Discussion Papers No. 490, December 2006
Statistics Norway, Research Department

John K. Dagsvik, Astrid L. Mathiassen and Bengt J. Eriksson

Quality adjusted price indexes for discrete goods

Abstract:
This paper discusses the construction and computation of a quality adjusted price index when the commodities are differentiated products, such as different brands of automobiles and refrigerators. The method we focus on is an extension of Trajtenberg's approach. A key result obtained in the paper is that the evolution of the quality adjusted price index depends crucially on the fraction of consumers that do not purchase a variant of the product. The method is applied to data on automobile demand in Norway from 1994 to 2002. Both the Laspeyres index and the index based on hedonic regression yield lower estimates of the pricees from 1999 to 2002 than does the quality adjusted price index. This is mainly due to variations in the fraction of persons who purchase new automobiles.

Keywords: Quality adjusted price index, Exact index theory, Hedonic price indexes

JEL classification: C25, C43

Acknowledgement: We are grateful for comments by Ådne Cappelen, Jørgen Aasness and Steinar Strøm.

Address: John K. Dagsvik, Statistics Norway, Research Department. E-mail: john.dagsvik@ssb.no

Discussion Papers comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers
in PDF are available on the Internet:
http://www.ssb.no
http://ideas.repec.org/s/ssb/dispap.html

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
NO-2225 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1. Introduction

In the standard textbook setting, consumer preferences are assumed to depend on different types of goods solely through their respective quantities. Thus, this theory of consumer behavior has been developed under specific and well-known assumptions about preferences and choice restrictions (budget constraints), and the Paasche and Laspeyres price indexes, up to a first-order (Taylor) approximation, have been derived. For this approximation to be valid, certain conditions need to be fulfilled (in addition to standard mathematical regularity conditions). These conditions are: (i) the set of goods available in the market does not change over time; (ii) the inherent properties of the goods also remain unchanged over time, i.e., the notion of quality changes is absent; and (iii) the goods are infinitely divisible. If these conditions are not fulfilled, one cannot use the standard theory of consumer behavior to justify the indexes mentioned above without further argument.

A typical feature of modern markets is that consumers face a variety of products that are differentiated with respect to sets of characteristics, which for convenience we shall call "quality" attributes. Some of these attributes are tangible and observable, whereas others are not, and such attributes may represent fashion or popularity. In addition, for many products, the product variants are chosen mutually exclusively, in the sense that only one is selected from the set of feasible variants.

In this paper, we extend the traditional price index theory by allowing for indivisible goods (discrete goods) characterized by product-specific attributes. The set of variants that appear in the market will typically vary from one period to the next so that, in practice, it is hard to observe the prices of the same good over time. What complicates the situation further is that even if observable product attributes (assuming these are available) do not vary—or change slowly over time—the popularity of the product may vary considerably. For example, many products, such as clothes, follow popularity cycles of the fashion industry. This variation is a consequence of the fact that the average preferences in the population for the product in question vary from one period to the next.

Some years ago, the so-called Boskin report (Boskin et al., 1996) focused on the need to take "quality" into account in price indexes. However, the notion of quality in this context is not new. In the 1960s and 1970s, economists started to consider how the empirically oriented theory of price indexes should account for changes in quality for markets with differentiated products. Rosen (1974) proposed a method for estimating demand and supply functions in markets with differentiated products, which, in principle, can be used to estimate supply and demand relations and subsequently derive price indexes. However, Rosen's method has proven to be intractable for applying in practical empirical analysis. Furthermore, this method is based on rather stylistic assumptions. It assumes, for example, that the variety of product variants is so rich that practically all ("continuous") combinations of attributes characterizing the product variants exist in the market simultaneously. Accordingly, the choice setting is no longer treated as a discrete one, as, under this presumption, one can acquire a product variant with any desired attribute combination. Other contributions in this tradition are Bartik (1987) and Epple (1987).[1]

In contrast, Trajtenberg (1990) takes the theory of discrete choice as his point of departure to establish a true quality adjusted price index. His approach is based on the multinomial logit discrete choice demand model. Our contribution in this paper, is to extend Trajtenberg's approach by (i) allowing for latent quality attributes that may vary over time, (ii) accounting for the possibility that prices may be endogenous, and (iii) demonstrating that the fraction of consumers that do not purchase a variant of the discrete good is, together with the mean population expenditure of the variants purchased, sufficient statistics for the quality adjusted price index (given prices of the outside divisible goods and a parameter that characterizes the demand of the differentiated product). The theoretical basis assumed for the price-setting regime is the assumption of oligopolistic competition, as adapted to the case with probabilistic discrete choice demand by Anderson et al. (1992). Other related works in this area include Crawford (1997), Feenstra (1995), Jonker (2002) and Song (2005). However, the approaches taken by these authors differ from Trajtenberg (1990) and ours.

The empirical application aims at estimating quality adjusted price indexes for new automobiles in Norway from 1994 to 2002. The main findings are that the standard Laspeyres index underestimates the "true" quality adjusted index for some years and overestimates it in other years. Moreover, we find that the conventional hedonic price index more or less yields the same figures as the Laspeyres index.

The paper is organized as follows. In the next section, we discuss the approach proposed by Trajtenberg (1990). In Section 3, we review some problems with the interpretation of the hedonic regression method. Finally, in Section 4, we discuss an empirical application in a discrete choice setting, namely the market for new automobiles in Norway. To compute the quality adjusted price index for new automobiles, it is necessary to estimate a key parameter in the demand relation. Here, we apply recent likelihood-based methods (Vitorino, 2004) to estimate an equilibrium model under the assumption of oligopolistic competition.

[1] There does not seem to be a generally accepted use of the label "hedonic methods". For example, some authors use the terminology hedonic method to mean regression models with prices as the dependent variable and product attributes as independent variables (hedonic regression), whereas others use hedonic in a much more general sense. In this paper, we shall use the term "hedonic regression" in the same way as Trajtenberg (1990) to mean regression models with log prices as dependent variables and product attributes as independent variables.

2. Construction of quality adjusted price indexes for discrete goods

In this section, we discuss the construction of exact quality adjusted price indexes for differentiated products. We start by discussing the assumptions about preferences and proceed by investigating their implications in the context of price index theory.

2.1. The case with multinomial logit demand

We consider a market with a differentiated product (for example, automobiles). Each consumer purchases at most one variant of the product in each period, or alternatively, in addition to quantities of divisible goods. To make the exposition consistent with the empirical application below, we assume that the variants are classified into separate groups indexed by $g = 1, 2, ..., S.$ That is, the product variants are classified along two dimensions; first they are divided into separate groups, and second each group contains different group-specific variants. For example, in the automobile market the groups may be different body gr|oups such as Sedan, Station wagon, etc. Let $B_t(g)$ be the set of variants within group *g* that are available in the market at time t. Consumer *i* has utility function $U_{itj}(g)$ of variant *j* in group *g* at period *t*, which is assumed to have the form

$$(2.1) \qquad U_{itj}(g) = \frac{\left(m_{it} - w_{tj}(g)\right)\theta}{p_t} + v_{tj}(g) + \varepsilon_{itj}(g),$$

where m_{it} represents income, $w_{tj}(g)$ is the price of variant *j* in group *g*, $v_{tj}(g)$ is a function of attributes of variant *j* in group *g*, p_t is a price index for the divisible (outside) goods, θ is a positive constant and $\varepsilon_{itj}(g)$ are random variables that represent unobserved heterogeneity in tastes. The alternative of not buying is indexed by $j = g = 0$, where $U_{it0}(0) = \theta m_{it} / p_t$. The terms $v_{tj}(g)$, $t = 1, 2, ...$, $s = 1, 2, ..., S$, can be interpreted as a quality indicator that is supposed to capture such effects as the fluctuations in average population popularity of variant *j* in group *g*. In this paper, we assume that the random error terms are i.i.d., independent of the other terms of the utility function and with extreme value c.d.f.

$$(2.2) \qquad P\left(\varepsilon_{itj}(g) \le x\right) = \exp\left(-e^{-x}\right).$$

It is well known (cf. McFadden, 1984) that this implies that

$$(2.3) \qquad P\left(U_{itj}(g) = \max_r \max_{k \in B_t(r)} U_{itk}(r)\right) = Q_{tj}(g) \equiv \frac{\exp\left(v_{tj}(g) - \tilde{w}_{tj}(g)\theta\right)}{\sum_r \sum_{k \in B_t(r)} \exp\left(v_{tk}(r) - \tilde{w}_{tk}(r)\theta\right) + 1}$$

for $j \in B_t(g)$, whereas for $j = g = 0$, we have

$$(2.4) \qquad P\left(U_{it0}(0) = \max_r \max_{k \in B_t(r)} U_{itk}(r)\right) = Q_{t0} \equiv \frac{1}{\sum_r \sum_{k \in B_t(r)} \exp\left(v_{tk}(r) - \tilde{w}_{tk}(r)\theta\right) + 1},$$

where $\tilde{w}_{tj}(g) = w_{tj}(g) / p_t$. As is well known, the model given in (2.3) and (2.4) satisfies the Independence from Irrelevant Alternatives (IIA) property, which in some cases is known to be restrictive.

