

# A Generalised Nested-Logit Model of the Demand for Automobile Variants<sup>(1)</sup>

Øyvind Thomassen

This paper estimates the demand for car model *variants* instead of looking only at demand for models in terms of the ‘baseline’ variant of each model as done in the literature. The data has sex and age of the buyer for every car sold in Norway 2000-2004, in addition to characteristics of the cars. The demand model uses this information to estimate taste coefficients which depend on demographic characteristics. A nested logit model and a generalised nested logit model are used to induce correlation in the logit error between products with observable and unobservable similarities. Results indicate that it may be problematic to have different logit errors for every product when the number of products is very high, even when allowing for flexible correlation patterns.

**Keywords:** Discrete choice, Demand estimation, Differentiated products, Automobiles

## 1. Introduction

Recent work [Petrin(2002), Bajari and Benkard(2003), Akerberg and Rysman(2005)] shows that discrete choice models with a utility component that is iid across products for a given consumer may have unrealistic implications in markets with many products. The idiosyncratic (usually logit) utility term implies that each product is differentiated in a new dimension. This tends to overestimate the benefits of variety and underestimate price elasticities and markups. The larger the number of products, the worse the problem is likely to be. This paper investigates whether the problem can be solved by allowing for correlations between a consumer’s idiosyncratic tastes for products that are similar with respect to observable characteristics.

---

(1) Financial support from the Institute of Economic Research of Seoul National University is gratefully acknowledged.

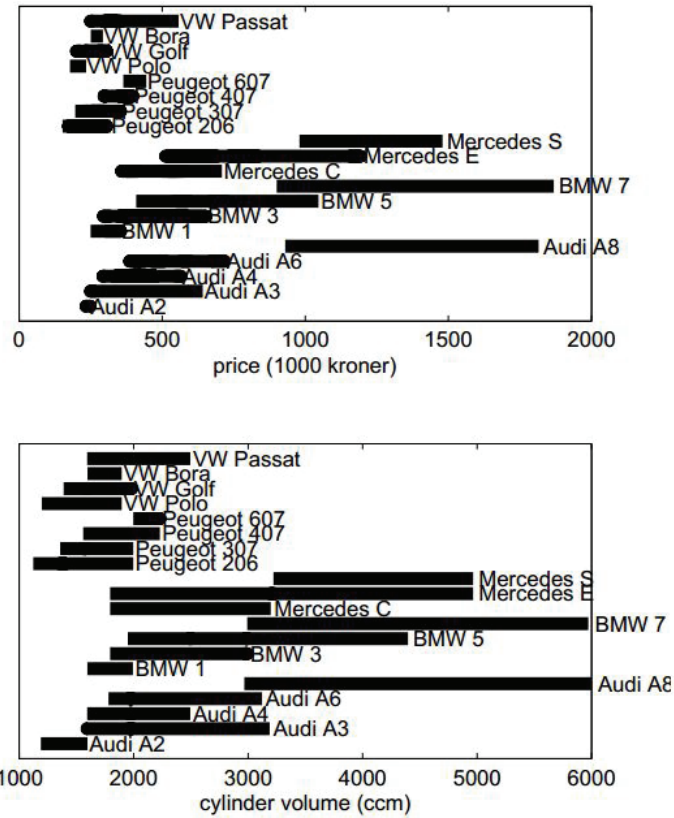
I do this by estimating a generalised nested logit model (GNL) [Wen and Koppelman (2001)], (or Principles of Differentiation nested logit model [Bresnahan, Stern, and Trajtenberg(1997)] using individual-level data on the sales of car model variants in Norway 2000-2004. The estimates indicate a very high correlation between the idiosyncratic tastes for variants of the same model. But implied price elasticities are unreasonably low, indicating that having a logit term for each variant of a model gives unrealistic results, even when flexible correlation patterns are in principle allowed for.

The next section discusses the data, the variants issue, and the link between demographics and car choices. The third section sets out the econometric model and gives an overview of the literature. The fourth section presents the results.

## **2. Data. Variants and demographics**

The data are yearly sales in Norway 2000-2004 of every car model variant by age and sex of the buyer. Sales of used cars are not included. Opplysningsrådet for Veitrafikken (The information council for road traffic) provided the data. The sales data contain characteristics of the cars but not prices, so prices have been determined by matching price lists and sales data according to the characteristics of the cars. Transaction prices would have been preferable but were not available. An informal look at discussion posts on the internet indicates that discounts on list prices are unusual for new cars, but that it is not uncommon to get extras (like winter tires) for free while paying the list price.

The total sales of new cars to individuals over the five years is 302,185. The numbers of car model variants offered in the five years were: 738, 747, 759, 782, and 818 respectively, when products were classified as having the same engine size if they had the same cylinder volume when rounded to the nearest 0.1 litres. To exclude products with an excessively small market share (some products had sales of just one or two) from the analysis, products were classified as having the same engine size if they had the same cylinder volume when rounded to the nearest 0.4 litres. Remaining variants with a market share below 0.0001 were then dropped. These changes reduced the number of cars in the data to 299,436, and the number of variants were now: 454, 443, 458, 467,



<Figure 1> Ranges of prices and cylinder volumes offered for sample models.

and 473. The descriptive discussion in the remainder of this section uses the data before these simplifications were made.

### 2.1. Car model variants

All the literature on discrete choice modelling of the demand for cars has treated the choice set as being the set of car models (“nameplates”).<sup>(2)</sup> In reality, most car models are offered in several variants, differing in such aspects as body style, engine size, fuel type and transmission. For the purposes of demand estimation, the existing literature has attributed sales of any variant of a given car model to the ‘base model’ - the cheapest

(2) Verboven(1999) considers variants in terms of engine, but uses a simple hedonic regression demand model.

variant of the model.

This subsection discusses the importance of the variants issue. As a first indication that variants of the same model can in fact be very different cars, <Figure 1> shows the ranges of prices and cylinder volumes offered for some well-known car models. The figure shows a selection of the model ranges of five different brands (Volkswagen, Peugeot, Mercedes, Audi and BMW). For each of the brands a model is usually regarded as an upgrade over the model which is located below it in the figure. The interesting point here is that - both in terms of price and engine size - there is a large degree of overlap between models which belong to different segments (compact, intermediate, large, luxury, *etc.*). For instance, the 'subcompact' VW Polo is offered with a top-end engine size that goes well into the range of cylinder volumes offered with the 'large' VW Passat. Similarly, the 'compact' Audi A3 can be purchased with an engine that is bigger than the low-end variant of the 'luxury' Audi A8.

These examples indicate that modelling demand only in terms of the 'baseline' model (*i.e.* picking a point on each range line in the figure) can be a strong simplification. <Table 1> shows the choice of characteristics offered for all the models marketed with at least ten different variants in 2004.

A natural question to ask is whether the wide range of variants is only available for a few models. The last column of <Table 1> shows the sales rank (out of 190 models) of the models with the largest number of variants. They are mostly high selling cars, ranking from 1st to 46th overall. <Figure 2> shows how many of the 190 models which offer the various numbers of options with respect to fuel types, the number of drive wheels, body styles (sedan, hatchback, station wagon *etc.*), cylinder volumes, lengths and prices (representing differences with respect to any characteristic). About 55 of the 190 models marketed in 2004 come in only one variant. The remaining 135 come in between 2 and 6 different body styles and between 2 and 8 different cylinder volumes.<sup>(3)</sup> Almost half of the models come with the choice of a diesel or petrol engine.

