depends, among all, on the parameters θ to be estimated.

Once we have computed the mean utilities $[\delta_{jt}(\cdot)]$ we can explicit our demand unobservables

$$\xi_{jt} = \delta_{jt}(\cdot;\theta_2) - x_{jt}\beta \tag{28}$$

and use the pricing equation (16) to explicit our supply unobservables

$$\omega_{jt} = \ln(mc_{jt}(\cdot;\theta_2)) - w_{jt}\gamma. \tag{29}$$

Eventually, we have all the tools to describe our computation procedure:

(I) Use simulations and compute the market shares (26);

(II) Use the contraction mapping (27) and determine the mean utility relation $\delta = x\beta + \xi$;

(III) Call up the pricing equation [the Augmented Lerner Index (13)] and derive the marginal cost relation $\ln(mc) = w\gamma + \omega$;

(IV) Use a linear GMM to simultaneously estimate the utility surface parameters (β) and the marginal costs parameters (γ) conditional on the θ_2 subset of parameters [$\theta_2 = \{\alpha, \sigma\}$];

(V) Get the unobservables ξ and ω and interact them, once corrected for the correlation between demand and supply, with a function of the product characteristics H (our instruments) as to get the moments $G(\cdot)$ to be minimized in (21);[26]

(VI) Known G, one is only left with the estimation of the parameters $\{\alpha, \sigma\}$. This requires to originally set the weighting matrix W to $(H'H)^{-1}$, as requested in a 2SLS estimator and use the Nelder-Mead simplex method to minimize (21).[27]

VII) Get the optimal parameters ($\widehat{\theta}_2$) and, eventually, $\widehat{\theta}_1\left(\widehat{\theta}_2\right)$.

The full procedure is summarized as follows (iteration numbers are denoted by subscript squared brackets):

[26] Notice that the price parameters α that enters our mean utility (δ_j) is also entering the markup function. Since we need to estimate the markup as to explicit the unobservables in the supply side, we have to include α in the subset of parameters θ_2 that enter non linearly in our GMM function.

[27] See Lagarias et al. (1998) for a description of the Nelder-Mead procedure.

1) Begin with an initial non linear parameters value $\theta_{2[0]} = \left(\alpha_{[0]}, \sigma_{[0]}\right)$ and an initial mean utility vector value $\delta_{[0]}$ [with $\delta_{[0]} = \left(\delta_{1[0]}, ..., \delta_{J[0]}\right)$] then, compute the function of the market shares $\left[\phi\left(x, p, \delta_{[0]}, \nu; \theta_{2[0]}\right)\right]$ and, subsequently obtain, using simulations, the market shares $\left[s\left(x, p, \delta_{[0]}, P_{ns}; \theta_{2[0]}\right)\right]$. Use the obtained market shares (26) in the contraction mapping (27) and derive $T_{\left(P_{ns}, \theta_{[0]}\right)}\left[\delta_{[0]}\right] \equiv \delta_{[1]}$.

2) Repeat step 1) (where θ_2 is always fixed at the starting value $\theta_{2[0]}$) until the contraction mapping converges. Let's suppose the value of its convergence is $T_{\left(P_{ns}, \theta_{1[0]}\right)}\left[\delta_{[0]}\right] \equiv \widetilde{\delta}$ then, simultaneously estimate $\{\widehat{\beta, \gamma}\}$ (see steps II-IV above) and get the residuals ξ and ω;

3) Apply the Nelder-Mead fixed point minimization to (21) and output $\widehat{\theta_{2[1]}}$.

Repeat steps 1-3 above until the Nelder-Mead procedure converges.

4) Minimize the function (21) twice to better ensure a global minimum. Then repeat the minimization a third time but this time replace the weighting matrix W with its efficient value

$$W = \left[\frac{1}{J}\sum_{j=1}^{J}\left(H_j(\cdot)T(\cdot)\begin{pmatrix} \xi_j(\cdot) \\ \omega_j(\cdot) \end{pmatrix}\right)'\left(H_j(\cdot)T(\cdot)\begin{pmatrix} \xi_j(\cdot) \\ \omega_j(\cdot) \end{pmatrix}\right)\right]^{-1}$$

This concludes the description of our estimation procedure.[28]

The next section presents our estimates when either the "*coarse*" or the "*fine*" exogenous income distributions are used.

[28] A random coefficient model with only demand side is well explained in Nevo (2000) and Nevo (2001). Nevo also offers in his homepage a Matlab version of the algorithms to estimate the demand side. In our estimates we have extended those original files to our demand and supply version.

7 Results

Tables VII and VIII show, respectively, the results of our estimates in case of exogenous income distributions drawn from a sample of the entire population (Table VII, based on the "*coarse*" income distributions), the refined income distributions drawn from a sample of the population of buyers (Table VIII, based on the "*fine*" income distributions).[29] As previously outlined, we pool our data and capture part of the potential correlation among *same-models* (section 4) by introducing a variable (*modrep*) that reports the number of periods a model stays in the market and allow for different fixed effects such as years, segments, parent houses and the cross effects year-parent house.

