A quality adjusted price index for discrete goods

by

John K. Dagsvik, Astrid L. Mathiassen

and

Bengt J. Eriksson

Research Department, Statistics Norway

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 prices 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 index

JEL classification: C25, C43

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

Address: John K. Dagsvik, Statistics Norway, Research Department, P.O. Box 8131, Dep. N-0033 Oslo, Norway. E-mail: john.dagsvik@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. 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. 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.

In this paper, we extend the traditional price index theory by allowing for indivisible goods (discrete goods) characterized by product-specific unobservable attributes which represent “quality”, suitably defined. 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 of course 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). Recently, there have been important advanced in the analysis of supply and demand of heterogeneous products with latent product attributes when consumers have heterogeneous preferences, based on an alternative approach, namely the theory of discrete choice. See for example Bajari and Benkard (2005), Berry, Levinsohn and Pakes (2004), Nevo (2003), Petrin, (2002). However, these contributions are not primarily focused on developing practical price indexes consistent with their representation of preferences they estimate.

In this paper we extend the approach proposed by Trajtenberg (1990) for calculating a true quality adjusted price index. His approach is based on a very simple demand model, namely the multinomial logit discrete choice demand model with no observed consumer specific variables. Our contribution in this paper, is to extend Trajtenberg’s approach to the construction of quality adjusted price indexes by (i) allowing the latent quality attributes to vary over time, (ii) accounting for the possibility that prices may be endogenous. However, and similarly to Trajtenberg, we also apply a simple multinomial (aggregate) demand model which enables us to obtain a very useful characterization of the quality adjusted price index. Specifically, we demonstrate 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). 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 a quality adjusted price index 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 extension of the approach proposed by Trajtenberg (1990). In section 3, we discuss an empirical application in a discrete choice setting, namely the market for new automobiles in Norway. To compute the quality

---

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
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. To the best of our knowledge, our analysis represents the first application in which these type of models have been estimated by a maximum likelihood procedure.

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, 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 \( g = 1, 2, \ldots, S \). 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 groups 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 \( U_{ij}(g) \) of variant \( j \) in group \( g \) at period \( t \), which is assumed to have the form

\[
U_{ij}(g) = \frac{(m_i - w_{ij}(g))\theta}{p_t} + v_{ij}(g) + \epsilon_{ij}(g),
\]

where \( m_i \) represents income, \( w_{ij}(g) \) is the price of variant \( j \) in group \( g \), \( v_{ij}(g) \) is a function of attributes of variant \( j \) in group \( g \), \( p_t \) is a price index for the divisible (outside) goods, \( \theta \) is a positive constant and \( \epsilon_{ij}(g) \) are random variables that represent unobserved heterogeneity in tastes. The alternative of not buying is indexed by \( j = g = 0 \), where \( U_{i0}(0) = \theta m_i / p_t \). The term \( v_{ij}(g) \) can be interpreted as a quality indicator that is supposed to capture such effects as the fluctuations in average population

---

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.
popularity of variant $j$ in group $g$. In this paper, we assume that the random error terms are independent, independent of the other terms of the utility function and with extreme value c.d.f.

\begin{equation}
  P(e_{ij}(g) \leq x) = \exp(-e^{-x}).
\end{equation}

Let $\tilde{w}_j(g) = w_j(g)/p_i$, and let $Q_{ij}(g)$ the probability that an agent chooses variant $j$ within group $g$ in period $t$ and $Q_{i0}$ the corresponding probability that no variant of the differentiated product is chosen by the agent in period $t$. It is well known (cf. McFadden, 1984) that the assumptions above imply that

\begin{equation}
  Q_{ij}(g) = P\left(U_{ij}(g) = \max_{k \in B_i(g)} U_{ik}(r)\right) = \frac{\exp(v_j(g) - \tilde{w}_j(g)\theta)}{\sum_{r_k \in B_i(r)} \exp(v_{ik}(r) - \tilde{w}_{ik}(r)\theta) + 1}
\end{equation}

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

\begin{equation}
  Q_{i0} = P\left(U_{i0}(0) = \max_{k \in B_i(0)} U_{ik}(r)\right) = \frac{1}{\sum_{r_k \in B_i(0)} \exp(v_{ik}(r) - \tilde{w}_{ik}(r)\theta) + 1}.
\end{equation}

As is also 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 quality adjusted price index

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). Assume that the utility structure is given by (2.1) and let $V_{i\theta}$ denote the indirect utility conditional on the choice set, attributes, prices and income, defined by

\begin{equation}
  V_{i\theta} = \max_{g \in B_i(g)} \max_{k \in B_i(g)} U_{ik}(g), U_{i0}(0)
\end{equation}

