Cross-country comparison of substitution patterns in the European car market

Paper submitted to obtain the degree of Master of Advanced Studies in Economics

Faculty of Business and Economics

K.U. Leuven, Belgium

Supervisor: Frank Verboven. Discussant: Patrick Van Cayseele

Candidate: Victoria Oliinik

January 2007

Abstract

AIMS. Using recent data on sales, prices and product characteristics of new passenger vehicles sold in Europe's seven largest markets during years 2000 through 2005, I estimate a demand function using two approaches to the differentiated product demand estimation: the Logit and the Nested Logit. For the Nested Logit model, I use nests that roughly correspond to segmentation used in the automobile industry. I then compare substitution patterns resulting from Logit and Nested Logit specifications across countries and segments. FINDINGS. Regression outcomes demonstrate strong evidence of segmentation. Consumer preferences are strongly correlated for the most basic as well as for the luxury and sports segments, whereas core segments display more heterogeneity. While car size preference varies across countries, Europeans typically like high and wide cars and dislike long cars. Engine performance appreciation appears to have a curved shape: consumers are willing to pay for a fast car but at a decreasing rate. Compared to their neighbors, Germans shop the most for premium brands yet display strongest price sensitivity, whereas most concerned about fuel consumption are the Brits. Aside from a few exceptions, all elasticities are within expected range, substitution patterns vary mainly across segments but follow a similar pattern across countries.

1 Introduction

The European car market is a phenomenon in its own right. While this is a mature market place with several well-established players, market leadership and dynamics differ to a large extent among countries. I study market demand in Germany, France, United Kingdom, Italy, Spain, Netherlands and Belgium. These seven markets account for approximately 85 % of EU-25 demand for new cars. Table 1 presents a snapshot of each market and its players.

Firstly, the top three car makes control between 29 and 56% of the market, with Netherlands being the least and France being the most concentrated market. The identity of main players varies from country to country, so does the concentration. Germany and France are countries with a strong presence of national brands, contrary to Netherlands and Belgium.

Table 1: LEADING BRANDS and MARKET SHARES - 2005 data

RANKING	GERMANY		FRANCE		UK		ITALY		SPAIN		NETHERLANDS		BELGIUM	
1	Volkswagen	19%	Renault	25%	Ford	14%	Fiat	20%	Renault	12%	Volkswagen	10%	Renault	11%
2	Opel/Vauxhall	10%	Peugeot	18%	Opel/Vauxhall	13%	Opel/Vauxhall	8%	Citroen	10%	Opel/Vauxhall	10%	Peugeot	11%
3	Mercedes	10%	Citroen	13%	Volkswagen	8%	Ford	7%	Peugeot	9%	Peugeot	9%	Volkswagen	11%
4	BMW	8%	Volkswagen	7%	Renault	7%	Volkswagen	6%	Ford	9%	Renault	9%	Citroen	10%
5	Audi	7%	Opel/Vauxhall	5%	Peugeot	6%	Renault	6%	Seat	9%	Ford	8%	Opel/Vauxhall	10%
6	Ford	7%	Ford	5%	Toyota	5%	Citroen	6%	Opel/Vauxhall	8%	Toyota	6%	Ford	6%
7	Renault	5%	Toyota	4%	BMW	5%	Toyota	6%	Volkswagen	7%	Citroen	5%	BMW	5%
8	Toyota	4%	Mercedes	3%	Citroen	4%	Lancia	4%	Toyota	4%	Volvo	4%	Audi	5%
9	Peugeot	4%	Fiat	2%	Honda	4%	Peugeot	4%	Audi	3%	Hyundai	4%	Toyota	5%
10	Skoda	3%	Audi	2%	Nissan	4%	Mercedes	3%	BMW	3%	Kia	3%	Mercedes	4%
11	Citroen	2%	Nissan	2%	Audi	3%	BMW	3%	Hyundai	3%	Audi	3%	Hyundai	3%
12	Mazda	2%	BMW	2%	Mercedes	3%	Alfa Romeo	3%	Mercedes	3%	Fiat	3%	Volvo	2%
13	Fiat	2%	Seat	2%	Mazda	2%	Audi	3%	Nissan	3%	BMW	3%	Seat	2%
14	Seat	2%	Hyundai	1%	Land Rover	2%	Nissan	2%	Kia	3%	Suzuki	3%	Skoda	2%
15	Hyundai	2%	Suzuki	1%	Fiat	1%	Chevrolet/Daewoo	2%	Fiat	2%	Seat	2%	Fiat	1%
MARKET SIZE (units sold)	3,342,122		2,067,789		2,439,726		2,263,782		1,643,496		465,149		480,088	
PREMIUM BRANDS SHARE	28%		8%		16%		11%		11%		14%		19%	
POPULATION (mio)	82.491		60.200		59.778		57.553		42.692		16.275		10.399	
GDP/CAPITA (thousands EUROS)	27.167		28.406		29.955		24.625		21.209		31.069		28.709	

This ranking is based on 2005 market shares. Mercedes excludes Smart, BMW excludes Mini. Premium brands include Audi, Mercedes, BMW, Volvo, Saab, Jaguar, Porsche and Lexus.

GDP/CAPITA: nominal GDP in EUROS/population.

Secondly, new car demand is the largest in Germany - with over 3.3 million annual new car sales - and the smallest in the Netherlands - covering 465 thousand units. Germany, France, United Kingdom, Italy and Spain are largest in terms of population and car demand size. Interestingly, demand for new cars, proportionally to the population size is largest in Belgium (5.1%). In the UK and Germany it corresponds to 4.1% of the population, followed by Italy (3.9%), Spain (3.7%), France (3.4%) and the Netherlands (2.9%). Furthermore, compared to their neighbors, Germany and Belgium are the markets where luxury brands have the largest market shares.

Considering the above disparity, I investigate to what extent there could be differences in demand elasticities between these markets and across product segments. In order to achieve this, I estimate a discrete choice model of demand for differentiated products using a Fixed Effects Instrumental Variable regression. I estimate Logit and two specifications of the Nested Logit model: one in which a common segmentation (nesting) parameter is imposed - the restricted preference case - and one where I allow segmentation parameter to differ for each consumer group - the flexible case. In the restricted case, nesting parameter estimates range from 0.42 to 0.79 and compare to findings of other researchers. Zooming in on the flexible case, nesting parameter estimates vary significantly from segment to segment. While for typical non-luxury products they fall within the expected range, for some products, contrary to initial theoretical expectation, they are outside the unit interval. This is mainly the case for niche products and smallest cars, indicating that consumers who shop in these segments will react only to a limited range of product and price incentives. For the Nested Logit model, I use twelve product clusters[1] in line with standard car industry practice and fine-tuned by personal experience. Elasticities estimated by both Logit and Nested Logit models differ between countries in absolute values, with Germany appearing to be the most and United Kingdom and Italy the least elastic markets. Differences in nesting parameters lead to substitution patterns that vary from segment to segment.

2 A snapshot of literature on this topic

Initially, Discrete Choice (DC hereafter) models had been developed to model choices among alternatives faced by decision makers(consumers, producers, etc.) and required individual-level data. Berry (1994) showed how to invert a non-linear individual choice model, which enabled estimation of individual

[1] I refer to clusters, nests and segments equivalently

demand models using aggregate data. Further in the text, I recur to his technique for my purposes. Following Berry (1994), came out the acclaimed article "Automobile Prices in Market Equilibrium" by Berry, Levinsohn, and Pakes (1995), hereafter referred to as BLP, which showed how to use demand side variables to address the issue of endogeneity of prices. The authors proposed a set of instruments that can be used in the absence of cost-side data. Furthermore, BLP proved validity of these instruments and demonstrated their use for the US car market demand. To estimate car demand, the authors used and Logit and the Random Coefficients Model. The latter is a generalization of the Nested Logit Model: it allows each consumer to differ in his valuation of characteristics and prices. Several subsequent papers by Frank Verboven have built around the above findings and used the Nested Logit model as the model of choice. My work is largely inspired by Brenkers and Verboven (2002), Brenkers and Verboven (2006) and Goldberg and Verboven (2001). Using Discrete Choice demand framework, these three articles focused on five out of the seven countries I study in this paper. The data they use cover more years and dates further back in time. I will refer to details of their findings further in the text.

3 Data

3.1 Segmentation

The available data cover all new automobiles sold in each of the seven European markets over six years. It is very unique in terms of detailed product description - every possible car characteristic has been carefully researched and coded for each and every version of car model offered on the market. The original data collected by JATO Dynamics, a private company that offers the most extensive coverage of the car market world-wide, has been regrouped into twelve segments. Products range from small utility cars to large prestigious automobiles. Table 2 provides an overview. The initial segmentation contained the E/E+ segment as two distinct segments. As these segments contained very few observations (about 4-7 models per year each), I grouped them together. Also, I excluded commercial (utility) vehicles from the analysis. Lastly, I separated the original SUV segment into the cheaper *volume* and the enhanced *luxury* segments.

In brief, all models are grouped by a combination of size (including the corresponding engine output to support the size) and product image. When considered along a product enhancement continuum, six of the segments are

Table 2: SEGMENTATION

DESCRIPTION	SEGMENT		TYPICAL PRODUCT (list not exhaustive)
MAINSTREAM	1	A	Ford Ka, VW Lupo, Peugeot 106
MODELS BY SIZE	2	B	Ford Fiesta, Peugeot 206, Opel Corsa , VW Polo
AND	3	C	Toyota Corolla, Ford Focus, Peugeot 307, VW Golf
PREMIUM	4	C+	Audi A3, Alfa Romeo 147, BMW 1-series
APPEAL*	5	D	Renault Laguna, VW Passat, Peugeot 407
	6	D+	Audi A4, BMW 3-series, Mercedes C-class, Lexus IS
PRESTIGE	7	E/E+	Mercedes E- and S-class, Peugeot 607, BMW 5-and 7-series
MULTI-PURPOSE	8	Mini MPV	Renault Scenic, VW Touran, Opel Zafira
	9	MPV	Renault Espace, Ford Galaxy, VW Sharan, Peugeot 807
NICHE	10	Sports	Audi TT, Ferrari Enzo, Opel Tigra, Porche 911, Toyota Celica
	11	SUV	Toyota Rav4, Nissan X-Trail, Suzuki Grand Vitara
	12	SUV+	Toyota Land Cruiser, BMW X-5, Jeep Cherokee, Porsche Cayenne

Premium appeal is designated by a plus sign

mainly composed of volume brands such as Ford, Renault, Opel/Vauxhall, Peugeot, Toyota and partly Volkswagen. At the opposite extreme are three segments that mainly include premium brands such as Mercedes, Lexus, Audi and BMW. The remaining three segments consist of a mix of luxury and volume brands that position their products somewhere in between to serve the D, Sports and SUV segments by offering both volume and premium highly differentiated products, that are assumed to appeal to the product-educated niche of consumers.

3.2 Data description

MARKETS. I use automobile specifications, prices and sales data for the years 2000-2005 in Europe's seven major markets. As I am mainly interested in a cross-country comparison, I study each country separately while applying the same econometric methods all across. Other facts call for considering these markets individually. Not so long ago, these were very disconnected car markets, in which certain car brands were favored by governments and consumers for historical and political reasons. A priori, one can suspect that this was not without consequence on consumer choices.

PRICES. All prices are manufacturer suggested retail prices (MSRPs) in home currency[2], which are basically dealer prices plus V.A.T. In the UK and Italy, MSRP's also include additional taxes and delivery charges, ranging from 400 to 800 euros depending on the car type. Retail prices in the Netherlands comprise BPM (belasting van personenauto's en motorrijwielen), a hefty additional "luxury" car tax. MSRPs have not been adjusted for

[2]As some markets adopted the euro during the period I analyse, I convert all prices to euros for consistency. The UK prices remain in British Pounds Sterlings

discounts or other promotional activities.

PRODUCT CHARACTERISTICS. The data made available for estimation contain the majority of items that are valued by a typical consumer. The dataset contains information on whether the following items had been fitted on the vehicle: audio system, ABS, cruise control, navigation system, leather seat material, refrigerated storage compartment, power windows, power (and remote) door locks, fog lamps, stability and traction control, air conditioning, automatic climate control, total number of airbags, transmission type, number of gears, four-wheel drive capacity and an indicator of product grade. Furthermore, there are measures of seating capacity, length, width, height and wheelbase. A number of engine specifications is also considered: number of cylinders, cylinder size (in cubic centimeters), maximum horsepower (DIN), maximum torque (Nm), maximum speed (km/h), acceleration (seconds needed attain 100km/hour speed), and combined fuel consumption (litres/100km).

4 Conceptual model

4.1 Discrete choice model of demand

Given that the automotive industry offers a large variety of differentiated products, estimating a simple aggregate demand function from standard cnsumer theory would be too parameter demanding. When there are J products in a market, $(J+1)^2$ own and cross price elasticities are to be estimated. Also, when accounting for consumer heterogeneity, it is rarely possible to apply the typical restrictions from consumer theory such as Slutsky symmetry[3] - hence the requirement to estimate additional parameters. In order to limit the number of parameters while imposing structure on demand that is as realistic as possible, I recur to the Discrete Choice framework. Among the several types of DC models available, I choose to work with a Logit and a Nested Logit models; the latter also being referred to as the Generalized Extreme Value (GEV) Model. DC models address the issues pertaining to choices of individual decision-makers (e.g. consumers) among alternatives (e.g. products). Train (2003) provides a very useful guide to DC models, where he also he specifies that the set of alternatives must satisfy the following three basic criteria:

1) The alternatives must be mutually exclusive,

[3] Slutsky symmetry stipulates that when prices change, demand change is a result of two effects: the substitution effect and the income effect.

2) The set of choices must be exhaustive, and

3) The number of alternatives must be finite.

The above are satisfied when considering a consumer's decision to buy a new car. Either the consumer buys car X or car Y[4] or both or nothing at all. All alternatives must be included: all cars available on the market are comprised in the model, including the option of not buying a car at all. The third and the most distinct specificity of DC models is that choices must be countable. In a regular regression model, the dependent variable is continuous, implying an infinite number of alternatives. To make a rough distinction between the standard regression models and DC models it can be convenient to think of the former as the choice of *how much*, and of the latter as the choice of *which*. My study of substitution patterns for the cars is in line with the latter. Also, most of the economic literature on automobiles addresses demand-side issues by using some type of DC model.

