KATHOLIEKE UNIVERSITEIT LEUVEN

Faculty of Economics and Applied Economics

Department of Economics

Bounds on willingness-to-pay in a pure-characteristics model of the demand for automobile variants

by

Øyvind THOMASSEN

Econometrics

Center for Economic Studies
Discussions Paper Series (DPS) 10.16
http://www.econ.kuleuven.be/ces/discussionpapers/default.htm

April 2010

DISCUSSION PAPER

Bounds on willingness-to-pay in a pure-characteristics model of the demand for automobile variants

Øyvind Thomassen*
Katholieke Universiteit Leuven

April 2010

Abstract

I estimate a pure-characteristics discrete choice model of the demand for automobile engine and body style variants, using market-level data. Revealed preference bounds and imposed bounds on the willingness to pay for characteristics as a percentage of product price are sufficient to identify nonparametric taste distributions. Substitution patterns as determined by closest substitutes appear very realistic. The model can be estimated on a single cross-section of data.

*Address: CES, Naamsestraat 69, 3000 Leuven, Belgium. Tlf: +32 1632 6640. Fax: +32 1632 6796
Email: oyvind.thomassen@econ.kuleuven.be

1 Introduction

I estimate a model of the demand for automobile variants based on the semiparametric pure characteristics model of Bajari and Benkard (2005), modified by imposing bounds on consumers' willingness to pay for a characteristic as a percentage of the price of the product they purchased. As in the random-coefficients logit literature, utility is linear in product characteristics and tastes for characteristics have a distribution in the population of consumers. The model differs from this literature in two respects: (i) taste distributions are estimated nonparametrically, and (ii) consumers do not have a taste for products (like the logit term) that is unrelated to observable characteristics.

A first stage estimates a scalar unobservable product characteristic for each alternative. A second stage finds the sets of coefficients that can rationalise the choice of each product. Under utility maximisation, the choice of a given alternative implies that this alternative's utility is greater than those of all the other alternatives. These inequalities imply bounds on the combinations of taste coefficients a consumer may have given that he chose the given alternative. In principle, as the number of products goes to infinity, the sets of implied taste coefficients for each product should become singletons. In practice, only bounds for each alternative are implied by the data. Bajari and Benkard (2005) proceed to aggregate the bounds derived from individual choices to get bounds on the aggregate taste distributions.

Any product which has less of a given characteristic than the consumer's chosen product provides a lower bound on that consumer's taste for the characteristic (conditional on his tastes for the other characteristics). Vice versa, any product with more of the characteristic provides an upper bound on the consumer's taste. Some characteristics take on only a few discrete values, and in many cases only two: a 0 or 1. For instance a car either has automatic or manual transmission.

For dummy (binary) variables we obtain either only lower bounds (if the chosen product has a 1) or only upper bounds (if 0). This means that the set of taste coefficients will not be closed.

Furthermore, since bounds on the taste for one characteristic depends on the tastes for other characteristics, this results in less tight bounds for those characteristics too. For instance, let product A be high in characteristic 1 and characteristic 2, where 2 is binary. If we have no upper bound on the taste for 2, a consumer may choose A because of an extremely strong preference for 2 regardless of its value of characteristic 1.

This problem can be mitigated by imposing conservative bounds on the distributions of tastes.[1] I propose to use bounds on the willingness-to-pay for characteristics, expressed as a percentage of the product chosen. This provides a way of choosing conservative bounds that are economically meaningful, and which vary with preferences.

I estimate the model using data on sales of new car model variants in Norway in 2004. I find that elasticities are unreasonably high. This confirms the finding from Bajari and Benkard's application to computer demand. Like them, I find that elasticities become much more reasonable when the number of products in the choice set is reduced. In this case there is an obvious way to do that, by removing all but the bestselling variant of each model. This leaves a choice set with just one set of characteristics for each model, like in BLP and most previous estimation of demand for cars on market level data. Substitution patterns appear to be very reasonable. Since I use a large number of product characteristics, whereas the number is very limited in BLP for computational reasons, the model appears to predict each product's closest substitutes very well. The results for the full choice

[1] Bajari and Benkard briefly mention this in the NBER version of the paper, but to not discuss how to obtain those bounds.

set indicate that it there may be reasons to hold on to the idiosyncratic (logit) taste term, contrary the commonly held view that tractability is the only reason to keep it. This is an issue for further work.

The next section provides some background from the literature on demand estimation for differentiated products. The third section sets out the theoretical model and discusses identification. The fourth section explains the estimation procedure and describes the data. The fifth section presents and discusses the results, while the last section concludes.

2 Literature

This section reviews parts of the literature on estimation of demand systems for differentiated products. Since the model in this paper can be viewed as a multi-dimensional extension of the vertical differentiation mode of Bresnahan (1987) I give and overview of that model. I then summarise a recent discussion about the desirability of idiosyncratic (e.g. logit) taste terms in discrete-choice models.

2.1 Bresnahan's model of vertical differentiation

Bresnahan (1987) estimates car demand using a vertical differentiation model like those in Mussa and Rosen (1978) and Shaked and Sutton (1982). In this model a consumer's utility is

$$u_{ij} = x_j \beta_i - p_j, \tag{1}$$

where the characteristic, x_j, is a scalar representing "quality".[2] The taste parameter β has an estimated density on a nonnegative support, so that all consumers have a positive marginal valuation of the characteristic. Consumers make different choices

[2] Quality is an estimated function of characteristics. The important thing in this context is that all consumers have positive marginal utility for the characteristic.

because they have different valuations of the characteristic relative to price.

Utility for each product can be pictured as a linear function of β, where $-p_j$ is the intercept with the vertical axis, and x_j is the gradient. Each consumer is located somewhere on the horizontal axis, and chooses the product with the utility line that is highest at this β-value. For all products to have positive demand, it is necessary that if one product has strictly lower quality than another, it must also have a strictly lower price. It follows that the utility lines are ordered in a pattern where the lowest-quality product has the highest-lying line close to the vertical axis, because it has the highest vertical intercept (lowest price). At some point this line is crossed by the product which is above it in the quality ranking (steeper slope). For high enough βs, this line is superseded by the third-lowest quality product, and so on. In general, the point where the utility lines for products j and $j+1$ cross is given by

$$\beta^{j,j+1} = \frac{p_{j+1} - p_j}{x_{j+1} - x_j}. \tag{2}$$

If products are indexed in order of increasing quality, product j's market share is then

$$F_\beta\left(\beta^{j,j+1}\right) - F_\beta\left(\beta^{j-1,j}\right), \tag{3}$$

where F_β is the cumulative distribution function of β. Bresnahan assumes a uniform density for β.

Demand for a product is then proportional to the length of the interval on the horizontal axis where this product's utility line is the highest:

$$q_j = \delta \cdot [\beta^{j,j+1} - \beta^{j-1,j}] = \delta \cdot \left[\frac{p_{j+1} - p_j}{x_{j+1} - x_j} - \frac{p_j - p_{j-1}}{x_j - x_{j-1}}\right], \tag{4}$$

where δ is the (constant) density function. The cross-price and own-price demand derivatives are $(\delta/(x_{j+1} - x_j))$ and $-(\delta/(x_{j+1} - x_j)) - (\delta/(x_j - x_{j-1}))$, so the more similar the products are in terms of quality, the higher the price elasticities. Graphically, price substitution happens in the following way: When the price of a product goes up, its utility line shifts down since the vertical intercept, $-p$, is lower. This means that the point where it rises above the utility of the lower-quality neighbour is shifted outwards, and the point where it is superseded by its higher-quality neighbour is shifted inwards, shrinking the interval where it is above the other lines.

2.2 Idiosyncratic tastes

Caplin and Nalebuff (1991) point out that including idiosyncratic error terms, like the logit term, 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 (Ackerberg 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.

Ackerberg and Rysman (2005) propose to let the distribution of the idiosyncratic term change with the number of products in the choice set to allow for congestion of product space. Berry and Pakes (2007) do away with the idiosyncratic term altogether. In this paper I estimate a model based on the pure characteristics model of Bajari and Benkard (2005). That model will be discussed in detail in the next section.

A central question is whether it is desirable to dispense with the idiosyncratic taste term. There is a feeling that the logit error lacks justification: "these approaches [pure characteristics models] are intuitively very attractive in the sense that there are no *ad-hoc* logit errors" (Ackerberg and Rysman 2005). Because the existing estimation procedures do not allow much variation in specifications, "it is harder for the authors to provide a sense for how their various maintained assumptions impact their results" (Reiss and Wolak 2006). This point holds for the inclusion of the logit error, as well as the parametric assumptions on the taste distributions. While most IO applications do not have sufficient data to dispense with functional form restrictions, recent work (Blow, Browning, and Crawford 2008) has formalised the data requirements for nonparametric identification of characteristics models.

3 Model and Identification

3.1 The model

There are J products defined as bundles of K characteristics (x_j, ξ_j), where $x_j \in R^{K-1}$ is observed, and $\xi_j \in R$ is not observed by the researcher. The unobserved characteristic represents such things as style, quality and service, collapsed into a scalar value. Each consumer chooses one product. This is the product which maximises his or her utility over the set of all products. Utility is a linear function of the product characteristics and price. The fact that consumers choose different products is uniquely accounted for by differences in their tastes for characteristics and in their price sensitivity. The final goal of the analysis is to estimate the joint distribution of the taste coefficients, i.e. the linear coefficients in the utility function.

A consumer's ranking of products is unaffected by the scale of utility. Utility can therefore be multiplied by an individual-specific constant, and an individual-specific constant can be added to it, without changing the consumer's utility maximising product. The following normalisation is therefore permitted: All price coefficients are set to -1 (multiply by individual-specific constant, the inverse of the price coefficient), and utility for the outside good of not buying a car is set to 0 (adding an individual-specific constant, minus the utility of the outside good). Utility is then given by

$$u_{ij} = x_j \beta_i - p_j, \tag{5}$$

where i indexes individuals and j indexes products. For simplicity of notation, the vector x_j includes the unobserved characteristic as well as all the other characteristics.

3.2 The hedonic price function

Using the assumption that utility is strictly increasing in the unobserved characteristic ξ for all products and for all consumers (and two mild regularity conditions), Bajari and Benkard (2005) show that for any two products j and j' with strictly positive demand, it must be true that

1. If $x_j = x_{j'}$ and $\xi_j = \xi_{j'}$, then $p_j = p_{j'}$.
2. If $x_j = x_{j'}$ and $\xi_j > \xi_{j'}$, then $p_j > p_{j'}$.

3. $|p_j - p_{j'}| \leq M(|x_j - x_{j'}| + |\xi_j - \xi_{j'}|)$ for some $M < \infty$.