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

\begin{equation}
  V_{i\theta} = EV_{i\theta} = E \max_{g \in B_i(g)} \max_{k \in B_i(g)} U_{ik}(g), U_{i0}(0) = \frac{m_i\theta}{p_i} + \log \left[ 1 + \sum_{g=1}^{s} \left[ \sum_{k \in B_i(g)} \exp(v_{ik}(g) - \tilde{w}_{ik}(g)\theta) \right] \right].
\end{equation}
The interpretation of (2.6) is that it expresses mean indirect utility given the choice set, observed and unobserved attributes, prices and income. Let \( v_t = (v_{1t}(1), v_{2t}(1), ..., v_{1t}(2), ...) \) and \( w_t = (w_{1t}(1), w_{2t}(1), ..., w_{1t}(2), w_{2t}(2), ...) \). From (2.6), it follows that the corresponding aggregate (mean) expenditure function, \( \bar{\tau}(v_t, w_t, p_t, B_t, u) \), is given by

\[
(2.7) \quad \bar{\tau}(v_t, w_t, p_t, B_t, u) = p_t \bar{\theta}^{-1} u - p_t \bar{\theta}^{-1} \log \left( 1 + \sum_{g=1}^{G} \left[ \sum_{k \in B_t(g)} \exp(v_{uk}(g) - \bar{w}_{uk}(g) \bar{\theta}) \right] \right),
\]

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

\[
(2.8) \quad \bar{\tau}(v_t, w_t, p_t, B_t, u) - \bar{\tau}(v_{t-1}, w_{t-1}, p_{t-1}, B_{t-1}, u) = \bar{\tau}(v_{t-1}, \delta_t w_{t-1}, p_t, B_{t-1}, u) - \bar{\tau}(v_{t-1}, w_{t-1}, p_{t-1}, B_{t-1}, u),
\]

which is equivalent to

\[
(2.9) \quad \bar{\tau}(v_t, \tilde{w}_t, 1, B_t, u) = \bar{\tau}(v_{t-1}, \delta_t \tilde{w}_{t-1}, p_t / p_{t-1}, 1, B_{t-1}, u).
\]

The interpretation of \( \delta_t \) 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 \((v_{t-1}, w_{t-1}, B_{t-1})\) to \((v_t, w_t, B_t)\). 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 mean welfare 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 \( \delta_t \). 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 \( \delta_t \). Note that \( \delta_t \) is a \emph{conditional} index that only captures the welfare change of the discrete good. The interpretation of \( \delta_t \) 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 \( \delta_t \) is determined by the equation

\[
(2.10) \quad \sum_{g=1}^{G} \left[ \sum_{k \in B_t(g)} \exp(v_{uk}(g) - \theta \tilde{w}_{uk}(g)) \right] = \sum_{g=1}^{G} \left[ \sum_{k \in B_{t-1}(g)} \exp(v_{uk}(g) - \theta \tilde{w}_{uk}(g) p_t / p_{t-1}) \right].
\]
The right-hand side of (2.10) is strictly decreasing in \( \delta \), and, consequently, equation (2.10) determines \( \delta \) uniquely. Once \( \{v_g(g)\} \) have been specified, and \( \theta \) and \( \{v_g(g)\} \) have been estimated, one can compute the price index by solving for \( \delta \) in the nonlinear equation (2.10).

A serious challenge one faces in the context of empirical application is how to specify the terms \( \{v_g(g)\} \). The approach of Trajtenberg and others is to specify \( v_p(g) \) as a function of observed variant-specific attributes and estimate the corresponding parameters. However, a serious drawback with this approach is that \( \{v_p(g)\} \) may vary over time in a way that is not captured by observable variant-or group specific attributes. To circumferent this problem we shall in the following section propose a particular "semiparametric" approach.

2.3. Further implications from the multinomial logit demand model

As mentioned above, the purpose of this section is to demonstrate how one can derive simple expressions for the quality adjusted price index, depending solely on \( \theta \), prices and observed demand.