DC models, as a whole can be summarized as a relationship between the choice outcome and the explanatory variables that motivate the choice. Applications in economic settings, namely the Random Utility Models[5], assume that the decision-maker is necessarily maximizing his utility (or benefit) when making the choice. Hence, in my investigation of consumer choice of different cars, I model the utility of consumer i considering J alternative products as U_{ij} for j=1,2...J. While each alternative j yields a certain individual-specific utility U_{ij}, each consumer chooses the product that gives him the largest utility: $U_{ij} > U_{ik}$ for $j \neq k$. Individual utility is then modeled in terms of preferences for product characteristics: differentiated products are projected onto a space of measurable characteristics[6], such as size, price or fuel efficiency from which consumers derive utility. This solves the multiple dimensionality of products problem (i.e. the number of parameters to be estimated reduces significantly). In the following section, I take advantage of the Berry (1994) inversion technique and explain how to aggregate the individual DC model to arrive at the market demand. I apply this aggregate demand system to

[4]The option of buying more than one car can ultimately be specified as well. As this is not the question of my investigation, I focus on the choice of a single product versus the option of not buying a new car at all.

[5]Random Utility Models explicitly address the fact that only a portion of individual utility is observed, the remainder is uncertain to the researcher.

[6]The idea was first formulated by Rosen (1974) in his theory of hedonic prices, based on the hypothesis that products are valued for their utility-bearing characteristics. Also, McFadden (1978) redefines products as bundles of characteristics in his Generalized Extreme Value Model.

the dataset for each of the markets[7].

4.2 Individual utility

For L consumers in the market, designating product specificities by subscript j and consumers by subscript i, the direct utility of individual i for a product j is formulated as:

$$u_{ij} = \beta_i x_j + \alpha_i f(z_i) + \xi_{ij} + \varepsilon_{ij} \tag{1}$$

In the above equation, x_j is a vector of observable product characteristics such as fuel consumption or engine performance. In the second term, z_i is the part of consumer revenue that is spent on other goods, so $f(z_i)$ can be thought of as utility derived from other goods such as clothing. The term ξ_{ij} encompasses unobserved by the researcher product characteristics such style, brand image, prestige, common demand shocks, etc, while ε_{ij} is the consumer-specific utility term that depends on personal factors. The utility of the so-called outside good, the option of not buying a car, is designated by u_{i0}. Assuming that consumers choose the alternative that maximizes utility, the budget constraint takes on the following form:

$$\sum_{k=0}^{J} p_k d_k + z_i = y_i \tag{2}$$

where product dummy d_k equals unity if out of (J+1) alternatives product j is chosen and zero otherwise; p_k is this product's price and y_i is consumer revenue. Prices of other products are normalized to one. Substituting $z_i = y_i + p_j$ into the utility function, I obtain the following indirect conditional utility:

$$u_{ij} = \beta_i x_j + \alpha_i f(y_i - p_i) + \xi_{ij} + \varepsilon_{ij} \tag{3}$$

Among the many conceivable specifications of $f(.)$, the following two are broadly used:

1) Linear

$$\alpha_i f(y_i - p_j) = \alpha_i y_i - \alpha_i p_j \tag{4}$$

and

[7] I therefore implicitly assume that consumers consider buying a car only in their home country

2) Logarithmic (Cobb-Douglas):

$$\alpha_i f(y_i - p_j) = \alpha_i ln(y_i - p_j) = \alpha_i ln(y_i(1 - p_j/y_i) \approx \alpha_i ln(y_i) - (\alpha_i/y_i)p_j \quad (5)$$

The first term of each specification is constant and drops out when utility is differentiated with respect to product j. Linear specification assumes that there are no income effects in utility - utility is a function of price only, whereas in the logarithmic case, utility is a function of price relative to income, (e.g. high income consumers are less price sensitive to high prices). As I will be comparing countries with different income and price levels, I have opted for the logarithmic specification.

The resulting indirect utility function, to which I will apply Logit and Nested Logit assumptions, is:

$$u_{ij} = \beta_i x_j - (\alpha_i/y_i)p_j + \xi_{ij} + \varepsilon_{ij} \quad (6)$$

4.3 From individual utility to market demand

For the specifications that follow, I assume that all consumers have a common price parameter a and do not differ in their willingness to pay for product characteristics such as horsepower or car size:

$$(\alpha_i/y_i) = \alpha/y$$

$$\beta_i = \beta$$

Consequently, the individual utility as specified in equation (6) becomes $u_{ij} = \beta x_j - \alpha/yp_j + \xi_j + \varepsilon_{ij}$, which can now be expressed as $u_{ij} = \delta_j + \varepsilon_{ij}$, with δ_j designating the mean utility of each product across consumers. Similarly, the mean utility of the outside good, δ_0, is normalized to zero, resulting in $u_{i0} = \varepsilon_0$.

4.4 The Logit Model

The Logit model is the simplest and the most popular. The simplicity of this model comes from the assumption that the unobserved part of utility is random. For each product j, consumer i's utility is modeled as $U_{ij} = V_{ij} + \varepsilon_{ij}$, where V_{ij} is (partially) observed by the researcher and ε_{ij} is random, distributed independently and identically (i.i.d.) extreme value, with cumulative distribution $F(\varepsilon_{ij}) = e^{-e^{-\varepsilon_{ij}}}$. The well-known property of Logit choice probability, namely Independence of Irrelevant Alternatives (the IIA property), stems from this specification of the error term. It means that for every alternative, the unobserved component of utility is independent of unobserved components of the other alternatives. In other words, consumers have uncorrelated preferences for similar products. That is, a consumer does not have a preference for cars within a certain segment. Even though he is planning to buy an Opel Corsa, he may also be considering a Ford Mondeo or VW Golf. Put differently, just because he is buying a B-segment Corsa, it does not mean that he likes other cars belonging to this segment, such as VW Polo. This model does not allow to group decision-makers by type or by taste for a specific segment. As shown by McFadden (1974), the Logit formula for choice probabilities implies that unobserved utility e_{ij} is extreme value distributed.

Bringing Logit distributional assumptions of ε_{ij} to the above specification of individual utility, the cumulative distribution function is $F(\delta_0, ..., \delta_j) = e^{-e^{-I}}$ with I being the "inclusive value" defined as

$$I = ln \sum_{k=0}^{J} e^{\delta_k} \tag{7}$$

and interpreted as the expected value of the maximum of all utilities[8]. The probability that individual i chooses product j is:

$$s_{ij}(\delta_j) = \frac{e^{\delta_j}}{e^{\delta_I}} = \frac{e^{\beta x_j - \alpha p_j + \xi_j}}{1 + e^{\beta x_j - \alpha p_j + \xi_j}} \tag{8}$$

As the utility of outside good is normalized to zero, $e^{\delta_0} = e^0 = 1$. The individual error, ε_{ij}, is integrated out. I assume that the individual choice probability function $\mathtt{s}_{ij}(\delta_j)$ is the same for all individuals, i.e.: $\mathtt{s}_{ij}(\delta_j) = \mathtt{s}_j(\delta_j)$. If δ is a vector of mean utilities of all products, and individual choice probability can be used as a proxy for market shares, then for each product j, the aggregate market share can be estimated by $s_j = \mathtt{s}_j(\delta) = \mathtt{s}_j(\delta_1, ...\delta_j)$, with the left hand side term representing the observed market share of product j, and the right hand side terms representing the choice probability of product j as derived above. Obviously, ξ_j, the econometric error term accounting for unobserved characteristics contained in δ, enters the model in a non-linear manner and renders estimation very complicated. Using a demand transformation technique proposed by Berry (1994), I divide both sides of the equation (8) by s_0, the observed market share of the outside good, take logs of both sides and substitute δ_j by its expression:

$$s_j/s_0 = e^{\delta_j}$$

[8]Normalizing I by a enables expressing utility in monetary units.

$$ln(s_j/s_0) = \delta_j = \beta x_j - \alpha p_j + \xi_j$$

This makes the model linear in parameters and facilitates econometric estimation with $ln(s_j/s_0)$ as explained variable, β and α as parameters to be estimated and ξ_j j as the econometric error term.

4.5 The Nested Logit Model

The assumption placed on ε_{ij} defines a DC model type. In the present case, it distinguishes the Logit from the Nested Logit model. While in the Logit case ε_{ij} is i.i.d., in the Nested Logit model, the error has a structure that allows for correlation inside groups of similar products. In this particular case, these groups refer to the segments of cars, as described earlier.

The Nested Logit model had been developed to avoid the IIA assumption inherent to the Logit model. Alternatives j that share some common characteristics are grouped into G+1 exhaustive and mutually exclusive sets (nests). The Nested Logit model allows for correlation of consumer preferences between alternatives within the same nest and independence between alternatives belonging to distinct nests.

While products may be grouped based on any criteria, in the case of the car market demand analysis, the groupings I am using correspond to specific segments of cars. This allows for some consumers to like a relatively cheap, compact car that is fuel-fuel efficient, and for others a large one with additional styling and luxury features. Consumers who like Opel Corsa, will also like VW Polo because both cars are small compact cars, offering personal mobility at a relatively low price. A consumer choosing, say between Opel Corsa and Mercedes S-class, will also look at attributes of alternatives within the B segment as well as within the E segment. The Nested Logit is set up to incorporate consumer tastes for similar products[9] into substitution patterns.

Using Berry's (1994) exposition of the Nested Logit case, the utility that consumer i obtains from each product j, belonging to a group g now becomes:

$$u_{ij} = \delta_j + \zeta_{ig} + (1 - \sigma)\varepsilon_{ij} \tag{9}$$

In addition to the previously defined Logit terms δ_j and ε_{ij}, there is now ζ_{ig}, an additional variable that's comming for all products within the group and whose distribution function depends on σ. The latter is bounded, $0 \leq \sigma \leq 1$ and represents the intensity of correlation of consumer tastes. As we continue to assume extreme value distribution for ε_{ij}, $[\zeta_{ig} + (1 - \sigma)\varepsilon_{ij}]$ inherits

[9]Similarity is determined by the way segmentation is made prior to estimation

this distribution. Theoretically, when σ equals unity, consumer tastes for products in the given segment are so highly correlated that they are perceived as perfect substitutes (the last term in equation (9) disappears). At the other extreme, σ equal zero indicates absence of correlation: consumers shopping in this segment do not have a common taste parameter. The way groupings (segments) are made sets the correlation structure prior to estimation.

The probability of choosing a product j belonging to a segment g now becomes a product of marginal and conditional probabilities. It is the probability of choosing product j given that the consumer first picked a specific segment g:

Marginal is the probability of choosing any of the segments is:
$\mathrm{s}_g(\delta,\sigma) = \frac{D_g^{(1-\sigma)}}{\sum_g D_g^{(1-\sigma)}}$

Conditional probability is the market share formula for product j in its segment g:
$\mathrm{s}_{j|g}(\delta,\sigma) = \frac{e^{\frac{\delta_j}{(1-\sigma)}}}{D_g}$, with $D_g \equiv \sum_{j\in J} e^{\frac{\delta_j}{(1-\sigma)}}$.

The resulting market share formula for product j becomes:

$$s_g(\delta,\sigma) = \mathrm{s}_{j|g}(\delta,\sigma)\mathrm{s}_g(\delta,\sigma) = \frac{e^{\frac{\delta_j}{1-\sigma}}}{D_g^{\sigma}\sum_g D_g^{(1-\sigma)}} \tag{10}$$

The aggregation procedure follows the same steps as in the Logit case[10], and the results in the following linear regression to estimate:

$$ln(s_j/s_0) = \beta x_j - \alpha p_j + \sigma ln(s_{j|g}) + \xi_j \tag{11}$$

The only difference compared to the Logit regression is the additional term $\sigma ln(s_{j|g})$, with σ being the parameter representing correlation of tastes between consumers choosing within a given segment and its corresponding covariate $ln(s_{j|g})$, the logarithm of product j's market share within its segment. Henceforth, I will refer to this variable as "segmentation variable". Compared to (6), equation (11) is structured so that mean utility of every product now also depends on correlation of consumer preferences vis-a-vis other products. In the following stages, once product attributes are accounted for, σ, the intesity of consumer tastes, together with α, the price parameter, will be used to explain market shares.

[10]In the Nested Logit case, the share of the outside good is defined as $s_0(\delta,\sigma) = \frac{1}{\sum_g D_g^{(1-\sigma)}}$

5 Estimation of the model

In summary of the above conceptual framework, I proceed to estimate the following two equations:

1) $ln(s_j/s_0) = \beta x_j - \alpha p_j + \xi_j$ for the Logit case, and

2) $ln(s_j/s_0) = \beta x_j - \alpha p_j + \sigma ln(s_{j|g}) + \xi_j$ for the Nested Logit case.

Market shares of each model (relative to the share of the outside good) are regressed on characteristics, prices, and, in the Nested Logit model, also on $ln(s_{j|g})$ (see below).

5.1 The endogeneity problem

Endogeneity of prices ensuing from simultaneity has been recognized in economic litterature for a number of years. Some researchers conjectured potential simultaneity issue for observed characteristics, but I believe that this is not problematic for automobiles.

In the case under stduy, endogeneity of prices stems from the simultaneity and from the omitted variable problems. Simultaneity is the result of market equilibrium: supply determining prices and prices affecting supply. Therefore, when used in the demand equation, price cannot be treated as exogenous. The omitted variable problem arises because researchers can rarely observe all the relevant factors from which consumers draw utility. When consumers value a certain characteristic, concept or product image and producers are aware of this, they charge more for this feature/service. Also, car companies run advertising campaigns that result in modification of consumer price sensitivity for their products. The positive correlation between prices and these unobserved effects leads to an upward bias in the price coefficient. We are in the differentiated goods context and as such competition in the automotive industry is not based on prices only. Companies invest heavily to distinguish themselves from the competition in terms of the reliability of their products, brand appeal, style etc. I can control for omitted variables by using a Fixed Effects regression, assuming that these variables are constant in time but are different among car models. Unfortunately, those unobserved variables that vary from year to year cannot be properly accounted for. As for simultaneity, consistent estimation requires instrumental variables (IV) techniques. IV techniques are also necessary for consistent estimation of $ln(s_{j|g})$ in the Nested Logit case.

If left unaccounted for, endogeneity issues discussed in the previous paragraph can translate into an upward slopping demand, leading to conclusion that the higher the price, the higher the market share.