2.2. The Trajtenberg approach

Now, suppose that the parameters of the demand model given in (2.3) and (2.4) have been estimated. Then, one can readily calculate quality adjusted price indexes by means of the expenditure function as proposed by Trajtenberg (1990). We shall now explain Trajtenberg's approach. Assume that the utility structure is given by (2.1) and let V_{it} denote the indirect utility conditional on the choice set, attributes, price and income, defined by

$$(2.5) \qquad V_{it} = \max\left(\max_g \max_{k \in B_t(g)} U_{itk}(g), U_{it0})\right).$$

It is well known (Trajtenberg, 1990) that the assumptions above lead to the following expression for the aggregate (mean) indirect utility

$$(2.6) \; V_t = EV_{it} = E\max\left(\max_g \max_{k \in B_t(g)} U_{itk}(g), U_{it0}\right) = m_{it}\theta + \log\left(1 + \sum_{g=1}^{S}\left[\sum_{k \in B_t(g)} \exp\left(v_{tk}(g) - \tilde{w}_{tk}(g)\theta\right)\right]\right).$$

The interpretation of (2.6) is that it expresses mean indirect utility given the choice set, observed and unobserved attributes, prices and income. Let $\boldsymbol{v}_t = (v_{t1}(1), v_{t2}(1), ...v_{t1}(2), ...)$ and $\boldsymbol{w}_t = (w_{t1}(1), w_{t2}(1), ..., w_{t1}(2), w_{t2}(2), ...)$. From (2.6), it follows that the corresponding aggregate (mean) expenditure function, $\bar{e}\left(\boldsymbol{v}_t, \boldsymbol{w}_t, p_t, B_t, u\right)$, is given by

$$(2.7) \qquad \bar{e}\left(\boldsymbol{v}_t, \boldsymbol{w}_t, p_t, B_t, u\right) = u - \theta^{-1}\log\left(1 + \sum_{g=1}^{S}\left[\sum_{k \in B_t(g)} \exp\left(v_{tk}(g) - \tilde{w}_{tk}(g)\theta\right)\right]\right),$$

where u is the given utility level divided by θ. We define the quality-adjusted price index (Trajtenberg), δ_t, as determined by

(2.8) $\overline{e}\left(\boldsymbol{v}_t,\boldsymbol{w}_t,p_t,B_t,u\right)-\overline{e}\left(\boldsymbol{v}_{t-1},\boldsymbol{w}_{t-1},p_{t-1},B_{t-1},u\right)=\overline{e}\left(\boldsymbol{v}_{t-1},\delta_t\boldsymbol{w}_{t-1},p_t,B_{t-1},u\right)-\overline{e}\left(\boldsymbol{v}_{t-1},\boldsymbol{w}_{t-1},p_{t-1},B_{t-1},u\right)$,

which is equivalent to

(2.9) $$\overline{e}\left(\boldsymbol{v}_t,\tilde{\boldsymbol{w}}_t,1,B_t,u\right)=\overline{e}\left(\boldsymbol{v}_{t-1},\delta_t\tilde{\boldsymbol{w}}_{t-1}p_{t-1}/p_t,1,B_{t-1},u\right).$$

The interpretation of δ_t is as follows: from period $t-1$ to *t*, there has been a possible change in the choice set, the attributes and the prices of the discrete goods from $\left(\boldsymbol{v}_{t-1},\boldsymbol{w}_{t-1},B_{t-1}\right)$ to $\left(\boldsymbol{v}_t,\boldsymbol{w}_t,B_t\right)$. The left-hand side of (2.8) expresses the actual change in the mean welfare in a money metric measure. The right-hand side represents the change in the mean value given that the choice set is kept fixed and equal to the initial choice set with variants that have initial attributes, but where the initial prices are rescaled by the same factor δ_t. This factor is determined so that the actual change in welfare becomes equal to the change in welfare caused solely by the scale transformation of the initial prices represented by δ_t. Note that δ_t is a *conditional* index that only captures the welfare change of the discrete good. The interpretation of δ_t is as an index that represents the welfare effect of the actual change that has taken place in the prices, choice set and attributes from period $t-1$ to *t*. From (2.4), (2.7) and (2.9), we find that δ_t is determined by the equation

(2.10) $$\sum_{g=1}^{S}\left[\sum_{k\in B_t(g)}\exp\left(v_{tk}(g)-\theta\tilde{w}_{tk}(g)\right)\right]=\sum_{g=1}^{S}\left[\sum_{k\in B_{t-1}(g)}\exp\left(v_{t-1,k}(g)-\theta\delta_t\tilde{w}_{t-1,k}(g)p_{t-1}/p_t\right)\right].$$

The right-hand side of (2.10) is strictly decreasing in δ_t and, consequently, equation (2.10) determines δ_t uniquely. Once $\{v_{tj}(g)\}$ have been specified, and θ and the parameters of $\{v_{tj}(g)\}$ have been estimated, one can compute the price index by solving for δ_t in the nonlinear equation (2.10).

2.3. Further implications from the multinomial logit demand model

A serious challenge one faces in the context of empirical application is how to specify the terms $\{v_{tj}(g)\}$. In general, these terms may vary over time in a way that is not captured by observable variant-or group specific attributes. To this end, we shall in this section explore a type of "semiparametric" approach, as we shall now explain. Note that it follows from (2.4) that (2.10) can be written as

(2.11) $$Q_{t0}^{-1}-1=\sum_{g=1}^{S}\left[\sum_{k\in B_{t-1}(g)}\exp\left(v_{t-1,k}(g)-\theta\delta_t\tilde{w}_{t-1,k}(g)p_{t-1}/p_t\right)\right].$$

Furthermore, (2.4) implies that

$$\text{(2.12)} \qquad \frac{Q_{tj}(g)}{Q_{t0}} = \exp\left(v_{tj}(g) - \theta \tilde{w}_{tj}(g)\right).$$

From (2.12) it follows that

$$\text{(2.13)} \qquad \sum_{k \in B_{t-1}(g)} \exp\left(v_{t-1,k}(g) - \theta w_{t-1,k}(g)\right) = \sum_{k \in B_{t-1}(g)} Q_{t-1,k}(g) \exp\left((1 - \delta_t p_{t-1}/p_t)\theta \tilde{w}_{t-1,k}(g)\right).$$

As a consequence, (2.11) and (2.13) imply that the price index is determined by

$$\text{(2.14)} \qquad Q_{t0}^{-1} - 1 = \sum_{g} \sum_{k \in B_{t-1}(g)} Q_{t-1,k}(g) \exp\left((1 - \delta_t p_{t-1}/p_t)\theta \tilde{w}_{t-1,k}(g)\right).$$

Thus, when θ has been estimated, one can use (2.14) to compute the quality-adjusted price index by solving for δ_t in (2.14) without relying on the specification and estimates of $\{v_{tj}(g)\}$.

One can often use a more simple formula derived from an approximation of the left hand side of (2.14). This approximation is close when $1 - \delta_t \, p_{t-1}/p_t$ is small, which is usually the case. If so, it follows by a first order Taylor expansion of the left hand side of (2.14) that the approximate solution for the price index determined by (2.14) is given by

$$\text{(2.15)} \qquad \delta_t \cong \frac{p_t}{p_{t-1}}\left(1 + \frac{1 - Q_{t-1,0}/Q_{t0}}{\theta y_{t-1}}\right),$$

where y_t is the mean deflated expenditure of the discrete good in period *t* given by

$$\text{(2.16)} \qquad y_t = \sum_{g} \sum_{k \in B_t(g)} \tilde{w}_{tk}(g) Q_{tk}(g).$$

The formula in (2.15) is quite interesting because it shows that, in addition to the previous period's expenditure on the discrete good, the change in the relative fraction of consumers who do not purchase a variant (or equivalently, the fraction of consumers who do purchase a variant) summarizes the welfare price effect of changes in tastes, prices and the choice set. As mentioned above, our approach to price index measurement accounts for the fact that people may prefer not to buy a new variant in a given period *t* (say), which implies an increase in Q_{t0}. This will happen if the unobserved quality attributes $\{v_{tj}(g)\}$ decrease in period *t*. Recall that this effect cannot be captured by the Laspeyres index because it gives no weight to individuals who do not purchase a car. Furthermore, recall that the index given in (2.15) takes into account that the choice set of available variants may vary from one year to the next. In addition, this effect is not captured by the Laspeyres index. If the latent quality attributes $\{v_{tj}(g)\}$ were not changing over time, the Laspeyres index would overestimate the price effect because it does not take into account changes in consumers' choice set and the possibility of no purchase. This is because, if the set of feasible variants increases from one period to the next, consumers will have more choices than before and therefore will be able to do better than before, and this welfare gain is unaccounted for in the Laspeyres index (Pakes et al., 1993). However, as the latent quality attributes may change over time, the sign of the difference between the Laspeyres price index and the quality-adjusted price index developed in this paper is ambiguous.

