Product differentiation with variants, and welfare effects of automobile engine options

Øyvind Thomassen*
Katholieke Universiteit Leuven

December 4, 2010

Abstract

I develop a demand framework for markets where products (models) have variants that differ only with respect to quantifiable product characteristics. There is an unobserved product characteristic and a consumer-specific logit term for models, but both are fixed across variants. The literature’s assumption of orthogonality between unobserved and observed product characteristics is not needed. A counterfactual where car models are restricted to have just one engine option shows that engine options increase profits in spite of tougher competition. Consumer surplus and tax revenue from an engine-size-related tax also increase.

*E-mail: oyvind.thomassen@econ.kuleuven.be. This is a revised version of chapter 1 of my 2009 Oxford DPhil thesis. A previous version was called Automobile engine variants and price discrimination. Thanks to my thesis supervisor Howard Smith, examiners Ian Crawford and Philipp Schmidt-Dengler, and Frank Verboven, Johannes Van Biesebroeck and Laura Grigolon for comments.
1 Introduction

It is a common practice for firms to offer variants of a given product which differ only with respect to one or a few characteristics. Some examples are the 16 and 32GB storage versions of the iPhone 4, different fat-levels of a milk brand, economy and business class seats on a British Airways flight from London to Paris, an internet service provider’s bandwidth packages, a Volkswagen Golf with a 79 or 158hp engine\footnote{The output of a given engine can be adjusted by software settings of the engine control unit.}, and four- or six-core AMD Phenom II processors\footnote{The four-core processor is physically the same as the six-core version, except with two cores disabled.}.

I define two goods as variants of the same product if they differ only with respect to product characteristics that are quantifiable and generic. By quantifiable I mean that there exists a generally accepted method for determining how much of the characteristic a product has. A feature is generic if it does not describe something like “being product X”. Horsepower is quantifiable and generic. Prestige is generic but not quantifiable. Toyota is quantifiable (with a 0-1 variable), but not generic.

The purpose of this article is (1) to point out that markets where products have variants do not fit into existing demand models, (2) to develop an alternative demand model for markets with variants, and (3) to use the model to analyse how engine options affect profits and welfare in the car market.

In the differentiated-products demand literature starting with Berry (1994) and Berry, Levinsohn, and Pakes (1995) (BLP), the utility a consumer gets from a product is a linear function of observable product characteristics. In addition, a product-specific constant captures unobserved product characteristics that affect consumers in the same way. Finally, utility includes a random shock that is iid across products and consumers. The shock repre-
sents likes and dislikes that are idiosyncratic in the sense that a consumer’s taste shock for product A provides no information about his taste shock for product B, or about another consumer’s taste shocks, either for product A or B.

These assumptions are not appropriate for markets with variants. Variants differ only by quantifiable generic characteristics, which in practice means observable variables. The effect of observable product characteristics is fully accounted for by directly including these characteristics in the utility function. Therefore, a difference in utility between two variants must be fully explained by the terms involving the observable characteristics. Letting either the taste shock or the unobserved product characteristic vary across variants of a product is therefore not justified, and imposes product differentiation that does not exist in reality. Accordingly, the model in this paper has unobserved characteristics and idiosyncratic taste shocks at the model-level, but not at the variant level.

I keep the substantive assumptions of the recent automobile demand literature\(^3\) that there are unobserved differences and idiosyncratic tastes for car models.\(^4\) As in many markets, it is plausible that consumers have real idiosyncratic tastes for models. When this is the case, removing the iid shock altogether, as in Bajari and Benkard (2005) and Berry and Pakes (2007), would underestimate product differentiation.

The taste for products introduces a new dimension of product differentiation for each product. Petrin (2002) discusses how this may overestimate the benefits of introducing new goods, since some consumers will have a very large idiosyncratic taste premium for every new product over existing ones, no matter how similar it is to existing products. In this paper shocks

\(^3\)BLP, Petrin (2002), Berry, Levinsohn, and Pakes (2004).
\(^4\)The literature abstracts away from the variant issue: each consumer chooses a car model whose price and engine characteristics are chosen by the researcher from among the model’s range of engine variants. This is likely to affect the estimated coefficients on price and engine characteristics.
enter only at the model level. Therefore the number of dimensions of product differentiation does not vary when the number of variants per model is changed.

The demand model permits identification with weaker assumptions than in the literature. Goldberg (1995) restricts the unobserved product characteristic to be the same across car models of a given brand. Brand-level fixed effects control for unobservable characteristics. BLP relax this assumption by including model-specific dummies to control for unobservables at the model level. But these dummies capture the mean effect of observed characteristics as well as unobserved characteristics. To separate the two effects, BLP need to assume orthogonality between unobserved and observed characteristics and instruments for price (since price is correlated with the unobserved characteristic). Any model that allows for alternative-specific unobservables needs this orthogonality assumption. However, the assumption is problematic because both types of product characteristics are choice variables of manufacturers.

I relax both the orthogonality assumption of BLP and Goldberg’s restriction that unobservables are fixed across models of a brand. Since unobservables do not vary across variants, I can control for unobservables with model-level fixed effects. My error term is an individual-level prediction error explained by sampling error, as is standard in micro-data discrete-choice models. Since all unobservable product characteristics are captured by the fixed effects, they are not in the error term. It follows that price is not endogenous, so no instrumental variables are needed. Price and product characteristics vary within models, so that their effects are separately identified.

In the car market, models, such as Volkswagen Golf, Toyota Camry, or BMW

---

5BLP micro (Berry, Levinsohn, and Pakes 2004) in principle need it too, but instead use calibration to find the unidentified parameters.

6E.g. Ackerberg, Benkard, Berry, and Pakes (2007) p. 4197 say: “There are plausible reasons to believe that product characteristics themselves are correlated with $\xi$ [the unobservable]”.

---
3-series, differ in the eye of the consumer with respect to design, prestige, dealership/service access, advertising exposure and other features that are usually both unobservable for researchers and partly uncorrelated across consumers and products. In addition to models there is extensive product differentiation through variants: options with respect to mechanical internals (engine, transmission), body style and trim levels increase the total number of products enormously. For instance a new Volkswagen Golf has 15 engine variants and Peugeot 407 has 13. Focusing on engine variants, I ask whether this differentiation is excessive from the point of view of car companies, and to what extent it benefits consumers.

Dixit and Stiglitz (1977) discuss how two opposite effects may mean that either too few or too many products are introduced: introducing a new product has a positive externality on consumers, because the firm cannot capture the entire consumer surplus, and a negative externality on other firms because of business stealing. In a model with vertical and horizontal differentiation similar to the empirical model in this paper, Gilbert and Matutes (1993) show that firms may offer several qualities each when they would be better off committing to offer only one, again because they do not take into account the business stealing externality on other firms.

I look at a simple counterfactual where I restrict each car model to have only one engine option. Comparing equilibrium prices and sales to those in the actual market, I find that going from one engine per model to the number in fact offered has the following results: (1) Welfare increases through (a) a 4.4% increase in consumer surplus, (b) an 8.9% increase in engine-size-related tax revenues and VAT, and (c) a 2.4% increase in profits. (2) Profits go up because of increased total sales (i.e. new consumers recruited from the outside good because of increased product variety), in spite of a small reduction in average markups.

---

7 I assume firms would choose to keep the variant currently accounting for the highest profits, and that there are no fixed cost savings in restricting the range of engine options.
The next section describes the utility model and derives the choice probabilities. Section 3 explains the identification strategy and compares it to the literature. Section 4 describes the data, and attempts to show that the assumptions of the model are satisfied by the data. Section 5 describes the estimator, section 6 the results. Section 7 gives the results from the counterfactual experiment.