5.1.1 Fixed Effects

The Fixed Effects method enables econometricians to control for the difference in levels of the unobserved product heterogeneity by tracking each (individual) observation over time, provided that panel data is available. Intuitively, the Fixed Effects method compares each car with itself over time and all cars between themselves. As such, this method is more powerful than the Ordinary Least Squares (OLS), which only does the latter. Operationally, the Fixed Effects procedure creates an average of all explanatory variables over time for each model, then removes this average from each observation in each year t and lastly runs an OLS on the demeaned variables. The potentially interesting yet unfortunate feature of the Fixed Effects method is that individual-specific effects (i.e. averages) cannot be estimated consistently. For the purpose of this paper, obtaining estimates of each car's constants is not a priority - my goal is to eliminate the unobserved heterogeneity.

5.1.2 Instrumental variables

While the ideal solution to the price endogeneity issue would be to instrument the demand equation with cost-side variables such as production factor prices (i.e. costs that affect the price decision and not demand), in a differentiated goods setting, it is very difficult. Automobiles are produced on more than one continent and differ in their composition of inputs and labor production costs. Product-specific data on costs is almost impossible to obtain. I therefore resort to demand-side instruments.

As for behavior of firms, I assume Bertrand-Nash oligopoly competition: firms take competitor characteristics and prices as given. This assumption enables me to recur to the BLP (1995) and Brenkers and Verboven (2006) instrument types. Their instruments are based on the following ideas:

- Each product's own characteristics are predetermined[11], and
- Prices set by firms, as well as the market shares are affected by the degree of closeness of competition faced by producers. In the context of

[11] This is the main identification assumption. Without it, the instruments mentioned are invalid

oligopolistic competition with differentiated goods, each product competes with other products made by the same firm and with competitor firm products from the same segment.

In view of the above propositions, for each product j, characteristics of competing products are exogenous - they are taken as given for each firm's pricing decisions. Also, everything else being equal, the more competing products in the segment, the more elastic the demand curve faced by by all models in this segment.

In the Nested Logit case, $ln(s_{j|g})$, the log of within group share, is endogenous because it is a function of the dependent variable: the larger the market share of product j in its segment, the larger will be its share in the whole market. Consistent estimation requires exogenous variables correlated with the within-group-share variable but not with the dependent variable to serve as instruments. As instruments, I use functions of characteristics of other products in the segment.

The instruments I use, on top of each product's own characteristics are the following:

1. Sums of characteristics of other products produced by the same firm,
2. Sums of characteristics of competitor products,
3. Number of products in the segment (in Nested Logit case).
4. In addition, 2) and 3) also interacted with group dummy variables to account for the degree of closeness among products.

In the BLP (1995) logic, all of these instruments are predetermined with respect to product j and therefore constitute valid candidates for instruments in the differentiated goods demand equation.

5.2 Data selection and summary

5.2.1 Data selection and adjustments

To simplify estimation, I use the most basic version for each model per market per year. The data covers the market extensively and all models are considered except for the outliers: I discard models whose sales are less than 0.015% of market demand. This corresponds to dropping observations whose yearly sales are less than 500 units in Germany or 80 units in Belgium. Potential market size is assumed to be population divided by four - which roughly corresponds to a family of four as a potential car consumer. The dependent variable in my regression, $ln(s_j/s_0)$, is derived in the following manner:

$L=$ total number of households,

$Q=$ sum of all cars sold in the market,

q_j = units of product j sold,

$s_j = (q_j/L)$, $s_0 = (q_0/L)$,
Putting all elements together and applying the ln function, we obtain

$$ln\left[\frac{s_j}{s_0}\right] = ln\left[\frac{(q_j/L)}{(L-Q)/L}\right] = ln\left[\frac{q_j}{(L-Q)}\right] \tag{12}$$

Taking Germany as example, for a population of roughly 80 million, the potential market size is 20 million. The actual car market demand (size) is about 3.3 million, which yields the size of the outside good, s_0, equal to 16.3 million cars. All prices and revenue figures are converted to their 2000 values using the CPI and the GDP deflator, respectively.

5.2.2 Summary Statistics

Table 3 presents sales-weighted summary statistics across countries as well as segment shares in each market, both in terms of numbers of models and percentage of sales. An average car sold in the seven markets studied, consumes about 9 liters of petrol per 100 km, takes about 8.5 seconds to reach the speed of 100km/hr from a standstill position. It is about 147 cm high, 172 cm wide, and 415 cm long. Germans buy the longest and the widest cars, the Italians buy the shortest and the narrowest. Compared to other countries, an average car bought in Italy is the most fuel-efficient and the slowest-to-accelerate. The car price relative to a consumer's average annual income varies across the markets from 0.4 in the UK to 0.8 in Spain. There are between 220 and 230 different models in each country, offered by over forty different brands.

The A segment represents about 6% of the total market demand. Its size varies significantly from country to country, with Spain consuming only 2% of this entry type of vehicles and Italy 14%. In view of this statistic, we can recall the legendary Fiat Seicento and Fiat Panda. Competition in

Table 3: SUMMARY STATISTICS (sales weighted)

1. Means and standard deviations (below)

VARIABLE	GERMANY	FRANCE	UNITED KINGDOM	ITALY	SPAIN	NETHERLANDS	BELGIUM
sales (units sold)	3,193,376	2,003,752	2,358,235	2,238,773	1,484,824	486,567	462,016
	167,693	106,150	137,658	83,244	74,471	47,640	20,538
fuel inefficiency (l/100km)	9.86	9.10	9.37	8.74	9.33	9.11	9.17
	2.2	1.8	1.9	2.3	2.3	1.8	2.0
length (cm)	421	401	415	396	415	415	418
	41.7	34.2	36.2	40.1	31.9	38.4	35.7
width (cm)	172	169	171	168	171	171	172
	6.9	6.8	7.5	8.5	7.5	18.9	7.7
height (cm)	148	146	146	148	146	148	148
	11.5	11.3	11.5	10.9	11.1	20.5	12.6
acceleration (seconds)	8.27	8.89	8.70	9.67	8.69	9.10	8.98
	10,671	10,671	6,472	6,652	6,685	8,600	7,595
price (real)	18,227	14,577	11,727	12,494	13,096	16,903	14,466
	0.40	0.40	0.25	0.29	0.36	0.29	0.29
price/income (real)	0.72	0.58	0.39	0.59	0.80	0.64	0.57
	0.40	0.23	0.26	0.29	0.37	0.30	0.29
number of models	231	217	226	221	219	221	217
number of brands	47	42	48	43	42	41	41
ttl number of obs.	1,385	1,151	1,357	1,203	1,201	1,241	1,244

2. Number of models per segment and market share of the segment (below)

VARIABLE	GERMANY	FRANCE	UNITED KINGDOM	ITALY	SPAIN	NETHERLANDS	BELGIUM
A segment	16	13	15	14	11	17	16
	6%	5%	4%	14%	2%	10%	3%
B segment	37	38	36	34	27	34	41
	19%	36%	29%	40%	29%	24%	28%
C segment	24	24	25	21	23	23	24
	22%	21%	23%	15%	32%	22%	24%
C+ segment	10	8	9	8	10	9	7
	4%	2%	3%	3%	3%	3%	3%
D segment	20	15	20	16	18	20	17
	12%	12%	12%	4%	12%	14%	11%
D+ segment	12	11	12	9	10	10	11
	12%	3%	6%	5%	4%	5%	7%
E/E+ segment	21	14	14	14	15	17	18
	9%	2%	3%	2%	2%	3%	4%
Mini MPV segment	18	20	15	14	13	15	20
	7%	12%	7%	6%	7%	10%	8%
MPV segment	24	16	15	15	15	18	16
	3%	3%	2%	1%	1%	3%	4%
Sports segment	27	21	27	16	12	13	16
	4%	1%	3%	1%	1%	1%	1%
SUV segment	23	12	24	21	26	15	16
	2%	2%	3%	3%	4%	2%	2%
SUV+ segment	25	25	28	21	22	20	27
	3%	2%	6%	5%	4%	4%	6%

this segment is quite fierce: about 15 car brands fight for 6% of this type of product demand.

The B segment covers about 30% of the total market demand with models like Peugeot 206, Opel Corsa and Toyota Yaris, just to name a few. Cars from this segment are the most popular in Europe. Out of all cars offered in each market, Germans care the least for the B-segment cars (segment size equals about 19%), while the French and the Italians love them (the respective segment shares are 36% and 40%). About 35 firms aim to cover B segment demand. Second largest is the C segment with popular models like the VW Golf, Toyota Corolla and Renault Megane. The average size of this segment is 23%, varying from 15% in Italy to 32% in Spain. Just over 20 brands serve this segment.

The C+ segment, mainly consisting of the smallest models of premium brands such as BMW and Audi, represents a stable 3% of the total European market. Only 7-10 brands serve this segment.

Third largest is the D segment with models such as Ford Mondeo and Renault Laguna, offered by about 18 brands to cover 11% of demand. D segment cars are the most popular in the Netherlands and least in the small-car-minded Italy.

About 11 brands focus on the D+ segment in Europe. This segment is dominated by German models such as BMW 3-series, Audi A4 and Mercedes C-class. No wonder why this segment covers 12% of market demand in Germany. While on average, 6% of total European sales go to this segment, the French shop the least in the D+ segment. A similar comment can be made for the E/E+ segments whose offer stretches to top luxury vehicles such as BMW 7 series and Mercedes E-class. In Germany, it represents about 9% of total vehicles sold, whereas in France only 2%, against the European average of 3%. Also, in Germany, a larger number of brands serve this popular segment: 21 versus the European average of 16.

As for the Mini-MPV segment, represented by small minivans such as VW Touran and Renault Scenic, in most markets it accounts for about 6% of the total car demand. Small MPV's, offered by 13 - 20 brands in Europe, are the most popular in France (12%) and in the Netherlands (10%). The larger MPV's such as VW Sharan, Ford Galaxy or Renault Espace are not as popular in Europe. Although a slightly larger number of brands (15 - 24) offers these vehicles, it represents 1-4% of total car demand in the markets of this study.

The Sports segment is the smallest of all, representing about 1% of market demand for five out of seven countries. Two countries differ from their neighbors in their love for fast cars: Germany and the UK with their respective Sports segment shares of 4% and 3%. While on average about 16 brands serve this niche car segment, in these two markets, 27 brands are present.

The remaining two segments, offer SUVs (Sports Utility Vehicles). The small and cheaper SUVs represent about 3% of demand in each of the markets. The hit models in this segment are Toyota RAV4 and Nissan X-Trail. The number of brands serving this segment varies from country to country, from 12 in the small-French-car-minded France to 26 in the SUV-loving Spain. The luxury SUV segment is slightly more popular with about 4% of market share in most countries. Compared to their neighbors, the British and Belgians buy a larger portion of the luxury SUVs like BMW X-5 or Lexus RX. About 25 brands aim at this niche segment.

5.3 Regression analysis

5.3.1 Car characteristics chosen for my regressions

Given the data available to me, I decided to make a rather parsimonious regression and run it for each market individually. Product attributes chose are acceleration (and acceleration squared), fuel consumption, length, width, and height. The reason is threefold:

a) Extensive market coverage - the data I use covers almost the total market for new cars and includes a much larger number of models than used in previous research. While experimenting which product characteristics to include as covariates in my regressions, I learned that in order to analyse the whole market, it is important to include product characteristics that are *relevant to each and every segment.* Take GPS navigation, for instance. Even though this item has been fitted on many new vehicles during 2002-2005, it has been mainly fitted on the pricier cars or in segments above C. Therefore, while this item may have been very important in some segments, it is irrelevant for segments A and B, which combined represent about 1/3 of total market demand. A regression analysis which intends to make a statement on the market as a whole may (correctly) indicate that GPS navigation variable is not statistically significant. The same explanation applies to the remaining product characteristics mentioned in Section 3.2.

b) I would like to make my results comparable to the papers that analyzed the same kind of data, only at earlier periods or in other markets.

c) A priori, there is some sense in taking only a few key items, around which consumers shopping in their respective segments would orientate his choice.

While my regression is parsimonious, my instruments are extensive. It turns out that all the remaining specifications mentioned in Section 3.2, as well as their functions described above, serve better as instruments. In a way, they also approximate for supply-side variables, which remedies endogeneity issues mentioned earlier.

The car characteristics I am using are acceleration, measured in the number of seconds it takes a car to reach 100km/hr speed from a stand-still position, fuel consumption, measured in the number of liters of fuel a car consumes per 100km, width, length and height in centimeters. I also add year dummies (not reported in regression tables) to account for year-specific macroeconomic fluctuations affecting the overall car demand. Because acceleration measure is expressed in a standard way, I expect the coefficient to have a negative sign: the longer it takes for the car to gain speed, the less appeal it will have to a typical consumer. Idem for fuel consumption and price. As for the size characteristics, the direction of the sign is apriori uncertain: is it big that's beautiful or is it small that's beatiful?

5.3.2 Estimates of car characteristics

Tables 4-10 present regression results for each country under analysis. In this subsection I will comment on product attributes only. In the next subsection I will discuss the segmentation parameter.

ACCELERATION: this is the most robust-across-specifications variable in all countries. As expected, the signs of the coefficients are negative, and in most cases, the p-value is 0.00 (see bolded coefficient estimates in the regression tables). In the expectation of acceleration to be valued by consumer in non-linear way, I also included the squared term of acceleration. This means that all else being equal, consumers dislike slow cars, but only up to a point. In other words, a typical consumer would not be willing to pay for a car that performs too well. Experience showed that this is a very convenient way to represent consumer valuation of a car's power. I also experimented with regressions that use horsepower, engine size (cc), torque, number of cylinders as well as with different combinations of these characteristics. Parameter estimates were either insignificant, had unexpected signs or proved not robust to the different demand specifications. This is somewhat dissonant to my experience in the automotive industry, where everyone talks in terms of engine size and horsepower. BLP (1995), whose dataset did not contain acceleration, used horsepower divided by vehicle weight as proxy for acceleration. For all countries except Spain, my estimates of acceleration compare to their Logit estimate of -0.12 for horsepower/weight. BLP (1995) did not estimate the squared term of the latter. My estimates of acceleration squared are positive and clearly significant, both across countries and specifications, suggesting that indeed, there is a stable non-linear pattern in consumers' valuation of acceleration.

FUEL CONSUMPTION: coefficient estimates are typically small, insignificant and in the zero range. The reason for this insignificance could lie in the fact that I have always selected the base version of each model, which did not display a significant variation in fuel consumption over time. The base versions selected already had the lowest fuel consumption in the model range. In the 6 years that I consider, the lowest levels of fuel consumption for petrol cars did not improve significantly.