2.4. Aggregation of subindexes for discrete goods

The Laspeyres and Paasche price indexes possess the property that one can conveniently combine subindexes to obtain aggregate indexes by adding the respective subindexes multiplied by the relevant budget shares. In this section, we shall discuss aggregation of subindexes for discrete goods.

Let $V_t(g)$ be the mean conditional indirect utility at period *t* given group *g*. Formally $V_t(g)$ is defined by

(2.17) $$V_t(g) = E\left(\max_{k\in B_t(g)} U_{itk}(g)\right).$$

Similarly to (2.6), it follows that

(2.18) $$V_t(g) = \theta m_{it} + \log\left(\sum_{k\in B_t(g)} \exp\left(v_{tk}(g) - \theta\tilde{w}_{tk}(g)\right)\right).$$

The corresponding conditional mean expenditure function is given by

(2.19) $$\bar{e}\left(\boldsymbol{v}_t(g), \boldsymbol{w}_t(g), p_t, B_t(g), u\right) = u - \theta^{-1}\log\left[\sum_{k\in B_t(g)} \exp\left(v_{tk}(g) - \tilde{w}_{tk}(g)\theta\right)\right].$$

Similarly to (2.9), it follows that the price index for group *g* is determined by

(2.20) $$\bar{e}\left(\boldsymbol{v}_t(g), \boldsymbol{w}_t(g), 1, B_t(g), u\right) = \bar{e}\left(\boldsymbol{v}_{t-1}(g), \delta_t(g)\tilde{w}_{t-1}(g)\, p_{t-1}/p_t, 1, B_{t-1}(g), u\right).$$

If we combine (2.18) and (2.20) it follows that the price index for group *g*, $\delta_t(g)$, is determined by the equation

(2.21) $$\sum_{k\in B_t(g)} \exp\left(v_{tk}(g) - \theta\tilde{w}_{tk}(g)\right) = \sum_{k\in B_{t-1}(g)} \exp\left(v_{t-1,k}(g) - \theta\tilde{w}_{t-1,k}(g)\delta_t(g)\, p_{t-1}/p_t\right).$$

In Appendix B, we prove that, to a first-order Taylor approximation, we have that

(2.22)
$$\delta_t \cong \frac{\sum_g \delta_t(g) y_{t-1}(g)}{\sum_g y_{t-1}(g)},$$

where $y_t(g)$ is the mean deflated expenditure within group *g* in period *t*, defined by

(2.23)
$$y_t(g) = \sum_{k \in B_t(g)} Q_{tk}(g) \tilde{w}_{tk}(g).$$

Equation (2.22) states that one can obtain the aggregate price index for the differentiated good, say cars, in the same way as for the conventional Laspeyres index, namely by adding the subindexes weighted by their respective budget shares as of the previous period.

Finally, let us derive an approximate closed form expression for $\delta_t(g)$, similarly to (2.15). Let $Q_t(g)$ denote the fraction of consumers that purchase a variant within group *g* in period *t*. By first-order Taylor expansion, we find that

(2.24)
$$Q_{t-1,0} Q_{t0}^{-1} Q_t(g) = Q_{t-1}(g) + \theta \sum_{k \in B_{t-1}(g)} Q_{t-1,k} \tilde{w}_{t-1,k}(g) \left(1 - \frac{\delta_t(g) p_{t-1}}{p_t}\right),$$

which implies that

(2.25)
$$\delta_t(g) = \frac{p_t}{p_{t-1}} \left(1 + \frac{\left(Q_{t-1}(g) - Q_t(g) Q_{t-1,0} Q_{t0}^{-1}\right)}{\theta \overline{y}_{t-1}(g)}\right).$$

Similarly to (2.15), the index formula in (2.25) depends crucially on the fraction of consumers that do not purchase any variant, and in addition on the fraction of the demand that is allocated to group g. Thus, this means that the welfare effect of changes in prices of the respective variants and changes in the choice sets are fully captured through these fractional consumption terms, provided the approximation based on the first order Taylor expansion is viewed as sufficiently accurate.

3. Empirical analysis of the market for new automobiles in Norway

In this section, we report empirical results based on the methods discussed above for index construction.

3.1. Data

The automobile sales data are obtained from the Information Council for Road Traffic, Inc. These data contain information about prices from each of the individual automobile import firms. The sales data on quantities contain information about the number of cars sold in each month, from 1993 until 2001, on the following disaggregate level: brand, make, body, number of doors, engine performance, engine volume and number of driveshafts. The set of automobile variants are all the combinations of body, make, model, engine performance and number of driveshafts. The price data are based on the prices set by the firms that import automobiles, and may therefore differ somewhat from the actual market prices. The prices include indirect taxes, but do not include the cost associated with registration and possible transportation costs associated with delivery. The cost of possible supplementary equipment is not included in the price. There are some problems associated with the merging of the price data file and the file on quantities. The reason for this is that different definitions of categories have been used sometimes for the price data and the quantity data. In addition, the price data and the quantity data are given on different aggregation levels. In particular, there seems to be problems with linking quantity and price data for those brands for which the demand is low. No information about possible supplementary equipment is recorded. We have chosen to estimate yearly prices as the average of the prices in January, June and December each year. Data on cars privately imported to Norway are not available. Summary statistics of the data are given in Appendix C.

3.2. Estimation of the multinomial logit demand model with endogenous prices

As mentioned in Section 2, in many cases, it is not possible to explain fluctuations in $\{v_{tj}(g)\}$ by observable attributes. As regards automobiles, Table C1 shows that the fractional demand for sedan cars decreases from 0.34 in 1994 to 0.13 in 2002, whereas the fractional demand for station wagons increases from 0.25 in 1994 to 0.51 in 2002. The prices (Table C3) do not change much during this period and Table C2 shows that, for both types of car, the increase in the choice sets of variants is large. Thus, neither price changes nor other observable attributes are capable of explaining these trends in the demand. In the empirical analysis, the groups of variants are the three body types "Combi", "Sedan" and "Station wagon". We assume that

(3.1) $$v_{tj}(g) = z_{tj}\beta + \xi_j(g) + \mu_t(g) + \eta_{tj}(g),$$

for $j \in B_t(g)$, where the z_j-variables we use are fuel consumptions (liters per km) and engine performance, and β is a vector of unknown parameters. The term $\mu_t(g)$ is the mean utility of the variants within body group g in period t, whereas $\xi_j(g)$ represents the deviation in mean utility of variant j from the mean utility $\mu_t(g)$ within a given body group g. Note that $\xi_j(g)$ is assumed not to depend on time. This restriction is a crucial for achieving identification. The terms $\{\eta_{tj}(g)\}$ are zero mean disturbances.

Next, consider the estimation procedure. From (2.3) and (3.1), it follows that the probability of purchasing variant j in period t, given that variant j and variant 1 belong to body group g, is equal to

$$\text{(3.2)} \qquad \log\left(\frac{Q_{tj}(g)}{Q_{t1}(g)}\right) = \xi_j^*(g) + (z_{tj}(g) - z_{t1}(g))\beta - \theta(\tilde{w}_{tj}(g) - \tilde{w}_{t1}(g)) + \eta_{tj}(g) - \eta_{t1}(g)$$

for $j,1 \in B_t(g)$, where $\xi_j^*(g) = \xi_j(g) - \xi_1(g)$.

