September 2002

Automobile Equilibrium Prices: an Empirical Study on the Italian Market

Franco Mariuzzo

Working paper 2002.06
Automobile Equilibrium Prices: an Empirical Study on the Italian Market

Franco Mariuzzo∗
Dipartimento di Scienze Economiche
Università degli Studi di Venezia
Cannaregio, 873
30121 Venezia, Italia†

September 11, 2002

Abstract

This paper provides an empirical analysis of own- and cross-price elasticities of substitution for the 1989-2000 period in the Italian automobile industry. To this end, we make use of product-level and aggregate consumer-level data consistent with a structural model of equilibrium in an oligopolistic industry. In order to get reliable elasticities of substitution we follow Berry Levinsohn and Pakes (1995) (BLP) and nest individual variability into a Cobb Douglas utility function. As a drawback, this does not deliver the nice closed form solutions of the logit demand functions and forces us to adopt econometrics procedures up to derive, by simulations and contractions, the equations that will enter our estimation methods. Furthermore, because of the endogeneity of prices, we estimate simultaneously demand and supply by using proper instruments in a GMM method.

Thanks to the availability in our data of a special section with information on individuals who bought a vehicle, we extend BLP by enriching our estimations with further moments relating individual prices and characteristics. This extension does not confine only to provide more efficient demand estimates but, it also adds relations between individual characteristics and price equilibrium.

Keywords: Differentiated Products, Discrete Choice, Automobile Industry, Structural Models.

JEL classifiers: JL11, JL62.

∗Corresponding author: tel. +39 041 2349105; fax +39 041 2349176; e-mail mariuzzo@unive.it
†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 of course my own.
1 Introduction

We use aggregate data and estimate own- and cross-price elasticities of substitution for the 1989-2000 Italian automobile industry. As to get reasonable values, we follow the random coefficient model in Berry Levinsohn and Pakes (1995) (BLP onwards) and provide an empirical analysis of the parameters of marginal cost, and demand.

Aware of the critics of poor demand estimation to the aggregate industry data [Berry Levinsohn and Pakes (2001) and Petrin (2001)], we extend BLP both by introducing in the utility function some specific dummies (such as sports car, vans and offroads) of which we can conjecture the distribution of individuals who are interested in, and by enriching our estimates with information from a special data section where households buying new vehicles (each year about 800, out of a sample of 8000 of which we know their characteristics) are asked the price they, respectively, paid for their purchases. This latter extension does not only confine to get better demand estimates but, it also adds direct relations between households characteristics and price equilibria. Computationally, we are only required to add further moments to our GMM estimations.

We can distinguish two main approaches to the empirical automobile literature; a first one that makes use of 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 an 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 a drawback, pays the price of neglecting the supply side and, therefore, all the subsequent equilibrium considerations. An answer to this critics is promptly offered in Goldberg (1995). Goldberg assumes the existence of a Nash equilibrium and, by a nested model, she provides an equilibrium analysis of demand and supply. Another attempt to address equilibrium analysis by using disaggregate data is proposed in Berry Levinsohn and Pakes (2001). On the other side, the stream of aggregate industry literature mainly addresses directly demand and supply and, under the assumption of the existence of a Nash equilibrium, estimates elasticities of substitutions (BLP) or,
going further, quantifies the benefit of product innovation Petrin (2001).

As well known in the homogeneous good literature, we face, also for the differentiated products market, the problem of correlation between prices and error terms (unobservable variables in our model) in demand and supply. The effects of ignoring this correlation, are that one obtains, other than inconsistent parameter estimates, counterintuitive results such as upward sloping demand curves. One solution to this endogeneity problem is, for the differentiated product market, suggested in Berry (1994). In our estimates we correct for this endogeneity by using proper instruments in a General Method of Moments (GMM) estimator Hansen (1982).

Among the empirical aggregate industry literature, BLP estimate demand and supply in the U.S. differentiated automobile markets and suggest, fine econometric tools to get more reliable own- and cross- price elasticities. They provide results using a GMM estimator and suggest simulators for determining market shares [Mac Fadden (1989) for details]. In order to get more efficient estimates, they enrich their product level data with exogenous information on consumers’ income characteristics. To be more specific, on the supply side, the authors estimate costs as a function of product characteristics while, on the demand side, they allow for heterogeneity in consumer tastes and estimate own- and cross-price elasticities by aggregating a discrete choice model of individual consumer behavior into market-level. BLP refer to Berry (1994) for solving the problem of correlation between prices and unobservable variables. In his paper Berry suggests a method for estimating a demand function when it is nonlinear in the unobserved demand component. His methodology mainly consists of transforming the observed quantity, price and characteristic data into a linear function of the unobserved demand components.

Berry Levinsohn and Pakes (2001) extends BLP by adding, this time, microdata enriched with consumers’ second choice information. The authors find that i) unobservable consumer attributes (our see ) are relevant to obtaining reliable substitution patterns; ii) the availability of information on consumers’ second choice is necessary to get better estimates.

Still in the path of providing more precise parameters’ estimates, Petrin

---

1 A main drawback in disregarding this simultaneity problem is we get unreasonably small (in absolute value) estimations of price elasticities.
(2001) suggests to augment the market level data with readily available data that relate the average characteristics of consumers to the characteristics of the products they purchase. This more precise estimation results are then used by the author to evaluate the welfare benefits of the minivans innovation in the U.S. auto market.

On the side of disaggregate consumer data equilibrium Goldberg (1995) suggests an equilibrium analysis. The author addresses the field of international competition and develops the demand side as a discrete choice model where, the process of buying an automobile is considered as a nested sequence of logit models with the choice between a car and the outside good at the highest level, and the choices of market segment, origin and model at subsequent nodes. The supply side offers estimations of parameter costs. Estimations of demand and supply are obtained by a pseudo-panel of household data that match information on the cars they buy.

Our dataset consists of three dimensions:

i) **Individuals** are households drawn from the Bank of Italy Survey of Households’ Income and Wealth (SHIW).

ii) **Products** include information on sales, list prices and physical characteristics for all new auto models sold in the Italian market.


See section 3 for data details.

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 the purchase of a new car. The static nature of the considered models (due mostly to a lack of data) fails other than to take intertemporal substitution effects into account, also into consider strategically firms entry and exit.

The paper is organized as follows. The next section summarizes the evolution of the Italian automobile industry and introduces the market. 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 computation mechanisms. In Section 7 we extend BLP by adding information on individuals who bought a vehicle (section to be written ...).
Section 8 we provide our estimation results. Finally, the paper concludes in Section 9. Appendix A describes a method for deriving the subset of individuals buying a new vehicle. Appendix B provides details on the Variance-Covariance matrices. Appendix C summarizes the implemented files.

2 The Automobile Italian Industry

Figure 1 shows the evolution of the Italian new car market for the period 1989-2000. We split the market in four different areas and ascribe a specific area to each parent house by virtue of her production country. We are therefore not considering the labyrinth of all possible subsequent cross mergers, acquisitions or other market transactions. From the graph one can notice how the increasing opening of the Italian market to foreign competition behaved, in the period, a large reduction (nearly 40%) of the Italian production for the Italian market. This is partly explained by the progressive elimination of import duties as requested by the European Union and, most, by the increasing effects of the globalization.