Define a mapping from observed and unobserved characteristics to price. Because of 1 the mapping is a function, since a point in the domain of the mapping maps to a unique point in its image set. Because of 2 the mapping is strictly increasing in the unobserved characteristic. Because of 3 the mapping defines a (Lipschitz) continuous surface. The price surface is denoted $p(x, \xi)$. In a logit demand model, such a surface does not necessarily exist, since two products with the same characteristics and different price both can have strictly positive demand because of the idiosyncratic taste term. The price function depends on the nature of competition, marginal costs, consumer preferences, and the products present in the market. If any of these primitives change, the shape of the price surface is also likely to change. The price function expresses the relationship between prices and characteristics in equilibrium in a particular market. See Bajari and Benkard (2005) pp. 1247-8 for a discussion of the price function.

3.3 Identification of the unobserved characteristic

The assumption used for identification is that the unobserved characteristic is independent of the observed characteristics. This assumption is slightly stronger than the mean independence assumption in BLP.[3] The unobserved characteristic has no inherent units, and so it is only identified up to a monotonic transformation. It is therefore assumed that it has been normalised so the marginal distribution of ξ is $\mathcal{U}(0,1)$. Bajari and Benkard (2005) use the proof of identification provided by Matzkin (2003). The identification result says that $\{\xi_j\}_{j=1,\ldots,J}$ is identified when the prices of many products are observed in a market, so that the joint distribution

[3] Manski (1994) discusses various orthogonality conditions.

$F(p, x)$ is known. The proof is:

$$
\begin{aligned}
F_{p|x=x_j}(p_j) &= Pr(p(x,\xi) \leq p_j | x = x_j) \\
&= Pr(\xi \leq p^{-1}(x, p_j) | x = x_j) \\
&= Pr(\xi \leq p^{-1}(x_j, p_j)) \qquad (6) \\
&= p^{-1}(x_j, p_j) \\
&= \xi_j
\end{aligned}
$$

The second line holds since the price function has an inverse for a given x since it is strictly increasing and continuous in ξ. The third line holds by the independence between x and ξ. The fourth line holds because the q-quantile of $\mathcal{U}(0, 1)$ equals q.

3.4 Identification of the taste coefficients

In the second stage, the unobserved characteristic recovered in the first stage is a given, and treated in the same way as the observed characteristics. It is therefore simply included as one of the elements of the x-vector, for notational convenience. For product j to maximise utility, it must be the case that

$$x_j\beta - p_j \geq x_l\beta - p_l, \quad \forall l \neq j. \qquad (7)$$

Let $\tilde{X}_j$ be the $(J-1) \times K$-matrix whose columns are the vectors $x_j - x_l$, $l = 1, \ldots, j-1, j+1, \ldots, J$, and let $\tilde{p}_j$ be the $J-1$ vector with elements $p_j - p_l$, running over the same indices as $\tilde{X}_j$. Then the condition that product j maximise utility can be written as a system of linear inequalities on standard matrix form. The set of taste coefficients permitted by the revealed preference condition (7) for product j is

$$A_j \equiv \{\beta \mid \tilde{X}_j\beta \leq \tilde{p}_j\}. \qquad (8)$$

This means that if a consumer chooses product j, his or her vector of taste coefficients must be inside the K-dimensional convex polyhedron A_j. Market share is the share of the population with taste vectors falling within each of the polyhedra.

It is possible to proceed to construct a demand system without making any further assumptions. Bajari and Benkard (2005) provide estimators for upper and lower bounds on the cumulative distribution function of tastes. The advantage of this approach is that features of demand functions can be estimated with only very weak assumptions. It seems difficult to obtain useful bounds on price elasticities in this way, however.

The alternative is to try to obtain a "point estimate" of the aggregate taste distribution. When there is only observations from one market, the distribution of probability mass inside each A-set is not identified. However, when the number of products goes to infinity in such a way that the A-sets are partioned ever more finely, in the limit these sets will be points (see Bajari and Benkard (2005) for a proof). Accordingly it can be expected that with a large number of products, the distribution of probability mass inside these sets will not be important. The estimation of the taste distribution can be regarded as entirely nonparametric, in which case it is possibly not fully identified. Alternatively, it can be assumed that the distribution is uniform over each A-set, but is otherwise unrestricted in the estimation, in which case it is fully identified.

4 Estimation

4.1 First stage

The unobserved characteristic is given in (6) as a quantile of a conditional distribution: $F_{p|x=x_j}(p_j)$. The nonparametric estimation of a conditional distribution functions and quantiles are well known problems. Matzkin (2003) suggests the following estimator, based on Nadaraya (1964):

$$\hat{F}_{p|x=x_j}(p_j) = \frac{\sum_{i=1}^{J} \tilde{k}_1(\frac{p_j - p_i}{h}) k_2(\frac{x_j - x_i}{h})}{\sum_{i=1}^{J} k_2(\frac{x_j - x_i}{h})}, \tag{9}$$

where $k_2(\cdot)$ is a multidimensional kernel, $\tilde{k}_1(u) = \int_{-\infty}^{u} k_1(s)ds$, and h is the bandwidth.

My data have around 900 products and 30 characteristics. It is not possible to estimate a quantile of a 30-dimensional density using just 900 data points. I follow Bajari and Benkard (2005) in assuming that the price function is additively separable in most of the characteristics, and then estimate the nonadditive part nonparametrically after removing the linear effects of the other variables. The price function is then $p(x, \xi) = p(x^A, \xi) + x^B\gamma$, where (x^A, x^B) is a partitioning of the vector x.[4] The γ was found by an OLS regression on the equation

$$p_j = x_j^A\theta + x_j^B\gamma, \tag{10}$$

and the price data used for the nonparametric estimation of the unobserved characteristic were

$$\tilde{p}_j = p_j - x_j^B\hat{\gamma}. \tag{11}$$

I used Epanechnikov kernels for the estimator in (9) (see for instance Martinez and Martinez (2002)):

$$k_1(\psi) = \frac{3}{4}(1 - \psi^2), \quad -1 \leq \psi \leq 1 \tag{12}$$

The bandwidths were chosen according to the Epanechnikov bandwidth rule (see Azzalini (1981)) $h = 1.3\hat{\sigma}n^{-\frac{1}{3}}$, where n is the number of observations and $\hat{\sigma}$ is the

[4] $x^A = (horsepower, cyl.vol., length)$.

empirical standard deviation of the data.

4.2 Second stage

Many of the characteristics used in the estimation are indicator variables. The revealed preference bounds in (7) give no upper bound on the coefficient for that characteristic for the people who buy the product with the highest value of that product, and conversely, no lower bound for the product with the lowest value of a characteristic. This problem is limited in the case of continuous characteristics, since there will usually only be one maximum and one minimum for each characteristic. For indicator variables, however, all products are either the (weak) maximum or (weak) minimum of that characteristic. This means that none of the A-sets (sets of coefficients that rationalise a given choice) will be closed. Leaving the coefficient for some characteristics unbounded also has repercussions in the sense that an extreme value for one characteristic often must be matched by an extreme value of another characteristic in order for the product to be the utility maximiser. I therefore imposed some conservative bounds on the coefficients. Since the price coefficient is normalised to -1, the coefficients of the characteristics have the convenient interpretation of willingness-to-pay for a one unit increase in the value of that characteristic. Accordingly, the bounds are formulated as bounds on the willingness-to-pay for characteristics, given as a percentage of the price of the product in question. (See table 3 for the bounds).

In the NBER version of their 2005 paper, Bajari and Benkard suggest a multi-stage Gibbs sampler to take random draws from the A-sets. The Gibbs algorithm is a general principle that can be used to draw from a multivariate density $f(x)$ which is difficult to draw from directly, but whose univariate conditional densities can be drawn from. Take a starting value $x^{(t)}$, and generate $X_1^{(t+1)} \sim f_1(x_1|x_2^{(t)} \ldots, x_K^{(t)})$, then $X_2^{(t+1)} \sim f_2(x_2|x_1^{(t+1)}, x_3^{(t)} \ldots, x_K^{(t)})$, and so on. This sequence converges to $f(x)$ (Robert and Casella 2004).

The revealed preference condition for the coefficient of characteristic 1 as given in (7) can be rewritten as, for all $l \neq j$,

$$\beta_1 \geq \frac{\sum_{k\neq 1} \beta_k(x_{l,k} - x_{j,k}) - (p_l - p_j)}{x_{j,1} - x_{l,1}} \quad \text{if} \quad x_{j,1} > x_{l,1} \tag{13}$$

$$\beta_1 \leq \frac{\sum_{k\neq 1} \beta_k(x_{l,k} - x_{j,k}) - (p_l - p_j)}{x_{j,1} - x_{l,1}} \quad \text{if} \quad x_{j,1} < x_{l,1}, \tag{14}$$

and similarly for the other coefficients. Denote the right hand side of the inequalities above $B(j, l, 1)$. This means that in general, for the product j, every other product provides either an upper or a lower bound on the coefficient values which could lead to the purchase of product j. The bounds based on willingness-to-pay, as described above, denoted as $\underline{b}_1$ and $\overline{b}_1$, provide additional bounds.

Suppose the distribution of probability mass inside each A-set is uniform. Even if this assumption is not maintained later, random draws under the assumption of a uniform distribution will reveal the support of the random vector β_j, and this support is precisely the set A_j. Given a starting value, $\beta_j^{(0)}$, which is inside A_j, it must be true that

$$\beta_{j,1}|\beta_{j,2}^{(0)}, \ldots, \beta_{j,K}^{(0)} \sim \mathcal{U}(\beta_{j,1,\min}, \beta_{j,1,\max}), \tag{15}$$

where the parameters of the univariate uniform depends on the conditioned-on betas in the following way:

$$\beta_{j,1,\min} = \max\{\underline{b}_{j,1} \ , \ \max\{B(j, l, 1) \ | \ l \neq j \ \text{ and } \ x_{j,1} > x_{l,1}\}\} \tag{16}$$

$$\beta_{j,1,\max} = \min\{\overline{b}_{j,1} \ , \ \min\{B(j, l, 1) \ | \ l \neq j \ \text{ and } \ x_{j,1} < x_{l,1}\}\}. \tag{17}$$

Given the starting value, $\beta_j^{(0)}$, the algorithm follows the Gibbs procedure described above, with equations (15-17) describing the conditional densities that are drawn from at each stage.

From a computational point of view, the Gibbs sampler is quick, taking about a minute to generate 3000 31-dimensional draws. What proved to be the difficult part was to find a starting point, i.e. any point satisfying (7). Bajari and Benkard (2005) report that they used as starting values coefficients derived from first-order conditions under the assumption that product space is filled up (so that consumers can pick a product anywhere in characteristics space). This method did not work for my data. The only method which turned out to be reliable was to use the centre of the Chebyshev ball (the largest K-dimensional ball wich can be fit inside the polyehedron), computed by Komei Fukuda's *cdd*[5] code, implemented for Matlab in *MPT*[6].