SIZE variables: in general, they are insignificant, especially in the case of length. Width tends to matter, but to a very small extent. Out of the three indicators of size, it is height that most often shows up as significant. This suggests that European consumers don't think in terms of "big is beautiful".

PRICES: My price estimates decrease in absolute value as I move from the Logit to and through the Nested Logit specification, suggesting that once accounted for segmentation, the price matters to a lesser extent. In the UK and Italy, the flexible Nested Logit regression produced a very small and positive price coefficient[12]. The UK, Italy and Spain show relatively smaller price coefficients, whereas Germany, France, Netherlands and Belgium indicate more realistic figures. I suspect that one reason why UK appears to be a relatively price-insensitive market is because of the discount culture specific to this country. The "price distortion" can vary from manufacturer to manufacturer and may not be fully captured when retail prices are used to indicate prices actually paid. Recall that Fixed Effects accounts only for the time-invariant aspects and does not capture things like limited-time incentives.

[12]In Italy, the coefficient estimate is not significantly different from zero.

Table 4: REGRESSION RESULTS - GERMANY

Parameter estimates, standard errors on line below

	LOGIT with Fixed Effects	NESTED LOGIT with Fixed Effects restricted	NESTED LOGIT with Fixed Effects flexible
acceleration	-0.24	**-0.24**	**-0.15**
	0.09	0.05	0.04
acceleration	0.01	**0.01**	**0.01**
	0.003*	0.00	0.00
fuel consumption	0.09	0.01	0.01
	0.051*	0.01	0.00
length	-0.0001	0.003	-0.002
	0.00	0.002**	.002***
width	0.02	**0.02**	0.002
	0.01	0.01	0.005***
height	0.06	**0.02**	**0.01**
	0.01	0.00	0.00
price/income (α)	-2.61	**-2.24**	**-0.63**
	0.35	0.20	0.15
constant	-25.97	**-17.25**	**-11.79**
	1.71	1.15	0.92
segmentation (σ all)	-	**0.54**	-
		0.03	
A segment	-	-	**0.99**
			0.046
B segment	-	-	**0.63**
			0.029
C segment	-	-	**0.61**
			0.024
C+ segment	-	-	**0.90**
			0.037
D segment	-	-	**0.94**
			0.067
D+ segment	-	-	**0.82**
			0.030
E segment	-	-	**0.79**
			0.058
Mini MPV segment	-	-	**0.76**
			0.038
MPV segment	-	-	**0.83**
			0.052
Sports segment	-	-	**1.03**
			0.030
SUV segment	-	-	**0.58**
			0.055
SUV+ segment	-	-	**0.90**
			0.056
Number of observations	1385	1385	1385

*Bolded if p-value 0.00, no asterics if significant at 5% or less, * if significant at 10%, ** if significant at 20%, ***if insignificant*

Table 5: REGRESSION RESULTS - FRANCE

Parameter estimates, standard errors on line below

	LOGIT with Fixed Effects	NESTED LOGIT with Fixed Effects restricted	NESTED LOGIT with Fixed Effects flexible
acceleration	**-0.29**	**-0.25**	**-0.19**
	0.093	0.065	0.052
acceleration	0.01	**0.01**	**0.01**
	0.004**	0.002	0.002***
fuel consumption	0.052	-0.015	0.009
	0.002	0.015	0.001
length	-0.0021	-0.0010	**-0.0004**
	0.0014**	.00105***	0.0010***
width	0.009	0.004	0.001
	0.007***	0.0049***	.0037***
height	0.002	0.001	**0.001**
	0.0029***	0.002***	0.002**
price/income (α)	**-1.27**	**-1.30**	**-0.48**
	0.272	0.170	0.117
constant	**-13.89**	**-11.38**	**-11.37**
	0.986	0.718	0.556
segmentation (σ all)	-	**0.51**	-
		0.038	
A segment	-	-	**0.99**
			0.069
B segment	-	-	**0.47**
			0.035
C segment	-	-	**0.62**
			0.031
C+ segment	-	-	**1.11**
			0.068
D segment	-	-	**0.85**
			0.114
D+ segment	-	-	**1.12**
			0.087
E segment	-	-	**1.09**
			0.147
Mini MPV segment	-	-	**0.66**
			0.044
MPV segment	-	-	**0.90**
			0.058
Sports segment	-	-	**1.17**
			0.039
SUV segment	-	-	**0.79**
			0.082
SUV+ segment	-	-	**0.84**
Number of observations	1151	1151	1151

*Bolded if p-value 0.00, no asterics if significant at 5% or less, * if significant at 10%, ** if significant at 20%, ***if insignificant*

Table 6: REGRESSION RESULTS - UNITED KINGDOM

Parameter estimates, standard errors on line below

	LOGIT with Fixed Effects	NESTED LOGIT with Fixed Effects restricted	NESTED LOGIT with Fixed Effects flexible
acceleration	**-0.25**	**-0.16**	-0.06
	0.064	0.033	0.024
acceleration	**0.007**	**0.005**	0.002
	0.002	0.001	0.001
fuel consumption	0.023	-0.039	-0.009
	0.029***	0.015	0.011***
length	0.003	-0.002	0.0021
	0.0025***	0.001**	0.001*
width	**0.010**	**0.010**	0.0002
	0.006**	0.003	0.002***
height	-0.011	**-0.017**	-0.0053
	0.005	0.003	0.002
price/income (α)	-0.42	-0.04	0.12
	0.221*	0.079***	0.051
constant	**-14.70**	**-9.20**	**-11.53**
	1.256	0.671	0.561
segmentation (σ all)	-	**0.79**	-
		0.03	
A segment	-	-	**1.00**
			0.029
B segment	-	-	**0.53**
			0.031
C segment	-	-	**0.71**
			0.022
C+ segment	-	-	**1.14**
			0.056
D segment	-	-	**1.06**
			0.078
D+ segment	-	-	**1.02**
			0.039
E segment	-	-	**0.74**
			0.082
Mini MPV segment	-	-	**1.13**
			0.053
MPV segment	-	-	**1.48**
			0.050
Sports segment	-	-	**1.19**
			0.029
SUV segment	-	-	**0.83**
			0.054
SUV+ segment	-	-	**1.00**
			0.052
Number of observations	1357	1357	1357

*Bolded if p-value 0.00, no asterics if significant at 5% or less, * if significant at 10%, ** if significant at 20%, ***if insignificant*

Table 7: REGRESSION RESULTS - ITALY

Parameter estimates, standard errors on line below

	LOGIT with Fixed Effects	NESTED LOGIT with Fixed Effects restricted	NESTED LOGIT with Fixed Effects flexible
acceleration	-0.52	**-0.42**	**-0.20**
	0.10	0.11	0.05
acceleration	**0.016**	**0.014**	**0.006**
	0.00	0.00	0.002
fuel consumption	-0.0001	-0.0004	-0.002
	0.010***	0.011***	0.005***
length	-0.006	-0.003	-0.0042
	0.003*	0.003***	0.003***
width	0.006	0.024	-0.0043
	0.014***	0.013*	0.006***
height	**0.038**	**0.023**	**0.022**
	0.01	0.01	0.00
price/income (α)	-0.14	**-1.69**	0.14
	0.673***	0.36	0.113***
constant	**-16.06**	**-15.99**	**-12.28**
	2.71	2.43	1.36
segmentation (σ all)	-	**0.41**	-
		0.101	
A segment	-	-	**0.58**
			0.083
B segment	-	-	**0.26**
			0.049
C segment	-	-	**0.64**
			0.041
C+ segment	-	-	**0.93**
			0.072
D segment	-	-	**1.00**
			0.268
D+ segment	-	-	**0.69**
			0.108
E segment	-	-	**0.86**
			0.239
Mini MPV segment	-	-	**0.97**
			0.097
MPV segment	-	-	**1.35**
			0.108
Sports segment	-	-	**1.18**
			0.056
SUV segment	-	-	**0.54**
			0.109
SUV+ segment	-	-	**0.87**
			0.139
Number of observations	1203	1203	1203

*Bolded if p-value 0.00, no asterics if significant at 5% or less, * if significant at 10%, ** if significant at 20%, ***if insignificant*

Table 8: REGRESSION RESULTS - SPAIN

Parameter estimates, standard errors on line below

	LOGIT with Fixed Effects	NESTED LOGIT with Fixed Effects restricted	NESTED LOGIT with Fixed Effects flexible
acceleration	**-0.56**	**-0.49**	**-0.20**
	0.108	0 .074	0.045
acceleration	**0.017**	**0.016**	**0.008**
	0.005	0.003	0.002
fuel consumption	0.004	-0.021	-0.012
	0.833***	0.015**	0.009**
length	0.002	-0.001	**-0.006**
	0.0028***	0.002***	0.002
width	**0.008**	**0.008**	0.002
	0.006	0.002	0.001**
height	**0.048**	**0.021**	0.009
	0.007	0.006	0.004
price/income (α)	**-0.72**	**-0.47**	-0.07
	0.147	0.087	0.033
constant	**-20.53**	**-14.38**	**-9.88**
	1.625798	1.30	0.866
segmentation (σ all)	-	**0.41**	-
		0.043	
A segment	-	-	**1.06**
			0.058
B segment	-	-	**0.53**
			0.037
C segment	-	-	**0.62**
			0.025
C+ segment	-	-	**1.12**
			0.065
D segment	-	-	**0.93**
			0.108
D+ segment	-	-	**0.75**
			0.061
E segment	-	-	**0.96**
			0.115
Mini MPV segment	-	-	**0.80**
			0.058
MPV segment	-	-	**1.16**
			0.069
Sports segment	-	-	**1.44**
			0.041
SUV segment	-	-	**1.01**
			0.061
SUV+ segment	-	-	**0.82**
			0.067
Number of observations	1201	1201	1201

*Bolded if p-value 0.00, no asterics if significant at 5% or less, * if significant at 10%, ** if significant at 20%, ***if insignificant*

Table 9: REGRESSION RESULTS - NETHERLANDS

Parameter estimates, standard errors on line below

	LOGIT witg Fixed Effects	NESTED LOGIT witg Fixed Effects restricted	NESTED LOGIT witg Fixed Effects flexible
acceleration	-0.17	**-0.17**	-0.07
	0.010*	0.063	0.0381*
acceleration	0.002	0.005	0.002
	.0041***	0.0026*	0.0015**
fuel consumption	0.119	0.042	0.017
	0.035	0.022*	0.0128**
length	0.004	-0.002	**-0.0055**
	0.003**	0.002***	0.00
width	-0.011	-0.006	-0.0013
	0.004	0.003	0.0016***
height	**0.011**	0.005	0.0010
	0.004	0.003	0.002***
price/income (α)	**-1.40**	**-1.01**	**-0.21**
	0.271	0.132	0.05
constant	**-15.76**	**-10.85**	**-9.41**
	1.27	0.84	0.68
segmentation (σ all)	-	**0.55**	-
		0.032	
A segment	-	-	**0.98**
			0.043
B segment	-	-	**0.57**
			0.028
C segment	-	-	**0.70**
			0.022
C+ segment	-	-	**1.14**
			0.045
D segment	-	-	**0.99**
			0.063
D+ segment	-	-	**1.01**
			0.043
E segment	-	-	**1.00**
			0.080
Mini MPV segment	-	-	**0.77**
			0.035
MPV segment	-	-	**0.94**
			0.054
Sports segment	-	-	**1.40**
			0.033
SUV segment	-	-	**0.54**
			0.045
SUV+ segment	-	-	**0.96**
			0.076
Number of observations	1241	1241	1241

*Bolded if p-value 0.00, no asterics if significant at 5% or less, * if significant at 10%, ** if significant at 20%, ***if insignificant*

Table 10: REGRESSION RESULTS - BELGIUM

Parameter estimates, standard errors on line below

	LOGIT with Fixed Effects	NESTED LOGIT with Fixed Effects restricted	NESTED LOGIT with Fixed Effects flexible
acceleration	**-0.17**	**-0.14**	**-0.10**
	0.092	0.060	0.036
acceleration	**0.000**	**0.002**	0.003
	0.0039 ***	0.0025***	0.002
fuel consumption	0.004	-0.014	0.005
	0.0253***	0.0165***	0.009***
length	0.004	0.003	0.0002
	.0025*	0.0016*	0.001***
width	**0.002**	**0.001**	-0.0005
	0.003***	0.002***	0.001***
height	**0.049**	**0.018**	-0.0004
	0.007	0.005	0.003
price/income (α)	**-1.06**	**-0.63**	**-0.16**
	0.236	0.507	0.055
constant	**-22.5**	**-15.5**	**-11.0**
	1.430	1.024	0.749
segmentation (σ all)	-	**0.51**	-
		0.036	
A segment	-	-	**1.11**
			0.042
B segment	-	-	**0.62**
			0.026
C segment	-	-	**0.75**
			0.023
C+ segment	-	-	**1.18**
			0.051
D segment	-	-	**0.98**
			0.083
D+ segment	-	-	**0.95**
			0.034
E segment	-	-	**1.05**
			0.099
Mini MPV segment	-	-	**0.94**
			0.035
MPV segment	-	-	**1.03**
			0.047
Sports segment	-	-	**1.54**
			0.035
SUV segment	-	-	**0.57**
			0.053
SUV+ segment	-	-	**0.94**
			0.069
Number of observations	1244	1244	1244

*Bolded if p-value 0.00, no asterics if significant at 5% or less, * if significant at 10%, ** if significant at 20%, ***if insignificant*

5.3.3 Interpretation of sigma coefficients

The estimates of segmentation parameters σ indicate the extent to which unobserved characteristics are correlated for products within a nest.

In the context of the car market, σ's estimates equal unity indicate perfect correlation of consumer preferences for models within the segment: these products are perfect substitutes. σ's close to zero indicate no correlation - a result that supports the Logit model or signals that segmentation structure requires revision. Empirical expectation is to obtain σ's somewhat between the two extremes, even though other studies found parameter estimates above unity. I estimate two specifications of the Nested Logit model - one with a unique σ for all segments (the restricted case) and one where I let σ's vary by segment (the flexible case). In the former, I impose the same correlation for all segments, whereas in the latter, I allow for correlation to differ from segment to segment. As presented in regression tables, the restricted specification shows that segmentation is meaningful: consumers do shop "in segments".