2 Model

2.1 Utility

I estimate the parameters of the conditional (on product choice) indirect utility functions of potential car buyers. Products are grouped into models. Products within a model, called variants, differ only with respect to observable characteristics. Products in different models differ with respect to characteristics that are not observable to the researcher, but known by market agents. Let \( \mathcal{J} = \{0, \ldots, J\} \) be the choice set, where \( j = 0 \) is the outside good. Each product is described by a double \((x_j, m(j))\) - a vector of observable product characteristics and an index for the model to which the product belongs. Let the vector \( x_j^0 \) be the subvector of \( x_j \) consisting of the variant characteristics (all characteristics that vary across variants of a model) and let \( x_{m(j)}^1 \) be the remaining characteristics (model characteristics), so \( x_j = (x_j^0, x_{m(j)}^1) \). Let \( p_j \) be price, \( y_i \) income, and \( z_{i} \) a vector of observable consumer characteristics. Conditional on choosing product \( j \) consumer \( i \) has

---

8 Individuals or products are not followed across time. I.e. they have different subscripts for different years of data.
indirect utility:

\[ u_{ij} = -(\bar{\alpha}_i + \tilde{\alpha}_i)p_j + x^0_j(\bar{\beta}_i^0 + \tilde{\beta}_i^0) + x^1_{m(j)}\bar{\beta}_i^1 + \delta_{m(j)} + \epsilon_{im(j)} \]  

(1)

\[ u_{i0} = \xi_0 + \epsilon_{i0} \]  

(2)

\[ \bar{\alpha}_i = \tilde{\alpha}(z_i, y_i) \]  

(3)

\[ \bar{\beta}_i^r = \tilde{\beta}^r(z_i, \nu_i), \quad r = 0, 1 \]  

(4)

\[ \epsilon_{im(j)} \sim \text{iid extreme value,} \]  

(5)

where \( \tilde{\alpha} \) and \( \tilde{\beta} \) are functions of consumer characteristics whose realisations \((z_i)\) or distributions \((y_i, \nu_i)\) are known.

The model mean effects, \( \delta \), can in theory be decomposed into the mean effect of model characteristics and unobservable characteristics: \( \delta_m = x^1_{m(j)}\tilde{\beta}_i^1 + \xi_m \). For most applications, the decomposition of \( \delta \) is not required, nor is it identified with the assumptions used in this paper. Unlike in the literature, \( \delta \) does not contain the mean effects of price and variant characteristics.

\( \xi_m + \epsilon_{im} \) captures the common and consumer-specific valuation of any aspect of model \( m \) that is not controlled for by the observable characteristics \( x \). Since variants of a model differ only with respect to observable characteristics \( x^0 \) these terms have \( m \)-subscripts.

\( \epsilon \) captures effects of unobservable characteristics which do not provide any information about the unobservable effect of another product on the same consumer, nor on the effect on a different consumer. An example is consumer A who likes the look of VW Golf, but dislikes the Ford Focus; consumer B who thinks the Golf is boring, and the Focus elegant; consumer C who likes the look of the Golf, and loves the Focus.
2.2 Choice probabilities

Denote the ‘observable’ part of utility $u - (\delta + \epsilon)$,

$$V_{ij} = -(\bar{\alpha} + \bar{\alpha}_i)p_j + x^0_j(\bar{\beta}^0 + \bar{\beta}^1_i) + x^1_{m(j)}\bar{\beta}^1_i.$$  

Conditional on the realisations of the random variables $(y, \nu)$ the probability that consumer $i$ chooses model $m$ is

$$Pr(m|i, y, \nu) = Pr\left(\max_{j \in J^m}\{V_{ij}\} + \delta_m + \epsilon_{im} > \max_{j' \in J^m}\{V_{ij'}\} + \delta_{m'} + \epsilon_{im'}, \ m' \neq m\right)$$

$$= \frac{\exp(\max_{j \in J^m}\{V_{ij}\} + \delta_m)}{1 + \sum_{m' \in M} \exp(\max_{j' \in J^{m'}}\{V_{ij'}\} + \delta_{m'})},$$

where $J^m$ is the set of variants of model $m$, and $M$ is the set of models in the market. Conditional on $(y, \nu)$, the $\max\{V\}$-terms are not random variables. The second equality therefore follows in a standard way by integrating over the distribution of the $\epsilon$-term to obtain a logit choice probability.

Still conditional on $(y, \nu)$, consumer $i$’s probability of choosing $j$ is equal to $Pr(m|i, y, \nu)$ if $j$ maximises $\{V_{ij}\}$ over $J^m$, and zero otherwise. Letting the indicator function $1[j|m, i, y, \nu]$ be one in the first case and zero in the second case, we can write the choice probability of product $j$

$$Pr(j|i, y, \nu) = 1[j|m, i, y, \nu] \frac{\exp(\max_{j \in J^m}\{V_{ij}\} + \delta_m)}{1 + \sum_{m' \in M} \exp(\max_{j' \in J^{m'}}\{V_{ij'}\} + \delta_{m'})},$$

Integrating over the distributions of $y$ and $\nu$, we now get consumer $i$’s unconditional choice probability for product $j$

$$Pr(j|i) = \int \int Pr(j|i, y, \nu) f(y|i) f(\nu) dy d\nu,$$

where $f(\nu)$ is the joint density of $\nu$, and $f(y|i)$ is the density of the em-
empirical income distribution in the population for consumers with observed characteristics $z_i$ in the time period $t(i)$ in which $i$ enters the data.

3 Identification

The model is estimated by assuming orthogonality between the individual-level prediction error (difference between observed choices and choice probabilities) and explanatory variables. This error term is distinct from the literature’s unobserved product characteristic. I control for unobserved product characteristics using fixed effects for car models in the utility function. Unobserved product characteristics therefore do not enter the error term. The error term is pure sampling error. Price and product characteristics are therefore exogenous. I do not need orthogonality between unobserved and observed product characteristics to identify the price parameter, since it is identified by within-model variation separately from model fixed effects.

3.1 Unobserved product characteristics and identifying assumptions

In discrete-choice demand models consumers choose among alternatives by maximising a utility function that depends on observed product characteristics. But typically consumers also care about product characteristics that are not observed by the researcher. Berry (1994) and BLP solve this problem by introducing product-specific constants in utility.

The product-specific constants ensure that utility includes the effect of unobservables. But the product-specific constants also capture the mean (across consumers) effect of price and observed characteristics. In order to separate the two effects, the literature assumes that unobserved and observed product characteristics are orthogonal. That is, the observed characteristics of a

---

9When observations are at the level of products, like in BLP, the orthogonality assump-
product provides no information that affects the expected value of the unobserved characteristic of that product. Since both unobserved and observed characteristics are choice variables of firms, this assumption is problematic.

An alternative identification strategy, which does not require orthogonality between unobserved and observed characteristics, is to use variation across products for which the unobserved characteristic can be assumed fixed. In this way Goldberg (1995) uses fixed effects for automobile brands to control for unobservable product characteristics.

My identification strategy uses the same principle. I restrict unobservable characteristics to be the same across engine variants of each car model. Since I observe products at the variant level, I can control for unobservables at the model-market level (rather than brand) with fixed effects. I therefore do not need either Goldberg’s restriction that unobservables are the same within a brand, or BLP’s orthogonality assumption.

10 Ackerberg, Benkard, Berry, and Pakes (2007) p. 4197 say: “There are plausible reasons to believe that product characteristics themselves are correlated with ξ [the unobservable]. After all the product design team has at least some control over the level of ξ, and the costs and benefits of producing different levels of the unobservable characteristic might well vary with the observed characteristics of the product.” Ackerberg and Crawford (2006) say: "Just like price, product characteristics are typically choice variables of firms, and as such one might worry that they are actually correlated with unobserved components of demand."
3.2 The error term

The model is estimated by assuming orthogonality between the individual-level prediction error (difference between choice probabilities and the indicator function for observed choices) and the explanatory variables (product and consumer characteristics).

When observed market shares are generated by a large number of consumers, sampling error (the difference between a market share generated by a small number of consumers and the choice probability of the true model) should go to zero, i.e. choice probabilities should exactly equal observed market shares. This paper uses individual-level data, but I observe only the sex and age of the buyer, so that the data can be regarded as market-share data for submarkets defined by the age and sex of the buyers. In each market there is only about 500 individuals distributing between approximately 300 (inside) alternatives. On average, therefore, each observation (group market share) is generated by less than two individuals. Since the true model’s choice probabilities for each submarket will not be attained with such a low number of draws, there will be a discrepancy between observed and predicted shares. The prediction error is therefore explained by sampling error.

The literature emphasises that price is endogenous because the error term is the unobserved product characteristic. In this paper, the error term is not the unobserved characteristic, but rather the individual-level prediction error. Unobserved product characteristics are controlled for by the model fixed effects. Therefore they do not enter the error term. I assume individual-level sampling error does not enter the structural pricing (supply) equation. Price is therefore exogenous.
4 Data

The first subsection describes the data. The assumption that there are no unobservable differences between variants is central to the paper’s identification strategy. Subsection 4.2 discusses how care has been taken to ensure that trim levels do not cause unobservable differences between engine variants. Subsection 4.3 discusses further issues concerning trim levels. The reader who does not care about this issue may skip subsections 4.2 and 4.3.