To draw from the full joint distribution of the betas, I used $ns = 1500$ draws for each product. The first 1500 of the 3000 draws are burn-in draws for the Gibbs sampler. Each draw from A_j is weighted by the market share of product j. The simulated market share, used to compute price elasticities, is

$$\check{s}_j = \sum_{l=1}^{J} \frac{s_l}{ns_l} \sum_{i=1}^{ns_l} \mathbf{1}(c_{li} = j), \tag{18}$$

where s_l is the observed market share of product l, i.e. the proportion of car buyers whose coefficients are in the set A_l. c_{li} denotes the product which maximises utility given the i-th draw of coefficient vector from A_l. $\mathbf{1}(\cdot)$ is the indicator function.

[5] Komei Fukuda, http://www.ifor.math.ethz.ch/~fukuda/cdd_home/cdd.html

[6] M. Kvasnica, P. Grieder and M. Baotić, Multi-Parametric Toolbox (MPT), http://control.ee.ethz.ch/~mpt/

4.3 Data

The data contain sales and product characteristics for all the 904 new car model variants sold in Norway in 2004. Previous demand analyses for cars have always treated a model (”name plate”) as one product, and have used the characteristics of the cheapest or most sold (”baseline”) variant as the characteristics of the model. In fact most models are marketed with a large number of different variants, varying in bodytype, engine size, transmission, or fuel type. Table 1 shows the 21 best selling models (arranged by price), some characteristics of the modal variant of each of these models, along with the mean, minimum and maximum of the characteristics across variants of this model. It also shows the number of variants of the model, total sales of the model, sales of the modal variant, as well as the model's rank in total sales. Table 2 does the same for a range of models chosen to represent the whole spectrum of car models, from the most expensive sports car through family cars to the smallest hatchback. Prices are list prices. It is known that transaction prices generally are lower than list prices, but these are not available apart from in smaller surveys which cover only a few products. The tables include the unobserved characteristic among the displayed characteristics. This will be discussed in the next section.

Table 3 shows the 30 characteristics used for estimation, along with their mean across available products, the sales weighted mean, and the minimum and maximum. It also shows the imposed bounds on the willingness to pay for one unit more of the characteristic. For example, a consumer's willingness-to-pay for an additional metre of length of a car is bounded above at 100% of the price of the car that the consumer actually bought. So if a consumer buys a car that costs 300.000 kroner (appr. £28,000), it is assumed that his willingness to pay for an additional 0.1 litre of cylinder volume on a car is bounded above to 30.000 kroner. For some characteristics it is assumed that willingness-to-pay is bounded below by zero, i.e. that nobody would pay a positive amount of money to have less of these characteristics. On the other hand, this possibility is allowed for many characteristics. For instance, if somebody does not like German cars, he or she may be willing to pay a little extra to have a car which is not German. The constraints are meant to be conservative, and it appears unlikely that they should be violated by the true distribution of taste coefficients. As discussed above, they can still contribute to identification of the model.

5 Results

5.1 The unobserved characteristic

The estimated unobserved characteristic ranges from 0.07 to 0.99 and has a mean of 0.52. Tables 1 and 2 show the value of the unobservable for the modal variant of a selection of models, as well as its mean, minimum and maximum across variants of each model. The unobserved characteristic is independent of all other characteristics (not including price). Generally speaking, a product with a high price relative to its characteristics will have a high value of the unobserved characteristic, since high quality, style or prestige is required for it to have positive demand in the presence of other, cheaper products with similar observed characteristics. This is illustrated by table 4. All Mercedes E-class variants have high values of the unobserved characteristic. This is consistent with the perception of Mercedes as a prestigious brand. The Peugeot 607 is a comparable car to the Mercedes E for similar engine sizes. A comparison reveals that the unobserved characteristic is lower for the Peugeot for similar specifications, presumably reflecting the higher prestige of the Mercedes. A similar pattern is found by comparing the Audi A4 with the Skoda Octavia, two similar models where the first is regarded as more prestigious than the second.

Variation within models is also illuminating. The low-end Mercedes E variants have extremely high values for the unobserved characteristic, reflecting some feature which gives them positive demand in spite of their very high prices relative to observed characteristics. It is pertinent to ask how the unobserved characteristic can vary so much within a given model. If the unobserved characteristic is taken to represent a specific feature, such as "prestige", it seems all Mercedes E should have the same value, or that the variants with bigger engines should have higher prestige. However, each variant is a package of characteristics, and the "prestige" derived from a big engine or a German manufacturer goes into those observed characteristics. The unobserved characteristic can therefore not simply be interpreted as "prestige" or "quality", but rather as the amount of prestige or quality that the car has beyond what is derived from its observed characteristics. Accordingly, a 5 litre top-of-the-range Mercedes E is indeed more prestigious than the bottom-of-the-range 1.8 litre version, but it is already clear from its observed characteristics that it is a very prestigious car. This is not the case with the 1.8 litre version however, and the model therefore assigns it a higher unobserved characteristic. Compared to a 2 litre Peugeot 607 which sells for 365,000 kroner, the 1.8 litre Mercedes must have substantial unobserved merit in order to warrant its 512,000 kroner price tag.

5.2 Taste coefficients

The draws generated in the second stage of the estimation are uniformly distributed inside the 30-dimensional revealed preference polyhedra. Figure 6 shows scatter plots of the joint densities of the taste coefficients for length and cylinder volume of consumers who buy three different variants of each of the models Audi A4, A6 and

A8. The variants are the bottom-of-the-range, a middle-of-the-range, and the top-of-the-range variants of each model in terms of price and engine size. The scatter plots are the projections of points distributed in a 30-dimensional space onto a 2-dimensional space. This explains why the sets do not look like polyhedra and the points do not look like they are uniformly distributed.

Taste coefficients for length and cylinder volume are bounded below by zero and above by the price of the product the consumer has chosen (cf. table 3). In the justification of the bounds I said that they are meant to be conservative bounds, and that they are unlikely to be violated by the true taste distribution in the population. At the same time, the model is not identified without these bounds. This means that the distribution of taste coefficient draws generated by the model *will* be constrained by the bounds to varying degrees. It does not follow from this that the bounds are too tight. Instead, this fact follows from the observation that revealed preference conditions from a cross-section of data are not sufficient to fully identify the taste distributions. The aim is to see whether the joint constraints of revealed preference and bounds on the willingness-to-pay are sufficient to yield useful predictions. It is clear that in the case of the top A8, the bounds severely restricts the area within which the draws fall, as points accumulate close to the upper bounds. Since the biggest A8 faces few or no competitors that are longer or have a bigger engine, revealed preference does not constrain the draws at all upwards. The mid A4, on the other hand, is located in a very densely occupied area of characteristics space. The large number of close competitors means that revealed preference provides upper constraints that are well below the upper bounds on willingness-to-pay.

Comparing the top or mid A4 with the bottom and mid A6 reveals an interesting pattern. The A4 is smaller than the A6, but otherwise the two models are similar with respect to design, quality and service. The different choices of buyers of the A4 and buyers of the A6, should therefore be due to a large extent to different tastes for length. This is indeed confirmed. Mid A4 buyers have a similar distribution of taste for volume to mid A6 buyers (cylinder volume being the same), but markedly lower tastes for length. The reason to pay 30.000 kroner to get an A6 instead of a very similar, but slightly shorter A4, is that the willingness-to-pay for length is high. Consumers who buy the top A4 have extremely high willingness-to-pay for cylinder volume, but do not care much about length. Conversely, bottom A6 buyers care very little about engine size, but have a very strong taste for length. The same pattern is confirmed by comparing the A6 and the A8.

To find the aggregate distribution of taste parameters, the draws for each of the 904 variants are aggregated, and weighted by the market share of each car, corresponding to the proportion of consumers represented by those particular draws. Since probability mass is uniformly distributed inside each taste coefficient polyhedron, it is essential for the approximation of the true aggregate taste distribution that each polyhedron is small in some sense. As discussed in the identification section, if the number of products goes to infinity in such a way that all polyhedra collapse to points, the distribution resulting from the model will equal the true distribution in the limit. The scatter plots in figure 6 give the impression that the polyhedra are quite chunky. However, if one imagines a 30-dimensional rectangle with sides similar to those formed by the scatter plot of the top A8, this 30-dimensional rectangle will contain all 904 taste coefficient polyhedra. These polyhedra are disjoint, and so much smaller than the rectangle which contains them all. Furthermore, most polyhedra are contained in a much smaller volume, with a few fringe products like the top A8 having much larger polyhedra.

Figure 6 shows kernel smoothed graphs of the aggregate marginal densities of some taste coefficients. Compared to the scatter plots in Figure 6, where points look almost uniformly distributed, these densities have much more probability mass concentrated at certain (low) levels. The products with high sales are concentrated in certain areas of characteristics space. In practice that means that many people have tastes leading them to prefer products in those areas. For the purposes of this discussion, that in turn means that draws in those areas are given much larger (market share) weights. All the marginal densities have peaks close to zero. This is most marked in the cases of length and cylinder volume. These coefficients are well constrained by revealed preference, because the corresponding characteristics are continuous and exhibit large variation in the data. The marginal densities for some dummy variables have less sharp peaks, because their taste coefficients are not identified as well by revealed preference, and therefore are more spread out.

The BLP model assumed independence between the taste coefficients for different characteristics. This assumption was made for reasons of feasibility in the estimation. Figure 3 shows kernel smoothed pairwise joint aggregate densities for some taste coefficients. The graphs exhibit some interesting examples of dependence between taste parameters that would be ruled out by BLP's assumption. First, there is a substantial proportion of the population with a relatively high taste for both cylinder volume and length, whereas very few people have that high tastes for only one of these characteristics and not the other. Secondly, there are many consumers with strong preferences for both four-wheel drive and a SUV body type. This is not surprising, but nevertheless a fact that should not be ruled out by distributional assumptions. Thirdly, an inverse relationship exists between taste for cylinder volume and disutility of fuel costs. Again a substantial group of consumers have a very low disutility of fuel costs and at the same time a very high preference for cylinder volume, whereas hardly anyone has such high tastes for cylinder volume while at the same time disliking fuel costs very much. Finally, the perhaps most interesting example shows that there is a strong inverse relationship between tastes for German cars and tastes for Japanese or Korean cars. Especially, consumers who have a high willingness to pay for their car being German, get a high disutility from an Asian car. Also, many consumers who value Asian cars dislike German cars. All these examples are intuitively appealing, and support the case that independence between taste parameters is an undesirable assumption.

5.3 Elasticities

Elasticities were computed by finding the numerical derivative of the simulated market shares given by (18) w.r.t. each product. I did this in two ways. Method I let consumers face a choice set containing all variants of every model, leaving them with 904 choices. Derivatives were computed with finite differences by letting the price increase for all variants of the relevant model, and then looking at how the joint market share of all variants of the model changed. To turn the derivatives into elasticities, they where multiplied by the price of the modal variant, and divided by the original joint market share. In method II, I simply removed all variants apart from the modal variant of each model from the choice set, and simulated demands with this new choice set.