Note first that it follows from (2.4) that (2.10) can be written as

\[
Q_{10}^{-1} - 1 = \sum_{g=1}^{G} \left[ \sum_{k \in \mathcal{R}_{g}(g)} \exp \left( v_{t,k} - \theta \delta \tilde{w}_{t,k}(g) \frac{p_k}{p_i} \right) \right].
\]

Furthermore, (2.4) implies that

\[
\exp(v_g(g)) = \frac{Q_{0}(g)}{Q_{10}} \exp(\theta \tilde{w}_g(g)).
\]

From (2.12) it follows that when (2.12) is inserted into (2.11) we obtain that

\[
\sum_{k \in \mathcal{R}_{g}(g)} \exp(v_{t,k} - \theta \delta \tilde{w}_{t,k}(g) \frac{p_k}{p_i}) = \sum_{k \in \mathcal{R}_{g}(g)} Q_{10}^{-1} Q_{t-1,k}(g) \exp((1 - \delta, \frac{p_k}{p_i}) \theta \tilde{w}_{t,k}(g)).
\]

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

\[
Q_{10}^{-1} - 1 = \sum_{k \in \mathcal{R}_{g}(g)} Q_{t-1,k}(g) \exp((1 - \delta, \frac{p_k}{p_i}) \theta \tilde{w}_{t,k}(g)).
\]

Thus, when \( \theta \) has been estimated, one can insert the respective observed fractions into (2.14) and compute the quality-adjusted price index by solving for \( \delta \) in (2.14) without relying on specifications of \( \{v_g(g)\} \), nor the corresponding parameter estimates.
One can often use a more simple approximate 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 right hand side of (2.14) that the approximate solution for the price index determined by (2.14) is given by

\[
\delta_t 
\approx \frac{p_t}{p_{t-1}} \left(1 + \frac{1 - Q_{t-1,0}/Q_{t0}}{\theta y_{t-1}} \right),
\]

where \(y_t\) is the mean deflated expenditure of the discrete good in period \(t\) given by

\[
y_t = \sum_{g} \sum_{k \in \mathcal{R}_t(g)} \tilde{w}_g(g) Q_{g}(g).
\]

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 fractions of consumers in period \(t-1\) and \(t\), 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. In other words, \(y_{t-1}, Q_{t-1,0}\) and \(Q_{t0}\) are sufficient statistics for the effects on \(\delta_t\) from changes in the demand induced by changes in prices and qualities. As mentioned above, our approach to price index measurement accounts for such a fact that more people may prefer not to buy a new variant in a given period \(t\) (say), than in period \(t-1\), which implies an increase in \(Q_{t0}\). This will happen if the unobserved quality attributes \(\{v_g(g)\}\) 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. This effect is not captured by the Laspeyres index either. If the latent quality attributes \(\{v_g(g)\}\) 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 so 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. Finally, recall that the computation of the index in (2.14/2.15) does not require that the analyst knows the structure of the quality indicator \(\{v_g(g)\}\). However, to identify and estimate the parameter \(\theta\) one needs to make some separability assumptions about the structure of \(\{v_g(g)\}\), to be discussed in section 3.2.
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, based on the index developed above.

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

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

Similarly to (2.6), it follows that

$$V_t(g) = \frac{\theta m_t}{p_t} + \log \left( \sum_{k \in B_t(g)} \exp \left( v_{ik} (g) - \theta \tilde{w}_{ik} (g) \right) \right).$$

The corresponding conditional mean expenditure function is given by

$$\bar{e}(v_{-i}(g), w_{-i}(g), p_{-i}, B_{-i}(g), u) = p_{i} u^{\theta - 1} - p_{i} \theta^{-1} \log \left( \sum_{k \in B_{-i}(g)} \exp \left( v_{ik} (g) - \tilde{w}_{ik} (g) \theta \right) \right).$$

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

$$\bar{e}(v_{-i}(g), w_{-i}(g), 1, B_{-i}(g), u) = \bar{e} \left( v_{-i}(g), \delta_i (g) \tilde{w}_{-i} (g), p_{-i}/p_{i}, 1, B_{-i}(g), u \right).$$

If we combine (2.19) and (2.20) it follows that the price index for group $g$, $\delta_i (g)$, is determined by the equation

$$\sum_{k \in B_{-i}(g)} \exp \left( v_{ik} (g) - \theta \tilde{w}_{ik} (g) \right) = \sum_{k \in B_{-i}(g)} \exp \left( v_{i-k} (g) - \theta \tilde{w}_{i-k} (g) \delta_i (g) p_{i-k}/p_{i} \right).$$

Similarly to (2.14) it follows from (2.12) that (2.21) can be expressed as

$$Q_{i,0}^{-1} Q_i (g) = Q_{i-1,0}^{-1} \sum_{k \in B_{i-1}(g)} Q_{i-1,k} (g) \exp \left( (1 - \delta_i (g) p_{i-1}/p_{i}) \theta \tilde{w}_{i-1,k} (g) \right).$$

