Econometric Research and Special Studies Department

CONSTRUCTING QUALITY ADJUSTED PRICE INDEXES: a comparison of hedonic and discrete choice models

Nicole Jonker

Research Memorandum WO&E no. 673
December 2001

De Nederlandsche Bank

CONSTRUCTING QUALITY ADJUSTED PRICE INDEXES: A COMPARISON OF HEDONIC AND DISCRETE CHOICE MODELS

Nicole Jonker

Research Memorandum WO&E no. 673/0033
December 2001

De Nederlandsche Bank NV
Econometric Research and
Special Studies Department
P.O. Box 98
1000 AB AMSTERDAM
The Netherlands

ABSTRACT

CONSTRUCTING QUALITY ADJUSTED PRICE INDEXES: a comparison of hedonic and discrete choice models

Nicole Jonker

The Boskin report (1996) concluded that the US consumer price index (CPI) overestimated the inflation by 1.1 percentage points. This was due to several measurement errors in the CPI. One of them is called quality change bias. In this paper two methods are compared which can be used to eliminate quality change bias, namely the hedonic method and a method based on the use of discrete choice models. The underlying micro-economic fundations of the two methods are compared as well as their empirical implementation. Although the discrete choice model has not often been used in order to calculate quality adjusted price indexes it seems to be quite promising to do so.

Keywords: logit models, cpi, consumenter behaviour, producer behaviour

JEL Codes: C25, C43, D11, D21

SAMENVATTING

CONSTRUCTING QUALITY ADJUSTED PRICE INDEXES: a comparison of hedonic and discrete choice models

Nicole Jonker

In het Boskin rapport (1996) werd geconcludeerd dat de Amerikaanse consumentenprijsindex (CPI) de werkelijke inflatie jaarlijks met 1,1 procentpunt overschat. Dit wordt veroorzaakt door een aantal meetfouten in de CPI. Een van betreft de vertekening van de CPI als gevolg van de kwaliteitsverandering. In dit artikel worden twee methoden met elkaar vergeleken die hiervoor kunnen corrigeren, namelijk de hedonische en een methode die gebaseerd is op discrete keuze modellen. Vergeleken worden de micro-economische funderingen van de twee methoden alswel aan hun empirische implementatie. Howel het discrete keuzemodel thans vrijwel niet gebruikt wordt voor het berekenen van kwaliteitsgecorrigeerde prijsindices, blijkt deze wel veelbelovend te zijn.

Trefwoorden: logit modellen, cpi, consumentengedrag, producentengedrag

JEL Codes: C25, C43, D11, D21

1 INTRODUCTION

The Boskin report (1996) discussed several types of measurement errors in the consumer price index (CPI). The CPI measures the cost of purchasing a fixed market basket of goods and services. All the goods and services in this basket can be divided over classification levels beginning with major groups like food and beverages, housing, transportation, etc. The CPI calculates the monthly changes (monthly inflation) in the total cost of this basket by aggregating the price indexes of its sub levels. Changes in the CPI are caused by changes in the prices of the products in this particular basket. The products and services in the basket and their expenditure share in the CPI are based on household surveys held in a certain base year t_0. The CPI is not a true cost-of-living index (COLI). A true COLI compares the minimum expenditure required to achieve a same level of welfare (or utility) across two points in time. The Boskin committee concluded that the CPI overestimated the true inflation in the US by 1.1 percentage point. The conclusions of the Boskin committee renewed the interest in the effect of measurement errors on the reliability of the CPI.

Central banks are highly interested in these issues because price stability, defined in terms of the CPI or the HICP, is their primary goal (e.g. ECB) or one of their primary goals (e.g. FED). At a national level an accurate measurement of inflation is very important because the CPI has a great impact on several topics like indexation in legal contracts, wages and benefits from governments and deflation of national accounts, wages and retail sales. With respect to the government budget, an upward bias in the CPI will result in a real increase of indexed government/social benefits. For the US for instance, correctly measuring the CPI is likely to result in lower future budget deficits and lower national debt. In the Boskin report it was shown how serious the consequences of overestimating the CPI may be. For the US an overestimation of the cost of living index by one percent point would after a dozen of years result in a USD 1 trillion higher national debt.

There are several measurement errors causing the overestimation of inflation in the CPI. The Boskin report mentions the product substitution bias, the outlet substitution bias, the new product bias and the quality change bias. This paper deals with quality change bias in price indexes.

Statistical agencies use different techniques to adjust price indexes for changes in product quality. Changes in products can be characterised as being *marginal or* as *non – marginal*. Marginal changes refer to small changes in product characteristics which have a continuous character whereas non-marginal changes in product characteristics refer to large changes in continuous product characteristics but can also refer to the inclusion of discrete product characteristics. In case of computers characteristics like the speed of the computer or the capacity of the hard disk are examples of continuous product characteristics and characteristics like having a DVD player or a CD-writer are examples of discrete product characteristics. Non marginal changes occur in products which have a standard variant product which can be extended by extra features, like cars or computers but also by houses. Products which are largely composed of a combination of continuous characteristics (strength, thickness, size, durability, efficiency) are e.g. food products and electronic household articles with little possibilities to have extra features like fridges, washing machines, etc.