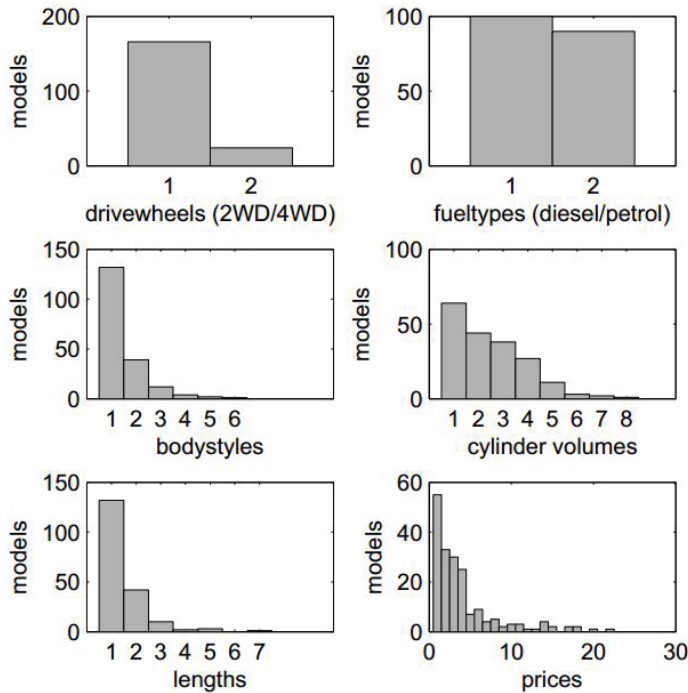
Finally, there is the question of whether most variants account for little of the model

---

(3) For the purpose of classifying products into variants, cylinder volumes have been rounded to the nearest centilitre so that negligible differences will not lead to classification as different variants.

<Table 1> Characteristics of sample car models

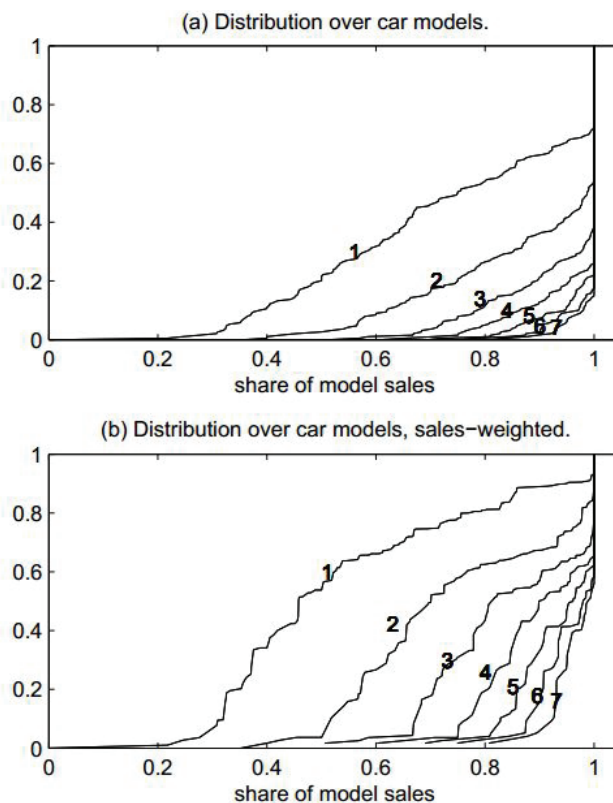
Make	Model	No. of vars.	No. of b.styles	Price		Cyl. Vol.			Lng.			Share			Share modal	Sales rank		
				median	min	max	median	min	max	min	median	max	4WD	Dies.			Auto	Airc.
Audi	A4	21	3	400	293	565	1896	1595	2496	458	454	458	0.43	0.38	0.57	1.00	0.25	18
Audi	A6	17	2	462	387	718	2393	1781	3123	480	480	492	0.53	0.41	0.71	1.00	0.28	36
Bmw	3	23	5	416	295	653	1995	1796	2993	448	426	449	0.13	0.30	0.83	1.00	0.41	28
Bmw	5	11	2	586	409	1046	2497	1951	4398	484	481	484	0.00	0.45	0.91	1.00	0.53	46
Citroen	C5	10	2	331	285	415	1997	1560	2179	475	475	484	0.00	0.60	0.30	1.00	0.37	39
Ford	FOCUS	16	5	259	179	318	1753	1388	1997	433	417	445	0.00	0.25	0.38	0.88	0.33	12
Ford	MONDEO	14	3	324	250	534	1999	1798	2967	473	473	480	0.00	0.36	0.00	1.00	0.54	10
Mazda	6	12	3	302	251	400	1999	1798	2261	469	468	470	0.08	0.25	0.08	1.00	0.44	11
Mercedes-Benz	C	12	3	456	358	709	2373	1796	3199	453	434	454	0.25	0.42	0.25	1.00	0.53	33
Mercedes-Benz	E	18	2	685	512	1193	2685	1796	4966	482	482	485	0.33	0.39	1.00	1.00	0.46	26
Nissan	PRIMERA	12	3	283	245	343	1820	1597	1998	457	457	468	0.00	0.25	0.25	1.00	0.22	32
Opel	ASTRA	20	4	229	195	310	1686	1364	1998	433	424	451	0.00	0.35	0.40	0.95	0.42	21
Opel	VECTRA	17	3	343	245	372	1998	1796	2198	460	460	482	0.00	0.47	0.29	0.94	0.52	9
Peugeot	206	12	4	194	152	320	1493	1124	1997	384	383	403	0.00	0.25	0.08	0.25	0.36	16
Peugeot	307	16	4	264	196	354	1587	1360	1997	439	420	442	0.00	0.44	0.25	0.88	0.31	4
Peugeot	407	10	2	355	297	398	1997	1560	2230	468	468	476	0.00	0.40	0.40	1.00	0.35	43
Renault	MEGANE	19	6	250	195	342	1598	1390	1998	435	421	450	0.00	0.32	0.00	0.79	0.36	22
Skoda	FABIA	10	3	190	151	269	1390	1198	1896	422	396	422	0.00	0.40	0.30	0.70	0.36	38
Skoda	OCTAVIA	11	2	249	191	327	1781	1390	1984	451	451	457	0.18	0.36	0.27	1.00	0.31	15
Toyota	AVENSIS	16	4	316	244	430	1995	1598	2362	463	463	470	0.00	0.25	0.38	1.00	0.32	1
Toyota	COROLLA	17	3	235	210	340	1598	1364	1995	436	418	439	0.00	0.47	0.18	1.00	0.38	2
Volkswagen	GOLF	15	2	275	201	306	1896	1390	1984	420	420	440	0.27	0.53	0.27	0.33	0.46	3
Volkswagen	PASSAT	11	2	312	250	558	1896	1595	2496	468	468	470	0.27	0.45	0.36	1.00	0.40	7
Volvo	S40	11	2	287	237	509	1870	1587	2521	451	447	451	0.00	0.27	0.18	1.00	0.47	25



\* Each diagram corresponds to a particular product characteristic. The height of the columns give the number of models with 1 variant with respect to the characteristic, 2 variants, *etc.*

<Figure 2> The number of car models offering each number of characteristics options.

sales. The second to last column in Table 1 gives a first hint at an answer to this question: it shows the share of model sales that come from the best-selling variant. For the about thirty cars with ten or more variants, this share ranges from 22% to 54%. <Figure 3> shows the empirical cumulative distribution function of the share of within-model sales accounted for by the best-selling variant ('1'), the two best-selling variants ('2'), and so on. For instance, panel (a) tells us that about 35% of car models on the market get 60% or less of their sales from the best-selling variant, whereas panel (b) tells us that about 70% of the cars sold are of models which get 60% or less of their sales from the best-selling variant. We can also see that about 20% of the cars sold are of models which get 80% or less of their sales from the four best-selling variants. Comparison of the two figures shows that the sales-weighted distribution have more mass at lower within-model sale shares. This means that high selling car models have their sales distributed more



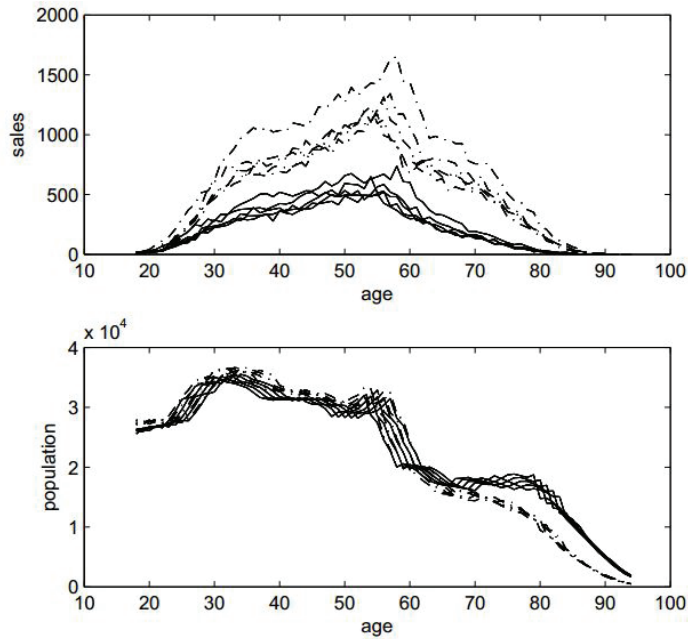
<Figure 3> Cumulative distribution functions of the share of within-model sales accounted for by the best-selling variant ('1'), the two best-selling variants ('2'), etc., where the distribution is over car models.

widely over a number of variants than the lower selling models.