Figure 2 provides, for the Italian market, the total sales of new cars in the period 1989-2000. We observe a sharp collapse of 20% in the sales in year 1992. That strong reduction is mainly explained by the exit of Italian Lira from
the European Monetary System (October 1992) which behaved a subsequent strong devaluation of the Italian currency. This explanation is strengthened by a strong reduction in the Italian market quotas (see Figure 1). The 1992 crisis was followed by a stagnation up to 1996 when the market fully recovered. If one compares figure 1 with figure 2, he can notice how the recovery is followed by some degree of substitution between Italian and foreign production. Although the effect is not clear from the graph, we should be aware that 1998 is the year of the “scrap-incentives”.

Figure 2: thousands of new cars sold in Italy

Table 1 shows the 1989-2000 trend for some major physical automobile characteristics. The table explains the evolution of the automobile Italian market. During the period we can observe how, an increase in cubic capacity and speed simultaneously go with a reduction in fuel consumption. Individuals seem to have modified their tastes versus faster and more fuel consuming cars. Finally, although the table doesn’t report information on airbag and ABS as standard, safety, proxied by the variable length and, somehow, trunk size. Another important information from the table is an increase in the variability of all characteristics but trunk size.
Table 1: Physical characteristics trend - means and standard deviations -

<table>
<thead>
<tr>
<th></th>
<th>cm³</th>
<th>length</th>
<th>trunk size</th>
<th>maximum speed</th>
<th>fuel consumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989</td>
<td>1255.323</td>
<td>3.827139</td>
<td>320.1572</td>
<td>153.945</td>
<td>6.453705</td>
</tr>
<tr>
<td></td>
<td>499.513</td>
<td>0.3836078</td>
<td>115.6272</td>
<td>21.54529</td>
<td>1.084095</td>
</tr>
<tr>
<td>1991</td>
<td>1325.539</td>
<td>3.894707</td>
<td>328.9476</td>
<td>160.7006</td>
<td>6.758563</td>
</tr>
<tr>
<td></td>
<td>398.9157</td>
<td>0.3972142</td>
<td>123.8283</td>
<td>24.87565</td>
<td>1.177617</td>
</tr>
<tr>
<td>1993</td>
<td>1333.786</td>
<td>3.862277</td>
<td>310.0216</td>
<td>160.2646</td>
<td>6.841448</td>
</tr>
<tr>
<td></td>
<td>398.2914</td>
<td>0.3980062</td>
<td>113.7599</td>
<td>21.25085</td>
<td>1.117779</td>
</tr>
<tr>
<td>1995</td>
<td>1408.926</td>
<td>3.934884</td>
<td>320.0149</td>
<td>166.5393</td>
<td>7.008699</td>
</tr>
<tr>
<td></td>
<td>416.0923</td>
<td>0.3881606</td>
<td>120.1359</td>
<td>21.25085</td>
<td>1.127497</td>
</tr>
<tr>
<td></td>
<td>441.2643</td>
<td>0.4046868</td>
<td>119.1506</td>
<td>22.4554</td>
<td>1.137888</td>
</tr>
<tr>
<td></td>
<td>509.1384</td>
<td>0.4374458</td>
<td>113.8058</td>
<td>22.38022</td>
<td>1.322383</td>
</tr>
</tbody>
</table>

Figure 3 offers a picture of the price-distribution trend. Its peculiarity is that we can clearly split the time period into two subperiods:

i) 1989-1993: a period where mean price movements are followed by distribution movements of the same sign;

ii) 1993-2000: a period of mean price stabilization associated with a process of convergence among different price models.

Figure 3: price-distribution (logarithm of thousands of Euro: base year 2000)
3 The Data

Unfortunately, due to a lack of Italian data matching individual characteristics to the product they purchase, we make use for the period 1989-2000, of product-level and aggregate consumer-level data.

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

A) Individuals are all households drawn from Bank of Italy Surveys on Households’ Income and Wealth (SHIW). Other than data on households characteristics such as disposable income, family components, area, age etc., the dataset holds a special section on vehicles’ purchase. Households are asked whether they bought/sold a vehicle in the year and, in case, the price they, respectively, paid/received for each transaction. We use these information both to derive our $s_0$ (quota of households preferring the outside alternative) and to obtain better demand estimates. Unfortunately, our data don’t let us to distinguish between used and new vehicles. We propose in Appendix A a method of minimum distances as to get the subset of households buying a new vehicle. As we can see from table 2, the sample size is about 8000 households a years (with a panel component).

Table 2. Households in the Bank of Italy SHIW dataset

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1987</td>
<td>8027</td>
<td>1206</td>
<td>350</td>
<td>173</td>
<td>126</td>
<td>85</td>
<td>61</td>
</tr>
<tr>
<td>1989</td>
<td>7068</td>
<td>1837</td>
<td>877</td>
<td>701</td>
<td>459</td>
<td>343</td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td>6001</td>
<td>2420</td>
<td>1752</td>
<td>1169</td>
<td>832</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td>4619</td>
<td>1066</td>
<td>583</td>
<td>399</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>4490</td>
<td>373</td>
<td>245</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>4478</td>
<td>1993</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>4128</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8027</td>
<td>8274</td>
<td>8188</td>
<td>8089</td>
<td>8135</td>
<td>7147</td>
<td>8001</td>
</tr>
</tbody>
</table>

B) Products include information on sales, list prices and physical characteristics such as, engine attributes (kilowatt, cubic capacity), dimensions (length), comfortability (number of doors, trunk size) and performance variables (fuel consumption, acceleration time, maximum speed). All these information are provided us with three different databases (two furnished by Editoriale Domus-Quattroruote and one by Fiat). To be more specific:

A special thanks to Andrea Battiston for having patiently added to the Fiat dataset the
i) A former Quattroruote database offers information on prices. We have, for the period 1989-2001, precisely 65715 quarterly prices. We, then, reduce the dataset to yearly prices (simply by averaging the quarterly prices).

ii) A latter Quattroruote database furnishes information on all the auto characteristics introduced above. The original dataset is of 16111 observations of which only 11125 are not the same (repeated) name plates. From 1996 on, the variable fuel consumption is marked by three different EU standards: 1) urban; 2) suburban; 3) mixed. Before that period the distinction was among: a) 90 km/h b) 120 km/h c) urban. From the average of a), b), c) it is possible to get a good approximation of the mixed fuel consumption in 3).

iii) Finally, the Fiat database (11246 different same models) offers information on market segments (28 exogenous different segments), quantities sold in each year, body, type of engine and, as in the Quattroruote database, characteristic variables such as kilowatt, cubic capacity and number of doors. We, then add to the original dataset a variable called area with the aim of describing each parent houses’ production location.

By merging ii) and iii) we obtain for the 1989-2000 period a database of 11055 name plates and, once we consider the name plates reported in different year to be different observations we have a cross-section of 46533 observations.

In order to provide parameter estimations, we restrict our database to model/years and different model/years (see next section for their respective definitions) as shown in the following table

Table 3. Model/years and parent houses

Editoriale Domus-Quattroruote code (Infocar-anno-mese) necessary to merge the two different databases.

3 To be more precise:
   i) the original dataset consists of 19997 records but only 16111 of them have information on prices;
   ii) it could either be the case that name plates that we consider repeated, differ from each other for some minor characteristics not reported in our data. However, we are not particularly worried about this details, for what we care more is that name plates in Editoriale Domus correspond to those in the Fiat database.