One possibility to adjust price indexes for changes in product quality is to use the ’matched model’ approach. With this method the price index is constructed only using the prices of products which are available in two adjacent periods. In this index the sample is confined to the products which do not change from one period to the next. However, this technique is not suitable for constructing price indices of product types involving rapid technical progress, like cars or computers, unless one uses chained price indexes. Another possibility is to use the overlapping method. This method is based on observing two different models of a particular good in a time span and use the ratio of the prices as a measure of quality adjustment. Yet another possibility focuses on price changes for a basic product specification. It adjusts these price changes for changes in the configuration of the basic model. For example if a DVD player becomes part of the standard specification of a certain PC model while it was not in the previous period the current price of this PC can be adjusted for the change in configuration by subtracting the price associated with buying the DVD player in the previous period (the associated price of the DVD player can be retrieved from the previous period accessories’ price list of the basic PC or can be retrieved by asking the manufacturer about the additional manufacturer’s costs associated with installing the DVD player standard) from the current PC price.

Nowadays, statistical agencies start using hedonic price indexes for products which undergo rapid technological changes. Hedonic methods refer to regression models in which product prices are related to product characteristics. They can be used to construct a quality adjusted price index of a good or a service. Berndt (1991) provides a very interesting ’historical’ overview on hedonic price equations. Waugh (1928) was the first to incorporate quality measures when explaining prices of vegetables. Court (1939) was the first to estimate a simple hedonic price equation for cars sold in 1925-1927. He included product characteristics as regressors in order to correct for product changes and he included year dummies reflecting price changes due to the passage of time. This type of regression is the basic hedonic price equation. This estimation technique became more popular in the seventies and the eighties through the work of

Griliches (1961) on car prices and Chow (1967) on computer prices. These early studies showed that quality adjusted price indexes of cars and computers decreased over time. For the Netherlands, Cramer (1966) estimated the first quality corrected price indexes of new passenger cars sold between 1950-1966. In his study, Cramer showed among others that the quality corrected price index for new cars was about 30% lower than the uncorrected price index. More recent studies on quality adjusted price indexes for cars are e.g. Raff and Trajtenberg (1995) and Blow and Crawford (1998).

Mid seventies also another branch of econometrics, started to develop, namely that of discrete choice modelling. McFadden (1974) showed that the conditional logit model (see e.g. Maddala, 1983) could be derived from random utility theory. In a conditional logit model the effect of choice characteristics on choice probabilities is estimated through the estimation of the so-called vector β. The elements of β characterise the utility function and are not directly tied to the marginal effects of changes in variables on choice probabilities. However, the monetary valuation of product characteristics can be based on this vector (see e.g. Chattopadhyay, 2000, Cropper et. al., 1993). This can be done for both continuous variables and discrete variables. In that sense the conditional logit model analyses a problem similar to the hedonic price index problem, although it has never been used as such.

Discrete choice models have been used, albeit not often. They have primarily been used to value housing characteristics. Mason and Quigley (1990) have compared the, as they call it, benefit estimates from the conditional logit and hedonic models using Monte Carlo simulation. They find that the hedonic method yields estimates of *marginal* changes in characteristics which are as good as those based on the conditional logit model. This has also been found in a simulation study of Cropper et al. (1993). However, they also find that the conditional logit model yields superior estimates of *non-marginal* changes compared to the estimates obtained using hedonic methods. These results suggest that there is a large group of articles which may benefit from the use of discrete choice models when deriving quality adjusted prices. Chattopadhyay (2000) models residential choice as a nested hierarchical choice process which may be more in line with the real choice behaviour of buyers. He uses the nested logit model for estimating preferences for housing attributes. He compares the nested logit estimates of the benefits of amenity changes with the estimates derived from the standard hedonic model. He finds that the nested logit benefit estimates are consistently lower.

Recently, some papers have appeared which paid attention to particular assumptions underlying the discrete choice models. The standard logit model has some disadvantages. The own and cross-price elasticities implied by this model are cumbersome (see e.g. Berry, Levinsohn and Pakes, 1995). Another problem is related to evaluating the welfare effects caused by economic changes (Petrin, 2001). Using a random coefficients discrete choice model overcomes these problems. One paper in which such a model is used is Nevo (2001). Nevo uses discrete choice models to produce a price index for ready-to- eat cereal that takes quality changes and the introduction of new products into account. His estimated price indexes depend heavily on the assumptions made and range between a 35% price increase over five years and a 2.4% price decrease. Another related paper is Bajari and Benkard (2001).

This paper proceeds as follows: section 2 gives a description of the theoretical models underlying hedonic methods and discrete choice models. Section 3 elaborates on the empirical implementation of these two approaches and how to derive a quality adjusted price index. Section 4 summarizes and concludes.

2 THEORETICAL MODELS

In this section the hedonic method and the discrete choice method on explaining consumer choices will be briefly described. Mason and Quigley (1990) describe the differences between the two models in more detail. The models have the same aim: estimating consumers' utility functions and retrieving the consumers' monetary valuation for particular goods or particular attributes of a specified good. However, the two models differ in the underlying assumptions, in the empirical implementation of the models and in the data requirements.