In what follows we only describe the results of Table VIII omitting, on purpose, those similar of Table VII. We start with the demand side. The parameter β_2 associated to *modrep* tells us the longer a product stays in the market, the lower consumers perceive its quality (although its effect is not highly significative). The other demand parameters show that individuals tend to prefer long and fast cars and to be partly adverse to the characteristic trunk size. If no explanation deserve the signs of the β parameters associated to length and speed, we need to draw a line on that associated to alimentation (β_3). Although our variable fuel consumption is expressed in monetary value and controls for cheaper diesel price, individuals still prefer diesel to leaded gasoline cars. Seemingly our consumers are aware that one of the positive particularities

[29] Since we want to compute the pure distorsion ensuing the two different income distributions, we base our comparisons on the same set of instruments, the same starting values for the non-linear parameters and the same (non-income) simulations.

of the diesel engine is to be more resistant (which comes to the price of higher marginal costs, γ_2) than his alternative leaded gasoline. Other interesting results, although not reported in the Table, are that Italian parent houses and small car producers generate (on average) higher market shares; reflecting individual tastes for national products and smaller cars.[30] Regarding the supply side, all the γ parameters are of the expected (positive) sign. A positive sign means that the higher the value of a k characteristic, the larger is its effect on the marginal cost. Particularly, faster and longer cars are more costly to produce. Concerning the non linear parameters the high value of the parameter σ_2, associated to cubic capacity is telling us that individuals buying a low cubic capacity car prefer to replace that car with a similar low cubic capacity car; whereas individuals buying a high cubic capacity car prefer to replace it with another high cubic capacity car. Finally, a particular description deserve the $\{\alpha, \sigma_y\}$ parameters. These parameters affect (directly) our equilibrium price elasticities of substitutions: higher absolute values of α and σ_y, imply higher elasticities of substitutions. A higher value of σ_y amplifies the price income effect: the higher σ_y, the more price sensitive are lower income individuals. Differences in the estimated $\{\widehat{\alpha}, \widehat{\sigma}_y\}$ resulting from the use of "*fine*" or "*coarse*" income distributions are the main source of discrepancy in the price elasticities of substitution which magnitudes are emphasized in Table IX. We observe higher

[30] An important determinant of individuals' purchases is the after sales services. Apparently we don't control for this variables, expecting, then, it to fall into our unobservables. However, one of the main component of the after sales services is the prompt availability of spare parts. Our cross dummies parent house-time capture also this effect. Consumers are aware that national products' spare parts are readily accessible and for this reason they feel more confident in purchasing Italian products. This effect although partly reduced in the latest years, is still a relevant determinant of individuals' choice.

absolute price elasticities when we base our estimations on the "*coarse*" income distributions. Differences in the estimated price elasticities are non negligible.

In the next subsection we report our elasticities of substitutions when either, "*fine*" income distributions, or "*coarse*" income distributions, are used.

7.1 Price Elasticities

Price elasticities are a fundamental picture of the understanding of a market. Parent houses compete with each others and react differently to exogenous idiosyncratic shocks. Which is the percentage change of the market share of model j to an exogenous shock that affects the production (the price) of model r? How elastic are the demands in a differentiated product market? All these answers are englobed in a matrix of price elasticity of substitutions which computation, in our case, requires to calculate each period a matrix $J_t * J_t$ of values. The matrix of price elasticities is therefore filled with

$$\begin{aligned} \epsilon_{jjt} &= \frac{p_{jt}}{s_{jt}} \int \left(\phi_{jt}(\cdot) \left(1 - \phi_{jt}(\cdot)\right) \left[\frac{\partial \mu_{jt}(\cdot)}{\partial p_{jt}} \right] \right) P_{0t}(d\zeta) \\ \epsilon_{rjt} &= -\frac{p_{jt}}{s_{rt}} \int \left(\phi_{jt}(\cdot) \phi_{rt}(\cdot) \left[\frac{\partial \mu_{rt}(\cdot)}{\partial p_{jt}} \right] \right) P_{0t}(d\zeta). \end{aligned} \tag{30}$$

Tables X and XI provide the price elasticities of substitutions (and their respective interval of confidence) for some selected automobiles marketed in year 2000. Table X refers to the "*coarse*" income distribution and Table XI to the "*fine*" income distribution. Table XI shows that an increase of 1% in the price of a BMW Serie 3 reduces its market shares of 11.30%, while an increase of 1% in the price of a Mercedes Class C produces a 0.6% increase in the market share of Mercedes Class C and the same increase of 1% in the price of a smaller car, such as a Fiat Panda, produces a negligible effect (0.007%) on the market shares of the same Mercedes Class C. The intervals of confidence offer us a measure of the reliability of the estimated elasticities.[31] Tables X and XI also highlight a lower price elasticity in the small car markets, which implies a higher market power in these particular segments. Eventually Table XII decomposes the price elasticities in i) own-price elasticities; ii) sum of cross-price elasticities coming from the other models owned by the same firm (a measure of the intensity of cannibalization); iii) sum of cross-price elasticities coming from models owned by the other firms. Again, at the parent house level comes out that estimations based on "*coarse*" distributions tend to consistently overestimate absolute price elasticities.