4 The Model

The number of competing parent houses is small enough to let us consider an underlying oligopolistic market structure with highly differentiated automobile models. The underlying game is, therefore, a differentiated product game with prices as strategic variables. We model firms as price-setting oligopolists and consumers as price takers. We follow the mainstream literature on aggregate data and assume the existence of Nash equilibrium in prices.

The following figure describes our consumer’s decisions. Each households (our consumer agents) decide first on whether to buy a new automobile. We denote as “Not Buy New” the alternative choice of either not buying an auto at all or, of buying a used car. Once the choice “Buy New” is taken, the agent further decides on the car that gives her the highest utility.
Figure 4: consumer choices.

On the other side of the market, firms simply decide on the prices that maximize their profits in a differentiated product market. We omit the game tree, for it is quite intuitive.

Let $F_t$ be the total number of firms in our market at time $t$ and be $J_t$ the total number of different models produced. Each firm $f_t = 1, \ldots, F_t$ produces a $J_{f_t} \subset J_t$ subset of models where

$$
\sum_{f_t=1}^{F_t} J_{f_t} = J_t
$$

For any $j_t \in J_{f_t} \subset J_t$ model and for any $f_t \in F_t$ firm, we have a $z \in \mathbb{R}^{|z|}$ vector of observed characteristics.\(^4\) We provide in Appendix A a definition of $j_t$ (model/years). Finally, $t = 1, \ldots, T$ is our time index.

In order to save some notation, we drop onwards subscripts $f$ and $t$.

We denote by $(x_j, \xi_j)$ and $(w_j, \omega_j)$ respectively, the observed, unobserved demand and the observed, unobserved marginal cost for product $j$.

We assume utility of consumers to depend on product characteristics, prices and individual taste parameters. By aggregating a discrete-choice model of consumer behavior we derive the market demand.\(^5\) The supply side is instead obtained by estimating the parameters on the cost function. Equilibrium price is a $J - \text{upta}$ determined by the first order conditions of the profit maximizing firms.

\(^4\)Where $|z|$ is the total number of observed characteristics.

\(^5\)For more details on discrete choice models see McFadden (1981).
We define now what is a model $j$ in our framework. We offer two alternative definitions. A former definition is

**Definition 1** i) (model/years): is the 2–upla $(\text{Parent House, Model})$ $j \in \tilde{J} \supset J$ with vector of observed characteristics the one associated with its most sold Name Plate in the year.

and an alternative definition

**Definition 2** ii) (model/years): is the 2–upla $(\text{Parent House, Model})$ $j \in \tilde{J} \supset J$ with vector of observed characteristics the one obtained by a yearly weighted (by sales) mean of the observed characteristics of its different Name Plates.

In our estimates we make use of definition ii). Furthermore, as our dataset is a rotating panel, we define when a model is the same over time. It follows:

**Definition 3** (same model/years): a model/years is, over time, the same if it has both the same 2–upla $(\text{Parent House, Model})$ and, none of its characteristics has changed more than $\pm 20\%$. It follows what is a different model/years.

From definition 3 let $\tilde{J}$ be the set of model/years and $J$ the set of different model/years.

In the following two subsections we follow closely the BLP notation and describe demand, supply and market equilibrium .

### 4.1 The Demand Side

We derive our demand by aggregating a discrete choice model of individual consumer behavior. We are aware that, when individuals choose among different models of cars, belonging to different classes, they do not restrict their decisions only on prices but, they also consider different product characteristics. This approach, suggested by Lancaster (1971, 1991), other than being quite immediate when one has to do with product differentiation, offers the possibility of moving from the product space to the characteristic space, which is quite useful when one has to do with many products and few characteristics (as in our case). By this approach we better explain why products may be physically similar but differ in consumers’ perception about quality, durability, status, or
Unfortunately, some characteristics such as style, reputation and past experience are unobservable to us but, they are rather frequent determinants of demand. For this reason, we consider their effects in our model.

We represent the utility derived by consumer \( i \in I \) from consuming product \( j \) to be \( U(\zeta_i, p_j, x_j, \xi_j; \theta) \). Where \( I \) is the number of individuals in the economy, \( \zeta_i \) is a vector of individual characteristics whereas \((p, x, \xi)\) are vectors and matrices of product characteristics. In this case \( p \) represents the products price vector and \( x \) and \( \xi \) are, respectively, observed and unobserved matrices of product attributes. Finally, \( \theta \) includes any parameters that determinate the distribution of consumer characteristics \( \{\alpha, \sigma\} \), as well as, conditional on these characteristics, the utility parameters that describe the utility surface \( \beta \) and the marginal costs \( \gamma \).

We avoid, onwards, representing vectors and matrices in bold letters.

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

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

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 know that consumers with different characteristics make different choices so, we define

\[
A_j = \{\zeta_i : U(\zeta_i, p_j, x_j, \xi_j; \theta) \geq U(\zeta_i, p_r, x_r, \xi_r; \theta), \text{ for } r = 0, 1, ..., J \}
\]

to be the set of values for \( \zeta_i \) that induces the choices of good \( j \in J \). Assuming ties occur with zero probability (which means the distribution function \( P \) of \( \zeta \)

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

\[\text{Manski (1977) argues that randomness in observed consumer behavior is mainly due to unobservable characteristics influencing consumer choice.}\]
is absolute continuous with respect to the Lebesgue measure), we obtain, by aggregation, the market share for good $j$

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

where $P_0 (d\zeta)$ is nothing but the density of $\zeta_i$ in the population and $0 < s_j < 1$ with $\sum_{j=0}^{J} s_j (\cdot) = 1$. In section 5.1 (infra) we describe in detail the $s_j$ computation. Finally, assuming $I$ to be the number of consumers in the market we derive the aggregated demand functions

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

which, as observed in Berry (1994), are nonlinear functions of $\xi$.

We now need to assume some shape to our utility function. As known in the literature (Anderson, De Palma and Thisse (1992)) a utility additively separable in a product characteristics and a consumer characteristics component provides poor substitutions effects. That is, conditional on market shares, elasticities of substitutions do not depend, in that case, on the observable characteristics of the product. We are in other words saying that, if, for example, Fiat 500 and Ferrari have close market quotas (the $s_j$); a change in the price of a Porsche will affect both the models in the same way. Which is obviously unbelievable. We do 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 and satisfy the just cited reasons, 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

$$U(\zeta_i, p_j, x_j, \xi_j; \theta) = (f (y_i) - p_j)^{\alpha} G (x_j, \xi_j, \nu_i) \exp (\epsilon_{ij})$$

where $f (y_i)$ is some function of the individual income that we reveal, further,
in subsection 5.1.8. Finally, $\epsilon_{ij}$ (assumed to be i.i.d. across products and consumers) includes the unobserved individual and product characteristics. We assume $G(\cdot)$ to be linear in logs, then

$$u_{ij} = \alpha \ln (f(y_i) - p_j) + \sum_{k=1}^{[z]} \beta_k x_{jk} + \xi_j + \sum_{k=1}^{[z]} \sigma_k x_{jk} \nu_{ik} + \epsilon_{ij} \tag{4a}$$

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

$$u_i0 = \alpha \ln (f(y_i)) + \xi_0 + \sigma_0 \nu_{i0} + \epsilon_{i0} \tag{4b}$$

where $(\nu_{i1}, ..., \nu_{[z]})$ is a vector of idiosyncratic consumer tastes that interact with product characteristics (alias marginal utility of characteristics); $x_{j1}$ is a dummy vector and $\sigma_k$ the standard deviations of the marginal utility distributions. This representation assumes individuals to have different preferences for each different observable characteristics. That means, the effect of $x_k$ units of characteristic $k$ on the marginal utility to be $(\beta_k + \sigma_k \nu_{ik})$. One may, from the distribution of tastes for characteristics $k$, notice how, higher values of $\beta_k$ (the mean) or $\sigma_k$ (the standard deviation) might explain an increase in the share of consumers buying cars with higher $k$ characteristic values. Moreover, the value of $\sigma_k$ is relevant in explaining the substitution effects. Let’s suppose an increase in the price of a high $k$ characteristic car. 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 $\sigma_k$), not tend to substitute towards other high $k$ cars; ii) whereas, in case of high $\sigma_k$, the opposite is true (similar products become better substitutes). This effect is simply explained by the marginal utility for the $k$ characteristic. 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, $\beta_k$ and $\sigma_k^2$.