Let's concentrate on the composite good case. In both approaches there is a set of consumers who have preferences over the n measurable characteristics of a composite good x and over m other goods $z_1, \ldots, z_m$. These preferences can be represented by a utility function U:

$$u = U(x_1, x_2, .., x_n, z_1, z_2, .., z_m) \tag{1}$$

It is assumed that the utility function U is concave with respect to the product characteristics of good x and of the other goods $z_1, z_2, .., z_m$. If relative prices of the other goods remain constant over time one can apply Hicks' aggregation theorem yielding a utility function representing preferences defined over quantities of characteristics $(x_1, x_2, ..., x_n)$ and a composite commodity, the quantity of which is denoted by z, i.e.

$$u = U(x_1, x_2, .., x_n, z) \tag{2}$$

There exist J variants of good x and the j^{th} variant is denoted by x_j. This product variant x_j can be described by a vector of n measurable characteristics, $x_j = (x_{1j}, x_{2j}, .., x_{nj})$. If consumer i chooses product variant x_j with price p_j, if y_i is this consumer's income and if we assume a constant unity price of one for z than the utility he attaches to consuming x_j becomes

$$u = U(x_{1j}, x_{2j,..,} x_{nj,}\ y_i - p_j) \quad (3)$$

So for both methods the basic theoretical model is one of consumers maximising their utility over the composite good x and the other goods subject to their budget constraint $y_i = p_j + z$.

2.1 **Hedonic method**

The hedonic price method is well described in e.g. Berndt (1991) and Triplett (2000). In hedonic price equations the observed price of a product is considered to be a function of its characteristics. Hedonic methods are based on the idea that a product is a bundle of characteristics and that consumers actually buy bundles of product characteristics instead of products themselves. The implicit value of these characteristics for the consumers can be estimated by means of hedonic price equations.

The theory behind the model is described by Rosen (1974). He analysed hedonic prices using a spatial equilibrium framework. He assumes that producers of a certain good operate in a competitive environment. Therefore single producers take product prices as given and can not influence them. The class of goods can be characterised by n measured characteristics and any location in the plane represents a vector $x=(x_{1,}\ x_{2,}\ldots,x_n)$ with x_k equal to the level of the k^{th} product characteristic. A price $p(x)=p(x_{1,}\ x_{2,}\ldots,x_n)$ is defined at each point in the plane. It is assumed that a large amount of product varieties exist to choose from. Both consumers and producers base their decisions with regard to consumption respectively on production of packages of characteristics on maximising behaviour. Consumers maximise utility and producers maximise profits. The observed prices p(x) are the market clearing prices matching consumers' and producers' choices perfectly and leading to a market equilibrium.

It is assumed that consumers maximise their utility subject to the non-linear budget constraint. This requires that consumers choose z and $(x_{1,}\ x_{2,..,}x_n)$ to satisfy their budget constraint and to meet the first order conditions. If the price function is continuous and differentiable then the following holds for each consumer:

$$\frac{\partial p}{\partial x_k} = \frac{\partial U}{\partial x_k} \Big/ \frac{\partial U}{\partial z} = \frac{U_{x_k}}{U_z},\ \text{for k=1},\ldots,n \quad (4)$$

Consumers buy the product variant which offers the desired combination of product characteristics.

For simplicity it is assumed that producers have separate plants each producing one possible configuration. The vector M denotes the number of units produced of all the firm's configurations. Within a firm there are no spill-over effects from plant to plant. The j^{th} element of M denotes the number of units produced by a plant offering configuration j. The total costs of a firm are given in the cost function C(M; γ) where the vector γ reflects the underlying variables in the cost minimisation problem like factor prices and production function parameters. C is assumed to be convex in M. Each plant maximises profit $\pi = M(j)p(j) - C(M(j), x_{1j}, x_{2j}, .., x_{nj})$ by choosing $M(j)$ and x optimally. The revenue of one product variant x_j is given by the implicit price function of product characteristics p(x). Optimality of the plant 's choice requires that the marginal revenue from additional attributes equals their marginal cost of production per unit sold. Furthermore, optimality requires that the number of quantities are such that the unit revenue p(x) equals marginal production cost evaluated at the optimum bundle of characteristics.:

$$\begin{aligned} \frac{\partial p}{\partial x_k} &= \frac{\partial C}{\partial x_k} / M, \text{ for k=1}, ..., n \\ p(x) &= \frac{\partial C}{\partial M} \end{aligned} \quad (5)$$

In the hedonic method it is assumed that the consumers have preferences for any conceivable configuration of a composite good. However, in practice the consumers are more limited in their purchases of the good's configuration, since not every conceivable configuration is also available.

2.2 Conditional logit model

The discrete choice model which is used is known as McFadden's conditional logit model (McFadden, 1974). In short, the idea of his model is as follows. Suppose that an individual wants to buy a particular good X in period t and can choose among J different variants X_j. To each variant j individual i attaches a level of indirect utility U_{ijt}. The variant which he likes most, i.e. the car type he thinks will give him the highest level of indirect utility is bought by this individual. So it is assumed that consumers are rational decision makers and actually choose the type which optimizes their perceived utility subject to budgetary constraints. The utility individual i attaches to variant j in period t U_{ijt} can be decomposed into a part originating from how individual i perceives characteristics of variant j x_{ijt}, the utility he gets from consuming y_i -p_j other goods and a residual ε_{ijt}. This residual captures errors made in this maximisation process which are due to imperfect perceptions about the product's utility as well as the inability of the researcher to measure all the relevant variables. It follows from the random utility model, which McFadden uses in his article, that the residuals are independently and identically distributed with the Extreme Value (EV) distribution. The model is easy to estimate but has as a drawback that it assumes that the odds of choosing between any pair of alternatives is independent of the other possible choices. This property is also known as the IIA (Independence of Irrelevant Alternatives) property and is quite restrictive.