The median own price elasticity from method I was -35, and -13 in method II. BLP report own price elasticities that mostly range from -3 to -6, and say that General Motors' own economists found these estimates plausible. Table 5 shows the own price elasticities, markups (price minus marginal cost) and markups as a percentage of price for the 50 best selling products in the market, computed using method II and the assumption of a Nash equilibrium in prices, with profit maximising entities being the 16 car manufacturing companies which produce the

197 products on the market. Table 6 reproduces implied percentage markups for a sample of products reported in BLP. BLP's markups are higher, although they only report a few examples. Their overall mean percentage markup is 23.9, compared to 11.4 in my results. I have not been able to find any accounting numbers on markups so far, but a mean of 11.4% is most likely too low. The executive/big family type cars at the top of Table 5 (Mercedes E to Saab 9.5) all have relatively high markups, ranging from 19 to 32%. This is consistent with the observation that these models are in a niche with few competitors of similar regard. Further down the list markups are generally lower, consistent with the fact that characteristics space is more crowded in that area.

In Tables 7-10 each row shows the elasticites for a car with respect to the price of the cars in each column, for the same sample of cars used previously to represent the spectrum of choices in the market. Tables 7 and 8 are elasticities computed with method I, and Tables 9 and 10 are elasticities computed with method II. In Tables 7 and 9, the models displayed are the ones used previously to represent the spectrum of choices in the market, while Tables 8 and 10 display the top selling 21 cars. Broadly the cars have been arranged with the most expensive in the top-left corner, and the cheapest in the bottom-right corner.

The elasticities for the representative sample of cars in Table 7 exhibit a pattern where cross-price elasticities of cars that are far away from each other are zero or very low. Accordingly the areas furthest away from the diagonal mostly mostly consist of zeros. On the diagonal are own-price elasticities, and broadly speaking, as one moves further away from the diagonal, one gets to cross elasticities with products that are more different. In the top left and the bottom right areas of the matrix, the belt of positive cross elasticities around the diagonal is thin. Since these extreme areas represent products at the fringes of characteristics space, they have fewer substitutes. Moving towards the middle of the matrix, where characteristics space is more densely filled with products, the belt around the diagonal gets thicker, since products have more substitutes. Moving away from the diagonal, it is generally true that cross elasticities gradually get lower, as they represent products that are gradually more different from the product at the diagonal. Cross elasticities higher than one are normally only found for adjacent or almost adjacent products.

In Table 8 the belt around the diagonal is much thicker. The products in this matrix are the best selling cars, and are located at the centre of characteristics space, which is where they appeal to the largest number of consumers. These products are therefore closer substitutes than those in the previous table, which included the models at the fringes of characteristics space. Accordingly, cross elasticities are positive for almost all products. It is still true, however, that elasticities are lower the further away they are from the diagonal.

Overall, the substitution patterns resulting from method I seem very reasonable. The size of the substitution effects is clearly too high, however. Implied markups from these elasticities are almost all well below one percent of price, many virtually zero. Bajari and Benkard (2005) also find that elasticities are unreasonably high in their application to demand for personal computers. They suggest that the result is due to the assumption of perfect information about all products on the part of consumers, and conclude that this is unlikely to hold with a choice set containing 700 products. To mitigate this problem they remove from the choice set any product with a market share of less than 0.75%. This left them with only 24 products, and median own price elasticity fell from -100 to -11. I follow their choice, but in way that is less ad-hoc. By removing all but the modal variants of each model, the choice set becomes that used by the previous literature, including BLP. All models are still included. The last column of Table 5 shows that the modal variant usually represents a very large proportion of the total sales of the model. I also tried to remove all models with market shares of less than 0.5%, leaving 40 products. This gave a median own price elasticity of -5, but the market share threshold is hard to justify.

As expected, elasticities are much lower with method II. This also means that many more cross elasticities are zero. In Table 9, the belt of positive cross elasticities around the diagonal is much narrower than in Table 7. Removing all but the modal variant of each model removes a model from areas of characteristics space which it does in fact cover, but which are not covered by the modal variant. In these areas of characteristics space it may be close neighbours with other models whose modal variants are quite different from its own modal variant. In this way, products which are in fact substitutes in variant-space (which is what the consumer faces) are not in model-space. The general features of the substitution patterns remain unchanged when moving from method I to method II. In table 9, products in the middle of the matrix have more substutes than the ones on the fringes, and the high selling products in Table 10 have many more substitutes than the ones in Table 9. The size of the elasticities is now much more reasonable, with all but three of the own elasticities being single digit for the top selling products. It appears that low market share products are more likely to have strangely high elasticities, such as the Mercedes S-class in Table 9, with -51. This is possibly because reavealed preference conditions constrain taste parameters less for special (low market share) products than for products in the more densely populated parts of characteristics space, leading to bad estimates of these consumers' preferences. For the higher-selling products, elasticities largely appear reasonable. Some examples of seemingly reasonable high cross elasticities between similar products are the two SUVs, Range Rover and VW Touareg (1.28 and 0.34) or Ford Mondeo and VW Passat (0.96 and

3.04).

The question of how a model's location in characteristics space when it changes from a set of points (many variants) to one point (only modal variant) affects the predicted best substitutes is important. Table 11 shows the ten best substitutes for four randomly chosen high-selling products in a densely populated area of characteristics space. The substitutes are ranked according to their elasticities with respect to the price of the sample car. (I also tried to rank substitutes using derivatives or displacement ratios, but this did not make much of a difference.) To the left are best substitutes for one-variant case and to the right is the many-variants case. The number of substitutes that are commmon to the two cases is 6 out of 10 for the VW Golf, 5 for the Toyota Avensis, 3 for the Ford Focus, and 2 for the Audi A4. As expected, the inclusion of all variants appears to make a difference.

Bajari and Benkard's pure characteristics model results in elasticities that are too high. This is confirmed by my results. Their suggestion is that consumers cannot perfectly process or obtain the information on all available products when their number is very high. It appears that in most cases this problem cannot really be solved by simply dropping products, since the researcher has to guess what products consumers actually do consider. Leaving out secondary variants of each model like here is one way to proceed, but clearly not ideal. To the argument about imperfect information, it could also be countered that consumers only need to collect detailed information on a relatively small number of products, since they can ascertain at almost no cost that most products are out of the question anyway. Another explanation for the high elasticities would be that consumers have idiosyncratic tastes for products. This would imply that after a given price change for product A, many people who originally would choose product A, still choose it even if there is a product B which is extremely similar in all objective respects, simply because there is a subjective reason that they find A better than B. Other consumers may feel the same way about B. This type of effect cannot be captured by unobserved characteristics like the one estimated in this model, because two products with similar characteristics and similar price also will have similar estimated unobserved characteristic. Finding a way of estimating a multidimensional unobserved characteristic could possibly go some way towards mitigating the problem. The objections to logit models discussed above have in common that they claim that characteristics space expands too much with the logit error, making products look more different than they really are. Possibly, though, it expands too little in a pure characteristics model.

6 Conclusion

I estimate the Bajari-Benkard model on recent car data from Norway, whereas previous applications of this model have been to personal computers and housing (Bajari and Kahn 2005). The model has several advantages over the BLP model. It can be estimated with data from one time period and one market, unlike BLP which requires many markets for identification. It can also accomodate data with a much larger number of characteristics than what is possible in BLP. I used 30, whereas the numerical properties of BLP are unstable even with a much smaller number. BLP assumes that taste coefficients are independently normally distributed. The Bajari-Benkard model makes no parametric assumptions on the taste distributions, and allows for dependence between the distributions. In BLP, simulation error becomes a big problem when the number of products is very large (Berry, Linton, and Pakes 2004). In the Bajari-Benkard model, a large number of products is an advantage for the estimation of taste distributions, although it appears that the number needs to be reduced afterwards to avoid excessive substitution effects.

Although the model as used here seems practically useful, especially since it only requires one cross section of data, the procedure of removing variants cannot be defended theoretically. Ideally the model should give reasonable elasticities with all variants left in, since this is the choice set actually faced by consumers. The cause of this problem is not known. Bajari and Benkard's argument of imperfect information is not fully satisfactory, especially as long as there are alternative explanations of the problem that are equally consistent with the observed facts, such as an idiosyncratic taste term in the true utility functions.

There is no conclusive evidence to suggest that the Bajari-Benkard style pure characteristics model employed in this paper is inferior to the BLP as a way to estimate demand for car models on market level data. There are also several regards in which this model appears to be superior to BLP. On the downside, however, this model does not seem to fulfill its promise of being able to estimate demand systems for a very large number of differentiated products. The need to reduce the number of products in the choice set, although defensible from a practical viewpoint, is ad-hoc and theoretically unsatisfactory. There is an unexplained problem of elasticities that are too high in this model, as this and the original application show. Further work should investigate whether that is a feature that is driven by the pure characteristics assumption.

References

Ackerberg, D. A., and M. Rysman (2005): "Unobserved Product Differentiation in Discrete Choice Models: Estimating Price Elasticities and Welfare Effects," *RAND Journal of Economics*.

Azzalini, A. (1981): "A Note on the Estimation of a Distribution Function and Quantiles by a Kernel Method," *Biometrika*, 68(1), 326–328.

Bajari, P., and C. L. Benkard (2003): "Discrete Choice Models as Structural Models of Demand: Some Economic Implications of Common Approaches," .

——— (2005): "Demand Estimation With Heterogenous Consumers and Unobserved Product Characteristics: A Hedonic Approach," *Journal of Political Economy*.

Bajari, P., and M. E. Kahn (2005): "Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities," *Journal of Business and Economic Statistics*, 23, 20–33.

Berry, S., O. B. Linton, and A. Pakes (2004): "Limit Theorems for Estimating the Parameters of Differentiated Products Demand Systems," *Review of Economic Studies*, 71(3), 613–654.

Berry, S., and A. Pakes (2007): "The Pure Characteristics Discrete Choice Model of Demand," *International Review of Economics*.

Blow, L., M. Browning, and I. Crawford (2008): "Revealed Preference Analysis of Characteristics Models," *Review of Economic Studies*, (75).

Bresnahan, T. (1987): "Competition and Collusion in the American Automobile Oligopoly: The 1955 Price War," *Journal of Industrial Economics*, 35, 457–482.

Caplin, A., and B. Nalebuff (1991): "Aggregation and Imperfect Competition: On the Existence of Equilibrium," *Econometrica*, 59(1), 25–59.

Manski, C. F. (1994): "Analog Estimation of Econometric Models," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, chap. 43. Elsiever.