This descriptive discussion of the data shows that the presence of several variants, often varying greatly in their specifications, is the case for most car models, and more so for cars with high sales than for those with lower sales.

## 2.2. Age, sex and the demand for cars

Some of the literature on discrete choice modelling of the demand for cars has used information about consumer attributes to model taste heterogeneity. This will be discussed in Section 3. Here I discuss a few issues regarding the connection between the choice of car on the one hand, and the demographic factors included in my data - sex and



\* Dotted line for men, solid line for women.

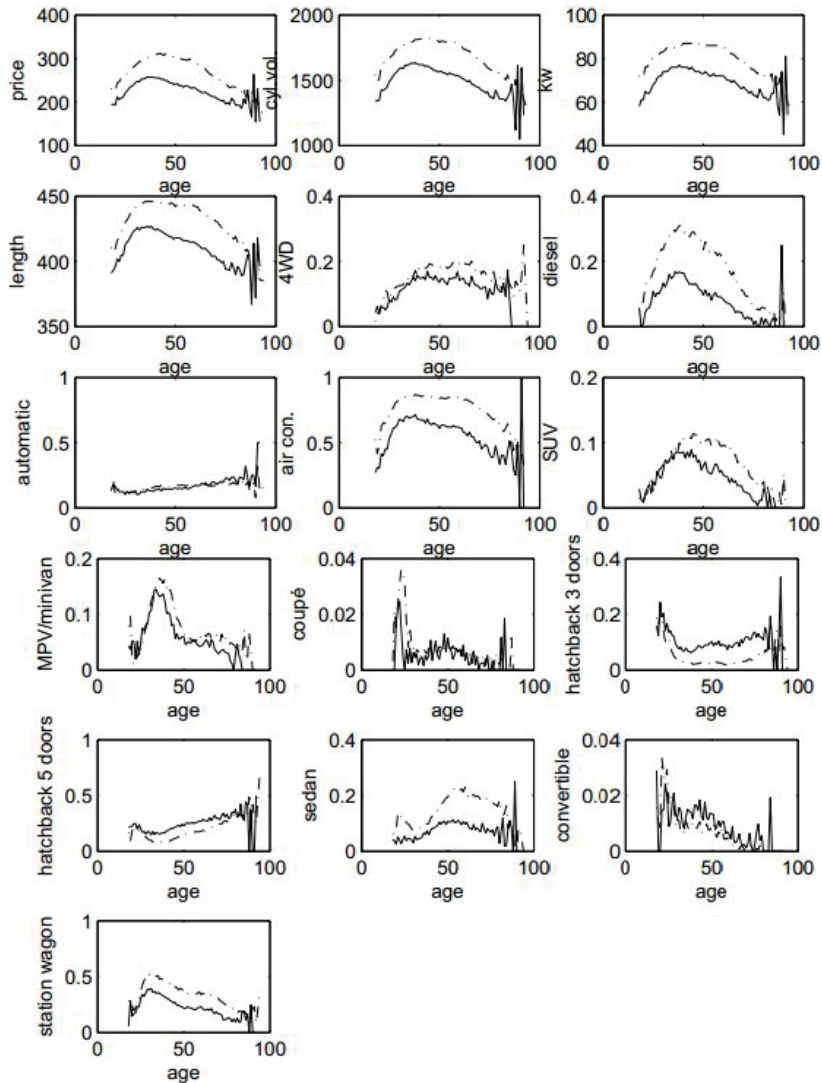
<Figure 4> Car sales in 2000-2004, and population between ages 18 and 94 by sex and age.

age - on the other.

People's sex and age influences their choice about whether to buy a car at all, and it influences their choice of car if they decide to buy one. The first panel in <Figure 4> shows the number of cars sold to women and men, respectively, of different ages in the years 2000-2004. The second panel shows the Norwegian population in the same years, by sex and age. The yearly graphs are not marked with the year, but in both the sales and the population it is possible to identify the years by following the outward movement of certain peaks. Two features of the car sales graphs stand out: Middle-aged people buy more cars than other age groups (also relative to their population share); and, men buy more cars than women.

The data contain the sex and age of the registered buyer. In some cases there may be a different person in the household who will be the primary driver, and who influences the choice of car. This means that the data may not accurately reflect the connection between





\* Dotted line for men, solid line for women.

<Figure 5> Mean characteristics of cars purchased, by sex and age

demographics and preferences in cars. As an objection to the procedure used in this paper, this point has limited validity, however. The reason for using demographics is to be able to explicitly model taste heterogeneity. Therefore, if there is systematic variation in the tastes of *buyers* according to their sex and age (whether these *buyer*-tastes are in some cases derived from *driver*-tastes or not), the demographic information contributes

to uncovering the substitution patterns. If the purpose of the analysis was to uncover *driver-tastes* this would be more of a problem. Additionally, the problem is irrelevant for the substantial proportion of people over 20 years of age live alone - 22% in 2001.

<Figure 5> shows the mean characteristics (across the years 2000-2004) of vehicles by the buyer's sex and age. Characteristics of women's purchases are shown as solid lines and men's as dotted lines. In certain cases a low number of observations make the mean behave in an erratic way (such as very high ages, where very few people buy cars, or the dummy for convertible). Mean values of almost all characteristics exhibit systematic variation according to sex and age. Some examples are: Men and middle-aged people buy bigger, more expensive and more powerful cars than women and young and old people; station wagons and multi-purpose vehicles (minivans) exhibit peaks for people aged 30-40 when most people have children living with them and need more space; conversely sales of coupés and convertibles peak at age 25, and then drop dramatically; hatchbacks (3 doors especially) are popular with the young and the elderly; sedans see a drop at child rearing ages and then rise steadily again from age 40 to reach a peak at 55-60.

### 3. Econometric model and estimation

#### 3.1. Literature

Certain features of the car market have been important in shaping the methodologies used for studying it: it has a large number of differentiated products and the available data have mostly been on the market level (with products as the unit of observation), not at the individual level (with buyers as the unit of observation). The large number of products means that it is impracticable to estimate the demand system directly as a system of equations expressing each product's demand in terms of all the prices of all the products. The literature has therefore concentrated on discrete choice methods which model product choice in terms of utility maximisation, where the parameters of a utility function with product characteristics as arguments are estimated. The simplest version of such a model is a multinomial logit model. Each individual chooses the alternative with

the highest utility. If utility is given by

$$(3.1) \quad u_{ij} = x_j \beta + \varepsilon_{ij},$$

where  $\varepsilon_{ij}$  is iid extreme value,  $x_j$  is a vector of characteristics of product  $j$ , and  $i$  indexes the individual buyer, the probability that the buyer will choose product  $j$  is:

$$(3.2) \quad P_{ij} = \Pr(u_{ij} > u_{il}, \forall l \neq j) = \frac{\exp(x_j \beta)}{\sum_{l \in J} \exp(x_l \beta)}$$

Notice that when only the logit error term contains an  $i$  index, the choice probability is the same for all individuals. When data are at the market level, then, a model of this kind could be estimated by matching the predicted market shares  $(1/n) \sum_{i=1}^n P_{ij} = P_j = P_j$  to the observed market shares, for instance by a maximum likelihood procedure, where the number of observations is the number of products. In this way an individual-level utility function is aggregated for estimation on market-level data. Cross-price elasticities in this model are given by:

$$(3.3) \quad \eta_{jl} = \frac{p_l}{P_j} \frac{\partial P_j}{\partial p_l} = \frac{p_l}{P_j} [-\beta_1 P_j P_l] = -\beta_1 p_l P_l,$$

where  $\beta_1$  is the coefficient on price, and  $p_l$  is the price of product  $l$ . This shows that in the multinomial logit model, the percentage demand effect of a price change in product  $l$  is the same on all other products. In the car context, this assumption - that all cars are equally good substitutes - cannot be sustained. In fact - the main concern of the literature has been to find a way of flexibly capturing the aspect of car demand that some cars are very similar, and therefore most likely close substitutes, while others are very different.