8 Conclusion

Table VI shows the income distributions drawn from all individuals ("*coarse*" income distributions) and the income distributions drawn from the subsample of buyers ("*fine*" income distributions) to be statistically different. These statistical differences are not only confined to a mean effect but also to a concentration effect raising, consequently, a specification issue. Furthermore, given the income distributions (their simulations) enter (directly) the computation of the price elasticities of substitutions, allowing for the right income distribution ends to be determinant for proper computations.

We showed that i) the "*fine*" income distributions produce richer estimates

[31] We have computed the inteval of confidence using a Bootstrapping technique based on 1000 replications.

(better fits); ii) the use of "*coarse*" income distributions overestimate individuals' sensitivity to a price change. The lesson we draw from this paper is that models like BLP, which make use of "*coarse*" income distributions, produce an upper bound to the price elasticities. Our estimations for the Italian automobile market stress (Table IX) average overestimations above 50%. Differences that are quite substantial.

Acknowledgments

This paper was presented at the Harvard University IO group and University of Venice. We thank Agar Brugiavini, Piero Gottardi, Julie Mortimer, Ariel Pakes, Laura Romeu, Roberto Violi and seminar participants for helpful comments. A special thank to P. Paul Walsh for the extremely useful conversations.

References

[1] Anderson S., De Palma A. and Thisse F. (1992). "Discrete Choice Theory of Product Differentiation," Cambridge: MIT Press.

[2] Ben-Akiva M. (1973). "Structure of Passenger Travel Demand Models," Ph.D. Dissertation, Department of Civil Engineering, MIT.

[3] Berkovec J. (1985). "New Car Sales and Used Car Stocks: a Model of the Automobile Market," *RAND Journal of Economics*, 16, 195-214.

[4] Berkovec J. and Rust J. (1985). "A Nested Logit Model of Automobile Holdings for One Vehicle Households," *Transportation Research*, 19B, 275-285.

[5] Berry S. T. (1994). "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 242-262, Vol. 25.

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

[7] Berry S., Levinsohn J. and A. Pakes, (2004), "Estimating Differentiated Product Demand Systems from a Combination of Micro and Macro Data: the New Car Market", *Journal of Political Economy*, Vol.112, 68-105.

[8] Goldberg P. K. (1995). "Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry," *Econometrica*, 891-951, Vol. 63, Issue 4.

[9] Hansen L. (1982). "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

[10] Hausman J. and Taylor W. (1981) "Panel and Unobservable Individual Effects," *Econometrica*, 49, 1377-1398.

[11] Hausman J., Leonard G. and Zona J. D. (1994). "Competitive Analysis with Differentiated Products," *Annales D'Économie et de Statistique*, 34, 159-180.

[12] Hayashi F. (2000). "Econometrics," Princeton University Press.

[13] Lagarias J. C., Reeds J. A., Wright M. H. and Wright P.E. (1998). "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions," *SIAM Journal of Optimization*, 112-147, Vol. 9.

[14] Lancaster K. J. (1971). "Consumer Demand: a New Approach," New York: Columbia University Press.

[15] Lancaster K. J. (1991). "Modern Consumer Theory," ed. Edward Elgar. Aldershot.

[16] Manski C. F. (1977). "The Structure of Random Utility Models," *Theory and Decision*, 229-254, Vol. 8.

[17] Mariuzzo F., Walsh P. P. and Whelan C. (2003). "Firm Size and Market Power in Carbonated soft Drinks," *Review of Industrial Organization*, Special Issue November-December.

[18] McFadden D. (1978). "Modelling the Choice of Residential Location," in *Spatial Interaction Theory and Planning Models*, ed. by A. Karlvist, et al. Amsterdam: North-Holland, 75-96.

[19] McFadden D. (1981). "Econometric Models of Probabilistic Choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. Manski and D. McFadden. Cambridge: MIT Press.

[20] McFadden D. (1989). "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," in *Econometrica*, 995-1026, Vol. 57, Issue 5.

[21] Nevo A., (2000). "A Practitioners Guide to Estimation of Random Coefficients Logit Models of Demand", *Journal of Economics & Management Strategy*, 513-548, Vol. 9, Issue 4.

[22] Nevo A., (2001). "Measuring Market Power in the Ready-to-Eat Cereal Industry", *Econometrica*, 307-342, Vol. 69, Issue 2.