It is important to notice how utility in (4a) can be decomposed into a com-

---

8 We suggest a function of income (rather than income directly) to enter our utility function. In this way, we neither need to restrict each models to be afforded by each consumers ($0 \leq p_j \leq y_i, \forall i \in I, j \in J$), nor, to restrict individual choices to depend on their yearly incomes: whenever ($y_i < p_j, \forall j \in J_1 \subset J, i \in I$) individuals could only choose among $j \in (J \setminus J_1)$ models.

9 As observed by BLP, the use of exogenous data on the income distribution, leads to more precise parameter estimates.

10 The $\nu_{i0}$ accounts for a possible higher unobserved variance in the idiosyncratic component of the outside alternative.
mon (to all consumers) mean component

\[ \delta_j \equiv \sum_{k=1}^{|z|} \beta_k x_{jk} + \xi_j \]  

(4a1)

and a deviation from that mean

\[ \mu_{ij} = \alpha \ln (f(y_i) - p_j) + \sum_{k=1}^{|z|} \sigma_k x_{jk} \nu_{ik} \]  

(4a2)

where \( \mu_{ij} \) (the heterogeneity in consumer tastes) depends on the interaction between consumer preferences (\( \nu_i \)) - included the exogenous income distribution information - and product characteristics (\( x_j \)). A decomposition that ends to be quite useful in our estimation procedure (see next section).

Our dataset (see appendix A) provides information on prices paid by different individuals in their respective purchases and some relevant observed product characteristics. In order to get more reliable elasticities of substitutions among different models we add our data exogenous information on individual preferences distribution (such as \( f(y_i) \) in (4)).

### 4.2 The Supply Side

We assume the following additive total cost function

\[ \bar{C}(q; \cdot) = [q_I \exp (w \gamma_I + \omega) + q_T \exp (w \gamma_T + \omega)] + F \]  

(5)

where subscripts \( I \) and \( T \) stand, respectively, for the Italian and international markets and the total production is simply \( q = q_I + q_T \). Whereas, \( w \) is the observed subset of cost characteristics, and \( \omega \) the unobserved one, and \( \gamma_I \) the coefficients to be estimated. \( F \) is a fixed cost. We need to distinguish between Italian and international markets, for we have only data on the Italian production (we know only \( q_I \)). We, subsequently, confine ourself to the following conditionally linear (in quantity) variable cost for the Italian market

\[ C(q_I; \cdot) \equiv \left( \bar{C}(q; \cdot) - q_T (w \gamma_T + \omega) - F \right) = q_I \exp (w \gamma_I + \omega). \]  

(5a)

We omit onwards in our notation subscript \( I \).

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

\[ \ln \frac{dC(q; \cdot)}{dq} \equiv \ln(me) = w \gamma + \omega, \]  

(5b)
We expect \( w \) to be inclusive of the relevant characteristics observed by consumers. The intuition behind (5b) is that larger cars or, cars with higher unobserved characteristic values, are expected to be more costly to produce.

Given the demand system in (3), the profits of firm \( f \) (relative to the sales on the Italian market) are

\[
\Pi_f = \sum_{j \in J_f} (p_j - mc_j) q_j
\]

Maximizing (6) we get, for every \( f \in F \) parent house, the common first order conditions

\[
s_j (\cdot) + \sum_{r \in J_f} (p_r - mc_r) \frac{\partial s_r (\cdot)}{\partial p_j} = 0, j \in J_f
\]

from which we get our price equilibria.

We define

\[
\Delta_{jr} = \begin{cases} 
-\frac{\partial s_r (\cdot)}{\partial p_j}, & \text{if models } r, j \in J \text{ are produced by the same firm} \\
0, & \text{(therefore } r, j \in J_f \text{) } r \neq j; \\
\end{cases}
\]

(8)
to be the vector notation.\(^{11}\)

If, as in the linear case, the own- and cross-price effects, \(-\frac{\partial s_r}{\partial p_j}\), are all constant with respect to prices, then, prices should be strategic complements but, Berry and Pakes (1993) notice that utility (4) is sufficiently rich to let prices to be either estimated as strategic complements or substitutes.

We can rewrite our first order condition in vector notation

\[
s (p, x, \xi; \theta) - \Delta (p, x, \xi; \theta) [p - mc] = 0
\]

\(^{11}\)We provide a simple example. Let’s suppose we have a parent house \( a \) producing two models labelled 1, 2 and a parent house \( b \) producing a model labelled 3. Such that \( F = \{a, b\} \) and \( J_a = \{1, 2\}, J_b = \{3\} \). It follows from the notation introduced

\[
\Pi_a = (p_1 - mc_1) q_1 + (p_2 - mc_2) q_2 \\
\Pi_b = (p_3 - mc_3) q_3
\]

The two parent houses maximize their profits deciding on the respective prices. Such as

\[
\max_{\Pi_a} = s_1 + (p_1 - mc_1) \frac{\partial s_1}{\partial p_1} + (p_2 - mc_2) \frac{\partial s_2}{\partial p_1} + (p_3 - mc_3) \cdot 0 = 0
\]

\[
\max_{\Pi_b} = s_2 + (p_1 - mc_1) \frac{\partial s_1}{\partial p_2} + (p_2 - mc_2) \frac{\partial s_2}{\partial p_2} + (p_3 - mc_3) \cdot 0 = 0
\]

\[
\max_{\Pi_b} = s_3 + (p_1 - mc_1) \cdot 0 + (p_2 - mc_2) \cdot 0 + (p_3 - mc_3) \frac{\partial s_3}{\partial p_3} = 0
\]

which explains the vector notation \( \Delta \).
and solving, for the price-cost markup

\[ p = mc + \Delta (\cdot)^{-1} s (\cdot) \]  \hspace{1cm} (9)

Finally, we define the markup vector to be

\[ b (\cdot) \equiv \Delta (\cdot)^{-1} s (\cdot) \]  \hspace{1cm} (9a)

Since \( p \) is a function of \( \omega \), \( b(\cdot) \) is, at his turn, a function of \( \omega \). This problem spreads in the following pricing equation

\[ \ln(p - b (\cdot)) = w \gamma + \omega \]  \hspace{1cm} (9b)

5 GMM Estimator

The fact that producers know the value of the unobserved (to us) product characteristics \( (\xi_j) \), arises correlation between prices and the unobserved product characteristics (cars with higher unmeasured quality should be sold at higher prices) and, therefore, one faces, as well-known from the homogeneous good literature, a simultaneity problem. Moreover, aggregate demand (3) is a non linear function of product characteristics. Aware of this critics, we assume, as common in the literature, \( \xi_j \) and \( \omega_j \) to satisfy the mean independency property for the supply and demand unobserved

\[ E [\xi_j (\cdot; \theta_0) | z] = E [\omega_j (\cdot; \theta_0) | z] = 0 \]  \hspace{1cm} (10)

with \( z = [x, w] \) and \( \theta_0 \) the true parameters value. Although relatively strong, we observe this assumption does not require prices to be uncorrelated with unobservables. Therefore, it only requires the observed product characteristics to be exogenous in our model.