In Appendix B, we prove that, to a first-order Taylor approximation, we have that
where $y_t(g)$ is the mean deflated expenditure within group $g$ in period $t$, defined by

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

Equation (2.23) 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 respective 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 of (2.22), we find that

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

where

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

Eq. (2.25) implies that

$$\delta_t(g) \approx \frac{p_t}{p_{t-1}} \left[1 + \frac{(Q_{t-1}(g) - Q_t(g))Q_{t-1,0} Q^{-1}_{t0}}{\theta \bar{w}_{t-1}(g)}\right].$$

Similarly to (2.15), the index formula in (2.27) 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, as defined in this paper, 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 sometimes been used 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 \( v_j(g) \) 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 we use are the three body types “Combi”, “Sedan” and “Station wagon”.

We assume now that

\[
(3.1) \quad v_j(g) = z_j \beta + \xi_j(g) + \mu_j(g) + \eta_j(g),
\]

for \( j \in B_i(g) \), where the \( z_j \)-variables we use are fuel consumptions (liters per km) and engine performance, and \( \beta \) is a vector of unknown parameters. The term \( \mu_j(g) \) 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 $\mu_i(g)$ within a given body group $g$. Note that $\xi_j(g)$ is assumed not to depend on time. This restriction is crucial for achieving identification. The terms $\{\eta_j(g)\}$ are zero mean random disturbances.

We shall now discuss our particular maximum likelihood 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

\[(3.2) \quad \log \left( \frac{Q_j(g)}{Q_1(g)} \right) = \xi_j'(g) + (z_j(g) - z_1(g)) \beta - \theta(\hat{w}_j(g) - \hat{w}_1(g)) + \eta_j(g) - \eta_1(g) \]

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

In the following 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 $c_j(g)$ 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

\[(3.3) \quad \pi_{tg}(w_t) = \left( w_t(g) - c_j(g) \right) Q_j(g) M_t - K_t(g), \]

where $M_t$ is the total number of consumers in year $t$ and $K_t(g)$ represents fixed costs. In the following, we assume that the quality indicators $\{v_j(g)\}$ are exogenously given to the firms, and that firm $j$ of type $g$ maximizes (3.3) with respect to its own price $w_j(g)$, taking the prices of other firms as given. The first-order conditions that correspond to this maximization problem are given by

\[(3.4) \quad \hat{w}_j(g) = \tilde{c}_j(g) + \frac{1}{\theta \left(1 - Q_j(g) \right)}, \]

for $j = 1,2,\ldots$, where $\tilde{c}_j(g) = c_j(g)/p_t$. Recall that $\{Q_j(g)\}$ 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 $\hat{w}_j(g)\theta\left(1 - Q_j(g) \right) > 1$. As regards the empirical specification of the price equation, we assume that

\[(3.5) \quad \tilde{c}_j(g) = b_j(g) + d_j(g) + \kappa_j(g), \]
with the normalization \(\sum_j b_j(g) = 0\), where \(b_j(g)\) and \(d_i(g)\) are unknown parameters and \(\{\kappa_j(g)\}\) 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 \(\{\eta_j(g)\}\) are independent and normally distributed with zero mean and variance \(s^2(g)\), depending on \(g\), and \(\{\kappa_j(g)\}\) are independent and normally distributed with zero mean and variance \(r^2(g)\), depending on \(g\).

Moreover, \(\kappa_j(g)\) and \(\eta_{ik}(g)\) are independent for all \(j, k, t, \) and \(\tau\). In addition, the error terms in different body groups are assumed to be independent. Note that in (3.2), \(\{\eta_j(g)\}\) are endogenous because they depend on \(\{\eta_j(g)\}\) through (3.4). Consequently, we cannot estimate \(\theta\) 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

\[
\log L = -\sum_g \sum_i \sum_{j \in M(g)} \left[ \log \left( \frac{Q_j(g)}{Q_i(g)} \right) - \xi_j^*(g) - \left( z_j(g) - z_i(g) \right) \beta + \theta \left( \hat{w}_j(g) - \bar{w}_i(g) \right) \right] + \frac{1}{2s^2(g)} + \log s(g)
\]

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, similarly to Vitorino, (2004). In the actual estimation procedure, \(\{Q_j(g)\}\) are replaced by their corresponding observed frequencies \(\{\hat{Q}_j(g)\}\). However, the errors \(\{\hat{Q}_j(g) - Q_j(g)\}\) 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 \(\{b_j(g)\}\). 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 \(\{b_j(g)\}\), 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 $\beta$, $\theta$, $\{r(g)\}$ and $\{r(g)\}$. More precise details of this procedure are given in Appendix B. The estimates of $\theta$ 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.