[23] Pakes A. and Pollard D. (1989). "Simulation and the Asymptotics of Optimization Estimators," in *Econometrica*, 1027-1057, Vol. 57, Issue 5.

[24] Petrin A. (2002). "Quantifying the Benefits of New Products: the Case of the Minivan," in *Journal of Political Economy*, 110, 705-729.

[25] Singh N., Vives X. (1984) "Price and Quantity Competition in a Differentiated Duopoly", *The Rand Journal of Economics*, 15 (4), 546-554.

1 The New Vehicles Dataset

As outlined in Section 3, our data don't let us to distinguish between used and new vehicles. To overcome this constraint we assume the market of used vehicle to clear each period: all vehicles sold by households at a price higher than $\underline{p}$ are, within the same year, purchased by other households. Vehicles sold by households at a price below $\underline{p}$ are assumed to be sold for scrap.[32]

We denote with superscripts "n" and "u" new and used vehicles' purchases whereas, "s" stands for vehicles' sales.

Let $I^u \subset I$ and $I^n \subset I$ be the subsets of individuals that, respectively, buy used and new vehicles.[33] As stated above, our data only provide information on the set of individuals who bought a vehicle $I^{u,n} \equiv (I^u \cup I^n) \subset I$ and sold a vehicle $I^s \subset I$. Since the order of individual $i \in I$ is completely random in the sample, we suggest the following recursive procedure as to separate the two subsets

$$\begin{aligned} I_r^u &= I_{r-1}^u \cup \left\{ \min_i \left\{ i : \min |p_i - p_h| \ \right\} \quad \right\} \text{ for } r = 1, ..., |I^s| \\ I_0^u &\equiv \{\emptyset\} ; \ i \in \left\{ I^{u,n} \backslash I_{r-1}^u \right\} ; h \equiv I^s [r] ; i \neq h. \end{aligned} \tag{31}$$

where $p_i \in P^{u,n}$ is the price paid by individual $i \in I^{u,n}$ whereas, $p_h \in P^s$ is the price received by individual $h \in I^s$. It follows

$$I^n = \left\{ I^{u,n} \backslash I_{|I^s|}^u \right\} \tag{32}$$

[32] We assume that value to be 3000 Euro in year 2000.
[33] Obviously for $i \in I^u$, $p_i^u \geq \underline{p}_i^u$.

We determine the shares of the outside alternative in our market to be

$$s_0 = \frac{|I| - |I^n|}{|I|} \tag{33}$$

Figure V plots the results of (33). Unfortunately, our results are not completely satisfactory. Our sample explains only 65-85% of the total sales in the period. This may be explained both by the fact that households can buy more than a car a year and by the number of new cars sold for commercial use. Alternatively, one could think of a sample underestimation.

To be coherent with the assumption that each household buys no more than a car a year, we recover our outside good market share s_0 from the following ratio

$$s_{0t} = 1 - \frac{\text{total sales of new auto in year } t}{\text{\# of households in the economy in year } t}.$$

which values are at their turn depicted in Figure V.

Figure I: Share of total unit sales by different area.

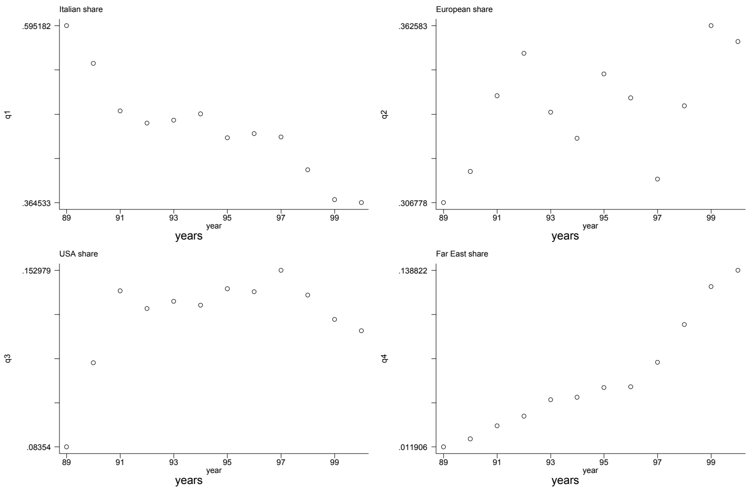

Figure II: Thousands of new car unit sales.

Figure III: Diesel and leaded gasoline prices per liter (in Euro).