4.1 Data

The data set is constructed by combining two data sources: new vehicle registrations and price lists from Norway 2000-2007. The registration data give the number of units sold by the sex and age of the buyer. I define models as products of the same brand (e.g. Toyota) and of the same nameplate (e.g. Corolla). Variants are defined by the engine, as specified by horsepower and whether it is diesel or not. The sales of a variant is the sum of sales over all products of a model which are the same in terms of horsepower and fuel type. (In the literature, for instance in BLP, sales are the sums over products with the same nameplate.) Prices are list prices, not transaction prices. However, car importers in Norway have a stated policy of resale price maintenance at fixed prices.

Both were provided by Opplysningsrådet for Vegtrafikken AS (the information council for road traffic).

In the sales data, products are defined by the following characteristics: brand, model (nameplate), body type, cylinder volume, horsepower, fuel type, number of seats and drive-wheels (two-wheel drive vs. four-wheel drive). Price lists include the further product characteristics length, weight, fuel consumption at mixed driving, airconditioning, number of collision bags, number of gears, automatic vs. manual gears, whether frontwheel or rear-wheel drive if 2WD, number of doors, and styling package: a set of features (predominantly aesthetic) summed up by a tag, like ‘sportline’ or ‘comfort’.

So-called “net prices” or “already-bargained prices” have been standard practice in the industry since around 1998-99, as referred in the country’s largest newspaper Verdens Gang in several articles, “More difficult to bargain”, 30.08.1998, http://www.vg.no/bil-og-motor/artikkel.php?artid=27325 and “Price drop on new cars”
Table 1: Descriptive statistics

<table>
<thead>
<tr>
<th>Year</th>
<th>Variable</th>
<th>Mean</th>
<th>s.d.</th>
<th>18-35</th>
<th>36-51</th>
<th>52-</th>
<th>18-35</th>
<th>36-51</th>
<th>52-</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>variants</td>
<td>241</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>models</td>
<td>82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>price(^b)</td>
<td>327</td>
<td>157</td>
<td>292</td>
<td>306</td>
<td>276</td>
<td>242</td>
<td>243</td>
<td>225</td>
</tr>
<tr>
<td></td>
<td>tax</td>
<td>97</td>
<td>60</td>
<td>88</td>
<td>91</td>
<td>79</td>
<td>68</td>
<td>67</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>unit sales</td>
<td>56300</td>
<td></td>
<td>7961</td>
<td>15433</td>
<td>15483</td>
<td>4156</td>
<td>8008</td>
<td>5259</td>
</tr>
<tr>
<td>2007</td>
<td>variants</td>
<td>375</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>models</td>
<td>94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>price</td>
<td>351</td>
<td>194</td>
<td>318</td>
<td>335</td>
<td>310</td>
<td>273</td>
<td>279</td>
<td>245</td>
</tr>
<tr>
<td></td>
<td>tax</td>
<td>154</td>
<td>127</td>
<td>123</td>
<td>133</td>
<td>119</td>
<td>100</td>
<td>103</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>unit sales</td>
<td>78139</td>
<td></td>
<td>10907</td>
<td>21499</td>
<td>23064</td>
<td>4444</td>
<td>10245</td>
<td>7980</td>
</tr>
</tbody>
</table>

\(a\)Not sales-weighted
\(b\)Price and tax in 1000s of 2004 kroner, adjusted by CPI.

Table 1 shows summary statistics for the first and last of the eight years in the data. The data have the sales of 491,853 units, spread over 2397 products, 661 models (giving an average of 3.6 variants per model), 48 age groups and two sex groups. There is considerable variation along dimensions useful for identifying the parameters: the number of products across years, average car tax in different years, and in the purchases of different consumer groups.

To allow consumers the choice of not purchasing a new car I need an estimate of the total market size. Since my data are by individual rather than by household, I could use the total population in each demographic (sex-age) group to estimate market size. However, on average 98.5% of people do not buy a new car in a given period. Since the goal is to analyse substitution patterns for cars, it seems reasonable to focus on the preferences of a more


14 Differences in average characteristics purchased between groups for a given choice set reveal the role of demographic variables. Correlations in sales across time within groups when the choice set changes (number of products, price, characteristics) reveal the effect of unobserved heterogeneity. Differences in (groupwise) sales within a model reveal the mean effects of price and engine characteristics.
closely defined group.\textsuperscript{15} I let the market size of a consumer group be twice the maximum (across periods in the data) number of people in that group who bought a new car.

The income distributions conditional on demographic group are from the population as a whole, not from the potential car buyers as I have defined them, but this is the best approximation available. I use data from Statistics Norway on the number of people in twelve sex-age groups belonging to each of nine income brackets to estimate a kernel-smoothed income probability density function for each population group.

4.2 Trim level: characteristics that vary within product units

In general, tractability concerns dictate that not all product characteristics can be included in the econometric model. Typically, for a given product unit as defined in the econometric model (e.g. nameplate/engine) the consumer faces a choice of other characteristics (e.g. transmission, leather interior), which affect the price of the product.

Denote the bundle of characteristics not included the ‘trim level’.\textsuperscript{16} To assign one price to each variant it is necessary to choose a trim level for each product. Denote this the ‘baseline’ trim level. A central identifying assumption in this paper is that there are no unobserved differences between variants. In the following I discuss trim levels in some detail in order to show that this issue does not cause any violation of the restriction on unobservables.

\textsuperscript{15} Also, for a given number of simulation draws, the larger the share of the outside good, the fewer simulation draws will result in a choice of one of the inside alternatives, making the simulation of the inside market shares less accurate. The problem could be reduced by oversampling draws that land on inside goods, like BLP do, but it requires an initial estimation with standard simulation techniques, which is difficult to do with any accuracy given the large computational burden of the model.

\textsuperscript{16} Depending on specific modelling assumptions, my definition of the expression ‘trim’ may be wider than common usage.
Trim levels do not cause any unobserved difference between variants as long as three conditions are satisfied: the same set of trim upgrades over the baseline is available for all variants of a model; the price of upgrades is the same for all variants; and variants within a model are assigned the same baseline trim level. Together with additively separable utility, this ensures that consumers choose trim upgrade independently of variant. The next three paragraphs look at how each condition is satisfied.

In some cases not all variant/trim-combinations show up in the price lists. But examination of car brands’ national web pages show that every engine variant is available with the full choice of trim levels in almost all cases. The few exceptions are mostly variants with engine sizes that are outliers relative to the model range, whose market shares are extremely small, which are not offered with the cheapest trim levels.

The price lists show that a given trim upgrade almost always costs the same regardless of engine size. That is, if you have to pay $321 to upgrade from “basic” to “super” in the 2.0litre variant, you also pay $321 for the same upgrade with the 2.4litre variant. I use this regularity to infer prices for variant/trim-combinations that are missing in the data. For the variants that still do not have a price at the baseline trim after imputing prices in this way, I reassign sales to nearest neighbour (according to horsepower, and if tied, fuel type).

The final condition, that all variants of a model are assigned the same baseline

---

17 In fact a weaker condition on availability is sufficient: let \( J_i^m \subset J^m \) be the subset of variants in model \( m \) such that consumer \( i \) prefers any variant in \( J_i^m \) to any variant in \( J^m - J_i^m \) regardless of trim level (even if the variants in the first set have the worst possible trim and those in the second the best possible). Then it is sufficient that the trim level that \( i \) prefers is available for all variants in \( J_i^m \), since the availability of trim will not change the choice of variant. It follows that for instance unavailability of low-end trim with top-end engine variants is unlikely to contaminate the estimated tastes for engine size.

18 Because of heterogeneity in price sensitivity, this framework is consistent with more expensive variants being sold with more expensive trim: people who do not mind paying extra for more horsepower may not mind paying extra for leather seats.
trim is up to the researcher as long as the corresponding prices are available. Since I do not have prices for every variant-trim combination, I chose as baseline the package for which I have prices for the variants corresponding to the highest sales (if tied, the cheapest package).\textsuperscript{19}

This discussion shows that the issue of baseline trim does not cause any unobserved differences between variants. Still the choice of baseline may be of some consequence. But as I discuss in the next subsection, these consequences are likely to be negligible - especially by comparison to the literature, where baseline trim also includes engine characteristics.