Next, consider the price-setting rule. We assume that prices are determined according to a setting with oligopolistic competition. It is assumed that each "producer" produces only one variant of automobile. Let $c_{tj}(g)$ denote the marginal cost of firm j (the firm that produces variant j) of type g in period t. Then, the expected profit of firm j of type g, conditional on prices, equals

$$\text{(3.3)} \qquad \pi_{tjg}(\boldsymbol{w}_t) = \left(w_{tj}(g) - c_{tj}(g)\right)Q_{tj}(g)M_t - K_{tj},$$

where M_t is the total number of consumers in year t and K_{jt} represents fixed costs. In the following, we assume that the quality indicators $\{v_{jt}(g)\}$ are exogenously given to the firms, and that firm j of type g maximizes (3.3) with respect to its own price $w_{jt}(g)$, taking the prices of other firms as given. In reality this assumption may not hold; it could be that firms take into account the demand for different "qualities"when setting prices. The first-order conditions that correspond to this maximization problem are given by

$$\text{(3.4)} \qquad \tilde{w}_{tj}(g) = \tilde{c}_{tj}(g) + \frac{1}{\theta\left(1 - Q_{tj}(g)\right)},$$

for $j = 1, 2, \ldots$, where $\tilde{c}_{jt}(g) = c_{jt}(g)/p_t$. Recall that $\{Q_{tj}(g)\}$ depend on prices, although this is suppressed in the notation. Anderson, Palma and Thisse (1992) have shown that there exists a unique price equilibrium determined by (3.4). Note that since marginal costs are positive the price equilibrium condition in (3.4) implies that $\tilde{w}_{tj}(g)\theta\left(1 - Q_{tj}(g)\right) > 1$. As regards the empirical specification of the price equation, we assume that

$$\text{(3.5)} \qquad \tilde{c}_{tj}(g) = b_j(g) + d_t(g) + \kappa_{tj}(g),$$

with the normalization $\sum_j b_j(g) = 0$, where $b_j(g)$ and $d_t(g)$ are unknown parameters and $\{\kappa_{tj}(g)\}$ are random error terms. Note that similarly to the specification of the latent quality attribute above, $\{b_j(g)\}$ does not depend on time. Next, we assume that the random error terms $\{\eta_{tj}(g)\}$ are independent and normally distributed with zero mean and variance $s^2(g)$, depending on g, and $\{\kappa_{tj}(g)\}$ are independent and normally distributed with zero mean and variance $r^2(g)$, depending on g. Moreover, $\kappa_{tj}(g)$ and $\eta_{\tau k}(g)$ are independent for all j, k, t and τ. In addition, the error terms in different body groups are assumed to be independent. Note that $\{\tilde{w}_{tj}(g)\}$ are endogenous because they depend on the endogenous fractional demands, $\{Q_{tj}(g)\}$, through (3.4). Consequently, we cannot estimate θ by OLS. For the same reason we cannot estimate the price relations in (3.4) by OLS. In Appendix B, we demonstrate that the likelihood function is given by

(3.6)

$$\log L = -\sum_g \sum_t \sum_{j \in B_t(g) \setminus \{1\}} \left[\left(\log\left(\frac{Q_{tj}(g)}{Q_{t1}(g)} \right) - \xi_j^*(g) - \left(z_{tj}(g) - z_{t1}(g) \right) \beta + \theta \left(\tilde{w}_{tj}(g) - \tilde{w}_{t1}(g) \right) \right)^2 \frac{1}{2s^2(g)} + \log s(g) \right]$$

$$- \sum_g \sum_t \sum_{j \in B_t(g)} \left[\left(\tilde{w}_{tj}(g) - c_j(g) - b_t(g) - \frac{1}{\theta\left(1 - Q_{tj}(g)\right)} \right)^2 \frac{1}{2r^2(g)} + \log r(g) \right] + \log |J|,$$

where J is the Jacobian associated with the transformation of the disturbances to the dependent variables, when the disturbances are viewed as functions of the dependent variables (prices and quantities sold) given by (3.2), (3.4) and (3.5). It turns out that this Jacobian does not depend on any of the unknown parameters of the model (Vitorino, 2004). In the actual estimation procedure, $\{Q_{tj}(g)\}$ are replaced by their corresponding observed frequencies $\{\hat{Q}_{tj}(g)\}$. However, the errors $\{\hat{Q}_{tj}(g) - Q_{tj}(g)\}$ are negligible. The loglikelihood function in (3.6) takes into account the fact that prices and fractional demands are *endogenous* variables. The estimation procedure now goes as follows. First, we maximize log$L$ with respect to the parameters $\{\xi_j^*(g)\}$, $\{c_j(g)\}$ and $\{b_t(g)\}$. The corresponding first-order conditions for this problem can be readily solved for these parameters. Second, we insert the formulas for the parameters $\{\xi_j^*(g)\}$, $\{c_j(g)\}$ and $\{b_t(g)\}$, obtained from the first-order conditions, into the loglikelihood function in (3.6) and we subsequently maximize the resulting loglikelihood function (given in (B.13) in Appendix B) with respect to the remaining parameters β, θ, $\{r(g)\}$ and $\{r(g)\}$. More precise details of this procedure are given in Appendix B. The estimates of θ and the variances *r*(*g*) and $s(g)$, $g = 1, 2, 3$, are given in Table 1 above.

From Table 1, we see that the observable attributes "fuel consumption" and "engine performance" are not significant. Thus, "price" is the only observable attribute that correlates significantly with demand.

Table 1. Estimates of structural parameters

Variable	Parameter	Estimate	*t*-statistics
Fuel consumption	β_1	0.2019×10^{-2}	1.4
Engine performance (kW)	β_2	0.2018×10^{-2}	1.4
Price $\times 10^{-5}$	θ	1.4985	16.9
Standard errors of tastes:			
Combi	$s(1)$	0.952	54.9
Sedan	$s(2)$	1.270	60.6
Station wagon	$s(3)$	1.009	61.4
Standard errors of marginal costs:			
Combi	$r(1)$	0.112	55.5
Sedan	$r(2)$	0.284	60.7
Station wagon	$r(4)$	0.218	62.1

We also carried out an estimation based solely on the demand relation in (3.2) and found that the parameters β_1, β_2, $r(1)$, $r(2)$, $r(3)$ and θ are practically equal to the estimates reported in Table 1. This means that OLS estimation based on (3.3) can be applied in this case. Therefore, we conclude that, without further knowledge or assumptions about marginal costs, the prices set by the firms in such a way that they are only weakly correlated with the disturbances, $\{\eta_{tj}(g)\}$. Alternatively, prices may be determined by some mechanism other than the simple oligopolistic price setting theory suggested above. Hence, the assumption of normally distributed error terms is not needed. However, an obvious weakness with our price-setting model in (3.4) is that only new cars are taken into account; the market for used cars is neglected.

3.3. Calculation of a quality adjusted price index for the nested multinomial logit model

In this section, we consider the calculation of price indexes based on the nested multinomial logit demand model for new automobiles. Let $N_{tj}(g)$ denote the number of variants of type *j* within body type *g* sold in year *t*. From (2.14) it follows that

(3.7) $$\sum_{g}\sum_{k\in B_{t-1}(g)}\frac{N_{t-1,k}(g)}{N_{t-1}}\exp\left(\left(1-\delta_t p_{t-1}/p_t\right)\theta\tilde{w}_{t-1,k}(g)\right)=\frac{\hat{Q}_{t0}^{-1}-1}{\hat{Q}_{t-1,0}^{-1}-1}.$$

The fraction p_t/p_{t-1} is estimated by the conventional Laspeyres index for the goods other than new cars. A simple version of the conventional Laspeyres index for new cars, δ_t^L, is calculated as

Table 2. Different of price indexes for all new automobiles (percent), multinomial logit model

All automobiles	1994	1995	1996	1997	1998	1999	2000	2001	2002
The Laspeyres index for new automobiles	100	102.5	98.2	100.8	102.2	101.9	103.1	106.5	108.2
The Laspeyres index for other goods (p_t)	100	103.6	104.4	107.1	109.7	112.6	116.4	119.9	121.4
Hedonic regression	100	102.1	96.3	98.6	98.7	97.8	97.8	104.3	106.0
Quality adjusted index, $\theta=1.5$ (estimated value)	100	101.4	94.1	97.7	97.9	106.0	111,2	115.2	116.2
First-order approximation of the quality adjusted index, (eq. 2.15), θ= 1.5	100	101.5	93.1	97.1	97.8	105.9	111.7	116.3	117.8
Quality adjusted index with other values of θ									
θ= 2.1	100	102.0	97.0	100.4	101.2	108.0	112.8	116.6	117.8
θ= 1.9	100	101.9	96.2	99.7	100.3	107.5	112.4	116.3	117.4
θ= 1.7	100	101.7	95.3	98.8	99.2	106.9	111.9	115.8	116.8
θ= 1.3	100	101.1	92.5	96.2	96.1	105.0	110.4	114.3	115.3
θ= 1.1	100	100.6	90.4	94.3	93.7	103.5	109.1	113.2	114.0
θ= 0.9	100	99.9	87.4	91.4	90.3	101.3	107.3	111.4	112.1
Fraction of persons 16–66 years of age that buy a new car	0.030	0.032	0.041	0.039	0.041	0.034	0.032	0.031	0.031

(3.8) $$\delta_t^L = \frac{\sum_g \sum_{j \in C_t(g)} w_{tj}(g) N_{t-1,j}(g)}{\sum_g \sum_{j \in C_t(g)} w_{t-1,j}(g) N_{t-1,j}(g)},$$

where $C_t(g) = B_t(g) \cap B_{t-1}(g)$. In Table 2, we report the calculation of different price indexes. We see that the Laspeyres index is higher than the quality adjusted price index up to 1997, whereas, from 1998 onwards, it yields lower figures than the quality adjusted index. The quality adjusted index drops from 101.4 percent to 94.1 percent from 1995 to 1996, and increases rapidly from 97.9 percent in 1998 to 111.2 percent in 2000. One important reason why the increase in the demand from 1995 to 1996 is so high is that the condemnation deposit was increased in 1996 in order to increase condemnation and stimulate the purchase of new and more environmentally efficient cars.