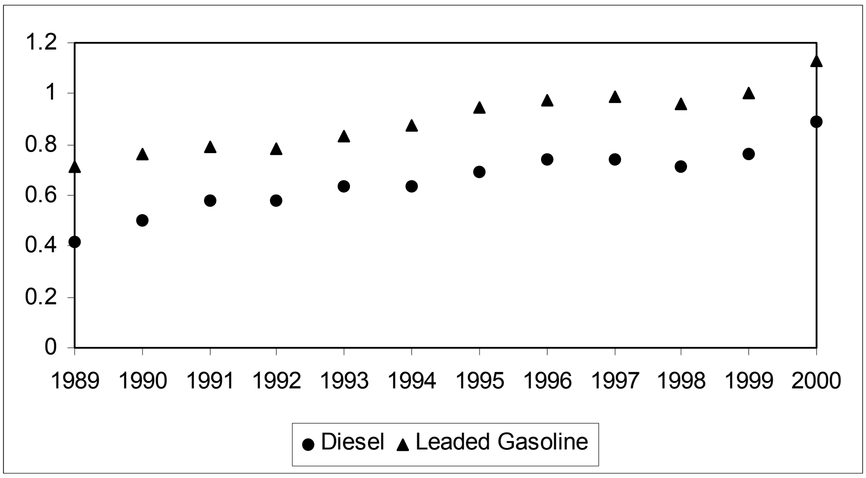

Figure IV: Real prices distribution (log thousands Euro: base year 2000).

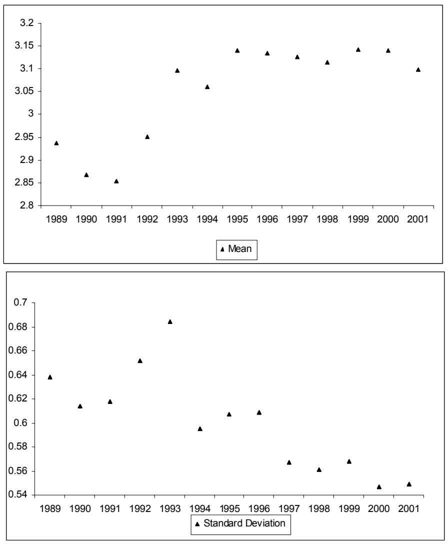

Figure V: Outside good market shares

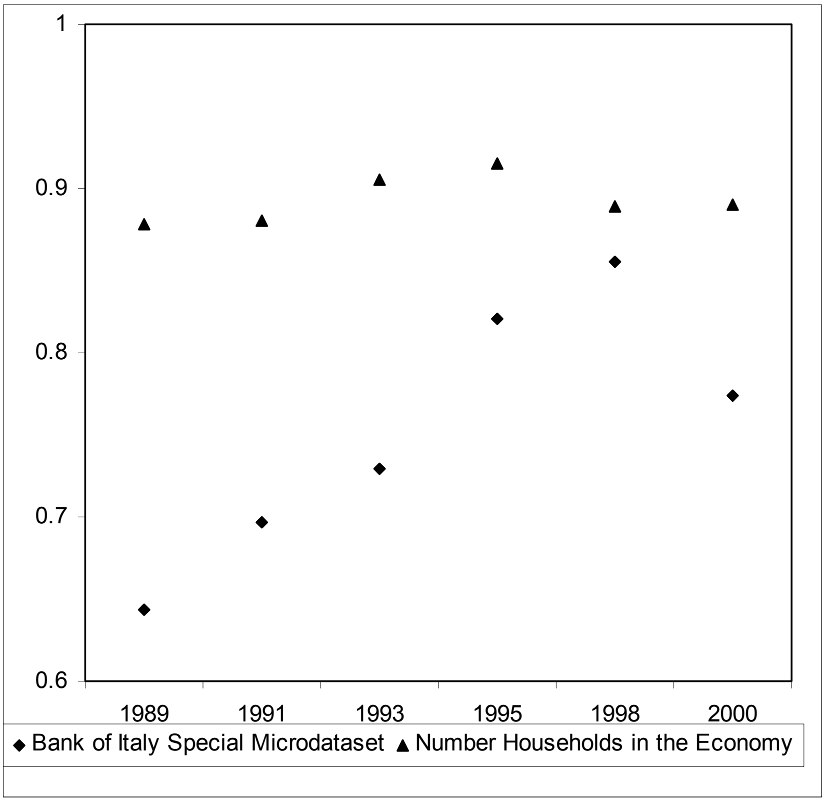

Table I: Concentration indexes.

Year	C4	HHI	GINI
89	0.69	2398	0.89
90	0.69	2058	0.89
91	0.69	1668	0.88
92	0.71	1696	0.88
93	0.67	1602	0.86
94	0.65	1628	0.84
95	0.66	1554	0.84
96	0.66	1631	0.86
97	0.65	1826	0.85
98	0.64	1522	0.84
99	0.65	1492	0.85
100	0.64	1431	0.84

Table II: Physical characteristics weighted by unit sales (means and standard deviations).