Within the logit framework, there are three ways to get around the problem of equal cross-price elasticities: making the logit error correlated for close substitutes (generalised extreme value (GEV) models); letting the taste coefficients ( $\beta$ s) depend on characteristics of the individual; and allowing the taste coefficients to have a distribution (random

coefficients). Early work focused on the first method, while more recent work has mostly been based on random coefficients and demographic interactions. The three methods can be combined.

GEV models in the form of the nested logit have been used to estimate car demand by Goldberg(1995), Ivaldi and Verboven(2004), Goldberg and Verboven(2001). The nested logit allows for a degree of correlation in the error term between alternatives that are members of the same nest, according to some predetermined nesting criterion. Cardell (1997) shows how the nested logit is equivalent to introducing a dummy variable for the nesting criterion in the utility function:

$$(3.4) \quad u_{ij} = x_j \beta + \zeta_{ig(j)} + \lambda \varepsilon_{ij},$$

where  $0 < \lambda \leq 1$ ,  $g(j)$  denotes the group or nest that  $j$  belongs to, and  $\zeta$  is a random variable with the unique distribution with the property that if  $\varepsilon$  is distributed extreme value, then  $[\zeta + \lambda\varepsilon]$  is also distributed extreme value. As  $\lambda$  goes to one, the model becomes a simple logit model, and as it goes to zero, within nest correlation goes to one. Although the nested logit model does not exhibit the equal cross-price elasticities of the logit, it is still inflexible because within-nest elasticities do have that property. Also - assuming, for the purposes of illustration, equal market shares for all products - cross-elasticities within nests are always higher than between nests. This is a serious problem since the nesting order is chosen by the researcher. For instance, by choosing a nesting hierarchy where class (compact, family, *etc.*) is at the top, followed by country of origin, the researcher imposes the constraint that all cars which share country of origin and class, will have higher cross-elasticities between them, than any two products which share only class. Bresnahan, Stern, Trajtenberg(1997), in a study of personal computers, relax this constraint within the GEV framework by letting each product belong to several nests. This is a version of the ‘generalised nested logit’ as proposed by Wen and Koppelman(2001).

Berry, Levinsohn, and Pakes(1995), Nevo(2001) and Petrin(2002) use a random-coefficients logit models, where the taste coefficients in (3.1) are normally distributed

random variables, with mean and standard deviation to be estimated. Each element, indexed by  $k$ , of  $\beta$  is then given by

$$(3.5) \quad \beta_k = \beta_k^0 + \sigma_k v_k$$

$$(3.6) \quad v_k \sim iidN(0,1).$$

In this model, the choice probabilities are

$$(3.7) \quad P_{ij} = P_j = Pr(u_{ij} > u_{il}, \forall l \neq j) = \int \frac{\exp(x_j \beta)}{\sum_{l \in J} \exp(x_l \beta)} dP(v),$$

and cross-price elasticities are now:

$$(3.8) \quad \eta_{jl} = \frac{p_l}{P_j} \frac{\partial P_j}{\partial p_l} = \frac{p_l}{P_j} \int [-\beta_l P_j P_l] dP(v).$$

Unlike the logit model, the random coefficients model has cross-price elasticities which depend on both products, and therefore can vary across products pairs. Berry, Levinsohn, and Pakes(1995) estimate this model using market level data. Petrin(2002) uses information on the mean demographic characteristics of buyers of different cars to better approximate the distribution of tastes. His model has taste coefficients like in (3.5), but in addition they depend on demographic characteristics. The integral in (3.7) is therefore over the distribution of demographic characteristics as well as over the normal distribution of the random part of the coefficients. Estimation of the taste parameters is then helped by matching the observed mean demographic characteristic for buyers of different products to the means predicted by the model.

Berry, Levinsohn, and Pakes(2004) go further in their use of demographic data. Unlike the previous auto demand literature, they have data with individuals as the unit of observation. Their utility specification of the same kind as Petrin's, but since their data are more informative, they are able to estimate a model which depends on demographics

in a more flexible way. On the other hand, they observe only one cross-section of car purchases (one year) and therefore have no variation in the choice set. They therefore have problems estimating the levels of the taste coefficients ( $\beta^0$  above), although they know how tastes vary with demographics. Goolsbee and Petrin(2004) estimate demand for satellite and cable tv, using a tastes that depend on demographics in a probit model. Brenkers and Verboven(2006) use a 3LNL with draws from the empirical income distribution to get a random price coefficient.

### 3.2. The model

The utility function is given as

$$\begin{aligned} u_{ij} &= v_{ij} + \varepsilon_{ij} \\ v_{ij} &= x_j^0 \beta^0 + \xi_j + \sum_k x_{jk}^1 \beta_{ik}^1 \\ \beta_{ik}^1 &= \sum_r z_{ir} \gamma_{kr}. \end{aligned}$$

where  $z_i$  is a vector of characteristics of the individual (sex, age and age squared).  $\varepsilon_{ij}$  is a logit error term, which can be either iid, or have some correlation across products. This will be discussed below. The vector  $x_j$  contains the following characteristics: price, fuel expenses (kroner per kilometer), horsepower, weight, length, the number of doors, dummies for diesel, 4WD, North European, South European, Asian, and a constant. The model is estimated in two stages. In the first stage only the parameters of the equation

$$(3.9) \quad u_{ij} = \delta_j + \sum_k x_{jk}^1 \sum_r z_{ir} \gamma_{kr} + \varepsilon_{ij},$$

are identified. From (3.9) it follows that

$$(3.10) \quad \delta_j = x_j \beta^0 + \xi_j.$$

This equation is then estimated in a second stage.<sup>(4)</sup> If it is assumed that the logit term is iid extreme value, the probability that the choice of individual  $i$ , denoted  $y_i$ , will be product  $j$  is given by:

$$(3.11) \quad P_{ij} = Pr(y_i = j | x_{t(i)}, \delta_{t(i)}, \beta_i^1) = \frac{\exp(v_{ij})}{\sum_{l \in J_t} \exp(v_{il})},$$

where  $t(i)$  denotes the time period where individual  $i$  is observed,  $x_t = \{x_l | l \in J_t\}$  and  $\delta_t = \{\delta_l | l \in J_t\}$  ( $J_t$  is the choice set in period  $t$ ).  $\delta_t$  is found by the contraction mapping suggested by Berry, Levinsohn, and Pakes(1995). This is the unique value of  $\delta_t$  which sets the predicted market shares of the model in period  $t$ , given the value of  $\beta$ , equal to the observed market shares in the period.<sup>(5)</sup>  $\delta$  is therefore a function of the  $\beta$ s. Conditional on  $x$  then,  $P_{ij}$  just depends on  $\beta_t = \{\beta_l^1 | \text{is.th.} t(i) = t\}$ . The likelihood function for the first stage is now:

$$(3.12) \quad L(\beta) = \sum_i \sum_j P_{ij}(\beta_{t(i)})^{1(y_i=j)},$$

where  $1(\cdot)$  is the indicator function. The choice set in each year includes an outside good, which has utility normalised to zero.<sup>(6)</sup> This means that we do in fact observe the choice of the entire population - the people who buy cars and the people who do not. Accordingly the likelihood function is the sum over the whole population. The number of  $y_i$ s within each age/sex group  $g$  which equal  $j$  is given by the car sales data, while the remaining people in the age/sex group  $g$  in the population are known to chose the outside good. The number of observations is therefore very high: 17,272,458, approximately 3.5 million (the population between the ages 18 and 94) in each of the five years. On the

---

(4) In addition to the characteristics mentioned above, the second stage estimation includes four year dummies to capture for instance macroeconomic effects. These are assumed to have the same effect on all consumers.

(5) The uniqueness of the  $\delta$  vector was demonstrated by Berry(1994).

(6) Since only differences in utility matter for the choice between alternatives, one good in each period can have its utility normalised.

other hand, the majority of these observations contain relatively little information, since there is no variation within each age/sex group choosing the outside good.