We have also used the approximation formula given in (2.15) to calculate the quality adjusted price index. From Table 2 we see that the figures produced by (2.15) are close to the exact index figures determined by (3.7).

Moreover, we have applied the hedonic regression method to calculate a hedonic price index. We refer to Appendix A for an explanation and critique of the hedonic method. The hedonic regression estimates are given in Table C7 in Appendix C. We note that from 1999 the hedonic index yields considerably lower figures than the quality adjusted price index and it is also somewhat lower than the Laspeyres price index.

Further down in Table 2, we have calculated the quality adjusted index for different values of θ. From the results, we can conclude that the index changes little when θ varies from 1.3 to 1.7. Even when θ varies from 1.1 to 1.9, the changes in the index are moderate in most cases.

From Table 2 we note that the fluctuations in the quality adjusted price index follow closely the fluctuations in the fraction of consumers that purchase (do not purchase) a car (last row in the table). The reason for this is apparent when we look at the index formula (2.15). Recall that this does not mean that the effects of changes in prices, choice sets and latent quality attributes are ignored, but simply that, under the assumptions of our demand model, these effects are captured by the fraction of consumers that do not purchase a variant (in the respective periods).

4. Conclusion

In this paper, we have developed a particular approach for calculating quality adjusted price indexes in markets with differentiated products. We have discussed how one can use the theory of discrete choice to derive exact price indexes that account for quality changes. Our approach is an extension of Trajtenberg’s method that explicitly takes into account the discrete choice setting, allowing for endogenous time-varying latent quality attributes and the option of not purchasing any variant of the discrete good. A key result established is that the fraction of consumers that do not purchase a variant of the differentiated product, and the mean population expenditure on the variants purchased, constitute sufficient statistics for the calculation of the quality adjusted price index, given the prices of the outside divisible goods and the price parameter θ in the demand function.

Our empirical application is based on data on sales of new automobiles in Norway. The results show that adjusting for quality implies a decrease in the corresponding price index until 1998 compared with the Laspeyres price index for automobiles, and an increase from 1999 to 2002. In parti cular, the fluctuation of the quality adjusted index parallels the fluctuations in the fraction of consumers that do not purchase a car (or equivalently, the fraction or persons that purchase a car). For example, from the last row in Table 2 we note that the fraction of persons that purchase a car increases a lot from 1995 to 1996 and decreases rapidly from 1998 to 1999. The corresponding quality adjusted price index decreases sharply from 1995 to 1996 and increases sharply from 1998 to 1999. This is due to the fact that, in addition to the mean population expenditure on the discrete product, changes in the fraction of consumers that do not purchase a car fully captures the cost of living effects of changes in choice sets, prices and latent quality.

We have also applied the hedonic regression method. The results show that the hedonic regression method produces lower estimated than the quality adjusted price index from 1999.

The methodology applied in this paper depends crucially on the specification of the demand model. In our model, the utility function is linear in income, and this property implies that the demand model does not depend on income. Researchers such as Pakes et al. (1993), Berry et al. (1995), Nevo (2003) and Vitorino (2004) have carried out empirical demand analyses based on a more general structure of the demand model with time-constant latent quality attributes. However, only Pakes et al. (1993) have calculated quality adjusted price indexes, under the restriction that the latent quality attributes are constant over time. Third, the modeling framework discussed in this paper is purely static, whereas automobiles are important durables that cannot be satisfactorily analyzed without an intertemporal modeling framework that incorporates consumers’ expectations and uncertainties. Unfortunately, however, this is a very demanding task and is far beyond the scope of the present analysis.

References

Anderson, S. P., A. de Palma and J.-F. Thisse (1992): *Discrete Choice Theory of Product Differentiation*. MIT Press, London.

Bartik, T. (1987): The Estimation of Demand Parameters in Hedonic Price Models. *Journal of Political Economy* **95**, 81–88.

Berry, S. (1994): Estimating Discrete Choice Models of Product Differentiation. *RAND Journal of Economics* **25**, 242–262.

Berry, S., J. A. Levinsohn and A. Pakes (1995): Automobile Prices in Market Equilibrium. *Econometrica* **63**, 841–890.

Boskin, M., E.R. Dulberger, Z. Griliches, R.J. Gordon and D. Jorgensen (1996): Toward a More Accurate Measure of the Cost of Living. Final Report to the U.S. Senate Finance Committee.

Crawford, G. S. (1997): New Products, New Programs and Prices: Measuring Consumer Benefits to Changes in Cable Television Choices, 1989–1995. Working Paper 97/32, Department of Economics, Duke University.

Dagsvik, J. K., and G. Liu (2002): A Note on Maximum Likelihood Estimation of Market Equilibrium in the Presence of Latent Attributes. Mimeo, Statistics Norway.

Epple, D. (1987): Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products. *Journal of Political Economy* **95**, 59–80.

Feenstra, R. C. (1995): Exact Hedonic Price Indexes. *Review of Economics and Statistics* **77**, 634–653.

Hulten, C. R. (2003): Price Hedonics: A Critical Review. *Economic Policy Review of the Federal Reserve Bank of New York*, September.

Jonker, N. (2002): Constructing Quality Adjusted Price Indices: A Comparison of Hedonic and Discrete Choice Models. Working Paper No. 172, European Central Bank, Frankfurt am Main.

McFadden, D. (1984): Econometric Analysis of Qualitative Response Models. In Z. Grilliches and M. Intriligator (eds.): *Handbook of Econometrics*, Vol. II, pp. 1395–1457, North Holland, Amsterdam.

Nevo, A. (2003): New Products, Quality Changes and Welfare Measures Computed from Estimated Demand Systems. *Review of Economics and Statistics* **85**, 266–275.

Opplysningsrådet for Veitrafikken AS (OFV), Bilpriser, pr. 15 August 2001. (The Information Council for Road Traffic in Norway, Inc. Car Prices per 15 August 2001).

Pakes, A., S. Berry and J. A. Levinsohn (1993): Applications and Limitations of Some Recent Advances in Empirical Industrial Organization: Price Indexes and the Analysis of Environmental Change. *American Economic Review* (Papers and Proceedings) **83**, 240–246.

Rosen, S. (1974): Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *Journal of Political Economy* **82**, 34–55.

Song, M. (2005): The Quality Adjusted Price Index and Consumer Heterogeneity. Mimeo, School of Economics, Georgia Institute of Technology.

Trajtenberg, M. (1990): *Economic Analysis of Product Innovation*. Harvard University Press, London.
Vitorino, M. A. (2004): Estimation of Equilibrium Models Using a Likelihood Based Approach. Mimeo, Business School, University of Chicago, Chicago.

Appendix A

The hedonic regression approach

For the sake of relating the theoretical development above to the conventional literature on hedonic price regressions, we now consider the deterministic case with no random error term in the utility function given in (2.1). In this case, we shall review a possible theoretical motivation for the hedonic regression approach. The discussion is similar to Trajtenberg (1990, pp. 35–37). Let

$$U\left(x_0, \sum_k x_k \mu_{tk}\right)$$

be the utility function of the consumers where x_k is the quantity of product variant $k > 0$, x_0 represents the quantity of other goods, U is a quasi-concave and increasing function and $\mu_{tk} > 0$ are weights that are supposed to represent "quality". The form of the utility function means that there is perfect substitution between quality and quantity. The utility function above is equal for all consumers. In this case, it follows that only a single variant will be demanded unless the quality-adjusted prices are equal

(A.1) $$\frac{w_{tj}}{\mu_{tj}} = \frac{w_{t1}}{\mu_{t1}}.$$

The quality price index in this case can thus be expressed as

(A.2) $$\delta_t = \frac{w_{t1}/\mu_{t1}}{w_{t-1,1}/\mu_{t-1,1}}.$$

Assume now that

(A.3) $$\log\left(\frac{\mu_{tj}}{p_t}\right) = z_{tj}\beta,$$

where z_{tj} is a vector of suitable observed attributes of variant *j*. Then, we can write (3.1) as

(A.4) $$\ln w_{tj} = \alpha_t + z_{tj}\beta + \log p_t,$$

where

$$\alpha_t = \ln w_{t1} - \mu_{t1}.$$

Hence, the period specific intercept α_t can be used to compute δ_t because (A.2) implies that

(A.5) $$\delta_t = \frac{p_t \exp(\alpha_t - \alpha_{t-1})}{p_{t-1}}.$$

There are several shortcomings with the conventional hedonic regression approach outlined here. First, as population heterogeneity is ignored, eq. (A.1) must hold in order to be consistent with the observed fact that there is positive demand for each variant in the market. However, this implies that consumers will be indifferent with respect to the different variants, which seems rather unrealistic. Second, the hedonic approach ignores the effect of variations in the choice set. Specifically, in a dynamic market, some variants disappear whereas others emerge as a result of innovations. The quality adjusted index approaches discussed in Section 2 explicitly take the choice set B_t into account, whereas the hedonic approach fails in this respect. We refer readers to Hulten (2003) and the references therein for a critical review of additional aspects of the hedonic regression method. Computed from Estimated Demand Systems.