We provide in what follows a proper estimator aimed to correct for this endogeneity problem and estimate simultaneously demand and supply parameters. Although, for efficiency reasons one would have preferred a maximum likelihood estimator, its computational bundersome let us, by virtue of assumption (10), use a general method of moments (GMM) Hansen (1982) to estimate our parameters ((14) infra).
Before introducing the respective moments, let us refresh our relations between demand and supply unobserved and prices and product characteristics

\[ \xi = \delta(p, \cdot; \theta) - x\beta \]
\[ \omega = \ln(p - b(\cdot)) - w\gamma \]  \hspace{1cm} (11)

with \( b(\cdot) \) and \( \delta(p, \cdot; \theta) \) defined respectively in (9a) and (4a1).

Notice however, our dataset is, in fact, a rotating panel. Let \( \mathcal{J} \) be the full set of all model/years and be \( J \subset \mathcal{J} \) the subset of models that are different to each other in the considered period-lag. In order to avoid problems of autocorrelation we suggest to aggregate the same-models over time such that we define our moments

\[ G^J(\theta) = \mathbb{E}_t \left[ H_{ji}(z)T_{ji}(z) \begin{pmatrix} \xi_{ji}(\theta, s_0, P_0) \\ \omega_{ji}(\theta, s_0, P_0) \end{pmatrix} \right] \]  \hspace{1cm} (12)

where \( s_0 \) are the market quota and \( P_0 \) the population distribution whereas \( H_{ji}(z) \) and \( T_{ji}(z) \) are, respectively, the matrix of functions of unobservable variables and the normalization of the variance-covariance matrix. Given the above assumption of independence we have \( G^J(\theta_0) = 0 \). Unfortunately, \( s_0 \) and \( P_0 \) are unknown, so, the function that we actually observe is

\[ G(\theta, s_n, P_{ns}) = \frac{1}{J} \sum_{j=1}^{J} \frac{H_{j}(z)T_{j}(z)}{2x2} \begin{pmatrix} \xi_{j}(\theta, s_n, P_{ns}) \\ \omega_{j}(\theta, s_n, P_{ns}) \end{pmatrix} \]  \hspace{1cm} (13)

in this case \( s_n \) are the observed market quota and \( P_{ns} \) is the empirical distribution of \( ns \) drawn by a \( P_0 \) population.\(^{12}\) Our objective function is just

\[ \text{Min}_{\theta} (G'(\cdot) G(\cdot)) \]  \hspace{1cm} (14)

We describe in next subsection the different components of our \( G \) function of moments (13).

\(^{12}\)The knowledge of \( s^0 \) would require to know, for each consumer, the model/years \( j \) he bought.
5.1 Moments’ Components

i) We define $H_j(z)$ a $2L \times 2$ matrix of functions of the standardized errors $(\xi_j, \omega_j)$.\(^{13}\) Where $L$ is the number of used instruments. BLP, through the restriction that in equilibrium errors are symmetric, suggest the following efficient instruments to enter $H_j(z)$ matrix

$$z_{jk}, \sum_{r \neq j, r \in J_f} z_{rk}, \sum_{r \neq j, r \notin J_f} z_{rk}$$

(15)

where $J_f \subset J$ is the set of models offered by the same $f^{th}$ parent house.\(^{14}\) Therefore, while product $j$ characteristics are valid instruments for themselves ($z_{jk}$), instruments for prices are both the cost side variables excluded from the demand side and, those deriving from the equilibrium first order condition (7) where parent house $f$’s choice of product $j$ price is determined both by its proximity in the characteristics space of competing products and by its own-products ($\sum_{r \neq j, r \in J_f} z_{rk}, \sum_{r \neq j, r \notin J_f} z_{rk}$).

ii) The conditional Variance-Covariance matrix for product $j \in J$ is of the type

\(^{13}\)We provide a simple example of an $H$ matrix. Let’s suppose we use instruments for cubic capacity (cm$^3$) and kilowatt (Kw) where, while the former (cubic capacity) enters both demand and supply, the latter (kilowatt), enters only the supply side. Then, the matrix $H_j(z)$ may be represented as

\[
H_j(z)_{2L \times 2} = \begin{pmatrix}
\sum_{r \neq j, r \in J_f} cm_{3j} & 0 \\
0 & \sum_{r \neq j, r \notin J_f} cm_{3j} \\
0 & 0 \\
0 & 0 \\
0 & \sum_{r \neq j, r \in J_f} cm_{3r} \\
0 & \sum_{r \neq j, r \notin J_f} cm_{3r} \\
Kw_{j} & 0 \\
0 & Kw_{r} \\
0 & Kw_{r}
\end{pmatrix}
\]

\(^{14}\)Where symmetry means that permuting the order of the variables, the value of the function does not change.
with $\Omega(z_j)$ a 2x2 finite for almost every $z$. Then, the full sample Variance-Covariance matrix

$$\Omega(z) = \begin{pmatrix} \Omega_1 & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & \Omega_J \end{pmatrix}$$

(17)

can be normalized to

$$T(z)T'(z) = \Omega^{-1}(z).$$

(18)

See Appendix B for more details.

Finally, we represent the asymptotic variance of $\sqrt{n}(\hat{\theta} - \theta_0)$ by

$$(\Gamma\Gamma')^{-1} \Gamma'VT(\Gamma\Gamma')^{-1}$$

(19)

with

$$\Gamma_{LxL} = E \left[ \frac{\partial G^*}{\partial \theta} (\theta_0) \right]$$

(19a)

(the gradient) and

$$V_{LxL} = E \left( H_j T_j \Omega_j T'_j H'_j \right)$$

(19b)

(the moments’ variance-covariance $E(GG')$). See Appendix B for its representation in vector notation.

6 Computation

Before describing how to compute $G$ in (??), we need first to introduce the model we actually use in our estimates. Aware of the critics in the literature about the poorness of substitution effects of the logit model (which is a model without interactions $\mu_{ij}(\cdot) = 0$ and with a closed form solution) we address, in next subsection, directly to a model with interactions between consumers and product characteristics.
6.1 Model with Interactions

We follow the BLP notation and rewrite our utility model (4a-b) as

\[ u_{ij} = \alpha \ln \left( e^{\bar{m} + \bar{\sigma} \nu_{iy}} + k \right) - p_j + x_j \beta + \xi_j + \sum_k \sigma_k x_{jk} \nu_{ik} + \epsilon_{ij} \]  

(20)

where \( p \) stands for the maximum price on the price set and \((e^{\bar{m} + \bar{\sigma} \nu_{iy}})\) for the minimum value of the function \( e^{\bar{m} + \bar{\sigma} \nu_{iy}} \) for any \( \nu_{iy} \) simulated and, \( \varepsilon \) a small number greater than zero. Therefore \( k \) is nothing but a constant that we add to avoid problems both with the above logarithm and with the partial derivative in (8a).

\[ u_{i0} = \alpha \ln \left( e^{\bar{m} + \bar{\sigma} \nu_{i0}} \right) + \xi_0 + \sigma_0 \nu_{i0} + \epsilon_{i0} \]  

(20a)

Since each individual’s choices are invariant to i) multiplication of utility by each person specific positive constant and ii) addition to utility of any person specific number (i.e. to affine transformations), we can normalize \( u_{i0} = 0 \) (that is equivalent to subtracting \( u_{i0} \) from all \( u_{ij} \)).