$$U_{ijt} = V_{ijt} + \varepsilon_{ijt} \tag{6}$$

$$F(\varepsilon_{ijt}) = \exp(-e^{-\varepsilon_{ijt}}) \tag{7}$$

Assume that

$$V_{ijt}(x_{ijt},\, y_i - p_j\,) = e^{\beta_t' x_{ijt} + \delta(y_i - p_j)} \tag{8}$$

where β_t and δ are unknown parameters which have to be estimated. The elements of the vector β_t reflect the relative valuation of attributes in period t. Under the assumption of independently and identically distributed residuals ε_{ijt}, which is questionable in this case, having the EV distribution the probability P_{ijt} that individual i chooses type j at period t equals

$$P_{ijt} = P(U_{ijt} \geq U_{ij't}) = \frac{e^{\beta_t' x_{ijt}}}{\sum_{j=1}^{J} e^{\beta_t' x_{ij}}},\ j' \neq j \tag{9}$$

Another specification of utility V_{ijt} is given in equation 10. This specification comes from the random coefficients logit model. It is an extension of the one given in equation 8. The use of the random coefficients logit model instead of the standard logit model is, as already mentioned in the introductory section of this paper, recommended by Berry, Levinsohn and Pakes (1995) and by Petrin (2001).

$$V_{ijt}(x_{ijt},\, y_i - p_j\,) = e^{\bar{\beta}_t' x_{ijt} + \delta(y_i - p_j) + \xi_j + \sum_{k=1}^{n} \sigma_k v_{ik} x_{jk}} \tag{10}$$

In this specification ξ_j denotes the (mean) utility of the unobserved (for the econometrician, not for the consumers) characteristics of good j. Also new is $\sum_k \sigma_k v_{ik} x_{jk}$. Each consumer i has idiosyncratic tastes for the n observed characteristics $v_i = (v_{i1}, v_{i2}, ..., v_{in})$. The parameter σ_k measures the heterogeneity in tastes for characteristic k in the population. It is assumed that v_i is a mean zero factor of random variables with a known distribution function. The interaction $\sigma_k v_{ik}$ yields consumer i' s personal taste for characteristic k. The elements of the vector β_t reflect the mean relative valuation of attributes in period t. After integrating out ε_{ijt} the probability that consumer i buys good j at time t becomes

$$P_{ijt} = P(U_{ijt} \geq U_{ij't}) = \frac{e^{\bar{\beta}_t' x_{ijt} + \delta(y_i - p_j) + \xi_j + \sum_{k=1}^{n} \sigma_k v_{ik} x_{jk}}}{\sum_{j=1}^{J} e^{\bar{\beta}_t' x_{ijt} + \delta(y_i - p_j) + \xi_j + \sum_{k=1}^{n} \sigma_k v_{ik} x_{jk}}},\ j' \neq j \tag{11}$$

Define I_{ijt} to be a dummy variable equal to one if individual i buys type j in period t. The general form of the loglikelihood function of period t, for both described specifications of V_{ijt}, is:

$$\log L_t = \sum_{i=1}^{n} \sum_{j=1}^{J} I_{ijt} \log(P_{ijt}) \tag{12}$$

3 EMPIRICAL IMPLEMENTATION

3.1 **The hedonic method**

There are various ways of estimating an hedonic price equation and consequently there are also a number of ways to construct price indexes. Three related methods are presented here. In the first method equation 1 is estimated. This equation shows the basic form of a hedonic price equation. The price of variant X_j at time t is assumed to depend on n product characteristics (both discrete and continuous) stored in the vector x_{jt}, a constant term c and the random disturbance term ε_{jt}. The function f describes the functional form of the price equation. Diewert (2001) describes some frequently used functional forms and discusses their pros and cons. Furthermore, he discusses the use of flexible functional forms. Commonly used specifications for f are the log-log specification, the log-linear specification and the linear-linear specification. Sometimes, economic theory offers an indication which functional form should be used. However, the choice of the functional form is usually an empirical matter. Using Box-Cox transformations can help when making this choice (see e.g. Berndt, 1991, p. 127-128).

$$p_{jt} = f(c, x_{jt}) + \varepsilon_{jt} \tag{13}$$

With the second method one assumes that the implicit values of product characteristics do not change over the estimation period $t_0 \ldots$ T. Then one can pool the data from different periods and estimate equation 12 using period dummies D_t. Here, the implicit values of the continuous and discrete product characteristics are stored in the vector β. The parameter α is an intercept term and α_t $(t \neq t_0)$ acts as an intercept shift in log prices for period t compared to period t_0, once controlled for product characteristics.

$$\ln(p_{jt}) = \alpha + \alpha_{t_0+1} D_{t_0+1} + \alpha_{t_0+2} D_{t_0+2} + .. + \beta x_{jt} + \varepsilon_{jt} \tag{14}$$