4.3 Trim level here and in the literature

A consumer’s valuation of the baseline trim level will enter the unobserved characteristic. Choosing “super” instead of “basic” as baseline for model $m$ will increase the prices of all variants in $m$ by an equal amount. To ensure that $m$ keeps its market share $\delta_m$ will adjust upwards. For any given consumer the effect is exactly the same on all variants, and therefore does not create unobserved differences between variants. But the choice of baseline may still have an effect: consumers with a low price sensitivity will substitute into model $m$ (which has become more attractive since they exchange money against $\delta$ at a higher rate than the average consumer; and an equal amount with high price sensitivity will substitute out of the model). This may in principle affect the estimated distribution of price sensitivity. But the effect of these trim differences is likely to be negligible compared to the effects of engine characteristics, and model characteristics such as segment and size.

\textsuperscript{19}The raw data have total sales of 523,702 over eight years. For 499,167 of these matches are found between price lists and sales data. Failures may be due to privately imported obsolete products. Observations that cannot be matched are discarded. The number of sales for which there is a match between the two data sources, and for which I can find a price (at a baseline trim level) is 440,821. Remaining sales are reattributed to nearest neighbours. Finally I discard models with sales in a given year of less than 100 units. This reduces the number of products from 3,588 to 2,397 and total sales to 491,853.
The choice of baseline is of less consequence in this paper than in the model-level literature. There are three reasons for this: first, in model-level studies, trim includes engine upgrades over the baseline engine. Since this makes trim account for a larger share of utility, the magnitude of the problem is greater. Secondly, I do not estimate tastes for components of trim, while model-level studies do, since engine characteristics are in trim. Increasing baseline engine size therefore changes components of $x$. This makes consumers with low tastes for engine size substitute out of the model, which may affect estimates further. Thirdly, engine upgrades may be correlated (in the population of car models) with baseline engine sizes (in ways that may depend on the choice of baseline). Since the engine upgrades enter the unobserved characteristic, $\xi$, they may be an additional source of dependence between $\xi$ and $x$, violating the orthogonality assumption used in the literature.

5 Estimation

5.1 The objective function

Parameters are estimated by GMM, with moments

$$g_{ij}(\theta) = Z_{ij}[d_{ij} - Pr(j|i, \theta, \delta(\theta))],$$  

where $Z$ is a vector of functions of the explanatory variables. $d_{ij}$ is one if consumer $i$ chooses alternative $j$, zero otherwise. $Pr(j|i, \theta, \delta(\theta))$ is consumer $i$’s choice probability for product $j$ as defined in the model section (equations (7-9)). The model-market fixed effects, $\delta$, are a function of the parameters, found by setting the aggregate (across consumers) model choice probabilities equal to model market shares: $s_m = P(m|\theta, \delta)$.

---

20 parameters to estimate, 22 instruments: price, price$^2$, kw, kw$^2$, kw$^3$, age, age$^2$, age$^3$, wom, fuelcost, fuelcost$^2$, fuelcost$^3$, age, fuelcost, age$^2$, fuelcost, wom, diesel, diesel$^2$, age, diesel, age$^2$, diesel, wom, length, length$^2$, age, age$^2$, wom, constant.
The objective function is
\[
\left[ \sum_{i=1}^{n} \sum_{j \in J^{t(i)}} g_{ij}(\theta) \right]' W \left[ \sum_{i=1}^{n} \sum_{j \in J^{t(i)}} g_{ij}(\theta) \right], \text{ where (11)}
\]
\[
W = [(n)^{-1} \sum_{i=1}^{n} \sum_{j \in J^{t(i)}} Z_{ij}' Z_{ij}]^{-1},
\]
where \( J^{t(i)} \) is the choice set of consumer \( i \), depending on the time period where \( i \) is observed, and \( n \) is the number of individuals observed (sum of market sizes over groups and years). Efficiency could be improved by approximating the ideal instruments using initial estimates. But because of the large computational burden, and because the size of the data set makes efficiency relatively less of a worry I choose to use estimates from (11) as final.

The vector \( \delta \) that sets predicted and observed model market shares equal is found by the BLP contraction mapping.\(^{21}\) Conditional on the parameters, the model choice probabilities are an average over logit choice probabilities, like in BLP.

5.2 Simulation and asymptotic properties

The integral with respect to the density of \( \nu \) and \( y \) in the choice probabilities is computed by a frequency simulator.\(^{22}\) The simulator is not smooth in the parameters. I analytically integrate over the distribution of the logit term, but the choice probabilities still contain discontinuous functions of the parameters: the indicator function \( 1[j|m] \), and a max function for every

\(^{21}\delta^{t+1}_m = \delta^{t}_m + \log s_m - \log P(m|\theta, \delta^{t})\)

\(^{22}\)Using 30 quasi-random draws (scrambled Halton, generated by Matlab’s haltonset) for each of the 96 consumer groups in each of the 8 years. See Train (2003) for a discussion of Halton draws.
Indicator functions or max functions may change in jumps or not at all for a given change in their argument, making the same true for the objective function that depends on them. McFadden (1989) (Theorem 1, p. 1014) shows consistency and asymptotic normality for an estimator like (11) where $g$ is simulated by a function allowed to have jumps. A consistent estimator for the asymptotic covariance matrix is

$$A\text{var}(\hat{\theta}) = (\hat{G}'W\hat{G})^{-1}G'W\hat{\Lambda}W(\hat{G}'W\hat{G})^{-1}/n$$  \hspace{1cm} (12)

$$\hat{G} = n^{-1} \sum_{i=1}^{n} \sum_{j \in J(i)} \nabla_{\theta} g_{ij}(\hat{\theta})$$  \hspace{1cm} (13)

$$\hat{\Lambda} = n^{-1} \sum_{i=1}^{n} \sum_{j \in J(i)} \sum_{j' \geq j} g_{ij}(\hat{\theta})g_{ij'}(\hat{\theta})'.$$  \hspace{1cm} (14)

Berry, Linton, and Pakes (2004) show that under assumptions typical in the IO demand literature, consistency requires that the number of simulation draws must grow as the square of the number of products. There are two reasons that the result does not apply to the model in this paper (so that I can rely on the results of McFadden (1989)). First, they analyse a situation of dependence between observations in market-level data, because one market share depends on the shares of other products. In this paper observations are individual purchases which do not depend on each other. Secondly, they

---

23I did not want to use a smoothed simulator (such as a logit-smoothed accept-reject) as discussed in McFadden (1989), because it would effectively remove the feature of my model that variants differ only with respect to observed characteristics. In principle one could create a smooth simulator by decomposing the unconditional probability into a sum of probabilities conditional on which variant is chosen in each model, times the probability of the conditioned-on event. If there are $M$ models and each model has $V$ variants, the sum will have $M^V$ terms, i.e. approximately $661^{3.6} = 1.4E+10$ in this paper. The computational burden is roughly proportional to the number of separate choice probabilities which need to be computed, which is proportional to the number of draws and to the number of terms in the sum above. Therefore the computational burden goes down with the smooth simulator relative to the one I use only if we can reduce the number of draws by a factor of $M^V$ to obtain an equally good simulator. A second alternative would be to use a.

24See his assumption A12 p. 1018 for a regularity condition on the behaviour of the simulator.
look at a model where the error term (the unobservable characteristic) is a nonlinear transformation of the predicted market shares, so that simulation errors in the choice probabilities will cause a bias. In this paper the error term is a linear function of the choice probabilities, so that simulation error cancels across observations and disappears as the number of observations goes to infinity with the number of simulation draws held fixed.

Because of the irregularity of the objective function I need an optimisation algorithm that does not require continuity and is robust to local optima. I use the differential evolution genetic algorithm.\textsuperscript{25}

6 Results

6.1 Parameter estimates

Table 2 shows the estimates from the full model, and table 3 gives variable definitions and units. The mean taste coefficients, obtained by interacting with the mean values of the demographic variables, are: horsepower -7.7, fuel cost 18.7, diesel 0.6, and price -16.6. (For length and the constant for the outside good the means are contained in $\delta$ and therefore unknown.)