We assume that the vectors \((\nu_{iy}, \nu_{i0}, ..., \nu_{i|z|})\), which are fixed over time, are random draws from a multivariate normal distribution with mean zero and an identity variance-covariance matrix independent of the level of consumer’s income \((y_i)\).\textsuperscript{15} The income distribution is assumed to be lognormal and we estimate its parameters from the Bank of Italy Survey on Household’s Income and Wealth for each waves of our panel (we denote the estimated mean by \( \bar{m} \) and the estimated standard deviation - independent of time - by \( \bar{\sigma} \)). This allows us to use the exogenously available information on the income distribution, as to increase the efficiency of our estimation procedure.

As in BLP we obtain the market shares function in two stages. In the first stage we integrate out over the distribution of \( \epsilon_{ij} \) (assumed to be a type 2 extreme value distribution) and we obtain the (conditional on \( \nu_{ij} \)) market share

\textsuperscript{15}In our empirical procedure we take 40 draws for each of the 6 years, for a total of 240 draws.
functions

\[ f_j(x, p, \delta(\cdot), \nu; \theta) = \frac{e^{\delta_j + \mu_{ij}()} + \mu_{ij}()}{1 + \sum_{j=1}^{J} e^{\delta_j + \mu_{ij}()}} \]  

(21)

which also define the partial derivatives in (8)

\[
\frac{\partial s_j(\cdot)}{\partial p_j} = \int f_j(\nu, \cdot) (1 - f_j(\nu, \cdot)) \left[ \frac{\partial \mu_{ij}}{\partial p_j} \right] P_0(d\nu)
\]

\[
\frac{\partial s_j(\cdot)}{\partial p_r} = \int -f_j(\nu, \cdot) f_r(\nu, \cdot) \left[ \frac{\partial \mu_{ij}}{\partial p_r} \right] P_0(d\nu)
\]  

(8a)

where \( \mu_{ij} \) is (4a).

In the second stage, we integrate out over \( \nu_i \) as to obtaining the market shares

\[ s_j(x, p, \delta(\cdot), P_0; \theta) = \int f_j(x, p, \delta(\cdot), \nu; \theta) P_0(d\nu). \]  

(22)

Unfortunately, the non closed solution of (22) requires a simulation procedure.\(^{16}\)

A simple and immediate simulation procedure is offered by simply replacing the population density with its empirical distribution obtained from a set of \( ns \) random normal draws from \( P_0(\nu_1, ..., \nu_{ns}) \)

\[ s_j(x, p, \delta(\cdot), P_{ns}; \theta) = \int f_j(x, p, \delta(\cdot), \nu; \theta) P_{ns}(d\nu) \equiv \frac{1}{ns} \sum_{i=1}^{ns} f_j(x, p, \delta(\cdot), \nu; \theta) \]  

(23)

We are, now, only left with the determination of the \( \delta(\cdot) \) (the mean component of our utility function). We overcome its lack of analytical solution by a numerical solution. In that, we follow BLP and provide a contraction mapping operator

\[ T_{(s, p, \theta)}[\delta_j] = \delta_j + \ln(s_j) - \ln[s_j(p, x, \delta, P; \theta)] \]  

(24)

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

We have now all the tools to describe in next subsection our computation procedure.

\(^{16}\)The integral of (22) is difficult to calculate when the number of product characteristics is larger than three.
6.2 Computation procedure

We summarize how to determine $G(\theta, s_n, P_{ns})$ in the following steps:

(i) We calculate (by simulation) the market quota implied by the models (see (23));

(ii) Next, through the contraction mapping in (24), we determine $\delta(\cdot)$ and, subsequently, we derive the relation $\delta(\cdot) = x\beta + \xi(\{\alpha, \sigma\}, \cdot)$. Note that there is an analytic form for the $\beta$ parameters of the utility surface - conditional on $\{\alpha, \sigma\}$ (parameters of the distribution of consumer characteristics and of the price-income effect) - from where we get the unobserved demand characteristics $\xi(\{\alpha, \sigma\}, \cdot);

(iii) We derive the relation $\ln(p - b(\cdot)) = w\gamma + \omega(\cdot)$ which requires to calculate (8) - which depends itself by $\delta(\cdot)$ through $f(\cdot)$ - and to use relations (9a) and (9b). Note that there is an analytic form for the parameters costs $\gamma$ from where we get the unobserved supply characteristics $\omega(\cdot);

(iv) Finally, by interacting the efficient instruments $H(\cdot)$ and the variance-covariance normalization $T$ with $\xi(\{\alpha, \sigma\}, \cdot)$ and $\omega(\cdot)$ we get our $G$ objective function to be minimized in (??).

(v) Known $G$, we are only left to estimating our parameters $\{\alpha, \sigma\}$. By minimizing (??) through the Nelder-Mead simplex method we get our optimal parameters $\{\alpha, \sigma\} = \{\hat{\alpha}, \hat{\sigma}\}$.\[17\]

In what follows loop numbers are identified by subscript square brackets.

The full procedure is therefore the following:

1. Begin with an initial parameters value $\theta_{[0]} = (\alpha_{[0]}, \sigma_{[0]})$ and an initial mean utility vector value $\delta_{[0]}$ (with $\delta_{[0]} = (\delta_{1[0]}, ..., \delta_{J[0]})$) then, calculate (21) $[f_j (x, p, \delta_{[0]}(\cdot), \nu_i; \theta_{[0]})]$ and, subsequently, simulate (23) $[s_j (x, p, \delta_{[0]}(\cdot), P_{ns}; \theta_{[0]})]$. Use the simulation in (23) for the contraction mapping (24) and get $T_{\{s, P_{ns}, \theta_{[0]}\}}[\delta_{[0]}] \equiv \delta_{[1]}$.

2 Repeat step 1) with $\theta = \theta_{[0]}$ until the contraction mapping converges. Let’s suppose the value of its convergence be $T_{[s,F_n,\theta_{[0]}]}[\theta_{[0]}] \equiv \breve{\delta}$ then,

3 estimate by OLS the parameters $(\beta, \gamma)$:

$$\breve{\delta}(\cdot) = x\beta + \xi(\cdot; \{\alpha, \sigma\})$$
$$\ln(p - b) = w\gamma + \omega(\cdot)$$

from where we get the residuals $\xi(\cdot; \{\alpha, \sigma\})$ and $\omega(\cdot)$

4 apply the Nelder-Mead fixed point minimization of $(G'G)$ and, for each new parameter values, until convergence to $\{\alpha, \sigma\}$, repeat steps 1), 2) and 3) above (where the initial value at each new parameters set is, of course, its convergence value at the previous parameters set).