Maximising the likelihood yields estimates of the parameters  $(\delta_i, \beta_i)$ . The remaining parameters,  $\bar{\beta}$  can now be estimated by equation (3.10), with the unobserved characteristic as the error term. However, the characteristics vector,  $x_j$ , contains price. Since price is likely to be higher if the unobserved characteristic has a high value, price is endogenous in the regression. I therefore use a two stage least squares estimator, instrumenting for price with a special tax levied on all cars. The tax is a strictly increasing, convex, nonlinear function of cylinder volume, weight and horsepower.<sup>(7)</sup> Since price is the sum of the tax and the pre-tax price, the requirement of nonzero partial correlation of the tax with price is satisfied (this will be shown in the results section). Like the existing literature, I assume that the unobserved characteristic is uncorrelated with the observed characteristics (apart from price). The tax is then uncorrelated with the unobserved characteristic because it is a function of the observed characteristics, but since it is a nonlinear function it does not cause collinearity problems in the 2SLS regression.

### 3.3. The distribution of the logit error

This paper uses two different models which share the framework described above but differ in the assumptions made about the distribution of the error term. The first is a three-level nested logit (3LNL) model and the second a generalised nested logit (GNL) model.

The 3LNL places the alternatives in nests according to two criteria. At the top level there are two nest: one for the outside good, and one for all the cars. At the bottom level there is one nest for each car model, so that only variants of the same model are in the same bottom level nest. These assumptions imply that the logit error of the outside good is uncorrelated with those of the cars for every individual, and that the logit errors of variants of the same model have higher correlation between themselves than with the

---

(7) The tax is a sum of functions of cyl.vol., weight and horsepower, where each function is linear and increasing on subintervals, but changes slope four times.



logit errors of cars of different models. It should be noted that because of the consumer-specific taste coefficients, two alternatives in different nests could in principle still have utilities which are more highly correlated than alternatives in the same nest. Whether this is the case depends on the relative importance of the nonlogit part *versus* the logit error in the utility function. The choice probabilities for the 3LNL model are given by

$$(3.13) \quad P_{ij} = \frac{\exp\left(\frac{V_{ij}}{\lambda_2}\right) \exp\left(\frac{I_{ihg}}{\lambda_1}\right) \exp(I_{ig})}{\exp\left(\frac{I_{ihg}}{\lambda_2}\right) \exp\left(\frac{I_{ig}}{\lambda_1}\right) \exp(I_i)},$$

where  $V_{ij}$  and the inclusive values or logsums,  $I_{ihg}$ ,  $I_{hg}$  and  $I_i$ , are defined by

$$\begin{aligned} V_{ij} &= \delta_j + x_j \beta_i \\ I_{ihg} &= \lambda_2 \ln \sum_{j=1}^{J_{hg}} \exp[V_{ij}/\lambda_2] \\ I_{ig} &= \lambda_1 \ln \sum_{h=1}^{H_g} \exp[I_{ihg}/\lambda_1] \\ I_i &= \ln \sum_{g=1}^G \exp(I_{ig}), \end{aligned}$$

and where  $G$  is the number of nests at the top level (in this model  $G = 2$ , ‘car’ and ‘outside good’), and  $H_g$  is the number of subnests under nest  $g$  (in this model  $H = 1$  for the outside good nest, and it is equal to the number of car models for the ‘car’ nest), and finally  $J_{hg}$  is the number of products in top-level nest  $g$  and bottom-level nest  $h$ .<sup>(8)</sup> For reasons of parsimony, the logsum parameters have been restricted to be the same for all nests at each level, although this is not necessary.

The GNL model used in this paper places products in nests according to five criteria: model, length, horsepower, body style and car/not car. For the continuous characteristics

---

(8) For simplicity, the taste coefficient,  $\beta$ , has only been indexed by  $i$  instead by consumer group, and the time period notation has been suppressed.

length and horsepower, products are placed in nests according to which of seven length brackets and nine different horsepower brackets they belong to.<sup>(9)</sup> There is one nest for all alternatives, including the outside good, and one for all except the outside good. This approximates a top-level nest with choice between ‘car’ and ‘outside good’ [Wen and Koppelman(2001)]. The choice probabilities are

$$(3.14) \quad P_{ij} = \sum_g (\alpha_{d(g)} \frac{\exp(\frac{V_{ij}}{\lambda_{d(g)}}) \exp(I_{ig})}{\exp(\frac{I_{ig}}{\lambda_{d(g)}}) \exp(I_i)})$$

where  $g$  denotes nests (each length brackets, each horsepower bracket, each body style, etc.), and  $d(g)$  denotes the nesting criterion (length, body style, etc.), and

$$\begin{aligned} V_{ij} &= \delta_j + x_j \beta_i \\ I_{ig} &= \lambda_g \ln \sum_{j=1}^{J_g} \exp[V_{ij}/\lambda_g] \\ I_i &= \ln \sum_{g=1}^G \exp(I_{ig}). \end{aligned}$$

The logsum parameters have been restricted to be the same for all nests under the same nesting criterion for reasons of parsimony, although this restriction is not necessary. Furthermore, I follow Bresnahan, Stern, and Trajtenberg(1997) in letting  $\alpha_d = \frac{1-\lambda_d}{\sum_{d'}(1-\lambda_{d'})}$  for all  $d$ , so that  $\sum_d \alpha_d = 1$  and  $\alpha_d$  goes to zero as  $\lambda_d$  goes smoothly to one. The rationale for this is that when the logsum parameter is one, this criterion is not important for differentiation.

---

(9) For length, the dividing points between the brackets are, in cm: 380,400,420,440,460,480. For horsepower, the dividing points are, in kilowatts: 50,70,80,90,100,120,140,160.

&lt;Table 2&gt; Estimation results from 3-level nested logit model

	woman		age		age <sup>2</sup>		interc.	
	est.	std.err.	est.	std.err.	est.	std.err.	est.	std.err.
'price'	0.8116	0	35.3999	0.0233	-3.4328	0.0056	-10.0482	0.2227
'fuel expenses'	-0.1491	0.0009	-6.4587	0.002	0.37	0.0004	1.0484	0.1665
'horsepower'	-0.4603	0.0009	-7.719	0.0022	0.7725	0.0002	1.9181	0.0893
'weight'	-3.0162	0.0002	29.9743	0.0114	-4.5554	0.0034	0.3906	1.0444
'length'	-5.3069	0.0036	31.9824	0.0018	-3.0091	0.0004	-3.2844	0.4046
'diesel'	-0.1761	0.0003	-1.0398	0.0002	-0.0015	0.0001	0.1548	0.0492
'4WD'	-0.0022	0.0001	0.962	0.0001	-0.0371	0.0003	-0.2117	0.0255
'doors'	0.0039	0.0006	0.0155	0.0003	-0.0091	0.0003	0.0967	0.0102
'constant'	1.9764	0.0003	4.9334	0.0015	0.0133	0.0001	10.131	0.154
'north european'	0.0249	0.0001	-0.9898	0.0005	0.08	0.0002	0.4722	0.0237
'south european'	0	0.0015	-0.0245	0.0002	-0.0242	0.0003	0.0795	0.027
'asian'	0.016	0.0007	1.7232	0.0004	-0.1391	0.0003	-0.3271	0.0252
2001							-0.2744	0.0288
2002							-0.2841	0.0345
2003							-0.257	0.0361
2004							0.0428	0.0331
	est.	std.err.						
Model nests	0.2793	0.0016						
Car/not car nests	0.3293	0.0007						