The first two do not have the expected signs. Clearly the analysis does not succeed in disentangling these highly correlated effects. However, since the variables represent aspects of the same underlying feature, engine size, this does not necessarily compromise the model's ability to predict substitution patterns. A consumer usually cannot change one of these characteristics without changing the other, and so the effect of having a different engine in practice works through both characteristics. Gramlich (2009) discusses this

\textsuperscript{25}Developed by Kenneth Price and Rainer Storn, implemented for Matlab in the \textit{devec3} code, modified to compute population members in parallel. The code can be found on www.icsi.berkeley.edu/~storn/code.html.
Table 2: Parameter estimates from the full model. 491,853 observations.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Interacted with</th>
<th>est.</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>horsepower</td>
<td></td>
<td>4.767</td>
<td>0.007</td>
</tr>
<tr>
<td>fuel cost (kr/km)</td>
<td></td>
<td>4.932</td>
<td>0.007</td>
</tr>
<tr>
<td>diesel</td>
<td></td>
<td>-0.011</td>
<td>0.001</td>
</tr>
<tr>
<td>horsepower</td>
<td>age</td>
<td>-12.303</td>
<td>0.024</td>
</tr>
<tr>
<td>age sq.</td>
<td></td>
<td>-24.685</td>
<td>0.033</td>
</tr>
<tr>
<td>woman</td>
<td></td>
<td>1.425</td>
<td>0.010</td>
</tr>
<tr>
<td>fuel cost (kr/km)</td>
<td>age</td>
<td>12.172</td>
<td>0.021</td>
</tr>
<tr>
<td>age sq.</td>
<td></td>
<td>30.062</td>
<td>0.042</td>
</tr>
<tr>
<td>woman</td>
<td></td>
<td>-1.604</td>
<td>0.013</td>
</tr>
<tr>
<td>diesel</td>
<td>age</td>
<td>0.044</td>
<td>0.003</td>
</tr>
<tr>
<td>age sq.</td>
<td></td>
<td>2.158</td>
<td>0.004</td>
</tr>
<tr>
<td>woman</td>
<td></td>
<td>0.117</td>
<td>0.008</td>
</tr>
<tr>
<td>horsepower</td>
<td>std.norm.</td>
<td>0.435</td>
<td>0.005</td>
</tr>
<tr>
<td>fuel cost (kr/km)</td>
<td>std.norm.</td>
<td>-0.167</td>
<td>0.004</td>
</tr>
<tr>
<td>diesel</td>
<td>std.norm.</td>
<td>1.229</td>
<td>0.015</td>
</tr>
<tr>
<td>length</td>
<td>std.norm.</td>
<td>-0.102</td>
<td>0.007</td>
</tr>
<tr>
<td>inside good (const.)</td>
<td>std.norm.</td>
<td>7.719</td>
<td>0.010</td>
</tr>
<tr>
<td>price$^a$</td>
<td>$\alpha_1$</td>
<td>2.935</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>$\alpha_2$</td>
<td>47.082</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>$\alpha_3$</td>
<td>3.614</td>
<td>0.007</td>
</tr>
</tbody>
</table>

$^a$The price coefficient is $-[\alpha_1 + \alpha_2 \exp(-\alpha_3 \cdot income_i)]$.

Table 3: Variables and units

<table>
<thead>
<tr>
<th>Variable</th>
<th>Unit</th>
<th>mean</th>
<th>st.dev.</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td>price$^a$</td>
<td>kroner*1E-6</td>
<td>0.34</td>
<td>0.18</td>
<td>0.12</td>
<td>1.87</td>
</tr>
<tr>
<td>horsepower</td>
<td>kW*1E-2</td>
<td>0.98</td>
<td>0.39</td>
<td>0.37</td>
<td>3.75</td>
</tr>
<tr>
<td>fuel cost</td>
<td>(kr/litre)*(litres/km)</td>
<td>0.77</td>
<td>0.19</td>
<td>0.38</td>
<td>1.86</td>
</tr>
<tr>
<td>diesel</td>
<td>1 if diesel, 0 if petrol</td>
<td>0.34</td>
<td>0.47</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>length</td>
<td>metres</td>
<td>4.40</td>
<td>0.32</td>
<td>3.41</td>
<td>5.08</td>
</tr>
<tr>
<td>age$^b$</td>
<td>age*1E-2</td>
<td>0.50</td>
<td>0.23</td>
<td>0.23</td>
<td>0.70</td>
</tr>
<tr>
<td>age sq.</td>
<td>age squared *1E-4</td>
<td>0.27</td>
<td>0.05</td>
<td>0.05</td>
<td>0.49</td>
</tr>
<tr>
<td>woman</td>
<td>1 if woman, 0 if man</td>
<td>0.29</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>income</td>
<td>kroner*1E-6</td>
<td>0.34</td>
<td>1E-3</td>
<td>3.28</td>
<td>21</td>
</tr>
</tbody>
</table>

$^a$price, fuel cost and income in 2004 kroner, adjusted by CPI. 100 kroner, abbr. ‘kr’, is about 12 euros or 17 US dollars.

$^b$Persons of age <23 are assigned age 23 and with age >70 assigned age 70.
I regress horsepower on fuel cost over all products to obtain a rough measure of the technological connection between the two. The slope coefficient is 1.45. So if, starting from the average car, we move to a car with fuel cost 0.1 (kr/km) higher this car would typically have horsepower 0.145 (14.5 kW) higher. The average consumer would be willing to pay \((-7.7 \cdot 0.145 + 18.7 \cdot 0.1)/16.6 = 0.0454\), i.e. 45,540 kroner for this change. To see that this is roughly in line with market conditions, a regression of price on horsepower and fuel cost yields coefficients 0.43 and -0.13, respectively. Plugging in the changes in the two variables, we can expect price to go up by \(0.43 \cdot 0.145 - 0.13 \cdot 0.1 = 0.0493\), i.e. 49,300 kroner.

### 6.2 Price elasticities

I compute price elasticities for models by taking derivatives with respect to a change in the prices of all variants in a model as a percentage of the sales-weighted mean price. Own-price elasticities range from -3 to -7, and absolute value tends to increase with price. For semi-elasticities this pattern is reversed.

Table 4 shows own- and cross-price elasticities (multiplied by 100 for readability) for a sample of models in the 2007 market (found by ordering models by price and picking every fifth product). Products are ordered by price with highest price top right. As usual the table shows the elasticity of demand of the row entry with respect to the price of the column entry. There is a clear

---

26 Slope is 1.69 if I include length and diesel in the horsepower regression. Any characteristic in the regression is conditioned on, so in this case the slope is the technological trade-off between fuel cost and horsepower keeping length and fuel type constant. Willingness to pay is 34,300 if length and fuel type are kept constant. Including length and diesel in the price regression gives coefficients of 0.35 and 0.19, resulting in an expected price change of \(0.35 \cdot 0.169 + 0.19 \cdot 0.1 = 0.0781\), 78,100 kroner. All regression coefficients are significant to 95% except that (-0.13) on fuel cost in the first price regression.
pattern of higher cross-elasticities close to the diagonal (from top-right to bottom-left), and lower as we move away from the diagonal: cross-elasticities are higher among products in the same price category.

Variant own-price elasticities are larger in magnitude than those for models, since they include substitution within the model. Unlike model elasticities, variant-own price elasticities (and semi-elasticities) decrease in absolute value with price, presumably because intra-model substitution is larger in cheaper models.