7 Extending BLP

Consumer decisions depend on transaction prices which, differ from list prices, in that they include transportation costs, taxes, additional accessories and a dealer specific discount that, on his turn, may partly depend on the negotiating power of individual households. Unfortunately, our data include only list prices. Nevertheless, we provide in what follows a price analysis aimed to study the difference between the distributions of list prices (from our Editoriale Domus database) and transaction prices (from our Bank of Italy sample).
8 Results

Table 4 offers a first column of parameter estimates as in BLP. We omit the standard errors of the \( \alpha, \sigma \) parameters, for we need to review our Variance-Covariance matrix.

We start to describe the results in table 4 from the cost side. We observe that all the \( \gamma \) parameters but that of trunk size are of the expected (positive) sign. A positive sign means that the higher the value of a characteristic, the larger is its effect on the marginal cost. We are not worried by the negative sign of the parameter associated with the trunk size as we are considering all automobile therefore, also, auto such as mini-vans, pick-up which have bigger trunk size but lower production costs. Once we introduce in our estimates dummies for specific market segments (as in column 2) we expect to get rid of that negative sign.

Turning our attention on the demand side we observe that each characteristic has both a mean effect (associated with the \( \beta \)'s) and a standard deviation (associated with the \( \sigma \)'s) on individuals’ marginal utility. Again, we are not worried by the negative sign of the \( \beta \) associated with the length characteristic for a reason similar to that offered above. However, to give an idea of the effect of the parameters \( \beta, \sigma \) on individuals’ marginal utility, let’s focus on the \( \beta, \sigma \) associated with one of those characteristics, for example max speed. From the high values of both \( \beta_5 \) and \( \sigma_4 \) we expect a proportional increase in the number of consumers buying high speed cars and, even more, the high value of \( \sigma_4 \) tells us that, once the price of a high speed (i.e. Porsche) increases, consumers will move from that model to another high speed models (i.e. Ferrari). Finally, the \( \alpha \) parameter which explains the price-income effect is of the expected sign. All the \( \alpha, \beta, \gamma \) parameters but that associated with the trend variable (\( \gamma_7 \)) are significant.
Table 4: Estimated parameters of the demand and pricing equations (14).\(^{18}\)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Variable</th>
<th>Parameter Estimates</th>
<th>New Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta_1)</td>
<td>constant</td>
<td>-37.4800 (.66)</td>
<td></td>
</tr>
<tr>
<td>(\beta_2)</td>
<td>cm3</td>
<td>.0011 (.0002)</td>
<td></td>
</tr>
<tr>
<td>(\beta_3)</td>
<td>length</td>
<td>-5.9096 (1.4922)</td>
<td></td>
</tr>
<tr>
<td>(\beta_4)</td>
<td>trunk size</td>
<td>1.1913 (.673)</td>
<td></td>
</tr>
<tr>
<td>(\beta_5)</td>
<td>max speed</td>
<td>6.0823 (1.0414)</td>
<td></td>
</tr>
<tr>
<td>(\beta_6)</td>
<td>fuel consumption</td>
<td>-5.8766 (1.2922)</td>
<td></td>
</tr>
<tr>
<td>(\sigma_1)</td>
<td>cm3</td>
<td>10.4860</td>
<td></td>
</tr>
<tr>
<td>(\sigma_2)</td>
<td>length</td>
<td>6.7672</td>
<td></td>
</tr>
<tr>
<td>(\sigma_3)</td>
<td>trunk size</td>
<td>-23.589</td>
<td></td>
</tr>
<tr>
<td>(\sigma_4)</td>
<td>max speed</td>
<td>30.268</td>
<td></td>
</tr>
<tr>
<td>(\sigma_5)</td>
<td>fuel consumption</td>
<td>-15.404</td>
<td></td>
</tr>
<tr>
<td>(\gamma_1)</td>
<td>constant</td>
<td>-5.778 (.3130)</td>
<td></td>
</tr>
<tr>
<td>(\gamma_2)</td>
<td>cm3</td>
<td>.0011 (.0001)</td>
<td></td>
</tr>
<tr>
<td>(\gamma_3)</td>
<td>length</td>
<td>4.4343 (.6849)</td>
<td></td>
</tr>
<tr>
<td>(\gamma_4)</td>
<td>trunk size</td>
<td>-.6386 (.3099)</td>
<td></td>
</tr>
<tr>
<td>(\gamma_5)</td>
<td>max speed</td>
<td>2.8741 (.4780)</td>
<td></td>
</tr>
<tr>
<td>(\gamma_6)</td>
<td>fuel consumption</td>
<td>3.1715 (.6093)</td>
<td></td>
</tr>
<tr>
<td>(\gamma_7)</td>
<td>trend</td>
<td>.00851 (.0124)</td>
<td></td>
</tr>
<tr>
<td>(\alpha)</td>
<td>ln ((f(y) - p))</td>
<td>9.7875</td>
<td></td>
</tr>
</tbody>
</table>

Sports
Van, Off-Roads
Italy
Europe
USA

Table 5 and 6 will refer to the new parameter estimates which values express the percentage change in the market share of the \(j \in J\) model in response to a change of 1000 Euro in the \(r \in J\) model.

\(^{18}\)Product characteristic values are normalized to 1.
9 Conclusion

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 on the set of prices decided by the same (and competitors) companies on foreign markets. We faced such restriction by assuming a cost function additive in the Italian and abroad production. We are also aware that we are not directly modelling consumer decisions on after sales assistance and spare parts availability, but we are not afraid of that, for we expect equilibrium prices to be inclusive of this information.
References


A The New Vehicles Dataset

As pointed out in Section 3, our data don’t let us to distinguish between used and new vehicles. To overcome this data restriction we assume the used vehicle market to clear each period. In this case, all the vehicles sold by households at a price higher than $p$ are, within the same year, bought by other households.\(^\text{19}\)

Furthermore, there is a specific question regarding an estimation of households’ vehicle wealth which is, unfortunately, of no help to us.

We denote by superscripts $n$ and $u$ new and used vehicles purchases whereas, $s$ stands for sales.

Let $I^u \subset I$ and $I^n \subset I$ be the subsets of individuals that, respectively, buy used and new vehicles.\(^\text{20}\) 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 to separate the two subsets

\[
I_r^u = I_{r-1}^u \cup \left\{ \min_i \{ i : \min |p_i - p_h| \} \right\} \text{ for } r = 1, \ldots, |I^s|
\]

\[
I_0^u = \{ \emptyset \}; i \in \{ I^{u,n} \setminus I_{r-1}^u \}; h \equiv I^s [r]; i \neq h.
\]

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

\[
I^n = \left\{ I^{u,n} \setminus I_{|I^s|}^u \right\}
\]

(25a)

Since we are aimed to study the new car market we determine the quota of the outside alternative in our market

\[
s_0 = \frac{|I| - |I^n|}{|I|}
\]

(26)

The following figure describes the results from the above formulation. 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, moreover, by the number of new cars sold for commercial use. Although one could also think of a sample underestimation.

\(^{19}\)Vehicles sold by households at a price below $p$ are assumed to be sold for scrap. We assume that value to be 3000 Euro in year 2000.