	cc	length	trunk size	maximum speed	fuel consumption
1989	1255.32	3.83	320.16	153.95	6.45
	409.51	*0.38*	*115.63*	*21.55*	*1.08*
1991	1325.54	3.89	328.95	160.70	6.76
	398.92	*0.40*	*123.83*	*24.88*	*1.18*
1993	1333.79	3.86	310.02	160.26	6.84
	398.29	*0.40*	*113.76*	*21.26*	*1.12*
1995	1408.93	3.93	320.01	166.54	7.01
	416.09	*0.39*	*120.14*	*21.25*	*1.13*
1998	1424.01	3.94	310.14	167.56	6.80
	441.26	*0.40*	*119.15*	*22.46*	*1.14*
2000	1487.49	3.93	307.29	168.52	6.42
	509.14	*0.44*	*113.86*	*22.38*	*1.32*

Table III: Number of households in the Bank of Italy SHIW

	1989	1991	1993	1995	1998	2000
1989	8274	2187	1050	827	544	404
1991		6001	2420	1752	1169	832
1993			4619	1066	583	399
1995				4490	373	245
1998					4478	1993
2000						4128
	8274	**8188**	**8089**	**8135**	**7147**	**8001**

Table IV: Bank of Italy special data section on vehicles' purchase.

	1989	1991	1993	1995	1998	2000
Purchase	764	852	688	689	867	856
Sale	123	169	130	111	134	108

Table V: Automobile market sample size.

	# Parent Houses	# Model/Years	# Different Model/Years
1989	33	127	127
1991	36	156	53
1993	33	171	51
1995	34	188	60
1998	34	237	98
2000	37	260	62
		1139	**451**

Table VI: Difference in distributions.

	Purchased new car			All sample			t-stat (mean)	F-stat (Variance)
Year	**mean**	**sd**	**N**	**mean**	**sd**	**N**		
1989	35.57	18.56	579	26.45	16.15	8192	19.56	1.47
1991	36.61	17.41	625	25.46	14.89	8108	19.56	1.39
1993	38.02	19.69	454	25.25	16.81	7963	17.53	1.42
1995	34.67	17.38	486	24.74	16.33	8055	18.38	1.75
1998	36.79	20.39	616	26.73	17.4	7029	15.93	1.46
2000	38.02	19.19	657	26.64	17.39	7881	18.33	1.45

Table VII: Estimations based on the "*coarse*" income distribution.[34]

						R2	**0.62**
						Chi2(17)	**33.35**
						GMM	**0.011**
						N	**1138**
						ns	**100**
	LINEAR DEMAND PARAMETERS				**LINEAR SUPPLY PARAMETERS**		
			t-stat				**t-stat**
β1	**constant**	-18.737	-4.688	**γ1**	**constant**	-1.715	-1.712
β2	**modrep**	-0.043	-1.703	**γ2**	**alimentation**	0.323	2.804
β3	**alimentation**	0.800	3.076	**γ3**	**cubic capacity**	0.015	1.066
β4	**cubic capacity**	-0.218	-1.302	**γ4**	**length**	0.352	4.136
β5	**length**	0.692	1.910	**γ5**	**speed**	1.158	4.040
β6	**speed**	4.431	4.722	**γ6**	**fuel consumption**	0.048	3.029
β7	**fuel consumption**	0.001	0.017	**γ7**	**trunk size**	-0.005	-1.019
β8	**trunk size**	-0.032	-1.015				
alpha	**Prices**	-0.144	-2.381				
	NON LINEAR PARAMETERS						
σ1	**constant**	0.647	0.324				
σ2	**cubic capacity**	0.209	7.293				
σ3	**fuel consumption**	0.033	4.383				
σy	**Price-Income**	1.585	1.973				
	Dummies for segments	Yes			**Dummies for segments**	Yes	
	Dummies for firms	Yes			**Dummies for firms**	Yes	
	Dummy for years	Yes			**Dummy for years**	Yes	
	Cross dummies firms-years	Yes			**Cross dummies firms-years**	Yes	

[34]The variables cubic capacity, trunk size and speed (see Table II) have been divided by 100.

Table VIII: Estimations based on the "*fine*" income distribution.[35]

					R2		0.69
					Chi2(17)		33.69
					GMM		0.01
					N		1138
					ns		100
	LINEAR DEMAND PARAMETERS				LINEAR SUPPLY PARAMETERS		
			t-stat				t-stat
β1	constant	-20.395	-4.621	γ1	constant	-1.465	-1.135
β2	modrep	-0.028	-1.077	γ2	alimentation	0.304	3.449
β3	alimentation	0.470	1.518	γ3	cubic capacity	0.016	1.310
β4	cubic capacity	0.014	0.067	γ4	length	0.322	2.200
β5	length	0.912	2.486	γ5	speed	1.071	3.295
β6	speed	4.391	4.531	γ6	fuel consumption	0.047	3.075
β7	fuel consumption	-0.027	-0.300	γ7	trunk size	-0.007	-0.666
β8	trunk size	-0.055	-1.523				
alpha	Prices	-0.181	-2.825				
	NON LINEAR PARAMETERS						
σ1	constant	1.270	0.470				
σ2	cubic capacity	0.120	6.280				
σ3	fuel consumption	0.008	3.934				
σy	Price-Income	1.177	2.005				
	Dummies for segments	Yes			Dummies for segments	Yes	
	Dummies for firms	Yes			Dummies for firms	Yes	
	Dummy for years	Yes			Dummy for years	Yes	
	Cross dummies firms-years	Yes			Cross dummies firms-years	Yes	

Table IX: Estimated differences own and sum cross-price elasticities (product level).