Table 5 shows variant elasticities for several variants of the same models with the purpose of highlighting intra-model substitution patterns. The products were picked by taking a few arbitrary blocks of products adjacent in the data. Elasticities between variants of the same model can be high, but not all are, and some are zero. Restricting the idiosyncratic shock to be the same for all variants of a model permits, but does not impose, high intra-model elasticities. Substitution is spread less evenly across all products compared to demand models with idiosyncratic tastes for all products.
Table 4: Model elasticities (x100) - w.r.t. price change for all variants of a model

<table>
<thead>
<tr>
<th>Model</th>
<th>242</th>
<th>240</th>
<th>238</th>
<th>236</th>
<th>234</th>
<th>232</th>
<th>230</th>
<th>228</th>
<th>226</th>
<th>224</th>
<th>222</th>
<th>220</th>
<th>218</th>
<th>216</th>
<th>214</th>
<th>212</th>
<th>210</th>
<th>208</th>
<th>206</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mercedes-Benz E</td>
<td>-452</td>
<td>1.4</td>
<td>18.3</td>
<td>0.9</td>
<td>0.6</td>
<td>0.5</td>
<td>2.5</td>
<td>5.3</td>
<td>3.3</td>
<td>0.4</td>
<td>9.5</td>
<td>0.4</td>
<td>3.2</td>
<td>0.9</td>
<td>2.2</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saab 9-5</td>
<td>3.5</td>
<td>-510</td>
<td>20.3</td>
<td>1.0</td>
<td>0.7</td>
<td>0.5</td>
<td>3.1</td>
<td>6.5</td>
<td>4.1</td>
<td>0.6</td>
<td>12.3</td>
<td>0.6</td>
<td>4.6</td>
<td>1.3</td>
<td>3.4</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toyota RAV4</td>
<td>3.1</td>
<td>1.3</td>
<td>-455</td>
<td>1.2</td>
<td>0.9</td>
<td>0.6</td>
<td>3.9</td>
<td>8.3</td>
<td>5.0</td>
<td>0.7</td>
<td>16.3</td>
<td>0.8</td>
<td>6.6</td>
<td>1.9</td>
<td>5.2</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jeep PATRIOT</td>
<td>3.0</td>
<td>1.3</td>
<td>24.1</td>
<td>0.9</td>
<td>0.6</td>
<td>4.0</td>
<td>8.5</td>
<td>5.1</td>
<td>0.8</td>
<td>16.7</td>
<td>0.8</td>
<td>6.8</td>
<td>2.0</td>
<td>5.4</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyundai TUCSON</td>
<td>2.9</td>
<td>1.3</td>
<td>24.4</td>
<td>1.2</td>
<td>0.7</td>
<td>5.0</td>
<td>9.1</td>
<td>6.2</td>
<td>1.1</td>
<td>22.8</td>
<td>1.2</td>
<td>11.4</td>
<td>3.3</td>
<td>9.2</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kia SPORTAGE</td>
<td>2.5</td>
<td>1.1</td>
<td>19.8</td>
<td>1.0</td>
<td>0.8</td>
<td>-490</td>
<td>5.0</td>
<td>8.5</td>
<td>5.9</td>
<td>1.0</td>
<td>19.5</td>
<td>1.0</td>
<td>9.2</td>
<td>2.6</td>
<td>6.8</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mazda 6</td>
<td>2.0</td>
<td>1.0</td>
<td>18.3</td>
<td>0.9</td>
<td>0.7</td>
<td>0.7</td>
<td>5.3</td>
<td>9.1</td>
<td>-488</td>
<td>1.1</td>
<td>22.8</td>
<td>1.2</td>
<td>11.4</td>
<td>3.3</td>
<td>9.2</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VW TOURAN</td>
<td>2.3</td>
<td>1.1</td>
<td>21.0</td>
<td>1.1</td>
<td>0.8</td>
<td>0.7</td>
<td>5.0</td>
<td>4.0</td>
<td>1.2</td>
<td>23.4</td>
<td>1.2</td>
<td>12.1</td>
<td>3.4</td>
<td>9.2</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valvo V50</td>
<td>2.1</td>
<td>1.0</td>
<td>18.6</td>
<td>0.9</td>
<td>0.7</td>
<td>0.7</td>
<td>5.3</td>
<td>9.1</td>
<td>-488</td>
<td>1.1</td>
<td>22.8</td>
<td>1.2</td>
<td>11.4</td>
<td>3.3</td>
<td>9.2</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dodge CALIBER</td>
<td>1.7</td>
<td>0.9</td>
<td>18.0</td>
<td>0.9</td>
<td>0.7</td>
<td>0.7</td>
<td>6.0</td>
<td>9.6</td>
<td>7.1</td>
<td>-474</td>
<td>1.3</td>
<td>13.5</td>
<td>3.9</td>
<td>10.6</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VW GOLF</td>
<td>1.8</td>
<td>0.9</td>
<td>18.4</td>
<td>0.9</td>
<td>0.7</td>
<td>0.7</td>
<td>5.7</td>
<td>10.0</td>
<td>7.0</td>
<td>1.2</td>
<td>-453</td>
<td>1.3</td>
<td>13.2</td>
<td>4.0</td>
<td>11.4</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hyundai I30</td>
<td>1.6</td>
<td>0.9</td>
<td>18.0</td>
<td>0.9</td>
<td>0.7</td>
<td>0.7</td>
<td>5.9</td>
<td>10.0</td>
<td>7.2</td>
<td>1.2</td>
<td>-452</td>
<td>1.3</td>
<td>13.9</td>
<td>4.2</td>
<td>11.9</td>
<td>1.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toyota COROLLA</td>
<td>1.4</td>
<td>0.8</td>
<td>16.5</td>
<td>0.8</td>
<td>0.7</td>
<td>0.7</td>
<td>6.5</td>
<td>10.2</td>
<td>7.7</td>
<td>1.4</td>
<td>29.3</td>
<td>1.2</td>
<td>11.4</td>
<td>4.8</td>
<td>13.7</td>
<td>2.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kia CEE'D</td>
<td>1.3</td>
<td>0.7</td>
<td>16.5</td>
<td>0.8</td>
<td>0.7</td>
<td>0.7</td>
<td>6.3</td>
<td>10.5</td>
<td>7.7</td>
<td>1.4</td>
<td>30.0</td>
<td>1.6</td>
<td>16.4</td>
<td>4.3</td>
<td>24.5</td>
<td>1.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Toyota YARIS</td>
<td>1.2</td>
<td>0.7</td>
<td>16.6</td>
<td>0.9</td>
<td>0.7</td>
<td>0.7</td>
<td>6.3</td>
<td>11.2</td>
<td>7.9</td>
<td>1.4</td>
<td>32.1</td>
<td>1.7</td>
<td>17.4</td>
<td>5.4</td>
<td>40.7</td>
<td>2.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Honda JAZZ</td>
<td>0.9</td>
<td>0.6</td>
<td>13.0</td>
<td>0.7</td>
<td>0.7</td>
<td>0.7</td>
<td>6.8</td>
<td>9.7</td>
<td>8.2</td>
<td>1.5</td>
<td>32.6</td>
<td>1.7</td>
<td>19.2</td>
<td>5.8</td>
<td>17.0</td>
<td>4.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mazda 2</td>
<td>0.8</td>
<td>0.5</td>
<td>12.1</td>
<td>0.6</td>
<td>0.5</td>
<td>0.6</td>
<td>6.9</td>
<td>9.8</td>
<td>8.3</td>
<td>1.6</td>
<td>34.5</td>
<td>1.8</td>
<td>20.6</td>
<td>6.4</td>
<td>18.9</td>
<td>3.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kia PICANTO</td>
<td>0.8</td>
<td>0.6</td>
<td>15.9</td>
<td>0.8</td>
<td>0.7</td>
<td>0.6</td>
<td>6.3</td>
<td>14.0</td>
<td>8.2</td>
<td>1.4</td>
<td>39.5</td>
<td>2.0</td>
<td>21.8</td>
<td>7.1</td>
<td>24.4</td>
<td>3.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 5: Intra- and inter-model price elasticities (x100)