\(^{20}\)Obviously for $i \in I^n$, $p_i^u \geq p_i^s$. 
To be coherent with our assumption that each household buys no more than a car a year, we get our $s_0$ quota from the following ratio

$$s_0 = \frac{\text{total sales}}{\text{# of households in the economy}}$$

which values are expressed in the following figure.
B Variance-Covariance Matrices

We report an example of the full sample Variance-Covariance matrix

\[ \Omega(z) = E \begin{pmatrix} \xi^2_1(\cdot; \theta) & 0 & \ldots & 0 \\ 0 & \omega^2_1(\cdot; \theta) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \xi^2_J(\cdot; \theta) \end{pmatrix} \mid z \equiv \begin{pmatrix} \Omega_1 & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & 0 & \Omega_J \end{pmatrix} \]

and

\[ \Omega^{-1}(z) = \begin{pmatrix} \Omega^{-1}_1 & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & \Omega^{-1}_J \end{pmatrix} \equiv \begin{pmatrix} T' \Omega^{-1} T \\ 0 & \ldots & \ldots \\ \ldots & \ldots & 0 \end{pmatrix} = \begin{pmatrix} T^T_1 \Omega^{-1}_1 T_1 \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & T^T_J \Omega^{-1}_J T_J \end{pmatrix} \]

with \( \Omega(z) \) a \( 2J \times 2J \) matrix finite for almost every \( z \) and \( T^T T \) its normalization.

We reported in section 6.1 the following representation of the asymptotic variance of \( \sqrt{n} \left( \hat{\theta} - \theta_0 \right) \)

\[ (\Gamma \Gamma')^{-1} \Gamma' \sqrt{V} \Gamma (\Gamma \Gamma')^{-1}. \]

We offer now a vector notation for the \( V \) matrix. Let

\[ G(\theta) = \frac{1}{J} \begin{pmatrix} H_{12} & \ldots & H_{J2} \\ L_{2J} & \ldots & L_{2J} \end{pmatrix} \begin{pmatrix} T_1 & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & T_J \end{pmatrix} \begin{pmatrix} \eta_1 \\ \ldots \\ \eta_J \end{pmatrix} \]

with

\[ \eta_j = \begin{pmatrix} \xi_j \\ \omega_j \end{pmatrix} \]

then, we have

\[ V = \frac{1}{J} \begin{pmatrix} H_{12} & T_J & \Omega J & T_J^T & H_J' \\ L_{2J} & \ldots & \ldots & \ldots & \ldots \end{pmatrix} \]
C Estimation-Files (Description)

| Prices.do | The original file provided us by Editoriale Domus-Quattroruote (converted to the Stata file Prices.dta) contains yearly prices (quarterly reported) and a key variable infocar anno mese for all the different auto name plates. We, then, calculate the mean yearly prices and save them, for the period 1989-2001, in the file Price1.dta. In the obtained file we have very few years of missing prices (which means are missing all the quarterly reported prices of that specific year). This is obviously a problem once the Fiat dataset reports, for the corresponding name plate and year, a positive sold quantity. We get rid of this drawback by substituting the respective missing prices with the previous/subsequent price values (corrected for the sample-price yearly growth rate) and save all variables in the file Price2.dta. |
| Datitecnici.do | Editoriale Domus-Quattroruote provided us also with a file (converted to the Stata file Datitecnici.dta) containing the key variable infocar anno mese and information on the auto characteristics (for all the new auto name plates sold in the period 1989-2001). We, then, as explained in Appendix A, build the new variable consumo4 to match two different ways of reporting fuel consumption in the considered period and save all the variables in Datitecnici1.dta. |
| Clean.do | We use the Bank of Italy special section on the vehicles purchasing/selling. We save, each year, a Stata file [q(year).dta] containing, for all households, information whether they bought/sold a vehicle in the year and, in case, the price (in thousands of Euro 2000) they respectively paid/received. Notice anyway that these Stata file are inclusive both of new and old car purchasing! |
| Clean.m | From the files q(year).mat (which is q(year).dta converted in Matlab file) we use a minimum distance method (explained in Appendix A) as to identify the sub-sample of individual who bought a new vehicle in the year. We, then, calculate (each year) the quota of individuals who chose the alternative good and save all the variables in the file q(year)ecl.mat. |
| Clean.do | From the file q(year)ecl.dta (which is q(year)ecl.mat converted in Stata file) we use the second part of the file Clean.do to append in a unique file s0.dta the variables: households’ identifier, prices paid for the new auto purchase, time, and the alternative good quota s0. |
Simulaz.do

It generates the means and standard deviations (from a multinormal distribution with Identity Variance-Covariance matrix) to be associated with the product characteristics and the income distribution. It also computes, from the Bank of Italy data, the means and standard deviations for the income distribution (assumed to be log-normal) both for the case of a standard deviation depending, and not depending, on time. All the above variables (plus time) are saved in the file sim.dta.

Cardata.do

Fiat provided us with a file containing all the different new auto name plates sold in Italy each year for the period 1989-2001. Unfortunately, as the models were named differently to the Editoriale Domus-Quattroruote database we had to attach to each name plate in the Fiat database (about 12000 observations) the corresponding Editoriale Domus-Quattroruote key variable (infocar anno mese). We thank Andrea Battiston for having patiently created such common key variable.

We, then, merge the updated file (converted to the Stata file fiatdata.dta) with Datitecnici1.dta (where we have product information) and, subsequently, we create two variables to identify the different Parent Houses and Models. We save all the variables (reshaped in a way to have time as variable) in the file estm.dta. To this new file we merge the prices information Price2.dta, and, as explained in Definition ii) in Appendix A, we restrict the original dataset to years/model. Finally, we create all the instruments to be used in our estimations and save all the obtained variables in the file cardata.dta.

Automk.m

It loads Cardata.mat (the file cardata.dta converted to a Matlab file) and launches the file nelder.m for the Nelder-Mead Simplex minimization Routine (the matlab file fminsearch.m). This is the procedure required by the GMM to determine the parameter values (those that we cannot estimate by OLS!).

---

36
### Nelder.m
It is the file that gives us the function to be minimized in Automk.m. The file goes through the following procedure (given initial parameter and deltalj vectors):
1) It determinates the fij functions of the market quotas.
2) Uses the simulations (which values are in the file Cardata.mat) to evaluate the market quotas sj.
3) Uses a contraction mapping as to get the mean utilities deltaj.
4) Estimates the markup matrix.
5) Estimates by OLS the parameters (and standard errors) on supply side and part of those of the demand side.
6) From the OLS estimation it gets the supply and demand residuals (the unobserved product characteristics).
7) Gets also the Hj function of instruments (which instruments are in the file cardata.mat) and Tj (the Choleski product of the inverted variance-covariance matrix)
8) Builds the function G’G to be minimized.

### Automk.m
From the function G’G it applies The Nelder-Mead search method as to get the optimal parameters. Then, to the aim of calculating the standard errors of the estimated parameters, we use the files numjac2.m and Jacob.m as to get the Variance-Covariance matrix of the estimated parameters.

### Jacob.m
It is almost the same file of Nelder.m (the only difference is that it evaluates the G function instead of G’G) and save the file junk.mat with the variables necessary to determine the Variance-Covariance matrix of the estimated parameters.
Numjac2.m
It evaluates numerically the differentiation of the expected value of the G function (that we get from the file jacob.m) with respect to the estimated (by Nelder-Mead search) parameters.

Automk.m
From the file junk.mat it evaluates the standard errors of the parameters determined by Nelder-Mead search. Finally, it provides the matrix of own and cross-elasticity of substitution for a price change of 1000 Euro.

Quotas.do
From the file fiatdata.dta it builds (for the period 1989-2000) a graph of the market quotas by different areas.