Appendix B

Aggregation of subindexes; proof of eq. (2.22):

Let

$$Q_t(g)=\sum_{k\in B_t(g)} Q_{tk}(g).$$

The left hand side of (2.21) is equal to $Q_t(g)/Q_{t0}$ so that (2.21) can be expressed as

(B.1) $$Q_{t0}^{-1}Q_t(g)=\sum_{k\in B_{t-1}(g)} \exp\left(v_{t-1,k}(g)-\theta\tilde{w}_{t-1,k}(g)\delta_t(g)\,p_{t-1}/p_t\right).$$

We immediately note that when (B.1) is aggregated over *g* the left hand side of (B.1) becomes equal to $Q_{t0}^{-1}-1$. Hence (2.11) and (B.1) imply that

(B.2) $$\begin{aligned}&\sum_g\sum_{k\in B_{t-1}(g)} \exp\left(v_{t-1,k}(g)-\theta\tilde{w}_{t-1,k}(g)\delta_t(g)\,p_{t-1}/p_t\right)\\&=\sum_g\sum_{k\in B_{t-1}(g)} \exp\left(v_{t-1,k}(g)-\theta\tilde{w}_{t-1,k}(g)\delta_t\,p_{t-1}/p_t\right).\end{aligned}$$

Similarly to (2.14) it follows that (B.1) can be written as

(B.3) $$Q_{t-1,0}^{-1}\sum_{k\in B_{t-1}(g)} Q_{t-1,k}(g)\exp\left(\left(1-\delta_t(g)\,p_{t-1}/p_t\right)\theta\tilde{w}_{t-1,k}(g)\right)=Q_{t0}^{-1}Q_t(g)$$

and furthermore

(B.4) $$\begin{aligned}&\sum_g\sum_{k\in B_{t-1}(g)} Q_{t-1,k}(g)\exp\left(\left(1-\delta_t(g)\,p_{t-1}/p_t\right)\theta\tilde{w}_{t-1,k}(g)\right)\\&=\sum_g\sum_{k\in B_{t-1}(g)} Q_{t-1,k}(g)\exp\left(\left(1-\delta_t\,p_{t-1}/p_t\right)\theta\tilde{w}_{t-1,k}(g)\right).\end{aligned}$$

A first order Taylor expansion of both sides of (B.4) yields

(B.5) $$\begin{aligned}&\sum_g\sum_{k\in B_{t-1}(g)} Q_{t-1,k}(g)\left(1+\left(1-\delta_t(g)\frac{p_{t-1}}{p_t}\right)\theta\tilde{w}_{t-1,k}(g)\right)\\&\cong\sum_g\sum_{k\in B_{t-1}(g)} Q_{t-1,k}(g)\left(1+\left(1-\delta_t\frac{p_{t-1}}{p_t}\right)\theta\tilde{w}_{t-1,k}(g)\right)\end{aligned}$$

which implies that

(B.6)
$$\delta_t = \frac{\sum_g \delta_t(g) \sum_{k\in B_{t-1}(g)} Q_{t-1,k}(g) w_{t-1,k}(g)}{\sum_r \sum_{k\in B_{t-1}(r)} Q_{t-1,k}(r) w_{t-1,k}(r)} .$$

This completes the proof.

Q.E.D.

Maximum likelilhood estimation; proof of the claim that the likelihood function has the form given in (3.6):

Consider first the demand relations in (3.2). The part of the likelihood function that corresponds to (3.2) for body group g at time t (disregarding the Jacobian) equals

(B.7)
$$E \prod_{j\in B_t(g)\setminus\{1\}} \exp\left(-\left(\log\left(\frac{Q_{tj}(g)}{Q_{t1}(g)}\right) - \xi_j^*(g) - (z_{tj}(g) - z_{t1}(g))\beta + \theta(\tilde{w}_{tj}(g) - \tilde{w}_{t1}(g)) + \eta_{t1}(g) \right)^2 \frac{1}{2s^2(g)} - \log s(g) \right)$$

where the expectation is taken with respect to $\{\eta_{t1}(g)\}$. Let

$$R_{tj}(g) = \log\left(\frac{Q_{tj}(g)}{Q_{t1}(g)}\right) - \xi_j^*(g) - \left(z_{tj}(g) - z_{t1}(g)\right)\beta + \theta\left(\tilde{w}_{tj}(g) - \tilde{w}_{t1}(g)\right).$$

Then one can express (B.7) as

(B.8)
$$\begin{aligned} & E \prod_{j\in B_t(g)\setminus\{1\}} \exp\left(-\left(R_{tj}(g) + \eta_{t1}(g)\right)^2 \frac{1}{2s^2(g)} - \log s(g) \right) \\ & = E \exp\left(- \sum_{j\in B_t(g)\setminus\{1\}} \left(\left(R_{tj}^2(g) + 2\eta_{t1}(g) R_{t\cdot}(g) + \eta_{t1}^2(g)\right) \frac{1}{2s^2(g)} - \log s(g) \right) \right) \end{aligned}$$

where

$$R_{t\cdot}(g) = \sum_{j\in B_t(g)\setminus\{1\}} R_{tj}(g) .$$

Now observe that (3.2) implies that $R_{t\cdot}(g) \approx 0$, so that (B.8) reduces to

$$\exp\left(- \sum_{j\in B_t(g)\setminus\{1\}} (R_{tj}^2(g)/2s^2(g) - \log s(g)) \right) \cdot E \exp\left(-\eta_{t1}^2(g) n_t(g)/2s^2(g)\right)$$

where $n_t(g)$ is the number of variants in $B_t(g)\setminus\{1\}$.

Furthermore, we have that

$$E\exp\left(-\eta_{t1}^2(g)\,n_t(g)\big/2s^2(g)\right)$$
$$=\int \exp\left(-\frac{x^2 n_t(g)}{2s^2(g)}\right)\exp\left(-\frac{x^2}{2s^2(g)}\right)\frac{dx}{s(g)\sqrt{2\pi}}=\frac{1}{s(g)\sqrt{2\pi}}\int \exp\left(-\frac{n_t(g)+1}{2s^2(g)}\cdot x^2\right)dx.$$

By change of variable; $y=x\sqrt{1+n_t(g)}\big/s(g)$ we obtain that the last integral reduces to

$$\frac{1}{\sqrt{2\pi}\sqrt{1+n_t(g)}}\int \exp\left(-\frac{y^2}{2}\right)dy=\frac{1}{\sqrt{1+n_t(g)}}.$$

Thus, we have proved that (B.8) equals

$$\exp\left(-\sum_{j\in B_t(g)\setminus\{1\}}(R_{tj}^2(g)\big/2s^2(g)-\log s(g))\right)\frac{1}{\sqrt{1+n_t(g)}}.$$

Hence, we have proved that (B.7) can be written as

(B.9)
$$\frac{1}{\sqrt{1+n_t(g)}}\cdot\prod_{j\in B_t(g)\setminus\{1\}}\exp\left(-\left(\log\left(\frac{Q_{tj}(g)}{Q_{t1}(g)}\right)-\xi_j^*(g)-(z_{tj}(g)-z_{t1}(g))\beta+\theta(\tilde{w}_{tj}(g)-\tilde{w}_{t1}(g))\right)^2\frac{1}{2s^2(g)}-\log s(g)\right).$$

Consequently, it follows from (3.4), (3.5) and (B.9) that the loglikelihood function is given by

(B.10)
$$\log L=-\sum_g\sum_t\sum_{j\in B_t(g)\setminus\{1\}}\left[\left(\log\left(\frac{Q_{tj}(g)}{Q_{t1}(g)}\right)-\xi_j^*(g)-\left(z_{tj}(g)-z_{t1}(g)\right)\beta+\theta\left(\tilde{w}_{tj}(g)-\tilde{w}_{t1}(g)\right)\right)^2\frac{1}{2s^2(g)}+\log s(g)\right]$$
$$-\sum_g\sum_t\sum_{j\in B_t(g)}\left[\left(\tilde{w}_{tj}(g)-c_j(g)-b_t(g)-\frac{1}{\theta\left(1-Q_{tj}(g)\right)}\right)^2\frac{1}{2r^2(g)}+\log r(g)\right]+\log|J|-\frac{1}{2}\sum_g\sum_t\log(1+n_t(g))$$

where J is the Jacobian associated with the transformation of variables, from $\left\{\eta_{tj}(g)-\eta_{t1}(g),\mu_{\tau k}(r),t,\tau=1,2,...,T,\ j\in B_t(g),k\in B_\tau(r),g,r=1,2,3\right\}$ to $\left\{Q_{tj}(g),w_{\tau k}(r),t,\tau=1,2,...,T,\ j\in B_t(g),\tau\in B_\tau(r),g,r=1,2,3\right\}$, given by