Martinez, W. L., and A. R. Martinez (2002): *Computational Statistics Handbook with MATLAB*. Chapman & Hall/CRC.

Matzkin, R. L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71(5), 1339–1375.

Mussa, M., and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18, 301–307.

Nadaraya, E. A. (1964): "Some New Estimators for Distribution Functions," *Theory of Probability and Its Applications*, 15, 497–500.

Petrin, A. (2002): "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy*, 110(4), 705–729.

Reiss, P. C., and F. A. Wolak (2006): "Structural Econometric Modeling: Rationales and Examples from Industrial Organisation," in *Handbook of Econometrics, Volume 6*.

Robert, C. P., and G. Casella (2004): *Monte Carlo Statistical Methods*, Springer Texts in Statistics. Springer Science, second edn.

Shaked, A., and J. Sutton (1982): "Relaxing Price Competition Through Product Differentiation," *Review of Economic Studies*, 49, 3–13.

Table 1: Product characteristics of best-selling products

	Bodytype modal var.	No. of variants	Price (mode)	(mean)	(min)	(max)	Length (mode)	Cyl. vol. (mode)	Unobs. char. (mean)	(min)	(max)	Sales (modal var.)	(total)	Rank in total sales
Volvo V70	station	6	5.26	5.06	4.32	5.81	4.71	2.40	0.70	0.51	0.90	887	2349	13
Nissan X-Trail	SUV	4	3.90	3.89	3.55	4.33	4.46	2.18	0.64	0.46	0.81	1647	2039	15
Honda CR-V	SUV	1	3.55	3.55	3.55	3.55	4.64	2.00	0.25	0.25	0.25	1914	1914	18
Toyota RAV4	SUV	3	3.46	3.29	3.02	3.46	4.20	2.00	0.66	0.56	0.83	1305	2609	11
Subaru Forester	station	1	3.36	3.36	3.36	3.36	4.45	1.99	0.74	0.74	0.74	1978	1978	16
VW Touran	minivan	5	3.17	3.25	2.95	3.63	4.39	1.90	0.35	0.26	0.52	2015	3303	6
VW Passat	station	13	3.12	3.47	2.50	5.58	4.68	1.90	0.30	0.18	0.44	1525	3501	5
Ford Mondeo	station	16	3.12	3.63	2.50	5.42	4.80	2.00	0.42	0.21	0.62	1710	3239	7
Audi A4	station	24	3.01	4.55	2.93	10.34	4.58	1.60	0.60	0.27	0.81	639	2411	12
Toyota Avensis	station	17	3.00	3.29	2.44	4.30	4.70	1.79	0.54	0.35	0.73	2049	6301	1
Opel Vectra	station	17	2.78	3.28	2.45	3.72	4.82	1.80	0.52	0.40	0.68	1629	2996	8
Renault Megane	minivan	21	2.55	2.58	1.95	3.42	4.25	1.60	0.45	0.31	0.70	595	1884	20
Mazda 6	station	12	2.54	3.11	2.51	4.00	4.70	1.80	0.33	0.25	0.46	994	2205	14
Skoda Octavia	station	11	2.49	2.59	1.91	3.27	4.51	1.90	0.36	0.07	0.58	547	1866	21
Volvo S40	station	12	2.37	3.11	2.37	5.09	4.51	1.59	0.53	0.27	0.69	1161	1939	17
Toyota Corolla	hatchback	19	2.30	2.55	2.10	3.40	4.18	1.60	0.51	0.26	0.77	1787	5205	3
Ford Focus	station	21	2.30	2.70	1.79	3.18	4.45	1.60	0.65	0.36	0.77	917	2712	10
VW Golf	hatchback	14	2.29	2.65	2.01	3.06	4.20	1.60	0.39	0.17	0.62	2697	5662	2
Peugeot 307	hatchback	17	2.25	2.75	1.96	4.32	4.20	1.59	0.41	0.15	0.55	1272	4454	4
VW Polo	hatchback	5	1.78	2.01	1.78	2.34	3.90	1.20	0.49	0.31	0.76	1483	1901	19
Toyota Yaris	hatchback	9	1.60	1.91	1.60	2.13	3.61	1.00	0.54	0.43	0.67	1363	2914	9

Table 2: Product characteristics of sample products

	Bodytype modal var.	No. of variants	Price (mode)	(mean)	(min)	(max)	Length (mode)	Cyl. vol. (mode)	Unobs. char. (mean)	(min)	(max)	(modal var.)	Sales (total)	Rank in total sales
Porsche 911	coup	3	12.70	13.65	12.70	14.75	4.43	3.60	0.80	0.73	0.90	4	6	172
Ast. Martin DB9	coup	1	21.36	21.36	21.36	21.36	4.71	5.93	0.54	0.54	0.54	2	2	185
Range Rover	SUV	2	9.49	11.26	9.49	13.03	4.95	2.93	0.74	0.51	0.98	41	46	132
VW Touareg	SUV	4	7.12	9.28	7.12	11.47	4.75	2.46	0.40	0.07	0.94	84	116	99
Audi A8	sedan	5	9.30	12.86	9.30	18.15	5.05	2.97	0.69	0.55	0.89	13	37	137
Mercedes S-class	sedan	5	9.80	12.33	9.80	14.80	5.04	3.22	0.79	0.60	0.94	6	15	157
Mercedes E-class	sedan	21	5.41	8.47	5.12	16.88	4.82	2.15	0.71	0.38	0.97	607	1571	25
BMW 5-series	sedan	14	4.71	6.77	4.09	10.84	4.84	2.17	0.50	0.20	0.78	545	1049	39
Audi A4	station	24	3.01	4.55	2.93	10.34	4.58	1.60	0.60	0.27	0.81	639	2411	12
BMW 3-series	station	28	3.22	4.87	2.95	10.37	4.48	1.80	0.45	0.08	0.71	649	1569	26
Ford Mondeo	station	16	3.12	3.63	2.50	5.42	4.80	2.00	0.42	0.21	0.62	1710	3239	7
VW Passat	station	13	3.12	3.47	2.50	5.58	4.68	1.90	0.30	0.18	0.44	1525	3501	5
Volvo S40	station	12	2.37	3.11	2.37	5.09	4.51	1.59	0.53	0.27	0.69	1161	1939	17
Citron C5	station	11	2.90	3.63	2.85	5.49	4.84	1.75	0.47	0.16	0.89	341	968	41
Toyota Avensis	station	17	3.00	3.29	2.44	4.30	4.70	1.79	0.54	0.35	0.73	2049	6301	1
VW Golf	hatch-back	14	2.29	2.65	2.01	3.06	4.20	1.60	0.39	0.17	0.62	2697	5662	2
Opel Astra	hatch-back	21	2.10	2.44	1.95	3.66	4.25	1.60	0.51	0.16	0.92	731	1756	22
Toyota Yaris	hatch-back	9	1.60	1.91	1.60	2.13	3.61	1.00	0.54	0.43	0.67	1363	2914	9
Peugeot 206	hatch-back	12	1.52	2.13	1.52	3.20	3.83	1.12	0.50	0.27	0.72	636	1724	23
Fiat Punto	hatch-back	1	1.39	1.39	1.39	1.39	3.80	1.24	0.51	0.51	0.51	82	82	111
Daewoo Matiz	hatch-back	1	1.15	1.15	1.15	1.15	3.49	0.80	0.58	0.58	0.58	63	63	120

Table 3: Summary statistics

Variable	Unit	Mean	Sales mean mean	Min	Max -weighted	Lower bound on coef- on ficient in % of price	Upper bound on coef- on ficient in % of price
length	metres	4.44	4.4	3.49	5.19	0	100
cylinder volume	litres	2.1	1.76	0.8	6	0	100
fuel costs	kroner/km	0.68	0.63	0.27	1.42	-40	0
diesel	dummy	0.3	0.28	0	1	-20	60
kw*diesel	kw*dummy	0.28	0.24	0	2.3	0	100
kw*petrol	kw*dummy	0.78	0.61	0	3.68	0	100
doors squared	count/10	1.99	2.26	0.4	3.6	-60	60
doors	count	4.37	4.72	2	6	0	60
seats squared	count/10	2.54	2.65	0.4	8.1	-60	60
seats	count	4.97	5.11	2	9	0	60
air bags	count	5.29	5.44	0	9	0	20
4WD	count	0.19	0.21	0	1	-20	40
automatic	count	0.42	0.35	0	1	-20	60
weight	1000 kilogr.	1.4	1.32	0.78	2.52	-60	60
cylinders	count	4.53	4.06	2	12	-20	20
gears*automatic	count*dummy	2.13	1.78	0	7	0	40
gears*manual	count*dummy	3.05	3.36	0	6	0	40
new model this year	dummy	0.03	0.03	0	1	0	20
changed model this year	dummy	0.05	0.05	0	1	0	20
german	dummy	0.29	0.24	0	1	-20	20
french	dummy	0.15	0.12	0	1	-20	20
asian	dummy	0.21	0.34	0	1	-20	20
american	dummy	0.15	0.13	0	1	-20	20
swedish	dummy	0.06	0.08	0	1	-20	20
sedan	dummy	0.24	0.11	0	1	-40	40
hatch-back	dummy	0.23	0.29	0	1	-40	40
station wagon	dummy	0.25	0.37	0	1	-40	40
multi-purpose/minivan	dummy	0.11	0.11	0	1	-40	40
off-road/SUV	dummy	0.07	0.11	0	1	-40	40
convertible	dummy	0.06	0.01	0	1	-40	40
unobserved (estimated)	-	0.51	0.52	0.07	0.99	0	100
price	100.000 kroner	4.08	2.96	1.15	21.36	Coef- -ficient fixed at -1	