	Own-price elasticities			
	"coarse"		"fine"	
	mean	sd	mean	sd
89	-13.94	21.23	-8.50	15.86
91	-11.95	24.60	-8.98	21.13
93	-8.39	18.09	-9.83	16.77
95	-9.28	14.64	-9.58	19.33
98	-11.70	17.91	-8.14	12.25
100	-14.25	22.08	-9.80	16.10
	Sum Cross-price elasticities			
	"coarse"		"fine"	
	mean	sd	mean	sd
89	3.99	3.32	2.09	1.81
91	2.40	1.54	1.72	1.27
93	2.37	1.02	1.89	1
95	3.49	1.87	2.47	2.67
98	3.83	2.97	2.18	1.36
100	4.79	4.84	2.24	1.48

[35] The variables cubic capacity, trunk size and speed (see Table II) have been divided by 100.

Table X: A selected sample of the estimated own and cross-price elasticities (year 2000, "*coarse*" income distribution).

	Fiat Panda	Fiat Punto	Volkswagen Polo	Renault Clio	BMW Serie 3	Mercedes Class C	Audi A4
	-18.132	*0.000*	*0.000*	*0.000*	*0.002*	*0.003*	*0.002*
Fiat Panda	**-1.642**	**0.026**	**0.024**	**0.025**	**0.055**	**0.068**	**0.043**
	-0.082	*0.397*	*0.501*	*0.434*	*0.258*	*0.330*	*0.246*
	0.000	*-12.906*	*0.001*	*0.001*	*0.011*	*0.015*	*0.008*
Fiat Punto	**0.028**	**-3.556**	**0.054**	**0.062**	**0.172**	**0.205**	**0.123**
	0.438	*-0.471*	*0.340*	*0.332*	*0.309*	*0.360*	*0.228*
	0.000	*0.000*	*-14.274*	*0.000*	*0.002*	*0.002*	*0.001*
Volkswagen Polo	**0.005**	**0.009**	**-3.309**	**0.009**	**0.023**	**0.027**	**0.017**
	0.095	*0.057*	*-0.350*	*0.060*	*0.041*	*0.047*	*0.033*
	0.000	*0.000*	*0.000*	*-13.289*	*0.002*	*0.003*	*0.002*
Renault Clio	**0.006**	**0.013**	**0.011**	**-3.376**	**0.036**	**0.043**	**0.025**
	0.099	*0.072*	*0.075*	*-0.340*	*0.069*	*0.083*	*0.050*
	0.001	*0.004*	*0.003*	*0.003*	*-35.060*	*0.016*	*0.014*
BMW Serie 3	**0.014**	**0.042**	**0.033**	**0.040**	**-19.906**	**0.170**	**0.098**
	0.059	*0.083*	*0.066*	*0.081*	*-4.823*	*0.446*	*0.212*
	0.000	*0.002*	*0.002*	*0.002*	*0.006*	*-47.947*	*0.006*
Mercedes Class C	**0.006**	**0.018**	**0.014**	**0.018**	**0.069**	**-26.649**	**0.045**
	0.028	*0.037*	*0.029*	*0.037*	*0.182*	*-6.801*	*0.107*
	0.000	*0.000*	*0.000*	*0.000*	*0.002*	*0.002*	*-23.650*
Audi A4	**0.002**	**0.005**	**0.004**	**0.004**	**0.013**	**0.015**	**-13.770**
	0.009	*0.010*	*0.009*	*0.010*	*0.029*	*0.035*	*-3.194*

Table XI: A selected sample of the estimated own and cross-price elasticities (year 2000, "*fine*" income distribution).