Analogously, the exponent of α_t is an intercept shift in prices for period t compared to period t_0, once controlled for product characteristics. Equation 15 defines the quality controlled price index I_t of prices at t relative to prices in the base period t_0

$$I_t = \exp(\alpha_t) \tag{15}$$

However, if one thinks that the assumption of constant implicit prices of product characteristics is not valid then one can estimate separate hedonic price equations for each period in the sample and construct a price index. The estimated intercept terms $\tilde{\alpha}_t$ are now also period specific

$$\ln(p_{jt}) = \tilde{\alpha}_t + \beta_t x_{jt} + \varepsilon_{jt} \tag{16}$$

There are different product price indexes which can be used. Five common price indexes are the Laspeyres price index (LPI), the Laspeyres chain price index (LCPI), the Paasche price index (PPI), the Paasche chain price index (PCPI) and the Fisher ideal price index (FP). Their specifications are given below. With the LPI an index is calculated which indicates how much the product under investigation with the average base period characteristics would cost in period t relatively to what it cost in period t_0. The PPI does something similar, but uses the average period t characteristics instead of the average period t_0 characteristics. The LPI and the PPI are commonly used as approximations to the true cost-of-living indexes (COLI). COLI's indicate, roughly saying, how much money a consumer would need in period t relatively to the amount of money he needed in period t_0 to attain the same level of utility u in period t as in the base period t_0. It can be shown that under certain conditions the PPI_t underestimates the true increase of cost-of-living whereas the LPI overestimates it (see the discussion in Diewert, 1987). This is due to substitution effects in case of changes in the relative prices. This problem can be diminished by using chain indexes in which the period $t_0 - T$ is divided into sub-periods and for each subperiod an index is estimated. This reduces the problem of substitution bias. The price index at time t is then calculated by multiplying the subperiod price indexes covering the period from t_0 tot t. Another possibility is to take the geometric mean of the PPI and the LPI, which is known as the Fisher ideal price index P_F. This index is a superlative index number. Superlative index numbers meet certain reasonable criteria (Diewert, 1976) and give, in the case of retrieving a cost of living index, an excellent approximation (they provide better approximations than the indexes based on fixed weights which do not meet these criteria). Here it is used as a product price index.

$$\begin{aligned}
LPI_t &= \frac{\exp(\hat{\alpha}_t + \hat{\beta}_t \bar{x}_{jt0})}{\exp(\hat{\alpha}_{t0} + \hat{\beta}_{t0} \bar{x}_{jt0})} \\
LCPI_t &= LCPI_{t-1} * \frac{\exp(\hat{\alpha}_t + \hat{\beta}_t \bar{x}_{jt-1})}{\exp(\hat{\alpha}_{t-1} + \hat{\beta}_{t-1} \bar{x}_{jt-1})} \\
PPI_t &= \frac{\exp(\hat{\alpha}_t + \hat{\beta}_t \bar{x}_{jt})}{\exp(\hat{\alpha}_{t0} + \hat{\beta}_{t0} \bar{x}_{jt})} \\
PCPI_t &= PCPI_{t-1} * \frac{\exp(\hat{\alpha}_t + \hat{\beta}_t \bar{x}_{jt})}{\exp(\hat{\alpha}_{t-1} + \hat{\beta}_{t-1} \bar{x}_{jt})} \\
FP_t &= \sqrt[2]{LPI_t PPI_t}
\end{aligned} \tag{17}$$

Equation 18 presents the third method to calculate price indexes. This method is more straightforward than the second method. The assumption of constant implicit values of product characteristics is somewhat relaxed by estimating two-years regressions in which the intercept is allowed to shift between two adjacent years by means of including a dummy $D_{\tilde{t}}$ equal to one in year $\tilde{t}$ and equal to zero in year $\tilde{t} - 1$.

Table 1 Price indexes for new cars using traditional and hedonic methods

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Type of index										
CBS index	1.000	1.030	1.090	1.120	1.130	1.130	1.107	1.107	1.107	1.112
Weighted av. sample	1.000	1.032	1.146	1.165	1.198	1.220	1.206	1.227	1.231	1.207
Pooled regression	1.000	1.020	1.067	1.086	1.090	1.070	1.036	1.017	1.011	1.008
Two-years regression	1.000	1.019	1.061	1.082	1.085	1.065	1.034	1.018	1.010	1.004
Yearly regressions										
Laspeyres fixed base	1.000	1.019	1.064	1.077	1.077	1.073	1.041	1.038	0.985	0.933
Laspeyres chain index	1.000	1.019	1.064	1.081	1.083	1.064	1.036	1.020	1.011	1.011
Paasche fixed base	1.000	1.019	1.058	1.084	1.084	1.055	1.014	0.989	0.979	0.976
Paasche chain index	1.000	1.019	1.057	1.080	1.083	1.060	1.028	1.011	1.004	0.999
Fisher index	1.000	1.019	1.061	1.080	1.080	1.064	1.027	1.013	0.982	0.954

Then one assumes constant implicit values $\tilde{\beta}_{t-1t}$ only between two adjacent years $\tilde{t}-1$ and $\tilde{t}$ and not for the whole estimation period $t_0 \ldots T$.

$$\ln(p_{it}) = \alpha + \alpha_{\tilde{t}} D_{\tilde{t}} + \tilde{\beta}_{\tilde{t}-1\tilde{t}} x_{jt} + \varepsilon_{it} \quad (18)$$

An illustration of the use of hedonic methods is shown in table 1. There quality adjusted price indexes have been constructed for new passenger cars sold in the Netherlands between 1990-1999. The data set has been extensively described in Bode and Van Dalen (2001). We restrict ourselves here to showing the resulting quality adjusted price indexes originating from different hedonic methods but we do not show the underlying regression results because this is not the scope of this empirical illustration. However, they are available on request for the interested reader.

Table 1 shows that the official CBS index, which is used in the Dutch CPI, shows a rapid increase in car prices in 1991 and 1992 and a stabilization of the car prices in 1993-1999 at an about 11% higher level than in 1990. If the weighted average car prices of this sample are calculated a similar picture emerges but then the average car prices are in 1999 about 21% higher than in 1990.