(B.11) $$\eta_{tj}(g)-\eta_{t1}(g)=\log\left(\frac{Q_{tj}(g)}{Q_{t1}(g)}\right)-\xi_j^*(g)-(z_{tj}(g)-z_{t1}(g))\beta+\theta(\tilde{w}_{tj}(g)-\tilde{w}_{t1}(g))$$

and

(B.12)
$$\kappa_{tj}(g)=\tilde{w}_{tj}(g)-c_j(g)-b_t(g)-\frac{1}{\theta\left(1-Q_{tj}(g)\right)}.$$

It can be readily demonstrated (see Dagsvik and Liu, 2002) that the Jacobian J is independent of the unknown parameters of the model. See also Vitorino (2004) who has demonstrated this for the case with few variants. The Jacobian is therefore irrelevant for the solution of the maximization problem above, and it can be removed from the likelihood function. Let $N_{tj}(g)$ be the number of variants sold of type j within body type g, $N_t(g)$ the number of variants within $B_t(g)$, and N_t the total number of variants sold. For notational convenience let us now introduce the notation

$$Y_{tj}(g)=\log\left(\frac{\hat{Q}_{tj}(g)}{\hat{Q}_{t1}(g)}\right)=\log\left(\frac{N_{tj}(g)}{N_{t1}(g)}\right),\quad \bar{Y}_{.j}(g)=\frac{1}{T}\sum_t Y_{tj}(g),\quad \bar{\tilde{w}}_{.j}(g)=\frac{1}{T}\sum_t \tilde{w}_{tj}(g)$$

$$V_{tj}(g)=\frac{1}{1-\left(1-\hat{Q}_{t0}\right)N_{tj}(g)},\quad \bar{V}_{.j}(g)=\frac{1}{T}\sum_t V_{tj}(g),\quad \bar{V}_{t.}(g)=\frac{1}{N_t(g)}\sum_{j\in B_t(g)} V_{tj}(g),\quad \bar{\bar{V}}_{..}(g)=\frac{1}{T}\sum_t \bar{V}_{t.}(g),$$

where $\hat{Q}_{t0}$ denotes an estimate of Q_{t0} and T is the number of years we have observations for. If we use the first order conditions and solve for the intercepts and subsequently insert into the likelihood function we obtain that the remaining parameters θ, $s^2(g)$, $r^2(g)$, $g=1,2,3$, can be estimated by maximizing

(B.13)

$$\begin{aligned}
&\log L^*\\
&=-\sum_t\sum_g\sum_{j\in B_t(g)\setminus\{1\}}\left[\frac{\left(Y_{tj}(g)-\bar{Y}_{.j}(g)-\left(z_{tj}(g)-z_{t1}(g)-\bar{z}_{.j}(g)+\bar{z}_{.1}(g)\right)\beta+\theta\left(\tilde{w}_{tj}(g)-\tilde{w}_{t1}(g)-\bar{\tilde{w}}_{.j}(g)+\bar{\tilde{w}}_{.1}(g)\right)\right)^2}{2s^2(g)}\right]\\
&-\sum_t\sum_g\left(N_t(g)-1\right)\log s(g)\\
&-\sum_t\sum_g\sum_{j\in B_t(g)}\left[\left(w_{tj}(g)-\bar{w}_{.j}(g)-\bar{w}_{t.}(g)+\bar{\bar{w}}_{..}-\frac{1}{\theta}\left(V_{tj}(g)-\bar{V}_{.j}-\bar{V}_{t.}+\bar{\bar{V}}_{..}\right)\right)^2\frac{1}{2r^2(g)}+\log r(g)\right].
\end{aligned}$$

Q.E.D.

Appendix C

Summary statistics

Table C 1. Number of new cars sold by year and body of car

Year	Body of car								Total sales
	Combi		Sedan		Station wagon		Other carmakers		
	Fractions	Levels	Fractions	Levels	Fractions	Levels	Fractions	Levels	
1994	0.41	32 226	0.34	26 406	0.25	19 605	0.003	246	78 483
1995	0.43	36 043	0.33	28 307	0.23	19 790	0.004	369	84 509
1996	0.46	49 794	0.27	29 810	0.25	27 672	0.012	1 276	108 552
1997	0.40	40 983	0.28	28 495	0.31	32 465	0.014	1 414	103 357
1998	0.39	41 608	0.22	23 568	0.38	40 677	0.016	1 667	107 520
1999	0.36	32 229	0.20	18 064	0.42	37 736	0.015	1 354	89 383
2000	0.37	31 336	0.18	15 005	0.44	37 828	0.015	1 323	85 492
2001	0.35	29 496	0.17	14 514	0.46	38 583	0.013	1 081	83 674
2002	0.35	29 312	0.13	10 927	0.51	42 815	0.013	1 086	84 140

Table C 2. Number of variants of cars in the market each year

Year	Type of body								
	Combi	Variants entering	Variants dis-appearing	Sedan	Variants entering	Variants dis-appearing	Station wagon	Variants entering	Variants dis-appearing
1994	149			177			129		
1995	153			174			124		
1996	186			214			170		
1997	162			188			194		
1998	173			195			226		
1999	158			205			244		
2000	182			221			271		
2001	184			237			277		
2002	203			239			303		

Table C 3. Mean deflated prices across cars within type of body. NOK

Type of body/Year	min.	max.	mean	st.dev.
Combi				
1994	89 450	637 600	219 286	90 705
1995	84 336	583 398	215 598	85 218
1996	85 462	488 506	198 353	70 468
1997	92 904	463 658	193 107	66 530
1998	89 304	446 126	189 628	64 004
1999	95 826	468 561	187 250	56 987
2000	95 132	453 265	183 292	55 149
2001	97 262	453 628	194 480	60 224
2002	102 883	445 634	194 556	59 859
Sedan				
1994	110 533	1 267 650	327 336	179 339
1995	112 058	1 423 745	329 694	201 014
1996	109 642	1 577 267	323 329	213 100
1997	111 173	1 554 622	328 238	207 109
1998	101 094	1 312 671	330 469	190 127
1999	98 490	1 171 942	316 451	167 568
2000	127 513	1 262 887	323 045	176 836
2001	130 901	2 551 376	371 085	256 676
2002	133 773	2 606 575	371 043	244 822
Station wagon				
1994	117 756	765 800	342 997	140 346
1995	114 431	1 008 340	347 671	158 999
1996	125 239	871 648	319 025	127 045
1997	128 758	1 101 774	328 810	149 672
1998	121 263	935 886	302 236	140 331
1999	117 022	882 179	273 770	115 359
2000	107 732	932 131	291 965	143 046
2001	107 812	929 358	309 760	144 710
2002	112 109	1 093 081	320 761	156 064

Table C 4. Number of new combi cars sold by year and make of car

Year	Make of car					
	Alfa Romeo	Audi	BMW	Citroen	Daewoo	Daihatsu
1994	-	-	134	2370	-	219
1995	-	-	150	1817	130	58
1996	-	273	282	2035	251	19
1997	-	1142	228	1426	257	278
1998	-	948	106	1440	178	111
1999	-	853	47	929	72	229
2000	-	922	13	541	219	367
2001	113	731	100	935	262	180
2002	93	697	128	1125	58	204
Year	Make of car					
	Fiat	Ford	Honda	Hyundai	Kia	Lada
1994	637	3142	354	516	-	94
1995	1022	3465	695	2209	-	231
1996	651	4159	1259	3546	372	136
1997	871	4378	1236	2196	680	85
1998	715	3146	630	1866	716	57
1999	273	2973	205	1327	601	12
2000	306	2885	120	1336	274	-
2001	507	2476	407	933	308	-
2002	469	2477	639	844	632	-
Year	Make of car					
	Mazda	Mitsubishi	Nissan	Opel	Peugoet	Renault
1994	1947	928	2336	3611	2715	2117
1995	1334	741	2244	3495	2161	2446
1996	2849	1959	3091	4549	2528	2939
1997	2649	984	3538	3859	1872	2191
1998	1582	990	2763	4780	1176	1634
1999	1407	728	1657	3396	2282	1197
2000	1129	562	2076	2125	2345	900
2001	792	441	1309	2546	3027	1371
2002	701	381	1166	1697	3403	920

Table C 4 (cont.). Number of new combi cars sold by year and make of car

Year	Make of car				
	Rover	Saab	Seat	Skoda	Subaru
1994	-	949	402	288	133
1995	-	1844	637	724	135
1996	386	2209	621	844	560
1997	245	1959	739	977	848
1998	245	1568	765	1548	214
1999	213	1689	660	1265	159
2000	102	1436	655	1570	100
2001	42	1122	578	906	118
2002	29	963	441	722	119