In every single country, for both the restricted and the flexible specification, all estimates of segmentation parameters are statistically very significant: they all have p-values of 0.00. In the restricted specification, parameter estimates range from 0.4 (in Italy and Spain) to 0.79 (in the UK). According to Train (2003), this means that, for all possible values of characteristics that I considered, consumers display utility-maximizing behavior, overall.

In the flexible case, σ_j estimates range from values of about 0.5 to values greater than one. For segments where estimates that fall within the unit interval, the model I estimated is consistent with utility maximization for all possible values of characteristics and prices. This means that consumers shopping in these segments will always react when product competitiveness is improved either by enhancements of product characteristics or by a lower price. Regression tables show σ_j estimates above one, typically for the sports and luxury segments. This is also the case for the A (smallest car) segment in Belgium and Spain. Again, according to the litterature, estimates above unity indicate that, that in these segments, consumer behavior is consistent with utility maximization, but only for a certain range of values of product attributes. This implies for instance, that when a price of any of product in any of these segments is lowered, the probability that this product is purchased will actually drop. Equivalently, this can mean that when a product in this segment is enhanced, people will not necessarily continue buying it. What is then the link between an A-segment Ford Ka and the Sports-segment Ferrari? One possible scenario could be that if Ka's size significantly increases, say to the size of a C-segment car while other product attributes stay the same, people will not pay more or may actually refrain from buying it. They may find Ka's B-segment sister Fiesta more appealing because Fiesta will probably offer the acceleration that is more suitable for a larger car. Consumers will turn down a huge and powerless Ka. Similarly, imagine that a 300.000 Ferrari drops its price to say, 20.000. At this price, Ferrari may no longer be perceived as a status symbol.

It is worthwhile to put a note of caution on the interpretation of segmentation coefficients. Let us examine some specific cases. A σ_j coefficient equal to one means that on average, products in this segment are perfect substitutes. Take the Netherlands for instance. The segmentation parameter estimate for the A segment is 0.984, falling in the 95% confidence interval of [0.900, 1.068]. This is clearly the case of $\sigma_j \approx 1$, which will result in an-almost-perfect substitution pattern. Compare this to the Sports segment estimate of 1.404, which falls in the 95% confidence interval of [1.339, 1.469]. In this case, $\sigma_j > 1$, without a doubt. Dutch sports car shoppers are clearly far away from a rationally-minded consumer. But what about the following segments:

C+ with estimate of 1.141, belonging to the interval [1.052, 1.229], D+with estimate of 1.005, belonging to the interval [0.919, 1.090], E/E+ with estimate of 0.998, belonging to the interval [0.841, 1.153] and SUV+ with estimate of 0.963, belonging to the interval [0.815, 1.112]?

These extremes will imply a totally different substitution pattern: if I round these sigma estimates to 1, elasticities will show a perfect substitution pattern between the products in any one of these segments, ceteris paribus. These estimates will translate into negative elasticities, large in absolute value. If in my elasticities calculation I use the upper confidence interval figures, my elasticities will be positive, implying that consumers in this segment will like car price increases. On the other hand, this finding can be interpreted simply as evidence of large variation of consumer tastes for the cars in each of the segments discussed above. On one side of the "taste" spectrum, there are people who perceive some differences between luxury models, then there are those who view, say BMW and Mercedes products, are equivalent and at the other end there are those who will not accept if their BMW changes too much.

5.4 Substitution patterns

5.4.1 Definitions and formulas

In the discrete choice context, own price elasticities of demand measure by how much the probability that a given product is chosen changes in response to its price change. Similarly, cross price elasticities measure the change in choice probability provoked a competitor price change. In the context of my model, where choice probabilities are equated to market shares, the elasticities are simply $\frac{\partial s_j}{\partial p_j}\frac{p_j}{s_j}$ and $\frac{\partial s_j}{\partial p_k}\frac{p_j}{s_j}$, normalized for the variable's units, which enables their interpretation as a percentage change in the probability of being chosen (i.e. market share) associated with a one percent change in price.

For the Logit case, elasticities depend on market shares only. The formulas I used for elasticity calculations are the following[13]:

- Own elasticity: $\epsilon_{own} = -\alpha p_j(1 - s_j)$
- Cross elasticity: $\epsilon_{cross} = \alpha p_l s_l$

For the Nested Logit case[14], s_{jg}, designating models market share in its segment, and s_j, designating the share of the model in the total market, come into play. The formulas are the following:

- Own elasticity:

  $\epsilon_{own} = \frac{-\alpha}{1-\sigma} p_j(1 - \sigma s_{jg} - (1 - \sigma)s_j)$

- Cross elasticity vs. same segment:

  $\epsilon_{cross_{ss}} = \frac{\alpha}{1-\sigma} p_k(\sigma s_{k|g} + (1 - \sigma)s_k)$

  (cross elasticity for models of the same segment, model $k \neq j$), and

- Cross elasticity vs. different segment:

  $\epsilon_{cross_{ds}} = \alpha p_l s_l$

[13]See Appendix for the derivation of Logit and Nested Logit elasticities

[14]Restricted specification

(cross elasticity for models of different segments, for σ common to all segments).

Everything else being held constant, when σ=1, elasticities are high in absolute value. When σ=0, Nested Logit elasticities become Logit. Resulting from the independence assumption, in the Logit case, consumers can respond to a price increase of say, Toyota Yaris by choosing any other model in the market, be it a Renault Clio or Audi A4: switching is done with the same probability for models that have the same market share. This is an unrealistic implication and unfortunate implication of the Logit's IIA property. In the nested case, the probability of switching to a model in the same segment is higher than to a model in a different segment (same segment cross-elasticity versus different segment cross-elasticity). Everything else being equal, when the price of the B segment Yaris increases, consumers will substitute to the same segment Renault Clio with a higher probability than to say, the D+ segment Audi A4. In the restricted case, all segments share the same σ. Once accounted for the σ, only market shares come into play. In the flexible case, each segment has a different σ and the differences in σ_g will also matter.

5.4.2 Substitution patterns: Logit versus restricted Nested Logit specification

Elasticities I estimate are point elasticities evaluated at each model's price level in year 2005. Table 11 presents simple averages of elasticities for each segment. The pattern shows elasticities increasing in absolute value as one moves to the more expensive segments.

Likewise, the elasticities computed for the flexible case and shown in Table 6 and in the graphs that follow, indicate a pattern that increases in absolute value with price, but to a lesser extent than in the restricted specification. This result is consistent with findings of Brenkers and Verboven (2002).

Brenkers and Verboven experimented with a specification where consumers have a common price parameter (2002) and a specification where consumers are heterogeneous with respect to income (2006). The elasticities patterns they found were as follows:

Brenkers and Verboven	restricted specification (A)	flexible specification (B)
common price parameter (2002)	elasticities (absolute value) increase with price	same as (A 2002), flatter pattern
consumer-specific price parameter (2006)	elasticities (absolute value) increase with price	elasticities (absolute value) drop with price

According to the authors, the pattern where elasticities increase with price could be owing to the fact that consumers actually do not have a common price parameter - and therefore my assumption of common price parameter is inappropriate. This is perhaps what is seriously lacking in my specification and data.

5.4.3 Substitution patterns: flexible Nested Logit specification

Let us take three neighboring countries and see to what extent they differ in substitution patterns. Table 12 presents means of own price elasticities resulting from the flexible specification for year 2005.

Consumers in A-segment appear to be very price sensitive, especially in Germany. In Belgium, A-segment consumer behavior appears partially inconsistent with utility maximization - it's elasticities have not been computed. The segmentation parameter estimate was 1.11, falling in the 95% confidence interval with a lower bound of 1.02. If the lower bound value had been used, the resulting elasticity would resemble those of Germany and the Netherlands. As mentioned earlier, using the parameter estimate as it comes out from the regression analysis can be dangerous and misleading. In all three countries, the Sports segment displays the snob good property. Lower confidence bounds of the Sports segmentation parameter estimates were significantly above 1, resulting in positive elasticities and the above conclusion. Idem for the Dutch C+, D+, and E/E+ segments. A similar comment can be made for Belgium's C+ and E/E+ segments. Overall, my estimates show that shows that Sports and luxury segment cars display snob goods effects. Surprisingly, luxury SUV's do not follow the same pattern as luxury sedans. In all three markets, this appears to be a price-sensitive segment. Let us now have a second look at Tables 1 and 3: both indicate that Germans distinguish themselves from the other Europeans analyzed in my study in terms of their demand for luxury cars. Table 1 shows that in Germany, premium brands hold 24% of the market. Table 3 shows that a much higher portion of the "+" segment cars is bought in Germany compared to other countries. In Table 12,elasticities for the German luxury segments are quite high compared to those in other segments. I interpret this as follows: in Germany, the fiercest competition is on the luxury cars. Consumers want luxury cars, manufacturers offer a variety of luxury and this is reflected in price elasticities.

To place the above average elasticities in perspective, I include in the Appendix plots of own elasticities for Germany, based on calculations from the flexible model. These graphs give us an idea on which elasticities cars actually faced in 2005, given the price levels they charged for their base versions. The 2005 car prices relative to average income are plotted on the X-

Table 11: Cross country comparison of PRICE ELASTICITIES for year 2005, based on restricted specification

		GERMANY	FRANCE	UNITED KINGDOM	ITALY	SPAIN	NETHERLANDS	BELGIUM
		239 obs.	217 obs.	240 obs.	221 obs.	219 obs.	221 obs.	217 obs.
OWN	**ALL - logit**	**-2.29**	**-0.99**	**-0.22**	**-0.13**	**-0.82**	**-1.20**	**-0.79**
CROSS		0.000	0.000	0.000	0.000	0.000	0.000	0.000
OWN	**ALL - nested logit**	**-4.22**	**-2.01**	**-0.11**	**-2.57**	**-0.89**	**-1.91**	**-0.92**
SAME SEGMENT		-0.10	-0.06	-0.004	-0.05	-0.02	-0.06	-0.02
DIFFERENT SEGMENT		0.000	0.000	0.000	0.000	0.000	0.000	0.000
OWN	**A segment**	**-1.66**	**-0.83**	**-0.04**	**-1.08**	**-0.37**	**-0.76**	**-0.40**
SAME SEGMENT		-0.05	-0.03	-0.002	-0.03	-0.01	-0.02	-0.01
DIFFERENT SEGMENT		0.00	0.00	0.00	0.00	0.00	0.00	0.00
NUMBER OF MODELS		17	13	15	15	12	16	16
AVG. PRICE 2005		9,092	9,095	6,733	9,266	8,755	10,604	10,421
OWN	**B segment**	**-2.09**	**-1.09**	**-0.05**	**-1.37**	**-0.49**	**-0.96**	**-0.51**
SAME SEGMENT		-0.03	-0.01	0.00	-0.01	-0.01	-0.01	-0.01
DIFFERENT SEGMENT		0.00	0.00	0.00	0.00	0.00	0.00	0.00
NUMBER OF MODELS		38	38	36	42	32	39	41
AVG. PRICE 2005		11,874	11,836	8,465	11,608	11,445	13,251	11,670
OWN	**C segment**	**-2.51**	**-1.38**	**-0.07**	**-1.66**	**-0.58**	**-1.23**	**-0.61**
SAME SEGMENT		-0.07	-0.03	-0.002	-0.03	-0.01	-0.03	-0.01
DIFFERENT SEGMENT		0.00	0.00	0.00	0.00	0.00	0.00	0.00
NUMBER OF MODELS		21	23	25	23	24	21	23
AVG. PRICE 2005		14,613	15,163	10,414	14,124	12,943	17,223	14,219
OWN	**C+ segment**	**-2.99**	**-1.72**	**-0.08**	**-2.20**	**-0.70**	**-1.38**	**-0.65**
SAME SEGMENT		-0.23	-0.12	-0.01	-0.11	-0.03	-0.10	-0.06
DIFFERENT SEGMENT		0.00	0.00	0.00	0.00	0.00	0.00	0.00
NUMBER OF MODELS		8	8	9	8	10	9	7
AVG. PRICE 2005		17,318	18,935	12,788	19,355	20,393	20,162	16,356
OWN	**D segment**	**-3.47**	**-1.78**	**-0.09**	**-2.29**	**-0.73**	**-1.79**	**-0.78**
SAME SEGMENT		-0.11	-0.06	-0.004	-0.06	-0.02	-0.05	-0.02
DIFFERENT SEGMENT		0.00	0.00	0.00	0.00	0.00	0.00	0.00
NUMBER OF MODELS		18	15	20	17	17	21	17
AVG. PRICE 2005		20,516	20,237	14,629	19,622	16,010	25,085	18,743
OWN	**D+ segment**	**-4.54**	**-2.16**	**-0.11**	**-2.24**	**-0.89**	**-2.00**	**-0.92**
SAME SEGMENT		-0.20	-0.11	-0.01	-0.06	-0.04	-0.13	-0.05
DIFFERENT SEGMENT		0.00	0.00	0.00	0.00	0.00	0.00	0.00
NUMBER OF MODELS		11	10	12	10	9	9	10
AVG. PRICE 2005		27,278	25,308	18,757	19,265	19,367	29,027	21,563
OWN	**E/E+ segment**	**-7.18**	**-3.45**	**-0.26**	**-3.71**	**-1.43**	**-3.38**	**-1.55**
SAME SEGMENT		-0.22	-0.15	-0.01	-0.09	-0.03	-0.12	-0.04
DIFFERENT SEGMENT		0.00	0.00	0.00	0.00	0.00	0.00	0.00
NUMBER OF MODELS		16	11	14	11	14	14	14
AVG. PRICE 2005		47,103	40,969	41,108	31,751	27,728	47,841	41,992
OWN	**Mini MPV segment**	**-3.07**	**-1.62**	**-0.08**	**-1.96**	**-0.70**	**-1.46**	**-0.76**
SAME SEGMENT		-0.10	-0.05	-0.005	-0.05	-0.02	-0.05	-0.02
DIFFERENT SEGMENT		0.00	0.00	0.00	0.00	0.00	0.00	0.00
NUMBER OF MODELS		17	10	15	16	16	19	18
AVG. PRICE 2005		17,535	16,537	12,627	16,834	14,722	20,536	14,722
OWN	**MPV segment**	**-4.47**	**-2.46**	**-0.11**	**-2.89**	**-1.04**	**-2.19**	**-1.05**
SAME SEGMENT		-0.14	-0.08	-0.01	-0.09	-0.03	-0.08	-0.04
DIFFERENT SEGMENT		0.00	0.00	0.00	0.00	0.00	0.00	0.00
NUMBER OF MODELS		17	19	15	14	14	16	15
AVG. PRICE 2005		25,543	17,550	18,585	24,894	24,017	30,873	25,382
OWN	**Sports segment**	**-6.33**	**-2.76**	**-0.17**	**-4.30**	**-1.29**	**-2.83**	**-1.32**
SAME SEGMENT		-0.13	-0.06	-0.004	-0.08	-0.02	-0.09	-0.05
DIFFERENT SEGMENT		0.00	0.00	0.00	0.00	0.00	0.00	0.00
NUMBER OF MODELS		22	15	27	18	17	15	13
AVG. PRICE 2005		35,606	26,963	26,038	36,651	25,095	39,817	33,195
OWN	**SUV segment**	**-4.25**	**-2.14**	**-0.12**	**-2.90**	**-0.96**	**-2.14**	**-0.98**
SAME SEGMENT		-0.08	-0.05	-0.004	-0.05	-0.01	-0.06	-0.03
DIFFERENT SEGMENT		0.00	0.00	0.00	0.00	0.00	0.00	0.00
NUMBER OF MODELS		26	19	24	21	27	18	16
AVG. PRICE 2005		25,035	30,822	30,127	24,747	22,192	30,021	23,869
OWN	**SUV+ segment**	**-7.54**	**-3.56**	**-0.19**	**-4.81**	**-1.55**	**-3.66**	**-1.71**
SAME SEGMENT		-0.10	-0.07	-0.005	-0.06	-0.02	-0.07	-0.02
DIFFERENT SEGMENT		0.00	0.00	0.00	0.00	0.00	0.00	0.00
NUMBER OF MODELS		28	23	28	26	27	24	27
AVG. PRICE 2005		39,073	39,710	30,127	40,695	31,432	50,891	37,233