The price indexes based on the hedonic price equations follow a different pattern than the conventional price index of the CBS. Just as the CBS index they rise sharply during the first half of the 90's but unlike the CBS price index they decrease in the second half of the 90's to or even below the 1990 price level of new cars. This indicates that the CBS price index overestimates the price index for new cars, once controlled for quality changes, by at least 11%. The price index of new cars has a weight of 3.325% in the overall Dutch CPI. The overall Dutch CPI seems to be overestimated for the period 1990-1999 by 0.3 – 0.4 percentage points due to its overestimation of the price index of new cars.

The price indexes based on the pooled, the two-years and the yearly regressions are quite similar. They increase to about 1.085-1.090 in 1994 after which they decrease to slightly above the 1990 price level. Only the yearly fixed base period price indexes show an actual decrease in prices relative to the 1990 price level. Comparing the CBS index with these price indexes suggests that the major quality improvements of cars occurred during the second half of the 90's and to a lesser extent in the first half of the 90's.

As already mentioned in the former section the Fisher price index and the chain price indexes based on the yearly hedonic price regressions are expected to produce the best estimates of the true quality corrected price index. These three indexes lie between the Paasche and the Laspeyres index and the differences between them, especially between the two chain indexes are small. The Laspeyres chain index is equal or lies above the Paasche chain index. The difference between the three indexes is largest in 1999 when the difference between the Laspeyres chain price index and the Fisher price index amounts to 5.5 percentage points. The Laspeyres chain index indicates that the car prices in 1999 are 1.1 percentage point above the 1990 price level of new cars whereas the Fisher price index indicates that 1998 prices are 4.5 percentage point lower than the 1990 price of new cars once controlled for the quality improvements. The Paasche chain price index denotes a 0.4 percentage point price increase for cars between 1990-1998.

3.2 **The discrete choice method**

In this section the use of discrete choice models is discussed in the derivation of consumers' preference parameters and when making welfare comparisons. An approach is described aimed at constructing quality adjusted price indexes. Furthermore, three papers are discussed which focus on the identification and estimation problems when making welfare comparisons.

A possible way to estimate a quality adjusted price index based on discrete choice models is to derive the expenditure function. The idea is that by specifying a certain utility, or alternatively speaking standard of living level, $\tilde{u}$ one can derive the minimum amount of money needed to attain this utility level at different points in time. The ratio of the amount of money needed at time t and some base period t_0 serves as the quality adjusted price index.

The expenditure function is obtained by minimising the total expenditure necessary for the consumer to attain a specified utility level of $\tilde{u}$. An issue here is the choice of the utility level $\tilde{u}$. A possibility is to choose a level based on the choices of the product characteristics of the average consumer in the base period t_0 or the end period T (in that sense it is similar to the Laspeyres and the Paasche price index). In this context the minimisation problem is specified as follows

$$\min_{x_1,..,x_n,z} \sum_{k=1}^{n} p_k x_k + p(z)z \tag{19}$$

subject to

$$\begin{aligned} U(x_1,..,x_n,z) &> = \tilde{u} \\ x_{k,}\ z &> = 0\ (k=1,..,n) \end{aligned} \tag{20}$$

The optimal values of x^* and z^*depend on the prices and on the level of utility. The prices are derived by estimating the conditional logit model. H_j and H_z are known as Hicksian demand functions for the x_k and z.

$$\begin{aligned} x_k^* &= H_k(p_1,..,p_{nJ},\ p(z),\ \tilde{u}) = H_j(p_x,p(z),\tilde{u})\ ,(k=1,..,n) \\ z^* &= H_z(p_1,..,p_n,\ p(z),\ \tilde{u}) = H_z(p_x,p(z),\tilde{u}) \end{aligned} \tag{21}$$

Substituting the optimal values of the x_k's and z in $\sum p_k x_k + p(z)z$ gives the expenditure function.

$$\sum_{k=1,..,n} p_k x_k^* + p(z)z^* = \sum_{k=1,..,n} p_k H_k + p(z)H_z = m(p_x,p_z,\tilde{u}) \tag{22}$$

A quality adjusted price index $DCPI_t$ between two points in time t_0 and t is then achieved by deriving the expenditure functions m_{t_0} and m_t for t_0 respectively t and dividing m_t by m_{t_0} for a specified utility level $\tilde{u}$, with the vectors storing the prices p_{tx} and p_{tz} now being time dependent:

$$DCPI_t = \frac{m_t(p_{tx},p_{tz},\tilde{u})}{m_{t_0}(p_{t_0x},p_{t_0z},\tilde{u})} \tag{23}$$

As can be seen the price index of period t only depends on the specified utility level and prices (from both the base period and period t), but not on the quantities of goods (or parts of goods) consumed. This indicates that price indexes based on the discrete choice model do not suffer from lower level substitution bias.

There are a few studies in which discrete choice models are used to construct price indexes or, more generally, to identify and estimate consumers' preference parameters for making welfare comparisons. In these studies, the approach is different from the one described above. There, use is made of some welfare measures, like the equivalent variation (EV) and compensating variation (CV).