| Variant      | Citr. C4 1.4 l 87 hp P | Citr. C4 1.6 l 88 hp D | Citr. C4 1.6 l 107 hp P | Citr. C4 1.6 l 107 hp D | Citr. C4 2 l 134 hp D | Citr. C5 1.6 l 107 hp D | Citr. C5 1.7 l 1121 hp P | Toy. COROLLA 1.4 l 95 hp P | Toy. COROLLA 2 l 114 hp D | Toy. COROLLA 1.6 l 112 hp P | Toy. COROLLA 2 l 125 hp D | Toy. COROLLA 1.8 l 127 hp P | Toy. YARIS 1.3 l 86 hp P | Toy. YARIS 1.4 l 88 hp D | VW GOLF 1.4 l 79 hp P | VW GOLF 1.9 l 88 hp D |
|--------------|------------------------|------------------------|-------------------------|-------------------------|------------------------|-------------------------|---------------------------|---------------------------|-------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|---------------------------|
|              | -6669                  | -3680                  | 1004                    | 0.0                     | 0.0                    | 0.0                     | 0.0                       | 2.6                       | 0.0                     | 1.2                       | 0.0                       | 0.0                       | 1.7                       | 0.2                       | 0.0                       | 0.3                       | 0.3                       |
|              | 591                    | 220                    | 75.2                    | 5890                    | 90.0                   | 2.4                     | 0.7                       | 0.0                       | 0.0                     | 0.1                       | 0.0                       | 0.0                       | 0.0                       | 0.0                       | 0.0                       | 0.0                       | 0.0                       |
|              | 5729                   | 2698                   | -1704                   | -9381                   | 152                    | 0.2                     | 0.0                       | 0.0                       | 0.0                     | 1.5                       | 0.0                       | 0.2                       | 0.0                       | 0.0                       | 0.0                       | 0.0                       | 0.0                       |
|              | 0.0                    | 2.0                     | 1.8                     | 1.5                     | 0.8                    | 1.7                     | 0.2                       | 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                     | 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                   | 0.0                     | 0.0                       | 0.0                       | 0.0                     | 0.0                     | 0.0                       | 0.0                       | 0.0                       | 0.0                       | 0.0                       | 0.0                       | 0.0                       |
|              | 17.1                   | 0.0                     | 10.5                    | 4.2                     | 2.7                    | 0.1                      | 5.5                       | 0.0                       | 0.0                     | 0.0                     | 5.5                       | 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                     | 0.0                       | 0.0                       | 0.0                       | 0.0                       | 0.0                       | 0.0                       | 0.0                       |
|              | 14.5                   | 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                       |
7 The welfare effects of engine options

This section presents outcomes and welfare effects of a counterfactual experiment where firms are constraint to offer only one engine for each model.

7.1 The supply-side model

Since I have no data on transactions between car manufacturers, import companies and dealerships, I follow the literature and treat the supply side as consisting of vertically integrated car companies selling directly to consumers. The sales price can be decomposed as

\[ p_j = (\text{markup}_j + MC_j)(1 + \text{vat}) + tax_j \]  

(15)

where \( MC \) is marginal cost (the incremental cost of producing one more unit of product \( j \)). I assume constant marginal costs within the relevant output ranges (no economies of scale). \( tax \) is the engine tax (an increasing convex function of horsepower, weight and co2 emissions; in 2007 the tax on average accounts for 40% of price, ranging from 20 to 70%). \( \text{markup} + MC \) is the amount the car company is left with after paying the engine tax and the value-added tax, \( vat \), of 25%.

I follow the literature and assume that observed prices are a Nash equilibrium in a game where firms simultaneously choose prices to maximise profits

\[ \Pi_f = \sum_{j \in F_f} \left[ \frac{p_j - tax_j}{1 + vat} - MC_j \right] Ms_j(p) - C_f, \]  

(16)

where \( F_f \) is the set of products owned by firm \( f \) and \( M \) is the size of the market (summed over all consumer groups). In equilibrium the first-order

\[ 27 \text{The analysis in this section uses the demand system estimated on the full data set, but markups and the counterfactual are computed only for the 2007 market.} \]
conditions for profit maximisation of each product must hold:

\[
0 = \frac{s_j(p)}{1 + \text{vat}} + \sum_{j' \in F(j)} \left( \frac{p_{j'} - \text{tax}_{j'}}{1 + \text{vat}} - MC_{j'} \right) \frac{\partial s_{j'}(p)}{\partial p_j}, \quad j = 1, \ldots, J
\]

(17)

where \( F(j) \) denotes the set of products owned by the company which owns product \( j \). The only unknowns are the marginal costs, which I find by solving this system of \( J \) linear equations in the \( J \) unknowns. Once the system of equations is solved, we know markups and marginal costs.

### 7.2 Counterfactual

This subsection compares the actual market with equilibrium in a counterfactual market where each model is offered with only one engine size. I make two simplifying assumptions: first, that there are no fixed cost investments involved in expanding the range of engine sizes offered within a model (no economies of scope).\(^{28}\) Secondly, when companies offer only one engine variant of each model, they choose the one that generates the highest profits in the current market.\(^{29}\)

Table 6 compares outcomes for the counterfactual single-variant equilibrium with the outcomes in the actual multi-variant equilibrium.\(^{30}\) Profit increases, but only by a small amount, 2.4 per cent. In fact markups (average profit per unit) are marginally reduced, and profit goes up only because the in-

\(^{28}\)Often a new variant is created by taking the baseline engine of one of the brand’s higher segment models and putting it in the a lower segment model. Also, developing a more powerful version is less costly than developing an entirely new engine.

\(^{29}\)Endogenising the choice of baseline variant would be ideal, but is computationally too demanding. It would require the computation of price equilibria for every possible configuration of engine sizes, which is not feasible. Allowing two choices for each firm would require the computation of \( 2^{94} \) equilibria. Relaxing the assumption of perfect information about the choices of other firms could simplify the computation of equilibria, but creating a model of such a game is outside the scope of the paper.

\(^{30}\)The numbers for the actual market are based on the model’s predictions of sales. While these predictions are exactly correct for the sales of each model, and therefore aggregate sales, the model does not guarantee a perfect fit at the individual and variant level.
Table 6: Market outcomes with single-variant and multi-variant models

<table>
<thead>
<tr>
<th></th>
<th>i. single-variant models (counterfactual)</th>
<th>ii. multi-variant models (actual market)</th>
<th>change</th>
<th>% change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profits(^a)</td>
<td>4356</td>
<td>4460</td>
<td>104</td>
<td>2.4</td>
</tr>
<tr>
<td>Tax revenue</td>
<td>8341</td>
<td>9086</td>
<td>744</td>
<td>8.9</td>
</tr>
<tr>
<td>from Engine tax</td>
<td>5644</td>
<td>6290</td>
<td>646</td>
<td>5.3</td>
</tr>
<tr>
<td>from v.a.t.</td>
<td>2698</td>
<td>2796</td>
<td>98</td>
<td>3.6</td>
</tr>
<tr>
<td>Consumer surplus</td>
<td>63820</td>
<td>66597</td>
<td>2777</td>
<td>4.4</td>
</tr>
<tr>
<td>Unit sales</td>
<td>75618</td>
<td>77916</td>
<td>2298</td>
<td>3.0</td>
</tr>
<tr>
<td>Mean price(^b)</td>
<td>288.7</td>
<td>296.6</td>
<td>7.9</td>
<td>2.7</td>
</tr>
<tr>
<td>Mean markup</td>
<td>57.6</td>
<td>57.3</td>
<td>-0.3</td>
<td>-0.5</td>
</tr>
<tr>
<td>Mean horsepower (kW)</td>
<td>87.3</td>
<td>91.6</td>
<td>4.4</td>
<td>5.0</td>
</tr>
</tbody>
</table>

\(^a\)Profits, tax and consumer surplus in million kroner.  
\(^b\)Price, markup and horsepower are sales weighted.

increased product variety recruits some (3.0%) new consumers who otherwise would choose the outside good. Making the characteristics of the remaining engine variants endogenous may change the results, which should therefore be regarded with some caution. However, choosing the highest-profit variant of each model, as I currently do, means that chosen variants tend to be similar for models in the same segments. Models therefore compete head on in the current counterfactual. Allowing firms to choose characteristics would open up for more specialisation, and thereby increase markups even more, reinforcing my conclusion.

Consumers pay more, but are better off, implying that the non-price part of utility increases more than the disutility of price.\(^{31}\) It has been pointed out that logit models tend to overestimate the benefit to consumers of product variety (see Petrin (2002)), because each new product introduces a new dimension of differentiation for which some consumers have a very high id-

\(^{31}\)In my utility specification income enters the price coefficient linearly, but disposable income enters only linearly. Total consumer surplus is computed over all consumers, including those choosing the outside good in either case, but these do not experience any change in consumer surplus since their choice or its attributes do not change. See appendix for the calculation of consumer surplus.
iosyncratic taste. This is not the case here: no new dimension of product differentiation accompanies the introduction of new engine variants, because for each consumer, new variants have the same idiosyncratic shock as the baseline variant of the model. The crowding of characteristics space that results from multi-variant models is therefore taken into account.

8 Conclusion

I develop a new discrete-choice demand model suitable for a common class of products which do not fit in existing demand frameworks. I estimate the model on individual level data, but it could also be used for market level data, as long as a nonzero prediction error could be justified as sampling error or in another way.

The model is used to analyse welfare effects of the large number of engine options offered with car models. The question of what engine variants to offer if restricting the quality range is related to the recent product-choice literature, e.g. Mazzeo (2002) and Gramlich (2009). I simplify this question by assuming that firms would choose their current highest-profit engine variant.