Table 12: Price elasticities, based on flexible specification (2005)

SEGMENT	GERMANY	NETHERLANDS	BELGIUM
A	-18.1	-4.3	-
B	-0.7	-0.2	-0.30
C	-0.9	-0.4	-0.47
C+	-3.5	-	-
D	-7.1	-12.6	-14.47
D+	-3.2	-	-4.18
E/E+	-5.1	-	-
Mini MPV	-1.6	-0.6	-2.92
MPV	-3.3	-3.3	-
Sports	-	-	-
SUV	-1.4	-0.4	-0.59
SUV+	-8.9	-9.1	-7.47

Segment means of own price elasticities are calculated using σ exactly as they appear in regression results tables; otherwise elasticities not computed

Table 13: Comparison with previous studies

Data	**Verboven and Brenkers (2006)**	**My study**
Markets	D, F, UK, IT, B	D, F, UK, IT, B, NL, ES
Years covered	1970-1999	2000-2005
average number of models/market	+/-80	+/-220
number of segments	5	12
car characteristics considered	horsepower, size, fuel inefficiency	acceleration, size, fuel inefficiency
income distribution	taken into account	unaccounted for
Findings		
size	width and height have positive signs	same
horsepower and acceleration	expected sign only for flexible specification	always expected sign
fuel inefficiency	expected sign, significant	expected signs in half of the cases
price coefficient (restricted)	-1.422	-0.04 to -2.24
price coefficient (flexible)	-4.764	-0.48 to +0.12
segmentation parameter estimates (restricted)	0.41	0.4 to 0.7
segmentation parameter estimates (flexible)	-0.05 to 0.54	0.3 to 1.5
elasticities range (restricted)	-3.6 -5.6	-0.1 to -7.5
elasticities range (flexible)	-4.5 -10.9	positive to -18

axis and elasticities are on the Y-axis and each point on the graph corresponds to a model in a given segment.

5.5 Comparison of my analysis to previous work

As mentioned earlier, the demand specification I chose was inspired by the work of Brenkers and Verboven. In Table 13, I summarize the key similarities and the key differences.

6 Main Conclusions

6.1 Findings

1. Overall, the Logit model produces smaller elasticities versus the Nested Logit model.

2. Segmentation parameters are very important. Once segments are accounted for, price becomes less relevant and coefficients tend toward zero.

3. With restricted Nested Logit, my findings are in line with previous work: the data shows that segmentation matters and its extent varies from segment to segment.

4. Estimates from the flexible case indicate that segmentation parameters differ from segment to segment. They are higher in the flexible case, which suggests that imposing the same correlation structure on all segments may be misleading when looking at the whole market. However, if one is interested in the most commonly purchased vehicles, the restricted specification produces segmentation parameter values that are very similar to those of the flexible specification. This is the case for small-and-medium-sized cars. B and C segments combined cover between 43 and 61% of a country's total car market. Segmentation parameter estimates for these segments range from 0.4 to 0.7 in the restricted specification and from 0.3 to 0.8 in the flexible specification.

5. Typically the most expensive and the cheapest segments indicate either strong product substitution patterns or partial inconsistency with utility maximization assumed in the conceptual work.In all countries except Germany, luxury cars display snob goods properties. Sports segment shows this in all countries.

6.2 Limitations of my paper and scope for future research

1. I use the base versions of each model in order to make a simplifying assumption that is consistent across markets. Perhaps these are not the most representative grades. In reality, one can conceive a scenario in which firms do not price their products in a symmetric way - they intentionally distort the pricing of product line-up in order to create a low sticker price to attract consumers. Through my work experience

I learned that price walks between the different versions of the same model are not always symmetric across firms.

2. In addition, the base versions always correspond to the gasoline versions of the model. Market reality is that for the countries I analyze diesel versions account for between 15 and 70% of the sales volume. A higher order Nested Logit model can be left for future research to study the sensitivity of my results versus a model with fuel type sub-nests and/or grade sub-nests.

3. I do not account for consumer heterogeneity with respect to income. The idea of consumers having a common price parameter is not realistic: one would expect that the price parameter a is lower for the well-off consumers. This impacts price elasticities. This is perhaps the main shortcoming of this paper.

4. My segmentation is based on standard industry practice and experience. While it offers some insight from the price-setting point of view, one can argue that a portion of substitution is brand-specific. Future research can address this issue with controls for brand phenomena.

References

[1] Berry, S., 1994, "Estimating Discrete-Choice Models of Product Differentiation", Rand Journal of Economics, 25, 242-262.

[2] Berry, S., J. Levinsohn, and A. Pakes, 1995, "Automobile Prices in Market Equilibrium", Econometrica, 63, 841-890.

[3] Brenkens, R. and Verboven, F., 2002, Liberalizing a distribution system: the European car market, C.E.P.R. Discussion paper.

[4] Brenkers, R. and Verboven, F., 2006, "Liberalizing a Distribution System: the European Car Market," Journal of the European Economic Association, 4(1), 216-251.

[5] Goldberg, P. K. and F. Verboven, 2001, "The Evolution of Price Dispersion in the European Car Market," Review of Economic Studies, 68(4), 811-848.

[6] Train, Kenneth E., 2003. Discrete Choice Methods with Simulation, Cambridge University Press.

[7] McFadden, D. L., 1978, "Modeling the Choice of Residential Location" in Karqist (Ed.), Spatial Interaction Theory and Planning Models, North Holland.

[8] Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," Journal of Political Economy, 82, 34-55.

[9] Empirical Industrial Organization Summer School, organized by the Economics Network for Competition and Regulation in July 2005 at Universiteit van Amsterdam. Instructors: Prof. F.Verboven and Prof. P. Van Cayseele.

A Derivation of elasticities

A.1 Logit

A.1.1 Logit: own elasticity

$s_j = \frac{e^{\delta_j}}{\sum_{k=0}^{J} e^{\delta k}} = e^{\delta_j}\left[\sum_{k=0}^{J} e^{\delta_k}\right]^{-1},$

Label $e^{\delta_j} = A^j$ and $\left[\sum_{k=0}^{J} e^{\delta_k}\right]^{-1} = B$

Derive A^j and B with respect to p_j:

$\frac{\partial A_j}{\partial p_j} = -\alpha * e_j^{\delta}$

$\frac{\partial B}{\partial p_j} = \left[\sum_{k=0}^{J} e^{\delta k}\right]^{-2} * \alpha * e^{\delta_j}.$

Then

$B * \frac{\partial A^j}{\partial p_j} = -\alpha * e^{\delta_j} * \left[\sum_{k=0}^{J} e^{\delta_k}\right]^{-1} = -\alpha * A^j * B$

and

$A^j * \frac{\partial B}{\partial p_j} = e^{\delta_j}\left[\sum_{k=0}^{J} e^{\delta_k}\right]^{-2} * \alpha * e^{\delta_j} = \alpha * B^2 * A * A$

Combining the above and rearranging, obtain

$\frac{\partial s_j}{\partial p_j} = -\alpha * A^j * B + \alpha * A * A * B^2$

$=$ -$\alpha * e^{\delta_j}\left[\sum_{k=0}^{J} e^{\delta_k}\right]^{-1} + \left[\left[\sum_{k=0}^{J} e^{\delta_k}\right]^{-1}\right]^2 * \alpha * e^{\delta_j} * e^{\delta_j}$

Noting that $A * B = s_j$ and $A * A * B^2 = s_j^2$,

$\frac{\partial s_j}{\partial p_j} = -\alpha * s_j * \alpha * s_j^2 = -\alpha * s_j * (1 - s_j)$.

Then apply the formula for own price elasticity to obtain the Logit result:

$\epsilon_{own_L} = \frac{\partial s_j}{\partial p_j}\frac{p_j}{s_j} = -\alpha * s_j * (1 - s_j)\frac{p_j}{s_j} = -\alpha * p_j(1 - s_j)$.

A.1.2 Logit: cross elasticity

s$_k = e^{\delta_l}\left[\sum_{k=0}^{J} e^{\delta_k}\right]^{-1}$,

Label $e^{\delta_l} = A^l$ and $\left[\sum_{k=0}^{J} e^{\delta_k}\right]^{-1} = B$,
where for $k = l$: $e^{\delta_l} = e^{x_j\beta - \alpha p_l + \xi_l}$

so $s_k = A^l * B$,

$\frac{\partial A^j}{\partial p_l} = 0$

$\frac{\partial B}{\partial p_l} = \left[\sum_{k=0}^{J} e^{\delta_k}\right]^{-2} * \alpha * e^{\delta_l} = \alpha e^{\delta_l}\left[\sum_{k=0}^{J} e^{\delta_k}\right]^{-2} = \alpha * A^l * B^2$

Recognizing that $s_j = A^j * B$ and that $s_l = A^l * B$, combine and rearrange the above to obtain:

$\frac{\partial s_j}{\partial p_l} = 0 * B + \alpha * A^l * B^2 * A^j$, or equivalently:

$= 0 * \left[\sum_{k=0}^{J} e^{\delta_k}\right]^{-1} + \alpha * e^{\delta_l} * \left[\sum_{k=0}^{J} e^{\delta_k}\right]^{-2} * e^{\delta_j}$

$= \alpha * e^{\delta_l} \left[\sum_{k=0}^{J} e^{\delta_k} \right]^{-2} * e^{\delta_j}$

$= \alpha * s_j * s_l.$

Then apply the formula for cross price elasticity to obtain the Logit result:

$\epsilon_{cross_L} = \frac{\partial s_j}{\partial p_l} \frac{p_l}{s_j} = \alpha * s_l * s_j * \frac{p_l}{s_j} = \alpha * s_l * p_l$

A.2 Nested Logit

A.2.1 Nested Logit: basics

Recall that the nested formula for market share is the product of probability of choosing product j within nest g and probability of choosing segment g:

$s_j = s_{j|g} * s_g$, where

$s_{j|g} = \frac{e^{\frac{\delta_j}{(1-\sigma)}}}{D_g}$, $s_g = \frac{{D_g}^{(1-\sigma)}}{\sum_g D_g^{(1-\sigma)}}$, in which

$D_g = \sum_{j \in g} e^{\frac{\delta_j}{(1-\sigma)}}$ and $\delta_j = x_j \beta - \alpha p_j + \xi_j$.

A.2.2 Nested Logit: own elasticity

Label $e^{\frac{\delta_j}{(1-\sigma)}} = A^j$, $D_g^{-1} = B$, $D_g^{(1-\sigma)} = C$ and $\left[\sum_g D_g^{(1-\sigma)} \right]^{-1} = D$.