Table 4: Unobserved characteristics for sample models with variants

Make	Model	Cyl.vol.	Length	Price	Bodytype	Unobs. char.
'Mercedes-Benz	E	1.8	4.8	5.12	sedan	0.90
'Mercedes-Benz	E	1.8	4.8	5.60	station	0.96
'Mercedes-Benz	E	2.2	4.8	5.41	sedan	0.95
'Mercedes-Benz	E	2.2	4.8	5.94	station	0.97
'Mercedes-Benz	E	2.6	4.8	6.03	sedan	0.69
'Mercedes-Benz	E	2.6	4.8	6.40	sedan	0.78
'Mercedes-Benz	E	2.6	4.8	6.30	sedan	0.79
'Mercedes-Benz	E	2.6	4.8	6.50	station	0.77
'Mercedes-Benz	E	2.6	4.8	6.90	station	0.85
'Mercedes-Benz	E	2.6	4.8	6.81	station	0.85
'Mercedes-Benz	E	3.2	4.8	7.52	sedan	0.62
'Mercedes-Benz	E	3.2	4.8	7.87	sedan	0.72
'Mercedes-Benz	E	3.2	4.8	7.35	sedan	0.65
'Mercedes-Benz	E	3.2	4.8	8.35	station	0.78
'Mercedes-Benz	E	3.2	4.8	7.84	station	0.74
'Mercedes-Benz	E	4.0	4.8	10.07	sedan	0.38
'Mercedes-Benz	E	5.0	4.8	11.17	sedan	0.39
'Mercedes-Benz	E	5.0	4.8	11.50	sedan	0.44
'Mercedes-Benz	E	5.0	4.8	11.93	station	0.47
'Mercedes-Benz	E	5.4	4.8	16.30	sedan	0.60
'Mercedes-Benz	E	5.4	4.8	16.88	station	0.63
'Peugeot	607	2.0	4.8	3.65	sedan	0.59
'Peugeot	607	2.0	4.8	3.85	sedan	0.76
'Peugeot	607	2.2	4.8	4.25	sedan	0.66
'Peugeot	607	2.2	4.8	4.45	sedan	0.81
'Peugeot	607	3.0	4.8	5.85	sedan	0.22
'Audi	A4	1.6	4.6	2.93	sedan	0.62
'Audi	A4	1.6	4.6	3.01	station	0.65
'Audi	A4	1.8	4.6	3.80	sedan	0.51
'Audi	A4	1.8	4.6	4.00	sedan	0.53
'Audi	A4	1.8	4.6	3.33	sedan	0.43
'Audi	A4	1.8	4.6	3.81	sedan	0.66
'Audi	A4	1.8	4.6	3.98	station	0.62
'Audi	A4	1.8	4.6	4.19	station	0.63
'Audi	A4	1.8	4.6	3.52	station	0.53
'Audi	A4	1.8	4.6	4.00	station	0.75
'Audi	A4	2	4.6	3.39	sedan	0.63
'Audi	A4	2	4.6	4.76	sedan	0.59
'Audi	A4	2	4.6	3.81	sedan	0.73
'Audi	A4	2	4.6	3.54	station	0.70
'Audi	A4	2	4.6	4.76	station	0.66
'Audi	A4	2	4.6	4.01	station	0.81
'Skoda	OCTAVIA	1.4	4.6	1.91	hatchback	0.43
'Skoda	OCTAVIA	1.4	4.6	1.91	station	0.47
'Skoda	OCTAVIA	1.6	4.6	2.25	hatchback	0.29
'Skoda	OCTAVIA	1.6	4.6	2.28	station	0.43
'Skoda	OCTAVIA	1.8	4.6	2.71	hatchback	0.45
'Skoda	OCTAVIA	1.8	4.6	3.20	station	0.07
'Skoda	OCTAVIA	1.8	4.6	3.27	station	0.34
'Skoda	OCTAVIA	1.8	4.6	2.49	station	0.40
'Skoda	OCTAVIA	1.8	4.6	2.90	station	0.58
'Skoda	OCTAVIA	2.0	4.6	2.49	hatchback	0.28
'Skoda	OCTAVIA	2.0	4.6	3.07	hatchback	0.24

Table 5: Markups for best-selling products ordered by price

Make	Model	Cyl. vol.	Price	Markup (P-MC)	Markup as % of price	Own price elast- icity	Units sold modal var.	Units sold total	Sales modal % of total
'Mercedes-Benz	E	2.2	5.4	1.1	21	-5.4	607	1571	39
'Volvo	V70	2.4	5.3	1.1	21	-5.8	887	2349	38
'Bmw	5	2.2	4.7	1.5	32	-3.2	545	1049	52
'Audi	A6	1.8	4.4	0.8	19	-7.5	470	1614	29
'Saab	9.5	2.0	4.1	1.1	26	-3.8	812	1406	58
'Nissan	X-TRAIL	2.2	3.9	0.5	12	-9.0	1647	2039	81
'Mercedes-Benz	C	1.8	3.7	0.3	7	-14.9	515	1160	44
'Honda	CR-V	2.0	3.5	1.3	36	-3.0	1914	1914	100
'Subaru	LEGACY	2.0	3.5	0.3	9	-19.5	639	1050	61
'Toyota	RAV4	2.0	3.5	0.4	12	-8.5	1305	2609	50
'Mitsubishi	OUTLANDER	2.0	3.4	0.5	15	-7.1	721	1106	65
'Subaru	FORESTER	2.0	3.4	0.3	10	-10.6	1978	1978	100
'Bmw	3	1.8	3.2	0.2	7	-13.5	649	1569	41
'Volkswagen	TOURAN	1.8	3.2	1.0	32	-3.9	2015	3303	61
'Suzuki	VITARA	2.0	3.1	0.2	6	-16.4	1040	1365	76
'Saab	9.3	1.8	3.1	0.4	14	-7.2	736	1478	50
'Volkswagen	PASSAT	1.8	3.1	0.6	20	-5.1	1525	3501	44
'Ford	MONDEO	2.0	3.1	0.6	19	-5.2	1710	3239	53
'Peugeot	407	1.6	3.0	0.5	17	-9.9	355	953	37
'Audi	A4	1.6	3.0	0.3	10	-15.3	639	2411	27
'Toyota	AVENSIS	1.8	3.0	0.8	25	-4.8	2049	6301	33
'Citroen	C5	1.8	2.9	0.2	6	-22.6	341	968	35
'Volvo	V50	1.8	2.9	0.4	13	-15.6	705	1247	57
'Opel	ZAFIRA	1.8	2.8	0.7	26	-3.9	360	885	41
'Opel	VECTRA	1.8	2.8	0.5	19	-5.4	1629	2996	54
'Renault	MEGANE	1.6	2.5	0.6	25	-4.4	595	1884	32
'Mazda	6	1.8	2.5	0.2	7	-13.6	994	2205	45
'Audi	A3	1.6	2.5	0.1	5	-22.8	1009	1286	78
'Nissan	PRIMERA	1.6	2.5	0.1	5	-46.7	217	1025	21
'Renault	LAGUNA	1.6	2.5	0.2	8	-17.1	487	827	59
'Skoda	OCTAVIA	1.8	2.5	0.3	13	-7.8	547	1866	29
'Suzuki	LIANA	1.6	2.4	0.6	25	-5.1	1439	1446	100
'Volvo	S40	1.6	2.4	0.8	35	-3.7	1161	1939	60
'Citroen	XSARA	1.6	2.3	0.2	10	-10.3	225	768	29
'Toyota	COROLLA	1.6	2.3	0.6	25	-4.4	1787	5205	34
'Ford	FOCUS	1.6	2.3	0.3	14	-7.7	917	2712	34
'Mitsubishi	LANCER	1.6	2.3	0.1	6	-22.8	372	609	61
'Volkswagen	GOLF	1.6	2.3	0.1	6	-19.5	2697	5662	48
'Peugeot	307	1.6	2.2	0.5	22	-4.6	1272	4454	29
'Mazda	3	1.6	2.2	0.1	3	-56.1	576	841	68
'Opel	ASTRA	1.6	2.1	0.2	10	-10.4	731	1756	42
'Opel	MERIVA	1.6	2.1	0.6	28	-3.7	603	1107	54
'Suzuki	IGNIS	1.4	1.9	0.2	12	-9.1	474	932	51
'Citroen	C3	1.4	1.9	0.0	2	-51.1	254	623	41
'Volkswagen	POLO	1.2	1.8	0.6	33	-3.3	1483	1901	78
'Ford	FIESTA	1.4	1.8	0.1	5	-21.5	508	641	79
'Hyundai	GETZ	1.4	1.6	0.2	9	-11.1	1106	1284	86
'Toyota	YARIS	1.0	1.6	0.4	24	-4.2	1363	2914	47
'Skoda	FABIA	1.2	1.6	0.2	10	-10.7	519	1106	47

Table 6: Sample markups from BLP reproduced

Make	Model	Markup as % of price
BMW	7	29.27
Lexus	LS	32.78
Cadillac	Seville	30.8
Lincoln	Town Car	26.13
Acura	Legend	24.66
Nissan	Maxima	21.04
Buick	Century	23.87
Ford	Taurus	26.65
Honda	Accord	21.44
Chevy	Cavalier	22.46
Ford	Escort	19.02
Nissan	Sentra	15.54
Mazda	323	15.86

Table 7: Model price elasticities of sample products, computed with all variants 904 products

Porsche 911	-5	0.1	0.0	0.0	0.5	0.0	0.1	0.0	0.1	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Ast. Martin DB9	0.1	-15	0.0	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Range Rover	0.0	0.0	-12	3.2	0.3	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VW Touareg	0.0	0.0	1.5	-60	0.3	0.0	0.4	0.9	0.5	0.0	0.9	0.5	0.0	0.0	0.2	0.2	0.0	0.0	0.0	0.0	0.0
Audi A8	0.1	0.1	0.4	0.6	-84	1.7	20	3.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mercedes S-class	0.0	0.0	0.0	0.0	4.5	-88	40	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
Mercedes E-class	0.0	0.0	0.0	0.1	0.9	0.7	-19	1.8	4.4	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
BMW 5-series	0.0	0.0	0.0	0.2	0.3	0.0	2.9	-34	4.6	0.3	0.1	0.1	0.4	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.0
Audi A4	0.0	0.0	0.0	0.1	0.0	0.0	4.8	2.4	-57	4.3	0.8	3.4	0.6	0.2	1.3	1.4	0.4	0.3	0.1	0.0	0.0
BMW 3-series	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3	11	-34	0.8	4.7	1.0	0.0	3.1	3.9	0.6	0.0	0.0	0.0	0.0
Ford Mondeo	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.7	0.2	-43	1.7	0.3	0.9	1.1	0.3	0.9	0.0	0.0	0.0	0.0
VW Passat	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	4.0	1.2	2.5	-17	0.2	0.2	0.6	1.0	0.2	0.0	0.0	0.0	0.0
Volvo S40	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3	1.1	0.5	0.6	0.2	-21	0.1	1.4	0.2	0.4	0.0	0.1	0.0	0.0
Citron C5	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.8	0.0	2.7	0.4	0.1	-32	1.0	0.0	0.2	0.0	0.1	0.0	0.0
Toyota Avensis	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.6	0.5	0.6	0.2	0.3	0.1	-37	1.7	0.4	0.0	0.1	0.0	0.0
VW Golf	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.9	0.8	0.4	0.5	0.1	0.0	4.1	-27	2.6	0.3	0.8	0.0	0.0
Opel Astra	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	1.0	0.5	2.2	0.2	0.4	0.1	2.3	6.8	-71	0.3	0.8	0.0	0.0
Toyota Yaris	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.8	0.1	0.0	0.0	0.0	0.0	0.1	1.2	0.3	-12	0.3	0.0	0.0
Peugeot 206	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.1	0.0	1.0	2.4	1.1	0.5	-40	0.1	0.2
Fiat Punto	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.4	3.7	-14	1.0
Daewoo Matiz	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.3	0.3	0.0	4.0	0.6	-42