One of te studies is conducted by Nevo (2001). Nevo used a random effects discrete choice model to estimate the demand of ready-to-eat cereal. He continued by using the estimated demand system to evaluate the changes in welfare. In his demand system he introduces an outside good which consumers may choose if they decide not to purchase any of the ready-to-eat cereals which were under consideration by the researcher. For making welfare comparisons and for the construction of a price index Nevo used Hick's EV. The EV measures the change in consumer wealth that would be equivalent to the change in consumer welfare due to a change in prices (expressed in monetary terms). Nevo notes that two important assumptions have to be made during the estimation. The first one concerns what happens with the utility from the outside good in the period under investigation and the second is related to the precise specification of the error terms in the choice model. He decomposes the error term into two parts, i.e ξ_j^t reflecting the valuation of the unobserved (for the researcher) characteristics of the particular product variant j at time t and a mean zero stochastic term ε_{ijt}. He wonders what is actually captured by ξ_j^t and whether ξ_j^t changes over time. More specifically he asks himself whether ξ_j^t mainly reflects changes in the quality of the product (in which case it should be allowed to vary over time) or whether it reflects changes in consumer's tastes (in which case ξ_j^t should be kept constant over time). Nevo has adopted several combinations of assumptions (with respect to the outside good utility and the specification of ξ_j^t) in order to see by how much this would affect the resulting price indices. Changes in estimated price indices due to different assumptions concerning the utility of the outside good turned out to be quite large (about 35 percentage points) whereas the differences in estimated price indices due to differences in the specification of ξ_j^t turned out to be quite small.

A related paper is Petrin (2001). Petrin estimates the economic welfare effects of the introduction of the minivan for both minivan and non-minivan consumers in the US. Interesting is that he shows how aggregated consumer data can be used for welfare comparisons instead of consumer data on micro level. He demonstrates a technique for obtaining demand and supply curves with discrete choice methods using market level data augmented with aggragated consumer data of consumers buying different automobiles. Doing so he gets quite different, but much more reasonable, estimates of welfare benefits from the introduction of a new car type than with conventional methods for estimating demand and supply curves.

Bajari and Lenkard's (2001) theoretical paper deals with the identification and estimation of consumer preferences in hedonic discrete choice models of differentiated product demand. There, it is shown that the hedonic discrete choice model is generally not identified in case of unobserved product characteristics even though the entire demand function is observed. Furthermore, they state that choice data does not contain information on unobserved product characteristics. However, they show that under some (weak) conditions it is possible to recover the unobserved product characteristics using price information. They propose a two-stage Rosen like approach extended by incorporating product characteristics that are observed by consumers but not by economists. Then in the first stage, using nonparametric estimations, price data can be used under some weak conditions to recover the hedonic pricing function and the unobserved product characteristics. The idea behind this is that if two products have the same observable characteristics but differ in prices then the one with the highest price should have 'better' unobservable characteristics. Otherwise it would not have a positive demand. In the second stage they show that there is an inversion between consumer's choices and the preference parameters if some weak conditions hold, the product space is continuous and the specification of the utility function is known. In case of a discrete product space the authors suggest using a Gibbs sampling algorithm to simulate the population distribution of consumer's preference parameters.

3.3 A comparison of the two methods

Both approaches are based on the same theoretical concept namely consumers maximising their utility under a budget constraint. However, the two approaches differ in the further elaboration of the theoretical model. The hedonic approach is based on both consumers and producers utility maximising behaviour whereas the discrete choice model concentrates on consumer utility maximising behaviour. In the hedonic approach the resulting prices of product characteristics are market equilibrium prices in which each consumer's marginal rate of substitution between characteristics of the product and all other goods is equal to the marginal cost of producing this characteristic. In the discrete choice model only the consumer's behaviour is taken into account. The valuation of a product characteristic can be retrieved by calculating the marginal rate of substitution between that good and the other goods. Another theoretical difference is that the hedonic method is based on the idea that a product is a bundle of product characteristics and that consumers actually buy these characteristics instead of the products itself. Hence, it more or less assumes that a consumer has preferences over *any* conceivable configuration of a composite good and he purchases the one with the configuration which mathes is desrided configuration best. This is not the case with discrete choice models where consumers only have preferences over the offered existing configurations of a product. A drawback of the discrete choice model is that, due to the assumption of extreme value distributed error terms, it has the Independence of Irrelevant Alternatives property. This drawback may be (partly) overcome by using nested logit models.

A more pragmatic difference between these two methods lies in the data requirements for the empirical part. The hedonic method only requires aggregated market data like data from product prices, product characteristics and sales volume whereas for the discrete choice method also consumer data on income and probably also other consumer characterisitcs, whether or not aggregated, are needed. This is a drawback of the discrete choice method since information on individual consumers is, at least in the Netherlands and probably also in most other countries, not available. Usually, data sets only contain detailed information on product or on consumer characteristics. The extra costs incurred with collecting both types of data may be high. A solution suggested by Petrin (2001) may be to combine information from two sources, i.e. one with detailed product information and one with aggregated consumer information.

The performance of the two approaches has been compared by among others Mason and Quigley (1990), Cropper et. al. (1993). Mason and Quigley performed Monte Carlo experiments using both techniques on the same data-set in order to compare their willingness to pay for commodity characteristics estimates. Their results indicate that the hedonic method produces relatively good estimates when the size of the error term is small whereas the discrete choice model gives better estimates when the size of the error terms is medium or large. With respect to forecasting consumers' choices the hedonic model seems to perform relatively well when the size of the error terms is small whereas the discrete choice model does relatively well in case of medium and large sized error terms. Cropper et. al. compare, also by simulation, the performance of the multinomial logit model and the hedonic model in estimating consumer preferences for housing attributes. They ascribe preferences over the attributes of houses to a population of consumers and they calculate equilibrium prices by having them bid for a set of houses. With the resulting data set they estimate the two models. The estimation results show that marginal willingness to pay for an product attribute is estimated equally well by the two methods but that the logit model outperforms the hedonic method in valuing non marginal attribute changes.