We also carried out an estimation based solely on the demand relation in (3.2) and found that the parameters $\beta_1$, $\beta_2$, $r(1)$, $r(2)$, $r(3)$ and $\theta$ are practically equal to the estimates reported in Table 1.

### Table 1. Estimates of structural parameters

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter</th>
<th>Estimate</th>
<th>$t$-statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fuel consumption</td>
<td>$\beta_1$</td>
<td>$0.2019 \times 10^{-2}$</td>
<td>1.4</td>
</tr>
<tr>
<td>Engine performance (kW)</td>
<td>$\beta_2$</td>
<td>$0.2018 \times 10^{-2}$</td>
<td>1.4</td>
</tr>
<tr>
<td>Price $\times 10^{-5}$</td>
<td>$\theta$</td>
<td>1.4985</td>
<td>16.9</td>
</tr>
<tr>
<td>Standard errors of tastes:</td>
<td>$s(1)$</td>
<td>0.952</td>
<td>54.9</td>
</tr>
<tr>
<td></td>
<td>$s(2)$</td>
<td>1.270</td>
<td>60.6</td>
</tr>
<tr>
<td></td>
<td>$s(3)$</td>
<td>1.009</td>
<td>61.4</td>
</tr>
<tr>
<td>Standard errors of marginal costs:</td>
<td>$r(1)$</td>
<td>0.112</td>
<td>55.5</td>
</tr>
<tr>
<td></td>
<td>$r(2)$</td>
<td>0.284</td>
<td>60.7</td>
</tr>
<tr>
<td></td>
<td>$r(3)$</td>
<td>0.218</td>
<td>62.1</td>
</tr>
</tbody>
</table>

This means that OLS estimation based on (3.3) can be applied in this case. Therefore, we conclude that the prices seem to be set by the firms in such a way that they are only weakly correlated with the disturbances, $\{\eta(g)\}$. However, recall that 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 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 $N_t(g)$ denote the number of variants of type $j$ within body type $g$ sold in year $t$. From (2.14) it follows that
The fraction $\frac{p_t}{p_{t-1}}$ 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, $\delta_t^L$, is calculated as

$$
\delta_t^L = \frac{\sum \sum_{j \in G_t(g)} w_j(g)N_{t-1,j}(g)}{\sum \sum_{j \in G_t(g)} w_{t-1,j}(g)N_{t-1,j}(g)}.
$$

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>The Laspeyres index for new automobiles</td>
<td>100</td>
<td>102.5</td>
<td>98.2</td>
<td>100.8</td>
<td>102.2</td>
<td>101.9</td>
<td>103.1</td>
<td>106.5</td>
<td>108.2</td>
</tr>
<tr>
<td>The Laspeyres index for other goods ($p_t$)</td>
<td>100</td>
<td>103.6</td>
<td>104.4</td>
<td>107.1</td>
<td>109.7</td>
<td>112.6</td>
<td>116.4</td>
<td>119.9</td>
<td>121.4</td>
</tr>
<tr>
<td>Hedonic regression</td>
<td>100</td>
<td>102.1</td>
<td>96.3</td>
<td>98.6</td>
<td>98.7</td>
<td>97.8</td>
<td>97.8</td>
<td>104.3</td>
<td>106.0</td>
</tr>
<tr>
<td>Quality adjusted index, $\theta = 1.5$ (estimated value)</td>
<td>100</td>
<td>101.4</td>
<td>94.1</td>
<td>97.7</td>
<td>97.9</td>
<td>106.0</td>
<td>111.2</td>
<td>115.2</td>
<td>116.2</td>
</tr>
<tr>
<td>First-order approximation of the quality adjusted index, (eq. 2.15), $\theta = 1.5$</td>
<td>100</td>
<td>101.5</td>
<td>93.1</td>
<td>97.1</td>
<td>97.8</td>
<td>105.9</td>
<td>111.7</td>
<td>116.3</td>
<td>117.8</td>
</tr>
<tr>
<td>Quality adjusted index with other values of $\theta$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta = 2.1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>102.0</td>
<td>97.0</td>
<td>100.4</td>
<td>101.2</td>
<td>108.0</td>
<td>112.8</td>
<td>116.6</td>
<td>117.8</td>
</tr>
<tr>
<td>$\theta = 1.9$</td>
<td>100</td>
<td>101.9</td>
<td>96.2</td>
<td>99.7</td>
<td>100.3</td>
<td>107.5</td>
<td>112.4</td>
<td>116.3</td>
<td>117.4</td>
</tr>
<tr>
<td>$\theta = 1.7$</td>
<td>100</td>
<td>101.7</td>
<td>95.3</td>
<td>98.8</td>
<td>99.2</td>
<td>106.9</td>
<td>111.9</td>
<td>115.8</td>
<td>116.8</td>
</tr>
<tr>
<td>$\theta = 1.3$</td>
<td>100</td>
<td>101.1</td>
<td>92.5</td>
<td>96.2</td>
<td>96.1</td>
<td>105.0</td>
<td>110.4</td>
<td>114.3</td>
<td>115.3</td>
</tr>
<tr>
<td>$\theta = 1.1$</td>
<td>100</td>
<td>100.6</td>
<td>90.4</td>
<td>94.3</td>
<td>93.7</td>
<td>103.5</td>
<td>109.1</td>
<td>113.2</td>
<td>114.0</td>
</tr>
</tbody>
</table>
\[ \theta = 0.9 \]