For clarity, substitute the labels into s_j:

$s_j = e^{\frac{\delta_j}{(1-\sigma)}} * D_g^{-1} * D_g^{(1-\sigma)} * \left[\sum_g D_g^{(1-\sigma)} \right]^{-1} = A_j * B * C * D$

Derive s_j with respect to p_j follow this pattern:

if $x = A(p) * B(p) * C(p) * D(p)$, then

$$\frac{\partial x}{\partial p} = \frac{\partial A(p)}{\partial p} * BCD + \frac{\partial B(p)}{\partial p} * ACD + \frac{\partial C(p)}{\partial p} * ABD + \frac{\partial D(p)}{\partial p} * ABC$$

Now derive each of the terms one by one:

1) As $\frac{\partial A(p)}{\partial p_j} = \frac{-1}{1-\sigma} * \alpha * \frac{\delta_j}{e^{(1-\sigma)}} = \frac{-\alpha}{1-\sigma} * A^j$,

$$\frac{\partial A(p)}{\partial p_j} * BCD = \frac{-\alpha}{1-\sigma} * A^j * B * C * D$$

$$= \frac{-\alpha}{1-\sigma} * e^{\frac{\delta_j}{(1-\sigma)}} * D_g^{-1} * D_g^{(1-\sigma)} * \left[\sum_g D_g^{(1-\sigma)}\right]^{-1} = \frac{-\alpha}{1-\sigma} * s_j.$$

2) As $\frac{\partial B(p)}{\partial p_j} = \left[\sum_{j\in g}^{J} e^{\frac{\delta_j}{(1-\sigma)}}\right]^{-2} * \left[\frac{\alpha}{(1-\sigma)} * \frac{\delta_j}{e^{(1-\sigma)}}\right] = \frac{\left[\frac{\alpha}{(1-\sigma)} * \frac{\delta_j}{e^{(1-\sigma)}}\right]}{D_g^2} = \frac{\frac{\alpha}{(1-\sigma)} * A^j}{B^2}$,

then

$$\frac{\partial B(p)}{\partial p_j} * ACD = \frac{\left[\frac{\alpha}{(1-\sigma)} * \frac{\delta_j}{e^{(1-\sigma)}}\right]}{D_g^2} * \frac{\delta_j}{e^{(1-\sigma)}} * D_g^{(1-\sigma)} * \left[\sum_g D_g^{(1-\sigma)}\right]^{-1} = \frac{\alpha}{1-\sigma} * s_{j|g} * s_j.$$

3) As $\frac{\partial C(p)}{\partial p_j} = (1-\sigma) D_g^{-\sigma} * \left[\frac{-\alpha}{1-\sigma}\right] * \frac{\delta_j}{e^{(1-\sigma)}} = (1-\sigma) D_g^{-\sigma} * \left[\frac{-\alpha}{1-\sigma}\right] * A^j$,

$$\frac{\partial C(p)}{\partial p_j} * ABD = -\alpha * D_g^{-\sigma} * \frac{\delta_j}{e^{(1-\sigma)}} * \frac{\delta_j}{e^{(1-\sigma)}} * D_g^{-1} * \left[\sum_g D_g^{(1-\sigma)}\right] * D_g * D_g^{-1}$$

$$= -\alpha * A^j * B * C * D * s_{j|g}$$

$$= -\alpha * s_j * s_{j|g}$$

4) As $\frac{\partial D(p)}{\partial p_j} = -\left[\sum_g D_g^{(1-\sigma)}\right]^{-2} * (1-\sigma) * D_g^{-\sigma} * \left[\frac{-\alpha}{1-\sigma}\right] * e^{\frac{\delta_j}{1-\sigma}}$

$= \frac{-\alpha}{1-\sigma} * (1-\sigma) * D_g^{-\sigma} * A * D^2,$

$\frac{\partial D(p)}{\partial p_j} * ABC = \left[\sum_g D_g^{(1-\sigma)}\right]^{-2} * (1-\sigma) * D_g^{-\sigma} * \left[\frac{\alpha}{1-\sigma}\right] * e^{\frac{\delta_j}{1-\sigma}} * e^{\frac{\delta_j}{1-\sigma}} * D_g^{-1} * D_g^{(1-\sigma)} * D_g * D_g^{-1}$

$= \alpha * s_j^2.$

Finally, combine results 1)-4) to obtain $\frac{\partial s_j}{\partial p_j}$:

$\frac{\partial s_j}{\partial p_j} = \frac{-\alpha}{1-\sigma} * s_j + \frac{\alpha}{1-\sigma} * s_{j|g} * s_j - \alpha * s_j * s_{j|g} + \alpha * s_j^2$

$= -\alpha * s_j \left[\frac{1}{1-\sigma} - \frac{s_{j|g}}{1-\sigma} + s_{j|g} - s_j\right]$

$= -\alpha * s_j \left[\frac{1}{1-\sigma} + \left[\frac{1-\sigma}{1-\sigma} - \frac{1}{1-\sigma}\right] * s_{j|g} - s_j\right]$

$= -\alpha * s_j \left[\frac{1}{1-\sigma} - \frac{\sigma}{1-\sigma} * s_{j|g} - s_j\right]$

Then apply the formula for own price elasticity to obtain the Nested Logit result:

$\epsilon_{own_{NL}} = \frac{\partial s_j}{\partial p_j} \frac{p_j}{s_j} = -\alpha * s_j \left[\frac{1}{1-\sigma} - \frac{\sigma}{1-\sigma} * s_{j|g} - s_j\right] * \frac{p_j}{s_j}$

$= -\alpha * p_j \left[\frac{1}{1-\sigma} - \frac{\sigma}{1-\sigma} * s_{j|g} - s_j\right]$

$= \frac{-\alpha}{1-\sigma} * p_j (1 - \sigma * s_{j|g} - s_j * (1-\sigma)).$

A.2.3 Nested Logit: cross elasticity - same segment

(k = model in segment g). Label:

$$e^{\frac{\delta_j}{(1-\sigma)}} = A^j,\ e^{\frac{\delta_k}{(1-\sigma)}} = A^k,\ D_g^{-1} = B,\ D_g^{(1-\sigma)} = C,\ \left[\sum_g D_g^{(1-\sigma)}\right]^{-1} = D.$$

Now derive A^j, B, C and D with respect to p_k:

$\frac{\partial A^j}{\partial p_k} = 0$

$\frac{\partial B}{\partial p_k} = D_g^{-2} * \frac{\alpha}{1-\sigma} * e^{\frac{\delta_k}{1-\sigma}} = \frac{\alpha}{1-\sigma} * A^k * B^2$

$\frac{\partial C}{\partial p_k} = (1-\sigma)D_g^{-\sigma} * \frac{-\alpha}{1-\sigma} * e^{\frac{\delta_k}{1-\sigma}} * D_g * D_g^{-1} = -\alpha * A^k * B * C$

$\frac{\partial D}{\partial p_k} = \left[\sum D_g^{(1-\sigma)}\right]^2 * (1-\sigma) * D_g^{-\sigma} * \frac{\alpha}{1-\sigma} * e^{\frac{\delta_k}{1-\sigma}} * D_g^{-1} * D_g = \alpha * A^k * B * C * D^2$

Knowing that: $A^jBCD = s_j$, $A^kBCD = s_k$, $A^kB = s_{j|g}$ and $CD = s_g$, derive s_j with respect to p_k and plug in the results calculated above:

$\frac{\partial s_j}{\partial p_k} = B^2 * A^k \frac{\alpha}{1-\sigma} * A^jCD - \alpha A^kBC * A^jBD + \alpha A^kBCD^2 * A^jBC$

$= \frac{\alpha}{1-\sigma} * s_j * s_{k|g} - \alpha * s_j * s_{k|g} + \alpha * s_k * s_j$

$= \frac{\alpha}{1-\sigma} * s_j * s_{k|g} - \alpha * s_j * s_{k|g} + \alpha * s_j * s_k$

$= \frac{-\alpha\sigma}{1-\sigma} * s_j * s_{k|g} + \alpha * s_j * s_k$

$= \frac{\alpha}{1-\sigma} * s_j * \left[\sigma * s_{k|g} + (1-\sigma) * s_k\right]$

Plug in the above into the elasticity formula for Nested Logit:

$$\epsilon_{cross_{NLss}} = \frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} = \frac{\alpha}{1-\sigma} * p_k \left[\sigma * s_{k|g} + (1-\sigma)s_k\right]$$

A.2.4 Nested Logit: cross elasticity - different segment

(l =model in segment f,$f \neq g$)

$\frac{\partial A^j}{\partial p_l} = 0$

$\frac{\partial B}{\partial p_l} = 0$

$\frac{\partial C}{\partial p_l} = 0$

$\frac{\partial D}{\partial p_l} = -\left[\sum_g D_g^{(1-\sigma)}\right]^{-2} * (1-\sigma) * D_f^{-\sigma} * \frac{-\alpha}{1-\sigma} * e^{\frac{\delta_k}{1-\sigma}} * D_g * D_g^{-1}$

Again, set:

$e^{\frac{\delta_l}{1-\sigma}} = A^l$, $D_f^{-1} = B^l$, $D_f * D_f^{-\sigma} = C^l$, and replace

$= D^2 * (1-\sigma) * C^l * \frac{\alpha}{(1-\sigma)} * A^l * B^l$

$= \alpha * A^l * B^l * C^l * D^2$ (for σ common to all segments).

Combining the four results above:

$\frac{\partial s_j}{\partial p_l} = \alpha * A^l * B^l * C^l * D^2 * A^j * B * C = \alpha * s_l * s_j$

Plug in all of the above results into the elasticity formula:

$\epsilon_{cross_{NL_{ds}}} = \frac{\partial s_j}{\partial p_l}\frac{p_l}{s_j} = \alpha * s_l * s_j * \frac{p_l}{s_j} = \alpha * s_l * p_l$

B Plots of Nested Logit elasticities (flexible specification), GERMANY

Figure 1. A segment: in 2005, most cars faced elasticities strong and ranging from 13 to 18.

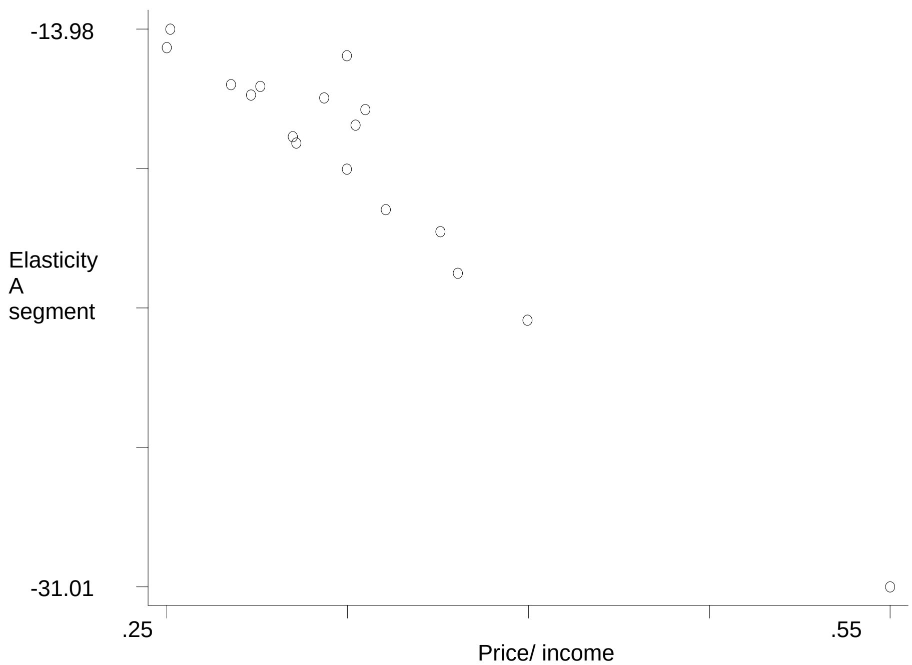


Figure 2. B segment: in 2005, elasticities varied between 0.6 and 0.9.

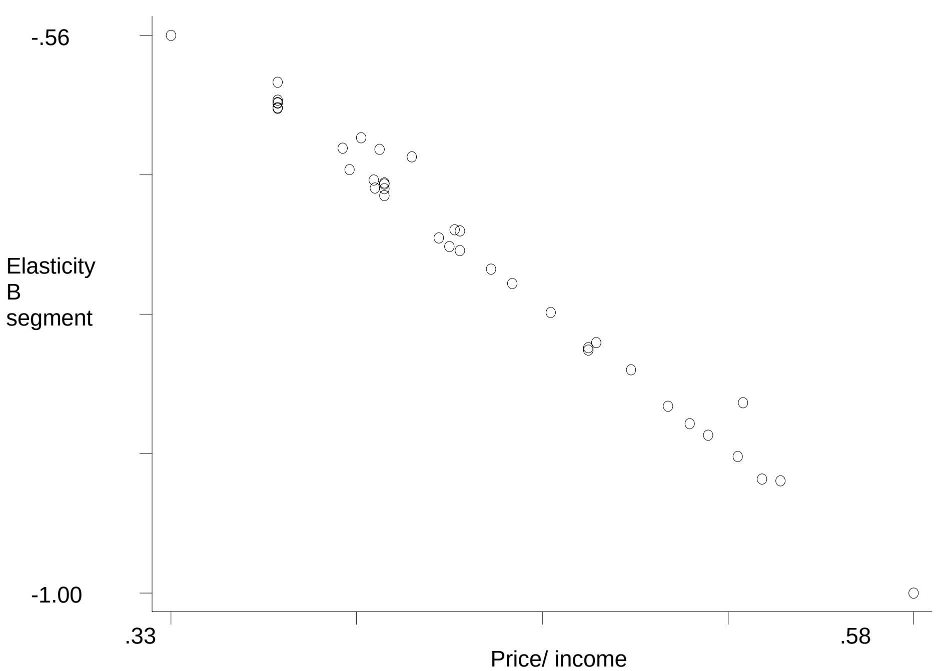


Figure 3. C segment: similarily to the B segment, in 2005, elasticities clustered in the very narrow interval (0.7, 1).