4 CONCLUDING REMARKS

After the publication of the Boskin (1996) report on cost of living indexes interest in correctly estimating such indexes renewed. The main result of the report was that conventional cost-of living indexes overestimate the true cost-of-living index. One of the reasons is that for certain products the product specific price index is overstated because quality improvements of the products have not been accounted for. An often used method in the academic world to construct quality adjusted price indexes is to use hedonic methods. However, discrete choice models may also be useful in this context.

The two methods differ both theoretically and empirically and they both have their pros and cons. The two approaches are based on the same theoretical concept namely consumers maximising their utility under a budget constraint. However, they differ in the further elaboration of the theoretical model. The most important difference is that the hedonic approach is based on both consumers and producers utility maximising behaviour whereas the discrete choice model concentrates on consumer utility maximising behaviour. A second main difference is that in the hedonic approach it is assumed that a consumer has preferences overe *any* conceivable configuration of a composite good whereas discrete choice models only focus on *existing* product variants.

With respect to differences in empirical work the first thing which is noteworthy is that price index figures obtained using discrete choice models do not seem to suffer from product substitution bias like the ones obtained through hedonic methods. Second, past empirical research has indicated that discrete choice models estimate the monetary value of non-marginal changes in product characteristics relatively well compared with the conventional hedonic estimation methods. This is also likely to result in a better quality of the price index figure. One of the main disadvantages of using discrete choice models is that the data requirements are higher; you need data on both product and consumer characteristics instead of only product characteristics. Furthermore, calculating the price indexes once when has the data is also more complex in case of the discrete choice models.

REFERENCES

Bajari, P and C. L. Benkard, 2001, Demand estimation with heterogenous consumers and unobserved product characteristics: a hedonic approach, *NBER* technical *working paper*, no. 272, Cambridge.

Berndt, E.R., 1991, The practice of econometrics: classic and contemporary,*Addison-Wesley Publishing Company*, Mass.

Berry, S., J. Levinsohn and A. Pakes, 1995, Automobile prices in market equilibrium, *Econometrica*, 63, 841-890.

Blow, L. and I. Crawford, 1998, A quality constant price index for new cars in the UK, 1986-1995, Working paper series W98/12, *The Institute for Fiscal Studies*, London.

Bode, B. and J. van Dalen, 2001, Quality-corrected price indexes of new passenger cars in the Netherlands, 1990-1999, working paper, *Erasmus University*, Rotterdam.

Boskin, M.J. et al., 1996, Towards a more accurate measure of cost of living, final report to the Senate Finance Committee from the Advisory To Study The Consumer Price, December 4, Washington DC, *Senate Finance Committee*.

Chattopadhyay, S., 2000, The effectiveness of McFadden's nested logit model in valuing amenity improvement, *Regional Science and Urban Economics*, vol. 30(1), 23-43.

Chow, G.C., 1967, Technological Change and the Demand for Computers, *American Economic Review*, vol. 57(5), 1117-1130.

Court, A. T., 1939, Hedonic Price Indexes with Automotive Examples, The Dynamics of Automobile Demand, *The General Motors Corporation*, New York, 99-117.

Cramer, J.S., 1966, Een prijsindex van nieuwe personenauto's, 1950-1965, *Statistica Neerlandica*, vol. 20(2), 215-224.

Cropper, M.L. et al., 1993, Valuing product attributes using single market data: a comparison of hedonic and discrete choice approaches, *The Review of Economics and Statistics*, vol. 75, 225-232.

Diewert, W.E., 1976, Exact and superlative index numbers, *Journal of Econometrics*, vol. 4, 115-145.

Diewert, W.E., 1987, Index numbers in The New Palgrave: A Dictionary of Economics, vol. 2, J. Eatwell, M. Milgate and P. Newman (eds.), *The Macmillan Press*, London, 767-780.

Diewert, W.E., 2001, Hedonic Regressions: A Consumer Theory Approach, Discussion paper no. 01-12, *The University of British Columbia*, Vancouver.

Griliches, Z., 1961, Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change, The Price Statistics of the Federal Government, General Series no. 73, Columbia *University Press for NBER*, New York, 137-196.

Maddala, G.S., 1983, Limited-dependent and qualitative variables in econometrics, *Cambridge University Press*, Cambridge.

Mason, C. and J.M. Quigley, 1990, Comparing the performance of discrete choice and hedonic models in: Fisher, M.M., Nijkamp, P. and Papageorgiou, Y.Y., (eds.), Spatial Choices and processes, *Elsevier*, North-Holland, Amsterdam, 219-246.

McFadden, D., 1974, Conditional logit analysis of qualitative choice behavior, in: P. Zarembka ed., Economic theory and mathematical economics, *Academic Press*, New York.

Nevo, A., 2001, New products, quality changes and welfare measures computed from estimated demand systems, *NBER working paper*, no. 8425, Cambridge.

Petrin, A., 2001, Quantifying the benefits of new products: te case of the minivan, *NBER working paper*, no. 8227, Cambridge.

Triplett, J.E., 2000, Draft copy Handbook on quality adjustment of price indexes for information and communication technology products, *OECD*, Paris.

Waugh, F.V., 1928, Quality Factors Influencing Vegetable Prices, *Journal of Farm Economics*,, vol. 10(2), 185-196.