<table>
<thead>
<tr>
<th>Fraction of persons 16–66 years of age that buy a new car</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.030 0.032 0.041 0.039 0.041 0.034 0.032 0.031 0.031</td>
</tr>
</tbody>
</table>

\[ C_i(g) = B_i(g) \cap B_{i,t}(g) \]. 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 116.2 percent in 2002. 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 \( \theta \). From the results, we can conclude that the index changes little when \( \theta \) varies between 1.3 to 1.7. Even when \( \theta \) varies between 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 a quality adjusted price index in markets with differentiated products. We have discussed how one can use the theory of discrete choice to derive an exact price index 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 prices and 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 of the variants purchased, constitute sufficient statistics for the calculation of the quality adjusted price index, given the prices of the outside divisible goods \( p_t \) and the price parameter \( \theta \) 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 particular, 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. They assume that the latent quality attributes are constant over time. Finally it should be noted that 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


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

\[ U\left(x_0, \sum_k x_k \mu_k\right) \]

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 \( \mu_k > 0 \) 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)

\[ \frac{w_{jt}}{\mu_j} = \frac{w_{jt}}{\mu_{jt}}. \]

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

(A.2)

\[ \delta_j = \frac{w_{jt}/\mu_{jt}}{w_{j,t+1}/\mu_{j,t+1}}. \]

Assume now that

(A.3)

\[ \log \left( \frac{\mu_j}{p_j} \right) = z_j \beta, \]

where \( z_j \) is a vector of suitable observed attributes of variant \( j \). Then, we can write (3.1) as

(A.4)

\[ \ln w_j = \alpha_t + z_j \beta + \log p_t, \]

where

\[ \alpha_t = \ln w_{jt} - \mu_{jt}. \]

Hence, the period specific intercept \( \alpha_t \) can be used to compute \( \delta_t \) because (A.2) implies that
(A.5)  
\[ \bar{\delta} = \frac{p_t \exp\left(\alpha_r - \alpha_{r-1}\right)}{p_{r-1}}. \]

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):

Note that it follows from (2.10) and (2.21) that

(B.1)  
\[ \sum_{g} \sum_{k \in B_{r-1}(g)} \exp(v_{r-1,k}(g) - \theta \tilde{w}_{r-1,k}(g) \delta_t) \frac{p_{r-1}}{p_t} \]

By using (2.12) it follows that (B.1) can be written as

(B.2)  
\[ \sum_{g} \sum_{k \in B_{r-1}(g)} Q_{r-1,k}(g) \exp\left(1 - \delta_t \frac{p_{r-1}}{p_t}\right) \theta \tilde{w}_{r-1,k}(g) \]

A first order Taylor expansion of both sides of (B.2) yields

(B.3)  
\[ \sum_{g} \sum_{k \in B_{r-1}(g)} Q_{r-1,k}(g) \left[1 + \left(1 - \delta_t \frac{p_{r-1}}{p_t}\right) \theta \tilde{w}_{r-1,k}(g)\right] \]

which implies that
This completes the proof. Q.E.D.