Year	Make of car			
	Suzuki	Toyota	Volkswagen	Volvo
1994	237	3797	5299	208
1995	248	3099	7199	100
1996	351	4806	9259	58
1997	305	5094	7382	-
1998	324	4661	10039	-
1999	188	4670	7019	-
2000	348	5443	6711	-
2001	754	5066	4543	-
2002	823	5373	5324	-

Table C 5. Number of new sedan cars sold by year and make of car

Year	Make of car					
	Alfa Romeo	Audi	BMW	Bentley	Buick	Cadillac
1994	-	1008	265	-	2	10
1995	-	1712	145	-	-	4
1996	-	2128	891	-	-	-
1997	-	2298	1561	-	-	-
1998	218	1772	1845	-	-	-
1999	341	1305	1560	-	-	-
2000	169	950	1355	-	-	-
2001	96	1735	1114	2	-	-
2002	42	944	979	1	-	-

Year	Make of car					
	Chevrolet	Chrysler	Daewoo	Daihatsu	Fiat	Ford
1994	5	230	-	3	34	1351
1995	-	2090	348	-	38	1018
1996	8	1730	408	-	5	1169
1997	17	1136	412	-	62	1280
1998	-	783	219	-	37	669
1999	57	755	90	-	17	603
2000	49	387	109	-	-	450
2001	12	97	121	-	-	878
2002	4	100	3	-	-	373

Year	Make of car					
	Honda	Hyundai	Jaguar	Kia	Lada	Lancia
1994	1180	2026	-	130	186	-
1995	727	1981	-	998	84	-
1996	1090	1658	-	787	33	1
1997	953	1374	-	577	10	11
1998	499	709	61	233	7	2
1999	490	576	54	57	1	10
2000	464	527	24	144	-	8
2001	365	287	19	33	-	-
2002	56	120	36	18	-	-

Table C 5 (cont.). Number of new sedan cars sold by year and make of car

Year	Make of car					
	Lexus	Maserati	Mazda	Mercedes Benz	Mitsubishi	Nissan
1994	-	-	941	-	2176	1828
1995	-	-	938	-	2639	2172
1996	-	-	1544	-	2383	2650
1997	-	-	1074	-	3109	2485
1998	-	-	1021	-	2706	1831
1999	82	-	709	-	1520	975
2000	129	1	731	-	929	986
2001	74	-	565	1231	509	623
2002	43	-	530	1378	536	238

Year	Make of car					
	Opel	Peugeot	Renault	Rover	Saab	Seat
1994	4111	1371	286	-	180	222
1995	3375	1156	129	-	230	525
1996	4686	2376	28	219	147	346
1997	2988	1449	488	248	174	330
1998	1868	684	220	254	674	187
1999	2102	667	21	441	434	136
2000	1378	809	17	311	247	182
2001	874	770	27	149	277	159
2002	1064	352	2	94	506	104

Year	Make of car					
	Skoda	Subaru	Suzuki	Toyota	Volkswagen	Volvo
1994	-	326	95	3552	1734	3292
1995	-	256	222	3396	1417	2897
1996	-	310	551	3995	1329	1254
1997	-	249	225	3628	4709	1654
1998	-	120	133	3278	2033	1620
1999	-	120	153	2022	2290	1401
2000	-	223	4	1871	1909	896
2001	-	199	-	1388	1612	1315
2002	134	227	-	1223	945	879

Table C 6. Number of new station wagons sold by year and make of car

Year	Make of car					
	Audi	BMW	Chevrolet	Chrysler	Citroen	Daewoo
1994	1292	31	-	106	360	-
1995	541	11	7	125	689	-
1996	1935	207	2	76	1465	-
1997	2028	629	3	473	1465	33
1998	2623	663	4	186	1196	6
1999	1951	697	2	97	1336	3
2000	1714	1052	2	326	1476	118
2001	1676	1017	1	122	1407	206
2002	2184	953	-	35	1615	15

Year	Make of car					
	Daihatsu	Dodge	Fiat	Ford	Ford USA	Honda
1994	-	-	44	5252	-	1
1995	-	-	36	5429	-	28
1996	-	6	16	3217	-	68
1997	-	2	189	4586	27	1137
1998	205	-	267	3275	25	2032
1999	213	1	104	3120	-	1809
2000	147	-	74	3077	-	1559
2001	95	-	76	4596	-	927
2002	148	-	1	3585	-	1486

Year	Make of car					
	Hyundai	Jeep	Kia	Lada	Land Rover	Lexus
1994	-	51	7	-	-	-
1995	-	59	38	17	-	-
1996	1	42	30	11	-	-
1997	28	29	19	2	48	-
1998	68	18	28	2	158	-
1999	1137	70	197	-	752	-
2000	2032	31	206	-	602	16
2001	1809	7	158	-	271	17
2002	1559	37	91	-	438	20

Table C 6 (cont.). Number of new station wagons sold by year and make of car

Year	Make of car					
	Mazda	Mercedes Benz	Mitsubishi	Nissan	Opel	Peugeot
1994	299	-	1044	843	3032	736
1995	68	-	1088	114	3201	719
1996	20	-	1394	220	3132	441
1997	1	-	2206	589	5410	1542
1998	1669	-	2005	1034	4385	1917
1999	1735	-	2453	1015	5260	2033
2000	1283	-	2040	894	4801	2400
2001	866	890	510	581	4303	1957
2002	833	786	1232	2160	2899	2921

Year	Make of car					
	Pontiac	Renault	Rover	Saab	Seat	Skoda
1994	10	10	-	-	-	88
1995	4	273	-	-	-	61
1996	3	1679	-	-	-	847
1997	-	2472	-	-	11	597
1998	-	2225	-	-	315	652
1999	-	1868	-	696	276	1617
2000	-	2246	-	479	156	1729
2001	-	1914	37	616	119	1949
2002	-	1812	41	756	94	1551

Year	Make of car					
	Sangyong	Subaru	Suzuki	Toyota	Volkswagen	Volvo
1994	1	134	20	1036	3629	2155
1995	4	130	43	1026	3199	3012
1996	1	207	649	3997	5964	2361
1997	1	409	602	4373	3277	3666
1998	1	1605	2439	3909	5105	3917
1999	8	1466	1690	3304	4591	1542
2000	1	1248	1350	4840	4610	1310
2001	3	990	994	5666	5615	1358
2002	1	1214	806	6986	4826	2519

Table C 7. Hedonic regression for new cars wagon[1)]

Variable	Parameter estimate	t-value
Intercept	11.5331	756.6
Fuel consumption$\times 10^{-4}$	0.37	0.4
Performance kw	0.0102	119.6
1994	0.1600	16.6
1995	0.1536	15.9
1996	0.0987	11.0
1997	0.0870	9.5
1998	0.0544	6.1
1999	0.0214	2.4
2000	-0.0045	-0.5
2001	0.0019	0.2
Alfa Romeo	0.0097	0.4
Audi	0.1708	14.0
BMW	0.1443	10.9
Bentley	0.2557	2.3
Cadillac	-0.0142	-0.1
Chevrolet	0.2440	3.1
Chrysler	-0.0009	-0.0
Citroen	-0.0020	-0.1
Daewoo	-0.2297	-10.1
Daihatsu	-0.1466	-4.9
Dodge	0.2076	2.6
Fiat	-0.1422	-6.5
Ford	-0.0272	-2.0
Ford USA	0.3675	3.3
Honda	-0.0298	-1.7
Hyundai	-0.1959	-10.3
Jaguar	0.0163	0.5
Jeep	0.3651	13.1
Kia	-0.1310	-5.1
Lada	-0.4280	-11.8
Lancia	0.0141	0.2
Land Rover	0.3767	12.7
Lexus	0.1761	3.3
Maserati	-0.1592	-1.0
Mazda	-0.0333	-2.2

[1)] Reference year: 2002. Reference make is Volvo and reference body group is Station wagon.

Table C 7 (cont.)

Variable	Parameter estimate	t-value
Mercedes Benz	0.2842	14.3
Mistubishi	0.0375	2.6
Nissan	0.0151	1.0
Opel	-0.0237	-1.9
Peugeot	-0.0091	-0.7
Pontiac	0.2011	2.8
Renault	-0.0614	-3.8
Rover	-0.0376	-1.6
Saab	0.0491	2.9
Seat	-0.0890	-5.6
Skoda	-0.1297	-7.8
Ssangyong	0.4815	6.7
Subaru	-0.0233	-1.4
Suzuki	-0.1775	-8.7
Toyota	-0.0269	-1.9
Volkswagen	0.0316	2.6
Combi	-0.1859	-32.2
Sedan	-0.0642	-11.7
# Observations		
R^2		

[1)] Reference year: 2002. Reference make is Volvo and reference body group is Station wagon.