## 4. Results

### 4.1. Parameter estimates

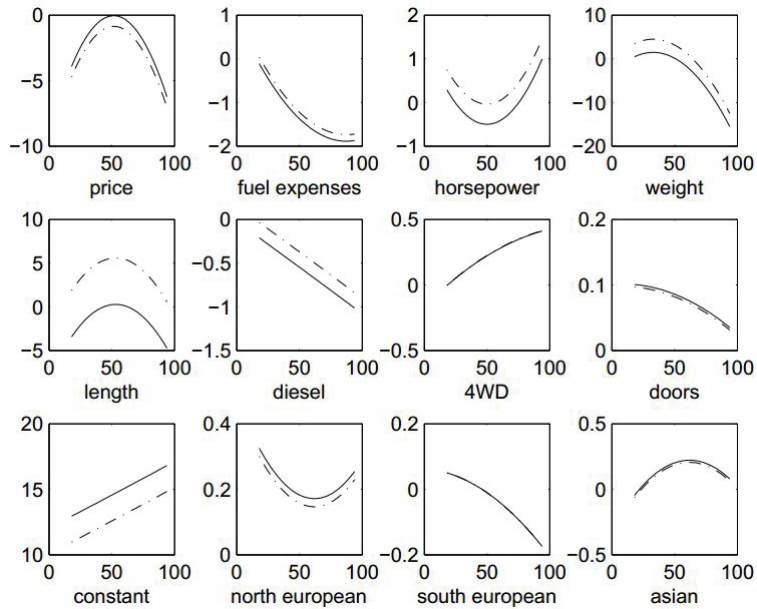
<Table 2> and <Table 3> presents the estimated parameters with standard errors. The columns marked 'woman', 'age' and 'age<sup>2</sup>' show the parameters determining the value of the group specific deviations,  $\beta_{g(i)}$ , of taste coefficients from the constant taste coefficients,  $\bar{\beta}$ , marked 'intercept' in the table. The intercept parameters are estimated by a 2SLS regression of the  $\delta$  vector (which results from the first stage) on the tax (the instrument for price) and the non-price characteristics. To test the requirement that the instrumental variable be partially correlated with price, I ran an OLS regression of price on the tax and the non-price characteristics. The coefficient on tax has a value of 16.3 for the t-statistic using heteroscedasticity-robust standard errors, and so is highly significant. The 2SLS estimation of the intercepts uses 2295 observations, the sum of the number

&lt;Table 3&gt; Estimation results from generalised nested logit model

	woman		age		age <sup>2</sup>		interc.	
	est.	std.err.	est.	std.err.	est.	std.err.	est.	std.err.
‘price’	0.2268	0	35.1525	0.0015	-3.7034	0.0001	-10.4604	0.4716
‘fuel expenses’	-0.0518	0.0001	-5.6566	0.0003	0.381	0	-0.5324	0.3267
‘horsepower’	-1.0049	0.0002	-7.0146	0.0007	0.6751	0.0001	1.428	0.1735
‘weight’	-3.0779	0.0005	26.1075	0.0013	-5.1584	0.0003	8.7833	2.1198
‘length’	-5.6496	0.0008	33.9622	0.0022	-2.879	0.0003	0.2666	0.7666
‘diesel’	-0.2696	0.0001	-0.8834	0	-0.0021	0.0001	-0.276	0.0966
‘4WD’	-0.0025	0.0003	1.3116	0.0001	-0.0359	0.0001	-0.3346	0.0481
‘doors’	0.0051	0.0002	-0.0089	0	-0.0069	0	0.2665	0.0206
‘constant’	2.6208	0.0005	4.9309	0.0001	0.0118	0	-10.7514	0.2929
‘north european’	0.0403	0.0002	-0.182	0.0001	0.0331	0.0001	0.4167	0.0464
‘south european’	0	0.0002	-0.0212	0.0001	-0.0247	0.0001	0.1837	0.0518
‘asian’	0.0295	0.0002	1.7302	0.0002	-0.0805	0	-0.2052	0.0482
2001							-0.3715	0.0553
2002							-0.4418	0.0663
2003							-0.4635	0.0688
2004							-0.1232	0.0638
	est.	std.err.						
Model nests	0.0004	0						
Length nests	0.3019	0						
Horsepower nests	0.7777	0.0002						
Body style nests	0.5233	0.0005						
All products nest	0.3435	0.0004						
Car/not car nests	0.723	0.0002						

of car model variants in each year. The first stage is the maximum likelihood estimation of the group specific parameters. This uses more than 17 million observations. About 300,000 of these are from the car sales data, giving information about age, sex and what car model variant they bought. Of the remaining observations only age, sex and the fact that they did not buy a new car in a given year is known. All the parameters from the first stage are highly significant for both models. In the second stage, only the 3LNL coefficients on the 2004 year dummy and the intercept for the characteristic ‘weight’ are not significantly different from zero.

In the 3LNL model, the logsum parameter for the model nests is lower than that for the car/not car nests. This means that the correlation in the logit errors for cars which are of the same model is higher than for products which are only in the same nest w.r.t. car/not

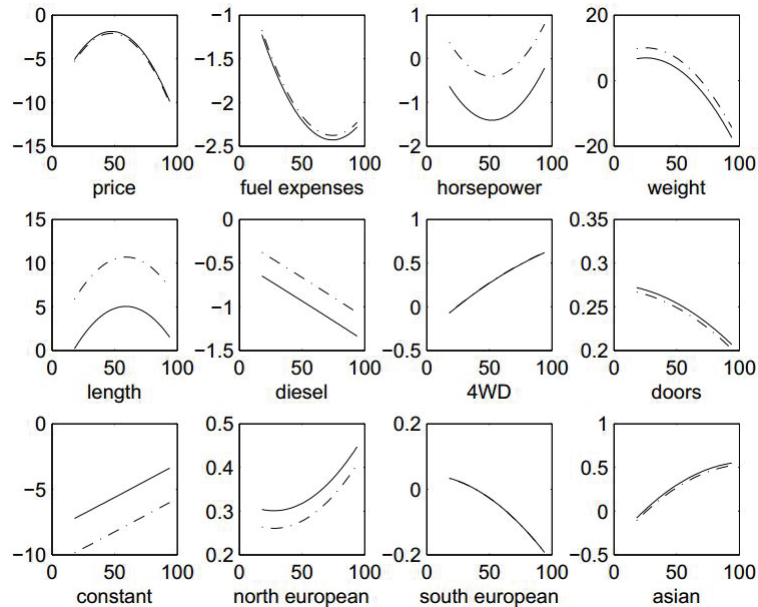


\*Solid line for women.

<Figure 6> Taste coefficients by age and sex, 3LNL model.

car. Both parameters are significantly different from 1, which would give uncorrelated errors, *i.e.* a simple multinomial logit. In the GNL model, the logsum parameter for model is extremely low, indicating an extremely high correlation in the logit term between variants of the same model. Being in the same length nest induces a lower but still high correlation. Having the same body style contributes slightly more to correlation in the error term than being in the same horsepower nest. I will discuss the logsum parameters further in relation to the implied elasticities below.

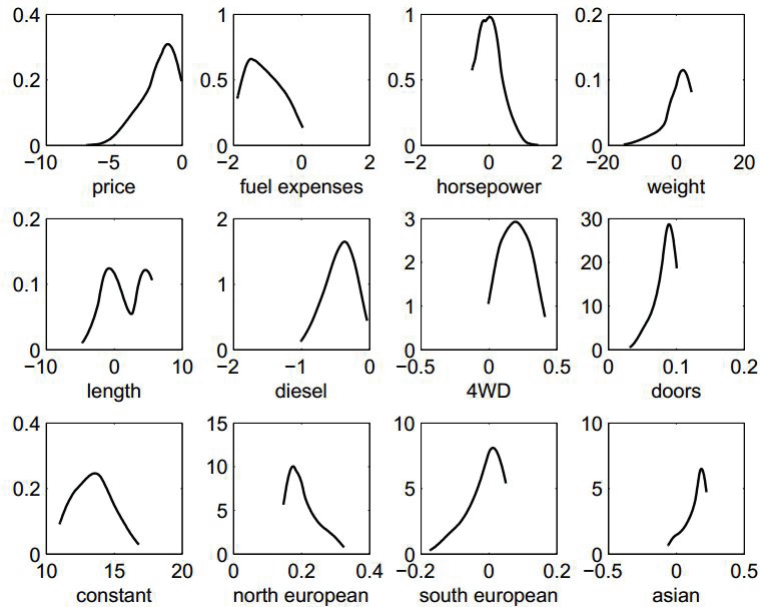
For easier interpretation the group specific taste coefficient, as determined by the parameters on ‘woman’, ‘age’ and ‘age<sup>2</sup>’, have been plotted for the 3LNL and the GNL models in <Figure 6> and <Figure 7>. Most coefficients have the expected signs for all groups. The price coefficient is negative for all, and has the lowest absolute value for middle-aged people. Since income has not been included in the analysis, it seems reasonable that the age groups with the highest disposable incomes have the lowest price sensitivity. The fact that women have a slightly lower price sensitivity than men in the



\*Solid line for women.