Maximum likelihood 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) \[
E \prod_{j \in B(g) \setminus \{1\}} \exp \left\{ - \left( \log \frac{Q_g(g)}{Q_{11}(g)} - \xi_j^*(g) - (z_y(g) - z_{j1}(g)) \beta + \theta(\tilde{w}_y(g) - \tilde{w}_{j1}(g)) + \eta_i(g) \right)^2 \frac{1}{2s^2(g)} - \log s(g) \right\}
\]

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

\[
R_y(g) = \log \left( \frac{Q_g(g)}{Q_{11}(g)} \right) - \xi_j^*(g) - (z_y(g) - z_{j1}(g)) \beta + \theta(\tilde{w}_y(g) - \tilde{w}_{j1}(g)).
\]

Then one can express (B.7) as

(B.8) \[
E \prod_{j \in B(g) \setminus \{1\}} \exp \left\{ - \left( R_y(g) + \eta_i(g) \right)^2 \frac{1}{2s^2(g)} - \log s(g) \right\}
= E \exp \left\{ - \sum_{j \in B(g) \setminus \{1\}} \left( R_y^2(g) + 2\eta_i(g)R_y(g) + \eta_i^2(g) \right) \frac{1}{2s^2(g)} - \log s(g) \right\}
\]

where

\[
R_y(g) = \sum_{j \in B(g) \setminus \{1\}} R_y(g).
\]

Now observe that (3.2) implies that $R_y(g) \approx 0$, so that (B.8) reduces to

\[
\exp \left\{ - \sum_{j \in B(g) \setminus \{1\}} \left( R_y^2(g)2s^2(g) - \log s(g) \right) \right\} \cdot E \exp \left(-\eta_i(g)\eta_i(g)2s^2(g)\right)
\]

where $\eta_i(g)$ is the number of variants in $B_1(g) \setminus \{1\}$.
Furthermore, we have that

\[
E \exp \left( -\frac{\eta_i^2(g) n_i(g)}{2s^2(g)} \right)
= \int \exp \left( \frac{-x^2 n_i(g)}{2s^2(g)} \right) \exp \left( -\frac{-x^2}{2s^2(g)} s(g) \sqrt{2\pi} \right) \frac{dx}{s(g) \sqrt{2\pi}} = \frac{1}{s(g) \sqrt{2\pi}} \int \exp \left( -\frac{\eta_i(g)}{2s^2(g)} x^2 \right) dx.
\]

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

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

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

\[
\exp \left( -\sum_{j=B_i(g) \cap \{t\}} (R_j^2(g)/2s^2(g) - \log s(g)) \right) \frac{1}{\sqrt{1 + n_i(g)}}.
\]

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

(B.9)

\[
\frac{1}{\sqrt{1 + n_i(g)}} \prod_{j=B_i(g) \cap \{t\}} \exp \left( -\log \left( \frac{Q_j(g)}{Q_i(g)} \right) - \frac{\xi_j^2(g) - (z_j(g) - z_i(g))\beta + \theta(\bar{w}_j(g) - \bar{w}_i(g))}{2s^2(g) - \log s(g)} \right)^2 \frac{1}{2s^2(g) - \log s(g)}.
\]

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

(B.10)

\[
\log L = \sum_g \sum_t \sum_{j=B_i(g) \cap \{t\}} \left[ \log \left( \frac{Q_j(g)}{Q_i(g)} \right) - \frac{\xi_j^2(g) - (z_j(g) - z_i(g))\beta + \theta(\bar{w}_j(g) - \bar{w}_i(g))}{2s^2(g) - \log s(g)} \right]^2 \frac{1}{2s^2(g) - \log s(g)} + \log |J| - \frac{1}{2} \sum_g \sum_t \log(1 + n_i(g))
\]

where \( J \) is the Jacobian associated with the transformation of variables, from

\[
\{\eta_0(g) - \eta_i(g), \mu_\tau(r), t, \tau = 1, 2, \ldots, T, j \in B_0(g), k \in B_1(g), g, r = 1, 2, 3\}
\]

to

\[
\{Q_0(g), w_\tau(r), t, \tau = 1, 2, \ldots, T, j \in B_0(g), \tau \in B_1(g), g, r = 1, 2, 3\},
\]

given by
\begin{equation}
\eta_{ij}(g) - \eta_{i}(g) = \log \left( \frac{Q_{ij}(g)}{Q_{i}(g)} \right) - \xi_{ij}(g) - (z_{ij}(g) - z_{i}(g)) + \theta(\bar{w}_{ij}(g) - \bar{w}_{i}(g))
\end{equation}

and

\begin{equation}
\kappa_{ij}(g) = \bar{w}_{ij}(g) - c_{ij}(g) - h_{ij}(g) - \frac{1}{\theta(1 - Q_{ij}(g))}.
\end{equation}