Table 8: Model price elasticities of best-selling products, computed with all variants 904 products

Volvo V70	-14	0.0	0.0	0.1	0.3	0.0	0.0	0.1	0.7	0.3	0.4	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Nissan X-Trail	0.0	-22	0.3	2.3	0.1	0.0	0.1	0.6	0.3	0.1	0.1	0.2	0.0	0.2	0.1	0.0	0.1	0.1	0.0	0.0	0.0
Honda CR-V	0.0	0.4	-9	2.2	0.0	0.0	0.0	0.0	0.3	1.0	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Toyota RAV4	0.1	2.2	1.6	-12	0.1	0.0	0.0	0.0	0.4	1.0	0.2	0.3	0.0	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.0
Subaru Forester	0.6	0.1	0.0	0.2	-15	0.0	0.0	0.0	2.6	0.1	0.0	0.1	0.3	0.3	0.0	0.0	0.8	0.1	0.0	0.0	0.0
VW Touran	0.0	0.0	0.0	0.0	0.0	-12	0.1	0.1	0.8	1.2	0.3	0.2	0.0	0.1	0.1	0.2	0.2	0.6	0.7	0.0	0.0
VW Passat	0.0	0.1	0.0	0.0	0.0	0.2	-17	2.5	4.0	0.6	1.5	0.1	0.2	1.7	0.2	0.1	0.0	1.0	0.0	0.0	0.0
Ford Mondeo	0.1	0.5	0.0	0.0	0.0	0.1	1.7	-43	0.7	1.1	24	0.1	0.7	0.6	0.3	0.0	0.7	0.3	0.1	0.0	0.0
Audi A4	0.8	0.2	0.2	0.4	1.3	0.9	3.4	0.8	-57	1.3	1.1	0.1	0.2	1.1	0.6	0.7	0.8	1.4	0.2	0.0	0.3
Toyota Avensis	0.2	0.0	0.3	0.4	0.0	0.5	0.2	0.6	0.6	-37	1.1	0.2	2.8	0.2	0.3	2.0	0.1	1.7	0.4	0.0	0.0
Opel Vectra	0.6	0.1	0.0	0.1	0.0	0.2	0.9	26	1.2	2.9	-50	0.1	0.6	0.4	0.4	0.1	1.1	0.6	0.1	0.0	0.0
Renault Megane	0.0	0.3	0.0	0.5	0.1	0.6	0.1	0.2	0.3	1.2	0.1	-56	0.1	0.8	0.2	0.3	1.7	2.2	18	0.2	0.0
Mazda 6	0.0	0.0	0.0	0.0	0.1	0.0	0.3	1.8	0.4	11.2	1.5	0.1	-29	0.4	0.1	0.1	0.0	0.1	0.0	0.0	0.0
Skoda Octavia	0.2	0.3	0.1	0.1	0.2	0.3	2.9	1.5	2.6	1.6	1.0	0.4	0.6	-34	0.2	0.7	0.4	2.4	0.5	0.0	0.2
Volvo S40	0.0	0.1	0.0	0.0	0.0	0.2	0.2	0.6	1.1	1.4	0.9	0.2	0.1	0.2	-21	0.2	1.0	0.2	0.1	0.0	0.0
Toyota Corolla	0.0	0.0	0.0	0.2	0.0	0.2	0.0	0.0	0.7	4.9	0.1	0.2	0.0	0.3	0.2	-18	0.7	1.2	0.4	0.1	1.5
Ford Focus	0.0	0.1	0.0	0.1	0.6	0.3	0.0	0.7	1.3	0.4	1.2	0.9	0.0	0.5	0.6	0.5	-30	0.1	0.1	0.0	0.2
VW Golf	0.0	0.1	0.0	0.1	0.0	0.7	0.5	0.4	0.9	4.1	0.5	1.1	0.0	1.1	0.1	0.6	0.1	-27	0.2	0.6	0.3
Peugeot 307	0.0	0.0	0.0	0.0	0.0	0.7	0.0	0.4	0.2	0.8	0.1	7.4	0.0	0.3	0.0	0.5	0.1	0.3	-17	0.0	0.0
VW Polo	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.2	0.1	0.0	0.4	0.0	0.1	0.0	0.1	0.1	2.8	0.1	-12	0.4
Toyota Yaris	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.8	0.1	0.0	0.0	0.0	0.2	0.0	3.1	0.3	1.2	0.0	0.2	-12

Table 9: Price elasticities of sample products, computed with modal variants only, 199 products

Porsche 911	-6.1	0.07	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
Ast. Martin DB9	0.04	-5.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
Range Rover	0.00	0.00	-6.4	1.28	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
VW Touareg	0.00	0.00	0.34	-12.0	0.27	0.02	0.51	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0
Audi A8	0.01	0.00	0.04	0.28	-16.6	0.18	2.68	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0
Mercedes S-class	0.00	0.00	0.00	0.07	1.60	-51.5	23.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
Mercedes E-class	0.00	0.00	0.00	0.13	1.21	1.06	-5.4	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
BMW 5-series	0.00	0.00	0.00	0.02	0.03	0.00	0.03	-3.2	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0
Audi A4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-15.3	5.34	0.52	0.19	0.13	0.32	0.52	0.10	0.00	0.00	0.00	0.00	0
BMW 3-series	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.5	-13.5	0.00	0.00	0.00	0.00	2.06	0.03	0.00	0.00	0.00	0.00	0
Ford Mondeo	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.09	0.00	-5.2	0.96	0.01	0.02	0.26	0.00	0.00	0.00	0.00	0.00	0
VW Passat	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	3.04	-5.1	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
Volvo S40	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.07	0.01	-3.7	0.04	0.37	0.00	0.00	0.00	0.00	0.00	0
Citron C5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.48	0.00	0.20	0.00	0.10	-22.6	0.83	0.00	0.03	0.00	0.00	0.00	0
Toyota Avensis	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.24	0.25	0.00	0.07	0.09	-4.8	0.00	0.01	0.00	0.00	0.00	0
VW Golf	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.01	0.00	0.00	0.00	0.00	0.00	-19.5	0.06	0.00	0.00	0.00	0
Opel Astra	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.10	-10.3	0.04	0.00	0.00	0
Toyota Yaris	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	-4.2	0.15	0.01	0
Peugeot 206	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.36	-6.7	0.23	0
Fiat Punto	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.01	0.00	0.25	1.66	-11.9	0
Daewoo Matiz	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.80	0.27	-1

Table 10: Price elasticities of best-selling products, computed with modal variants only, 199 products

Volvo V70	-5.8	0.10	0.08	0.06	0.17	0.00	0.00	0.32	0.00	0.04	0.00	0.00	0.00	0.21	0.00	0.01	0.04	0.00	0.00	0.00	0.00
Nissan X-Trail	0.16	-9.0	0.05	0.92	0.00	0.05	0.00	0.60	0.00	0.05	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Honda CR-V	0.09	0.04	-3.0	0.05	0.03	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.01	0.04	0.00	0.04	0.01	0.00	0.00	0.00	0.00
Toyota RAV4	0.20	1.64	0.12	-8.5	0.00	0.02	0.00	0.11	0.00	0.07	0.00	0.00	0.00	0.04	0.00	0.01	0.00	0.00	0.00	0.00	0.00
Subaru Forester	0.61	0.00	0.03	0.00	-10.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.77	0.00	0.00	0.00	0.00
VW Touran	0.00	0.05	0.00	0.02	0.00	-3.9	0.09	0.14	0.01	0.05	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.01	0.02	0.00
VW Passat	0.00	0.00	0.00	0.00	0.00	0.16	-5.1	3.04	0.10	0.00	0.00	0.00	0.00	0.64	0.01	0.00	0.00	0.00	0.00	0.00	0.00
Ford Mondeo	0.32	0.34	0.00	0.04	0.00	0.09	0.96	-5.2	0.09	0.26	0.44	0.00	0.07	0.16	0.01	0.00	0.03	0.00	0.00	0.00	0.00
Audi A4	0.01	0.00	0.00	0.00	0.01	0.04	0.19	0.52	-15.3	0.52	0.50	0.00	0.18	0.02	0.13	0.00	0.02	0.10	0.00	0.00	0.00
Toyota Avensis	0.04	0.03	0.04	0.02	0.00	0.03	0.00	0.25	0.09	-4.8	0.26	0.02	0.83	0.01	0.07	0.00	0.01	0.00	0.00	0.00	0.00
Opel Vectra	0.01	0.00	0.00	0.00	0.00	0.01	0.00	1.02	0.20	0.70	-5.4	0.01	0.05	0.00	0.07	0.00	0.03	0.00	0.00	0.00	0.00
Renault Megane	0.00	0.01	0.00	0.00	0.00	0.04	0.00	0.03	0.00	0.18	0.04	-4.4	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.00
Mazda 6	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.26	0.13	3.84	0.13	0.00	-13.6	0.04	0.02	0.00	0.02	0.00	0.00	0.00	0.00
Skoda Octavia	0.87	0.01	0.08	0.06	0.14	0.02	0.70	0.61	0.01	0.04	0.00	0.00	0.04	-7.8	0.01	0.07	0.79	0.00	0.00	0.00	0.00
Volvo S40	0.01	0.00	0.00	0.00	0.00	0.01	0.01	0.07	0.08	0.37	0.12	0.00	0.02	0.01	-3.7	0.01	0.02	0.00	0.00	0.00	0.00
Toyota Corolla	0.03	0.00	0.08	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.07	0.00	-4.4	0.01	0.00	0.00	0.09	0.68
Ford Focus	0.15	0.00	0.03	0.00	1.13	0.00	0.00	0.11	0.02	0.07	0.11	0.00	0.01	0.85	0.02	0.01	-7.7	0.00	0.00	0.00	0.01
VW Golf	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-19.5	0.01	0.02	0.00
Peugeot 307	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.01	-4.6	0.00	0.00
VW Polo	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.16	0.00	0.03	0.00	-3.2	0.00
Toyota Yaris	0.01	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	1.26	0.01	0.00	0.00	0.00	-4.2

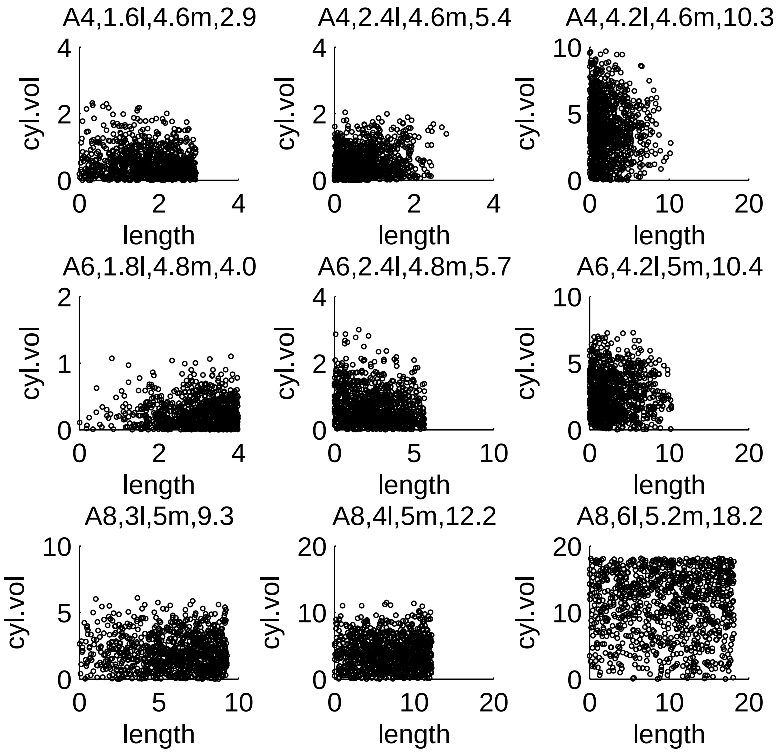

Figure 1: Scatter plots of joint densities of taste coefficients for sample products.