<Figure 7> Taste coefficients by age and sex, GNL model.

3LNL model, however, demonstrates that income is not the only factor to determine price sensitivity. The taste for fuel expenses is negative for all in both models. Perhaps unsurprisingly, the taste for horsepower is high among young people. It then dips for middle-aged people. The fact that it rises again for old people may appear strange. It must be kept in mind that fewer elderly people buy cars, so that the shape of the second order approximation is predominantly determined by the fit to the lower age groups. Concerning the length of cars, middle-aged people have the highest tastes, while the young and the elderly have a lower preference for length. The most striking aspect of the tastes for length is the fact that men have consistently higher valuations of it than women. Diesel engines are more valuable to young people. Young people also appreciate an extra door more than older people - probably because of family size. The constant is the value of simply having a car, irrespective of its characteristics - or, in a more meaningful interpretation: the negative of the constant is the value of not buying a car. Here we see that the value of the outside good is decreasing with age - perhaps reflecting



<Figure 8> Kernel smoothed densities of taste coefficients, 3LNL model.

a greater need for a car with age - and it is higher for men than for women. The last point is somewhat surprising, given that men buy more cars than women, but it has to be seen in conjunction with the fact that men value most characteristics more highly than women, and therefore also derive more utility from cars than women do. Northern European brands are valued more highly by all sexes and age groups than Southern European cars in both models. The tastes for Asian brands are just below those for North European for most groups.

<Figure 8> shows the (kernel smoothed) densities of of the taste coefficients for the 3LNL model. Contrary to <Figure 6> and <Figure 7>, this figure reflects the weight of certain age groups in the population. Notice the double-peaked density for length, because of the difference between men's and women's tastes for this characteristic.

#### 4.2. Elasticities

<Table 4> shows mean elasticities resulting from the two models for the 2004 market. For the 3LNL model the means over the following groups of elasticities are displayed:

<Table 4> Mean Elasticities. Own-price elasticities, and cross-price elasticities according to which nests are shared by products.

	Nesting criterion	Elasticities w.r.t. models which:	
		are in the same nest	are in different nests
3LNL model			
Own-price elasticities		-1.4765	
Cross-price elasticities	Model	0.0315	0.0017
GNL model			
Own-price elasticities		-1.4505	
Cross-price elasticities	Model	0.0186	0.0001
	Length	0.0011	0.0002
	Horsepower	0.0007	0.0003
	Body style	0.0007	0.0002
		shares all these nests	shares none of these nests
	Model, length, horsepower, body style	0.0516	0.0001
	Length, horsepower, body style	0.0045	0.0001
	Length, horsepower	0.0020	0.0001
	Length, body style	0.0027	0.0002
	Horsepower, body style	0.0015	0.0002

own-price elasticities, cross-price elasticities between variants of the same model, and cross-price elasticities between cars of different models. As could be expected, the cross-price elasticities are much higher between variant of the same models than between other cars. The mean own-price elasticity of -1.48 is quite low in absolute value. Other studies have found own-price elasticities in the range of -4 to -10 [see Berry, Levinsohn, and Pakes(1995) and Brenkers and Verboven(2006)] when the choice set is at the model, not variant, level. With variants, own-price elasticities should probably be higher, since there will be substitution away from a variant into other variants of the same models that is not captured when the analysis is in terms of models.

For the GNL model the mean own-price elasticity is about the same as for the 3LNL. The table then shows the mean cross-price elasticities between cars which are of the same model, in the same length nest, the same or horsepower nest, or which have the same body style, respectively. On average elasticities between variants of the same model are much higher (0.0186) than between cars which share other nests (0.0011 for



&lt;Table 5&gt; Sample elasticities from 3-level nested logit model

	Getz	Yaris	3	Focus	307	Golf	Avensis	Passat	RAV4	9-5
'Hyundai GETZ, 1.34l, Compact, 5doors, Petrol'	-0.6	0.004	0.003	0.003	0.002	0.003	0.007	0.005	0.007	0.002
'Volkswagen POLO, 1.2l, Compact, 4doors, Petrol'	0.006	0.004	0.003	0.003	0.002	0.004	0.006	0.005	0.007	0.002
'Toyota YARIS, 1.3l, Station wagon, 5doors, Petrol'	0.005	-0.7	0.003	0.003	0.002	0.003	0.006	0.005	0.007	0.002
'Opel MERIVA, 1.6l, MPV, 5doors, Petrol'	0.004	0.003	0.003	0.003	0.002	0.005	0.007	0.008	0.008	0.004
'Mazda 3, 1.6l, Compact, 5doors, Petrol'	0.004	0.002	-0.8	0.003	0.002	0.004	0.008	0.007	0.009	0.003
'Mitsubishi LANCER, 1.58l, Station wagon, 5doors, Petrol'	0.003	0.002	0.003	0.003	0.002	0.004	0.007	0.007	0.009	0.003
'Ford FOCUS, 1.6l, MPV, 5doors, Petrol'	0.003	0.002	0.003	-0.9	0.002	0.004	0.007	0.007	0.009	0.004
'Opel VECTRA, 1.8l, Sedan, 4doors, Petrol'	0.003	0.002	0.003	0.003	0.002	0.005	0.008	0.008	0.009	0.004
'Peugeot 307, 1.59l, Station wagon, 5doors, Petrol'	0.003	0.002	0.003	0.003	-1.1	0.005	0.007	0.009	0.009	0.004
'Mazda 6, 1.8l, Compact, 5doors, Petrol'	0.003	0.002	0.003	0.003	0.002	0.004	0.008	0.008	0.009	0.004
'Volkswagen GOLF, 1.9l, Compact, 4doors, Diesel'	0.003	0.002	0.003	0.003	0.002	-1.1	0.007	0.009	0.009	0.004
'Volkswagen GOLF, 1.97l, Compact, 5doors, Diesel'	0.003	0.002	0.003	0.003	0.003	0.033	0.008	0.010	0.010	0.005
'Toyota AVENSIS, 1.79l, Sedan, 4doors, Petrol'	0.003	0.002	0.003	0.003	0.002	0.004	-1.1	0.007	0.009	0.004
'Toyota AVENSIS, 1.79l, Station wagon, 5doors, Petrol'	0.003	0.002	0.003	0.003	0.002	0.004	0.040	0.007	0.009	0.004
'Volkswagen PASSAT, 1.9l, Station wagon, 5doors, Diesel'	0.002	0.002	0.003	0.003	0.002	0.005	0.007	-1.2	0.010	0.005
'Toyota AVENSIS, 2l, Station wagon, 5doors, Diesel'	0.002	0.002	0.003	0.003	0.002	0.004	0.036	0.008	0.010	0.004
'Toyota RAV4, 2l, SUV, 5doors, Diesel, 4WD'	0.002	0.002	0.002	0.002	0.002	0.003	0.007	0.007	-1.1	0.004
'Honda CR-V, 2l, SUV, 5doors, Petrol, 4WD'	0.002	0.002	0.003	0.003	0.002	0.004	0.008	0.008	0.009	0.004
'Saab 9-5, 1.99l, Station wagon, 5doors, Petrol'	0.002	0.002	0.003	0.003	0.002	0.005	0.007	0.009	0.010	-1.6
'Mercedes-Benz E, 2.15l, Sedan, 4doors, Diesel'	0.002	0.001	0.002	0.002	0.002	0.004	0.007	0.008	0.009	0.004