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 $N_{ij}(g)$ be the number of variants sold of type $j$ within body type $g$, $N(g)$ the number of variants within $B_{ij}(g)$, and $N_{i}$ the total number of variants sold. For notational convenience let us now introduce the notation

\begin{align*}
Y_{i}(g) &= \log \left( \frac{Q_{ij}(g)}{Q_{i}(g)} \right) = \log \left( \frac{N_{ij}(g)}{N_{i}(g)} \right), \quad Y_{i}(g) = \frac{1}{T} \sum_{t} Y_{i}(g), \quad \bar{w}_{i}(g) = \frac{1}{T} \sum_{t} \bar{w}_{i}(g) \\
V_{i}(g) &= \frac{1}{1 - (1 - Q_{io})N_{i}(g)}, \quad \bar{V}_{i}(g) = \frac{1}{T} \sum_{t} V_{i}(g), \quad \bar{V}_{i}(g) = \frac{1}{N_{i}(g)} \sum_{j \in B_{ij}(g)} V_{i}(g), \quad \bar{V}_{i}(g) = \frac{1}{T} \sum_{t} \bar{V}_{i}(g),
\end{align*}

where \( \hat{Q}_{io} \) denotes an estimate of $Q_{io}$ 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 $\theta$, $s_{i}(g)$, $r_{i}(g)$, $g = 1, 2, 3$, can be estimated by maximizing

\begin{align*}
\log L' &= -\sum_{t} \sum_{g} \sum_{j \in B_{ij}(g)} \left( \frac{Y_{i}(g) - \bar{Y}_{i}(g) - (z_{ij}(g) - z_{i}(g) - x_{ij}(g) + x_{i}(g)) + \theta(\bar{w}_{ij}(g) - \bar{w}_{i}(g))}{2s_{i}(g)} \right) ^{2} \\
&- \sum_{t} \sum_{g} (N_{i}(g) - 1) \log s(g) \\
&- \sum_{t} \sum_{g} \sum_{j \in B_{ij}(g)} \left( w_{ij}(g) - \bar{w}_{ij}(g) + \bar{w}_{i}(g) \right) ^{2} \left( \frac{1}{2r_{i}(g)} + \log r(g) \right).
\end{align*}

Q.E.D.
### Appendix C

**Summary statistics**

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

<table>
<thead>
<tr>
<th>Year</th>
<th>Combi Fractions</th>
<th>Sedan Fractions</th>
<th>Station wagon Fractions</th>
<th>Other carmakers Fractions</th>
<th>Total sales</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Levels</td>
<td>Levels</td>
<td>Levels</td>
<td>Levels</td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>0.41</td>
<td>32 226</td>
<td>0.34</td>
<td>26 406</td>
<td>0.25</td>
</tr>
<tr>
<td>1995</td>
<td>0.43</td>
<td>36 043</td>
<td>0.33</td>
<td>28 307</td>
<td>0.23</td>
</tr>
<tr>
<td>1996</td>
<td>0.46</td>
<td>49 794</td>
<td>0.27</td>
<td>29 810</td>
<td>0.25</td>
</tr>
<tr>
<td>1997</td>
<td>0.40</td>
<td>40 983</td>
<td>0.28</td>
<td>28 495</td>
<td>0.31</td>
</tr>
<tr>
<td>1998</td>
<td>0.39</td>
<td>41 608</td>
<td>0.22</td>
<td>23 568</td>
<td>0.38</td>
</tr>
<tr>
<td>1999</td>
<td>0.36</td>
<td>32 229</td>
<td>0.20</td>
<td>18 064</td>
<td>0.42</td>
</tr>
<tr>
<td>2000</td>
<td>0.37</td>
<td>31 336</td>
<td>0.18</td>
<td>15 005</td>
<td>0.44</td>
</tr>
<tr>
<td>2001</td>
<td>0.35</td>
<td>29 496</td>
<td>0.17</td>
<td>14 514</td>
<td>0.46</td>
</tr>
<tr>
<td>2002</td>
<td>0.35</td>
<td>29 312</td>
<td>0.13</td>
<td>10 927</td>
<td>0.51</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Year</th>
<th>Combi Variants entering</th>
<th>Variants disappearing</th>
<th>Sedan Variants entering</th>
<th>Variants disappearing</th>
<th>Station wagon Variants entering</th>
<th>Variants disappearing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>149</td>
<td>177</td>
<td>129</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td>153</td>
<td>174</td>
<td>124</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td>186</td>
<td>214</td>
<td>170</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>162</td>
<td>188</td>
<td>194</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>173</td>
<td>195</td>
<td>226</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td>158</td>
<td>205</td>
<td>244</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>182</td>
<td>221</td>
<td>271</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>184</td>
<td>237</td>
<td>277</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td>203</td>
<td>239</td>
<td>303</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table C 3. Mean deflated prices across cars within type of body. NOK

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