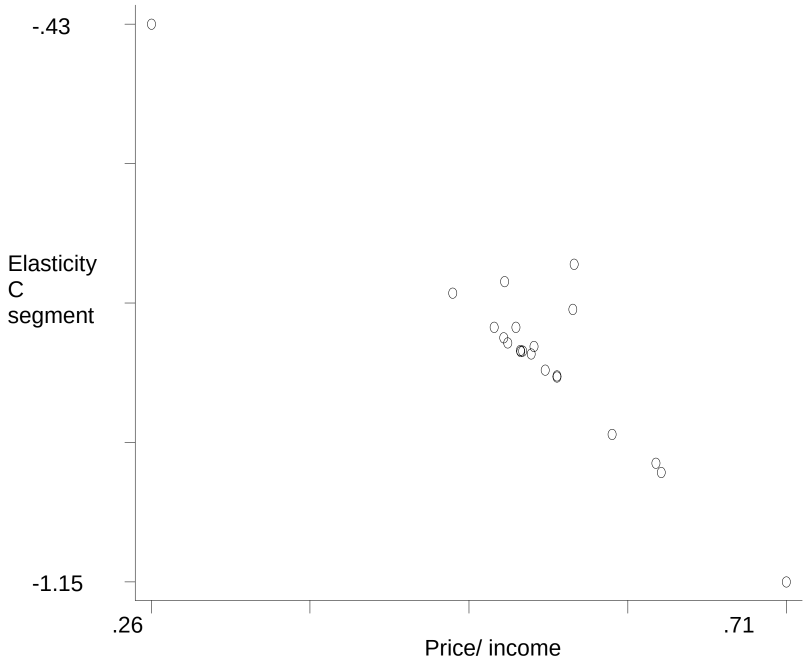
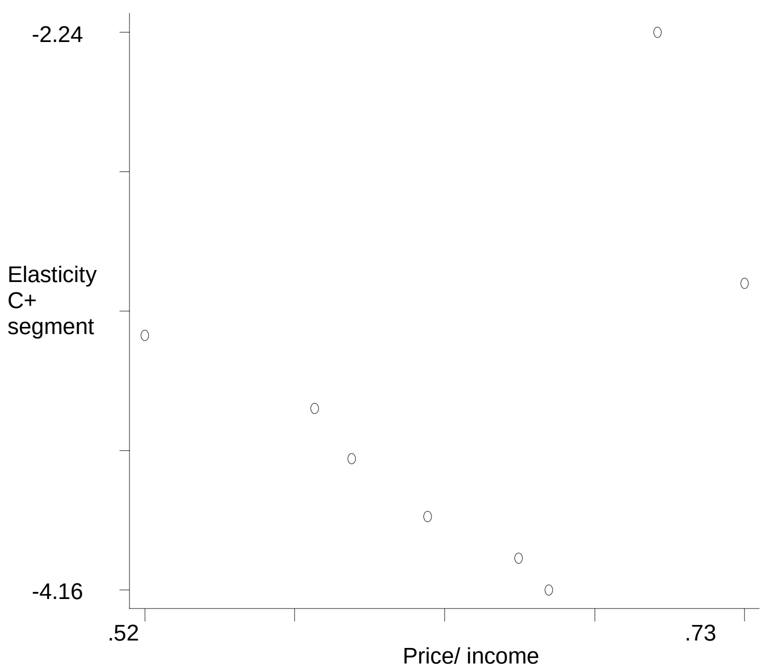
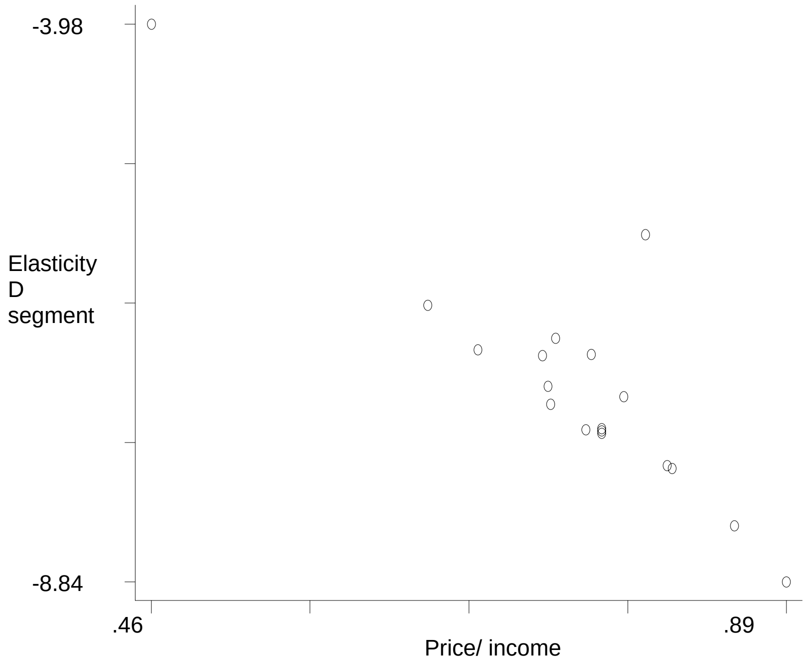


Figure 4. C+ segment : in 2005, most elasticities were in the range of 3.0 to 4.2



<u>Figure 5. D segment</u>: in 2005, most cars in this segment faced elasticities between 6.0 and 8.8.



<u>Figure 6. D+ segment</u>: in 2005, most elasticities were in the range 2.0- 3.5.

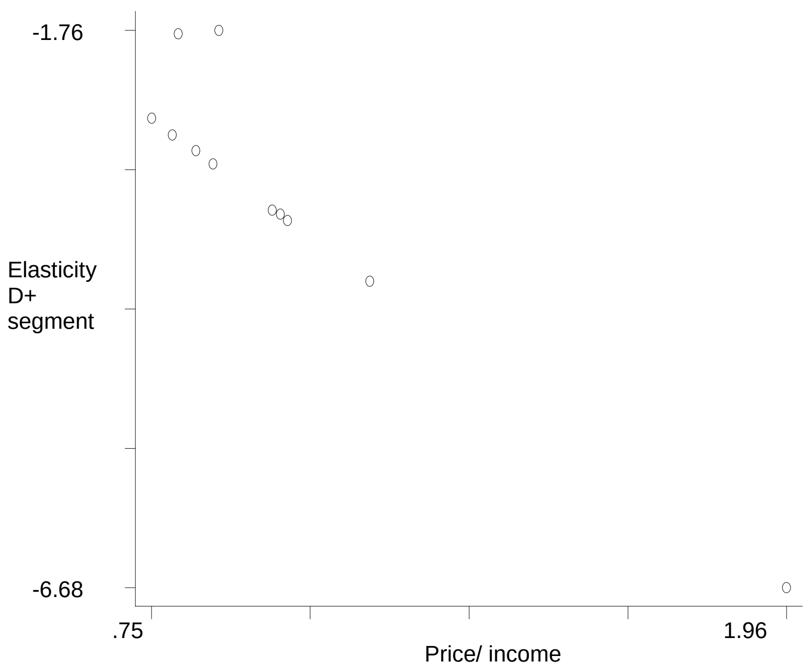


Figure 7. E/E+ segment: while the spread of elasticities in this segment is large, in 2005, most elasticities cluster around 2.0- 5.0.

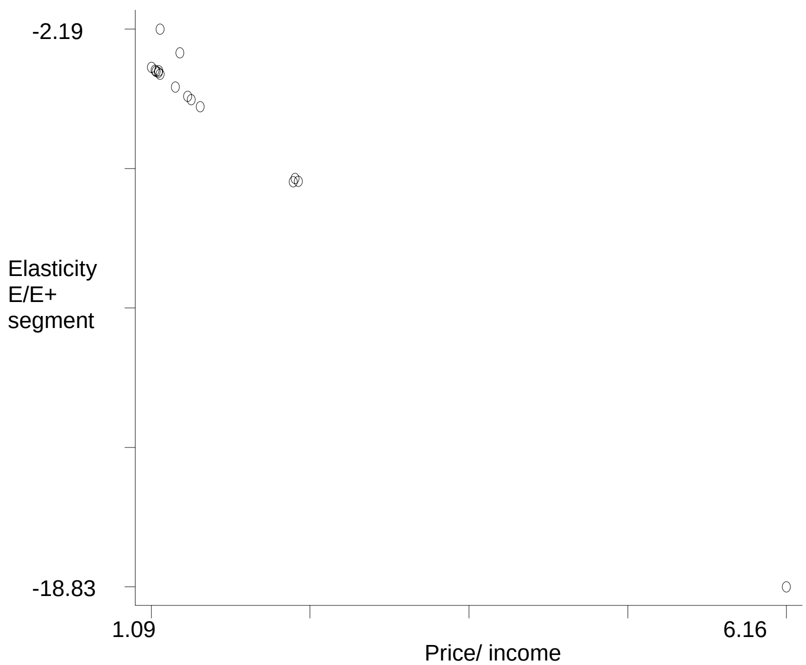


Figure 8. MiniMPV segment: in 2005, models sold in this segment faced elasticities between 1.2 and 2.0.

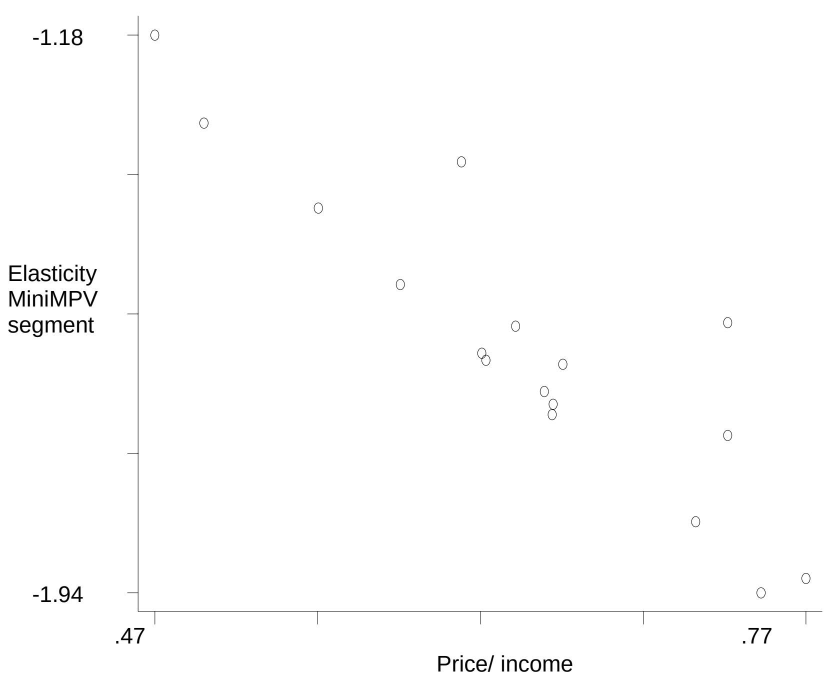


Figure 9. MPV segment: most cars sold in this segment faced elasticities of about 3.0

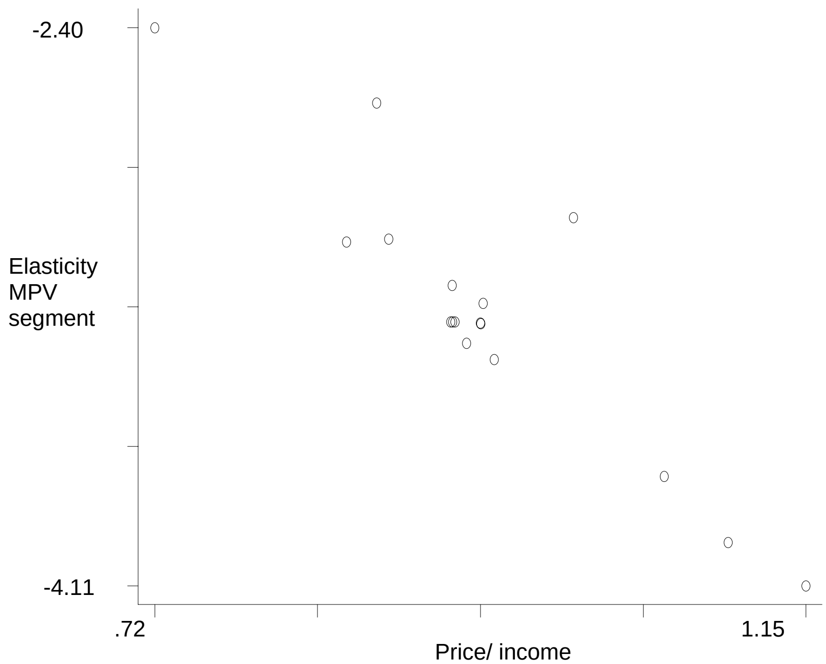
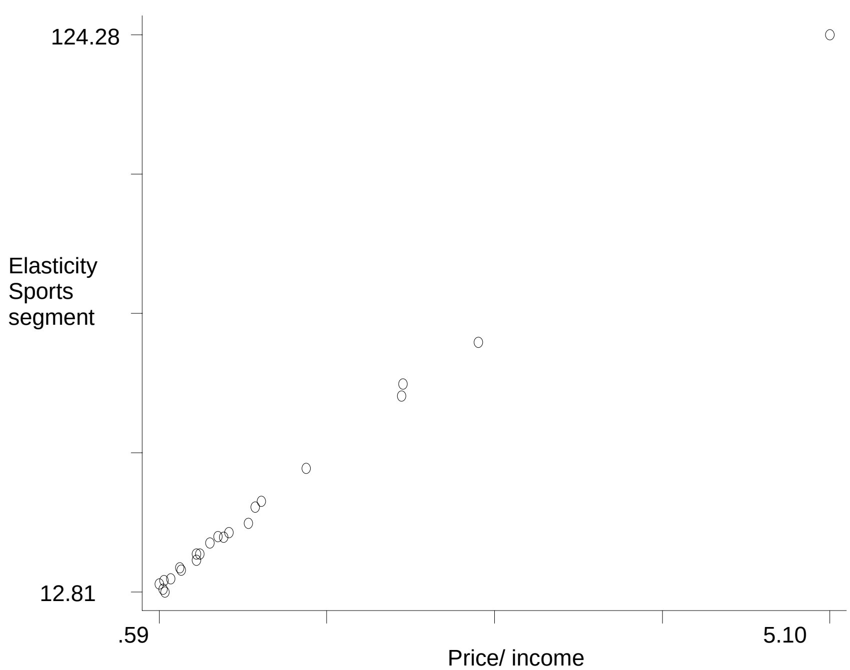
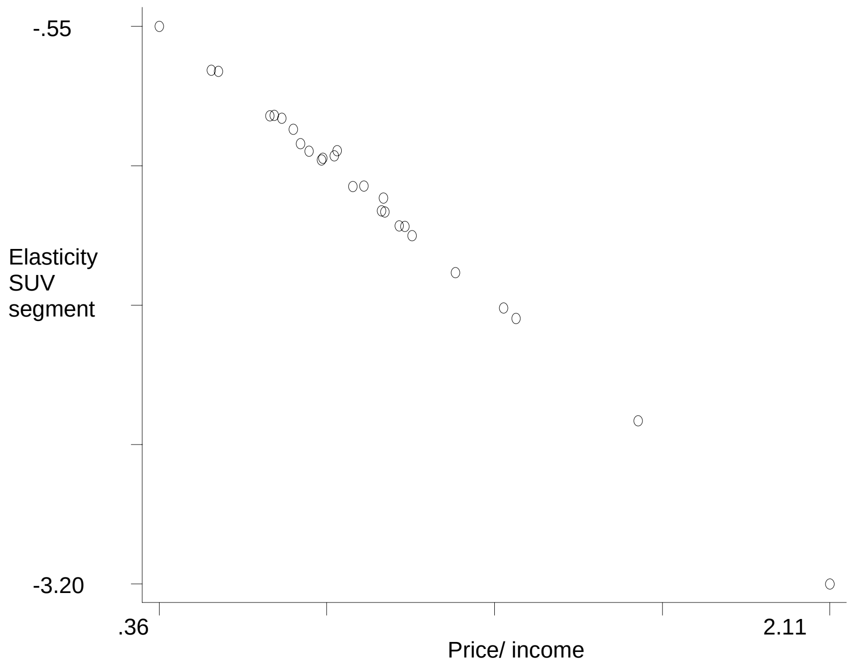


Figure 10. Sports segment: sports cars appear to be Veblen goods



<u>Figure 11. SUV segment</u>: in 2005, elasticities ranged from 0.5 to 2.0.



<u>Figure 12. SUV+ segment</u>: this segment displayed a wide range of elasticities: from 3.5 to15.