Table 11: Closest substitutes for four sample products, computed using highest cross-elasticities w.r.t. the sample products, when choice set has modal variant only

Make	Model	ltr	m	pri	Bodytype
Volkswagen	GOLF	1.6	4.2	2.3	hatchback
'Bmw	1	1.6	4.2	2.5	hatchback
'Audi	A3	1.6	4.2	2.5	hatchback
'Kia	CERATO	1.6	4.4	2.2	hatchback
'Daewoo	KALOS	1.2	3.8	1.3	hatchback
'Subaru	JUSTY	1.4	3.8	2.0	hatchback
'Daewoo	LACETTI	1.4	4.4	1.6	hatchback
'Audi	A2	1.4	3.8	2.3	MPV/minivan
'Subaru	IMPREZA	1.6	4.4	2.5	station
'Skoda	FABIA	1.2	4.2	1.6	station
'Volvo	V50	1.8	4.6	2.9	station
Toyota	AVENSIS	1.8	4.8	3.0	station
'Mazda	6	1.8	4.8	2.5	station
'Fiat	STILO	1.6	4.6	2.1	station
'Mitsubishi	LANCER	1.6	4.4	2.3	station
'Bmw	3	1.8	4.4	3.2	station
'Nissan	350Z	3.4	4.4	7.9	coup
'Volvo	V50	1.8	4.6	2.9	station
'Audi	A6	1.8	4.8	4.4	station
'Kia	CERATO	1.6	4.4	2.2	hatchback
'Kia	CARENS	1.6	4.4	2.1	MPV/minivan
'Citroen	C5	1.8	4.8	2.9	station
Ford	FOCUS	1.6	4.4	2.3	station
'Mitsubishi	SPACE	1.6	4	2.2	station
'Chrysler	PT	1.6	4.2	2.6	station
'Opel	AGILA	1	3.6	1.5	station
'Hyundai	ACCENT	1.4	4.2	1.7	hatchback
'Skoda	OCTAVIA	1.8	4.6	2.5	station
'Subaru	FORESTER	2	4.4	3.4	station
'Subaru	LEGACY	2	4.8	3.5	station
'Hyundai	ATOS	1	3.6	1.3	hatchback
'Opel	MERIVA	1.6	4	2.1	MPV/minivan
'Ford	FIESTA	1.4	4	1.8	hatchback
Audi	A4	1.6	4.6	3.0	station
'Bmw	3	1.8	4.4	3.2	station
'Volkswagen	CADDY	1.4	4.4	2.1	station
'Daewoo	NUBIRA	1.6	4.6	2.0	station
'Fiat	STILO	1.6	4.6	2.1	station
'Fiat	MAREA	1.6	4.4	2.2	station
'Audi	A3	1.6	4.2	2.5	hatchback
'Volvo	V50	1.8	4.6	2.9	station
'Skoda	FABIA	1.2	4.2	1.6	station
'Fiat	DOBLO	1.6	4.2	2.1	station
'Audi	A2	1.4	3.8	2.3	MPV/minivan

Table 12: Closest substitutes for four sample products, computed using highest cross-elasticities w.r.t. the sample products, when choice set has all variants

Make	Model	ltr	m	pri	Bodytype
Volkswagen	GOLF	1.6	4.2	2.3	hatchback
'Kia	CERATO	1.6	4.4	2.2	hatchback
'Daewoo	LACETTI	1.4	4.4	1.6	hatchback
'Daewoo	KALOS	1.2	3.8	1.3	hatchback
'Bmw	1	1.6	4.2	2.5	hatchback
'Fiat	STILO	1.6	4.6	2.1	station
'Audi	A3	1.6	4.2	2.5	hatchback
'Volkswagen	CADDY	1.4	4.4	2.1	station
'Audi	A2	1.4	3.8	2.3	MPV/minivan
'Opel	ASTRA	1.6	4.2	2.1	hatchback
'Jeep	WRANGLER	2.4	3.8	3.9	Off-road/SUV
Toyota	AVENSIS	1.8	4.8	3.0	station
'Mazda	3	1.6	4.4	2.2	hatchback
'Mitsubishi	CARISMA	1.6	4.4	2.3	hatchback
'Kia	CERATO	1.6	4.4	2.2	hatchback
'Fiat	STILO	1.6	4.6	2.1	station
'Nissan	350Z	3.4	4.4	7.9	coup
'Mitsubishi	LANCER	1.6	4.4	2.3	station
'Alfa Romeo	156	1.8	4.4	3.0	sedan
'Renault	LAGUNA	1.6	4.8	2.5	station
'Honda	CIVIC	1.6	4.2	2.3	hatchback
'Mazda	6	1.8	4.8	2.5	station
Ford	FOCUS	1.6	4.4	2.3	station
'Fiat	MAREA	1.6	4.4	2.2	station
'Alfa Romeo	147	1.6	4.2	2.6	hatchback
'Chrysler	PT	1.6	4.2	2.6	station
'Nissan	PATROL	3	5	7.1	Off-road/SUV
'Opel	ASTRA	1.6	4.2	2.1	hatchback
'Opel	AGILA	1	3.6	1.5	station
'Ford	FUSION	1.4	4	2.0	MPV/minivan
'Ford	FIESTA	1.4	4	1.8	hatchback
'Seat	LEON	1.6	4.2	2.1	hatchback
'Nissan	MICRA	1.2	3.8	1.7	hatchback
Audi	A4	1.6	4.6	3.0	station
'Alfa Romeo	166	2	4.8	4.5	sedan
'Nissan	PATROL	3	5	7.1	Off-road/SUV
'Jeep	WRANGLER	2.4	3.8	3.9	Off-road/SUV
'Bmw	3	1.8	4.4	3.2	station
'Volkswagen	BORA	1.6	4.4	2.5	sedan
'Volkswagen	CADDY	1.4	4.4	2.1	station
'Seat	TOLEDO	1.6	4.4	2.3	sed
'Bmw	5	2.2	4.8	4.7	sedan
'Mercedes-Benz	C	1.8	4.6	3.7	sedan
'Mercedes-Benz	E	2.2	4.8	5.4	sedan

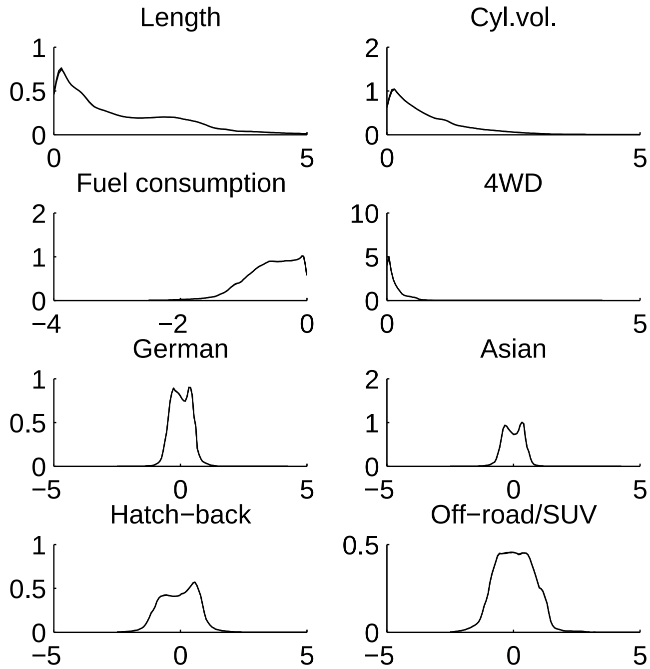

Figure 2: Aggregate (smoothed) marginal densities of some taste coefficients .

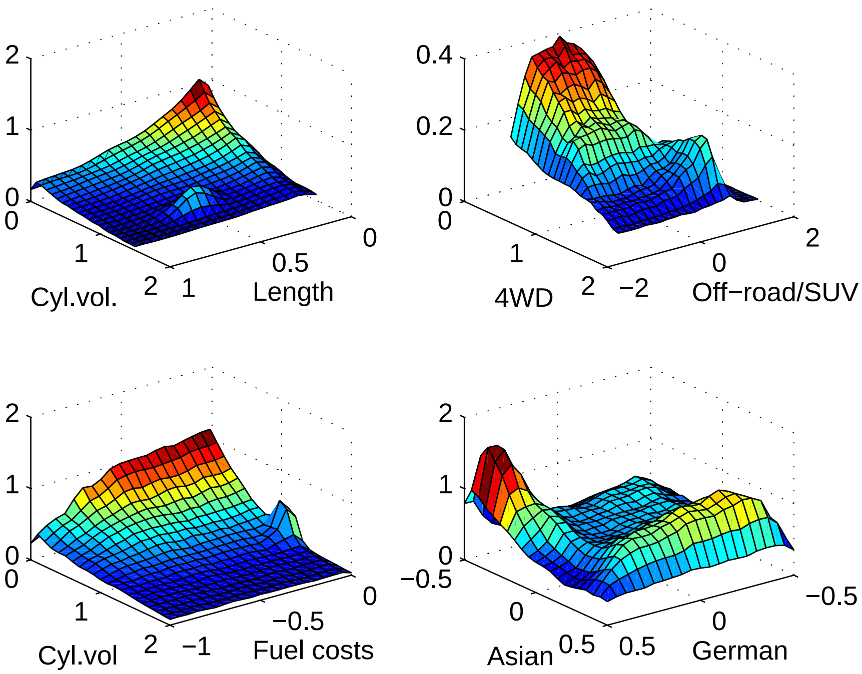

Figure 3: Aggregate (smoothed) pairwise joint densities of some taste coefficients .