	Fiat Panda	Fiat Punto	Volkswagen Polo	Renault Clio	BMW Serie 3	Mercedes Class C	Audi A4
	-14.901	*0.000*	*0.000*	*0.000*	*0.001*	*0.002*	*0.001*
Fiat Panda	**-1.690**	**0.020**	**0.020**	**0.020**	**0.028**	**0.031**	**0.027**
	-0.170	*0.430*	*0.487*	*0.454*	*0.290*	*0.330*	*0.292*
	0.001	*-16.636*	*0.001*	*0.001*	*0.005*	*0.005*	*0.003*
Fiat Punto	**0.025**	**-3.324**	**0.035**	**0.037**	**0.064**	**0.067**	**0.056**
	0.489	*-0.467*	*0.392*	*0.388*	*0.299*	*0.329*	*0.291*
	0.000	*0.000*	*-18.072*	*0.000*	*0.001*	*0.001*	*0.000*
Volkswagen Polo	**0.004**	**0.006**	**-3.183**	**0.006**	**0.009**	**0.010**	**0.008**
	0.092	*0.065*	*-0.422*	*0.067*	*0.049*	*0.056*	*0.049*
	0.000	*0.000*	*0.000*	*-16.722*	*0.001*	*0.001*	*0.001*
Renault Clio	**0.005**	**0.008**	**0.007**	**-3.157**	**0.013**	**0.013**	**0.011**
	0.107	*0.081*	*0.084*	*-0.431*	*0.063*	*0.069*	*0.060*
	0.001	*0.002*	*0.002*	*0.002*	*-29.192*	*0.008*	*0.005*
BMW Serie 3	**0.015**	**0.027**	**0.024**	**0.026**	**-11.297**	**0.057**	**0.046**
	0.120	*0.124*	*0.124*	*0.123*	*-3.306*	*0.234*	*0.190*
	0.000	*0.001*	*0.001*	*0.001*	*0.003*	*-35.566*	*0.002*
Mercedes Class C	**0.007**	**0.011**	**0.010**	**0.011**	**0.023**	**-13.469**	**0.019**
	0.057	*0.057*	*0.059*	*0.057*	*0.098*	*-4.275*	*0.090*
	0.000	*0.000*	*0.000*	*0.000*	*0.001*	*0.001*	*-25.967*
Audi A4	**0.002**	**0.003**	**0.003**	**0.003**	**0.006**	**0.007**	**-9.209**
	0.017	*0.016*	*0.016*	*0.016*	*0.025*	*0.028*	*-2.380*

Table XII: Estimated own and cross-price elasticities (year 2000).

Parent House	Coarse distribution: Own Price Elasticities	Coarse distribution: Sum Cross Price elasticities other models owned by the same firm	Coarse distribution: Sum Cross Price elasticities models owned by other firms	Fine distribution: Own Price Elasticities	Fine distribution: Sum Cross Price elasticities other models owned by the same firm	Fine distribution: Sum Cross Price elasticities models owned by other firms
1	-10.90	0.20	3.91	-7.28	0.12	1.97
2	-42.28	0.61	9.40	-27.86	0.16	3.60
5	-51.13	1.27	12.60	-32.23	0.27	4.43
6	-18.17	0.02	7.00	-10.33	0.01	2.76
7	-5.45	0.07	2.72	-4.57	0.05	1.68
8	-4.10	0.02	2.27	-3.64	0.02	1.50
9	-3.67	0.00	1.47	-3.64	0.00	1.27
11	-4.44	0.37	1.90	-3.93	0.26	1.28
12	-5.78	0.18	2.80	-4.66	0.11	1.65
13	-5.34	0.02	2.47	-4.64	0.01	1.64
14	-4.49	0.02	2.59	-3.86	0.02	1.60
16	-22.78	.	10.27	-12.32	.	3.59
17	-77.17	0.25	20.08	-56.11	0.05	7.24
18	-44.81	0.07	14.09	-32.27	0.02	5.36
19	-5.55	0.02	3.25	-4.34	0.01	1.82
21	-8.72	0.13	3.36	-6.30	0.08	1.83
22	-32.66	0.10	9.45	-21.45	0.04	3.63
23	-16.60	.	5.63	-10.32	.	2.54
26	-4.94	0.00	2.70	-4.37	0.00	1.72
27	-2.23	.	0.75	-2.32	.	0.89
28	-48.82	1.62	10.77	-27.78	0.29	3.71
30	-19.11	0.09	7.14	-10.64	0.03	2.69
31	-8.98	0.08	4.14	-6.01	0.04	2.00
32	-4.89	0.05	2.28	-3.87	0.04	1.43
33	-5.94	0.10	2.88	-4.73	0.06	1.69
35	-111.57	0.25	20.03	-91.36	0.06	7.81
36	-7.01	0.17	2.96	-5.22	0.11	1.67
37	-5.66	0.02	2.60	-4.69	0.02	1.65
38	-26.38	0.02	7.95	-13.37	0.01	2.84
39	-4.57	0.03	2.47	-3.96	0.02	1.57
40	-3.72	0.02	2.27	-3.50	0.01	1.55
41	-10.66	.	5.97	-7.25	.	2.73
42	-13.89	0.01	5.77	-8.99	0.00	2.63
43	-3.61	0.01	2.02	-3.37	0.01	1.42
44	-6.50	.	3.34	-5.48	.	2.05
45	-8.15	0.08	3.59	-5.82	0.05	1.87
47	-6.67	0.25	2.66	-5.43	0.17	1.63
48	-21.04	0.10	7.25	-11.30	0.04	2.69