<Table 6> Sample elasticities from generalised nested logit model

	Getz	Yaris	3	Focus	307	Golf	Avensis	Passat	RAV4	9-5
'Hyundai GETZ, 1.34l, Compact, 5doors, Petrol'	-1.3	0.003	0.002	0.001	0.000	0.001	0.003	0.002	0.003	0.000
'Volkswagen POLO, 1.2l, Compact, 4doors, Petrol'	0.002	0.001	0.001	0.001	0.000	0.001	0.001	0.001	0.002	0.000
'Toyota YARIS, 1.3l, Station wagon, 5doors, Petrol'	0.003	-1.4	0.001	0.001	0.000	0.001	0.002	0.001	0.002	0.000
'Opel MERIVA, 1.6l, MPV, 5doors, Petrol'	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.003	0.003	0.001
'Mazda 3, 1.6l, Compact, 5doors, Petrol'	0.002	0.001	-1.4	0.001	0.001	0.001	0.004	0.003	0.004	0.001
'Mitsubishi LANCER, 1.58l, Station wagon, 5doors, Petrol'	0.001	0.001	0.001	0.001	0.001	0.001	0.003	0.003	0.004	0.001
'Ford FOCUS, 1.6l, MPV, 5doors, Petrol'	0.001	0.001	0.001	-1.3	0.001	0.001	0.002	0.003	0.003	0.001
'Opel VECTRA, 1.8l, Sedan, 4doors, Petrol'	0.001	0.000	0.001	0.001	0.000	0.001	0.002	0.003	0.003	0.001
'Peugeot 307, 1.59l, Station wagon, 5doors, Petrol'	0.001	0.001	0.001	0.001	-1.4	0.001	0.002	0.004	0.003	0.001
'Mazda 6, 1.8l, Compact, 5doors, Petrol'	0.001	0.001	0.001	0.001	0.000	0.001	0.003	0.003	0.003	0.001
'Volkswagen GOLF, 1.9l, Compact, 4doors, Diesel'	0.001	0.001	0.001	0.001	0.001	-1.6	0.003	0.005	0.004	0.001
'Volkswagen GOLF, 1.97l, Compact, 5doors, Diesel'	0.000	0.000	0.001	0.001	0.001	0.032	0.002	0.003	0.003	0.001
'Toyota AVENSIS, 1.79l, Sedan, 4doors, Petrol'	0.001	0.001	0.002	0.001	0.001	0.002	-2.1	0.005	0.006	0.001
'Toyota AVENSIS, 1.79l, Station wagon, 5doors, Petrol'	0.001	0.001	0.002	0.001	0.001	0.002	0.145	0.007	0.007	0.002
'Volkswagen PASSAT, 1.9l, Station wagon, 5doors, Diesel'	0.001	0.001	0.001	0.001	0.001	0.002	0.005	-2.1	0.006	0.002
'Toyota AVENSIS, 2l, Station wagon, 5doors, Diesel'	0.001	0.000	0.001	0.001	0.001	0.002	0.088	0.007	0.006	0.002
'Toyota RAV4, 2l, SUV, 5doors, Diesel, 4WD'	0.001	0.001	0.001	0.001	0.001	0.002	0.004	0.005	-2.2	0.001
'Honda CR-V, 2l, SUV, 5doors, Petrol, 4WD'	0.001	0.000	0.001	0.001	0.001	0.001	0.004	0.004	0.005	0.001
'Saab 9-5, 1.99l, Station wagon, 5doors, Petrol'	0.000	0.000	0.001	0.001	0.001	0.001	0.002	0.004	0.003	-2.3
'Mercedes-Benz E, 2.15l, Sedan, 4doors, Diesel'	0.000	0.000	0.001	0.001	0.001	0.001	0.002	0.004	0.003	0.001

length). The final panel in <Table 4> shows elasticities between cars which share many nests. By far highest are the elasticities between cars which are of the same model, fall in the same length and horsepower categories, and have the same body style (0.0516). The elasticities between cars which share none of these nests are on average much lower (0.0001). Although the relationships between these numbers seem reasonable, the overall level seems too low. <Table 5> and <Table 6> show the elasticities between a sample of cars.<sup>(10)</sup> Here again the small magnitudes are evident. Another problem is that the elasticities of a given car (rows) w.r.t price changes of different cars (columns) do not vary much. The GNL does better in this respect than the 3LNL.

### 4.3. Discussion: logit with a large number of products

Caplin and Nalebuff(1991) point out that including idiosyncratic error terms in utility is equivalent to including a dummy for every product, and imposing draws from the extreme value distribution as the coefficients on these dummy variables. This implies that the introduction of a new product adds one dimension to unobserved characteristics space. Since the expected difference between the logit term of any two products is the same regardless of the number of products, there is no congestion in unobserved characteristics space [Akerberg and Rysman(2005)]. This is counterintuitive in the sense that one would expect products to become closer as their number increases, as in a Hotelling model. The congestion does occur in the observed part of characteristics space, but the additional dimension of unobserved characteristics space allows every new product to be differentiated in a new way. The lack of congestion appears to overestimate the benefit of variety to consumers [Petrin(2002)]. One would expect that as the number of products goes to infinity, every product should have a perfect substitute, *i.e.* that every consumer could substitute to some other product with zero utility loss. Bajari and Benkard(2003) show that in any logit model such utility losses are bounded away from zero in the limit.

The problems in the elasticities resulting from the models estimated in this paper

---

(10) The sample was taken by taking the 100 best selling cars, ordering them by price and choose every fifth car, in order to get a selection of different cars.

are consistent with the points made by Bajari and Benkard(2003) and Akerberg and Rysman(2005) about how the logit error term accounts for an excessively large share of utility when the number of products is large. This would have the consequences that a given price change has a relatively small impact on utility (hence the small magnitudes of elasticities) and the factors which differentiate products, *i.e.* observed utility, is relatively unimportant (hence the similarity of cross elasticities w.r.t. different products). The latter point also would explain why the GNL fares better at differentiating the cross elasticities, as it has correlation patterns in the logit error which depend more directly on the observed characteristics.

## 5. Conclusion

This paper estimates a demand system for car model variants using Norwegian data from 2000-2004. The data are individual level and therefore allow us to estimate taste heterogeneity very precisely in so far as it depends on age and sex. Two different flexible GEV models with demographic taste coefficients are estimated. Results indicate that it is problematic to have idiosyncratic random terms attached to every product when their number is as high as it needs to be to model demand for different variants of car models.

Assistant Professor, Department of Economics, Seoul National University

1 Gwanak-ro, Gwanak-gu, Seoul 151-746, Korea

Phone: (02) 880-6376

Email: oyvind.thomassen@gmail.com

## References

- Akerberg, D. A., and M. Rysman (2005): “Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects,” *RAND Journal of Economics*.
- Bajari, P., and C. L. Benkard (2003): “Discrete Choice Models as Structural Models

- of Demand: Some Economic Implications of Common Approaches,” Unpublished manuscript, Stanford University.
- Berry, S. (1994): “Estimating Discrete-Choice Models of Product Differentiation,” *RAND Journal of Economics*, **25**, 242–262.
- Berry, S., J. Levinsohn, and A. Pakes (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, **63**, 841–890.
- \_\_\_\_\_ (2004): “Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market,” *Journal of Political Economy*, **112**, **1**, 68–105.
- Brenkers, R., and F. Verboven (2006): “Liberalizing a Distribution System: The European Car Market,” *Journal of the European Economic Association*, **4**, **1**, 216–251.
- Bresnahan, T., S. Stern, and M. Trajtenberg (1997): “Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s,” *RAND Journal of Economics*, **28**, S17–S44.
- Caplin, A., and B. Nalebuff (1991): “Aggregation and Imperfect Competition: On the Existence of Equilibrium,” *Econometrica*, **59**, **1**, 25–59.
- Cardell, N. S. (1997): “Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity,” *Econometric Theory*, **13**, **2**, 185–213.
- Goldberg, P. (1995): “Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry,” *Econometrica*, **63**, 891–951.
- Goldberg, P., and F. Verboven (2001): “The Evolution of Price Dispersion in the European Car Market,” *Review of Economic Studies*, **68**, **4**, 811–848.
- Goolsbee, A., and A. Petrin (2004): “The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV,” *Econometrica*, **72**, **2**, 351–381.
- Ivaldi, M., and F. Verboven (2004): “Quantifying the Effects from Horizontal Mergers in European Competition Policy,” *International Journal of Industrial Organization*.
- Nevo, A. (2001): “Measuring Market Power in the Ready-to-Eat Cereal Industry,” *Econometrica*, **69**, **2**, 307–342.
- Petrin, A. (2002): “Quantifying the Benefits of New Products: The Case of the Minivan,” *Journal of Political Economy*, **110**, **4**, 705–729.

Verboven, F. (1999): “Product line rivalry and market segmentation - With an application to automobile optional engine pricing,” *Journal of Industrial Economics*, **47**, **4**, 399–425.

Wen, C.-H., and F. S. Koppelman (2001): “The Generalized Nested Logit Model,” *Transportation Research Part B*, **35**, 627–641.