References


Verboven, F. (1999): “Product line rivalry and market segmentation - With an application to automobile optional engine pricing,” *Journal of
Appendix

A. Hedonic regression

The assumption that there are no unobservable differences between models is central to the identification strategy of this paper. Results from a hedonic regression indicate that the assumption holds. Since any unobservable difference that affects the average consumer’s valuation of a good should show up in price, I look at how much of price variation can be explained by observed characteristics and model effects. Regressing prices (for all products) on model(-year) dummies (2397 products, 660 dummies), horsepower, fuel cost and diesel, cylinder volume and squares of horsepower, fuel cost and cylinder volume, and a constant gives an R-squared of 0.9913, leaving little scope for variant-specific unobservables to explain price variation.\(^{32}\) Results are given in table 7.

\(^{32}\)Including further engine-related characteristics (weight, CO2 emissions), interactions and third- and fourth-order terms increases R-squared to as much as 0.9972, but this is open to charges of overfitting.
Table 7: Regression of price on engine characteristics and model dummies. Unit of observation is variants. 2397 observations. Coefficients on the 660 model dummies not shown.

<table>
<thead>
<tr>
<th>explanatory variable</th>
<th>estimate</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>horsepower</td>
<td>-0.005</td>
<td>0.547</td>
</tr>
<tr>
<td>fuel cost</td>
<td>0.099</td>
<td>0.000</td>
</tr>
<tr>
<td>diesel</td>
<td>0.035</td>
<td>0.000</td>
</tr>
<tr>
<td>cylinder volume</td>
<td>0.023</td>
<td>0.002</td>
</tr>
<tr>
<td>horsepower squared</td>
<td>0.085</td>
<td>0.000</td>
</tr>
<tr>
<td>fuel cost squared</td>
<td>-0.026</td>
<td>0.031</td>
</tr>
<tr>
<td>cyl. vol. squared</td>
<td>0.009</td>
<td>0.000</td>
</tr>
<tr>
<td>constant</td>
<td>0.066</td>
<td>0.000</td>
</tr>
</tbody>
</table>

B. Consumer surplus

Consumer $i$’s surplus is the money value of the utility of his best choice:

$$\frac{1}{-\alpha_i} \max_{j \in J} u(p_j, x_j, i).$$  \hspace{1cm} (18)

Since $u(p_j, x_j, i)$ depends on random variables $\phi_i = (y_i, \nu_i, \epsilon_i)$ whose realisations are not known, I can only compute the expected consumer surplus:

$$E[CS_i] = E[\frac{1}{-\alpha_i} \max_{j \in J} u(p_j, x_j, i)],$$  \hspace{1cm} (19)

where the expectation is over $\phi_i$. I compute the values of the expectations by first conditioning on the draws of $(y_i, \nu_i)$. For each draw the expectation over $\epsilon$ is:

$$E[\max_{j \in J'} u(p_j, x_j, i)|y_i, \nu_i] = \log \left( \sum_{j' \in J'} \exp(-\alpha_i p_{j'} + x_{j'} \beta_i + \delta_{im(j')}) \right),$$  \hspace{1cm} (20)

where $J'$ is the choices remaining in the choice set after maximising over the products which are restricted to have the same realisation of $\epsilon$. The final expression in (19) is then found by integrating over $(y_i, \nu_i)$. The total

(expected) consumer surplus $CS$ in the market is found by summing over the consumers with characteristics $z_i$ for all $i$:

$$CS = \sum_i N_i E[CS_i], \quad (21)$$

where $N_i$ is the number of people in the market with characteristics $z_i$.

**C. Is there price discrimination over engine variants?**

This subsection looks at a slightly different, but related issue to that discussed in section 7.

In competitive markets second-degree price discrimination over quality may not be possible. Verboven (1999), Armstrong and Vickers (2001), Rochet and Stole (2002) find that depending on the intensity of competition, markups may be fixed across quality variants for each firm. Since the theory results depend on specific assumptions about differentiation and symmetry, I let the data determine these specifics. The theory models are special cases of the demand model used in this paper.

Taking the average of the smallest engine variant over all models (with at least two variants) gives a markup of 52,800 kroner, while the average of the biggest variant over all models gives a markup of 83,800 kroner. For comparison, overall sales-weighted mean markup is 57,300 kroner. Average percentage markups are 18.36 for smallest variants and 19.75 for biggest variants. Figure plots markups against horsepower for 9 of the 10 models in 2007 that have seven or more engine variants, with a line fitted by OLS. Models are arranged in order of increasing mean price, from left to right,

---

34My model has many products, each firm offers several models with quality variants, there are several dimensions of taste heterogeneity, and there is an outside good. In the theory literature there are two firms that each offer one model with quality variants, and there is one vertical (taste for quality) and one horizontal (taste for models) dimension of taste heterogeneity. Some theory models do not have an outside good.
then top to bottom. Markups clearly increase in horsepower for all models, with overall markups higher for the more expensive products.

Figure 1: Markups (in 1e-6 kroner, vertical axis) plotted against horsepower (kw*1e-2, horizontal axis) for models that have seven or more engine variants. Line fitted by OLS. Models arranged in order of increasing price. For the last two products some large engine variants do not show.

To find the average (over models) effect of engine size on markups within a model, I regress markups (and percentage markups) on horsepower, horsepower squared and cubed, a constant, and model dummies. The model dummies isolate the effect of horsepower, so that higher markups for models with overall higher horsepower does not contaminate the results. Table 8 shows the results. All higher order terms in horsepower are highly significant. Figure 2 plots the polynomial functions of markups and percentage markups obtained from the regression. Markups clearly increase in horsepower within
a model, while percentage markups are approximately constant.

Table 8: Results from regression of markups and percentage markups (markup/price) on horsepower, higher order terms of horsepower and model dummies in year 2007. 375 observations, 94 model dummies. Coefficients on model dummies not shown.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>hp</th>
<th>hp²</th>
<th>hp³</th>
<th>constant</th>
</tr>
</thead>
<tbody>
<tr>
<td>markup</td>
<td>-0.029</td>
<td>0.056</td>
<td>-0.007</td>
<td>0.038</td>
</tr>
<tr>
<td>(p-value)</td>
<td>0.205</td>
<td>0.001</td>
<td>0.047</td>
<td>0.000</td>
</tr>
<tr>
<td>markup/price</td>
<td>-0.214</td>
<td>0.129</td>
<td>-0.024</td>
<td>0.295</td>
</tr>
<tr>
<td>(p-value)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Figure 2: Within-model differences in markups and percentage markups as a function of engine size. Horsepower in the market ranges from 0.44 to 3.09, sales-weighted mean 0.92kW. 10th to 90th percentile (not sales-weighted) is (0.64,1.65)

Consumers who buy larger engine variants of a car model pay a higher premium over marginal cost than consumers who buy smaller engine variants of the same model. In this sense there is second-degree price discrimination. Percentage markups are approximately constant across engine variants.

Several theory papers find that pricing will be cost-plus-fixed-fee, i.e. no price discrimination. Since my empirical result is different, it is of interest to see
what assumptions in the theory literature generates the cost-plus-fixed-fee results. I simulate some simple duopoly models that relax the assumptions of the models of Verboven (1999) and Ellison (2005) (by including an outside good, making qualities and marginal costs asymmetric, and introducing heterogeneity in one or both the price and quality parameters). I find that constant absolute markups are the exception rather than the rule.

35 I compute two-firm two-quality Nash equilibria in models where the taste for firms is logit; the price coefficient is either 0.5, U(0,1) or N(0.5,0.5); quality coefficient either 1 or U(0.5,1.5); marginal costs are (3,4) for both firms, or (1,2)/(3,4); qualities are (7,9) for both, (6,8)/(7,9), or (3,5)/(7,9). If an outside good is included, it has utility -1.5 plus logit term. Equilibria are computed by iterating on the best-response functions. Best responses are computed using a genetic optimisation algorithm, constraining each firm’s high quality price to be weakly higher than its low quality price.

36 Increasing markups is the most common result, while large asymmetries between firms may give decreasing absolute markups for one of the firms. For instance, pricing is no longer cost-plus-fee in Verboven’s model if we include an outside good. Presumably, the participation constraint puts a downward pressure on prices, and more on the low quality product which is the closest substitute to the outside option.