

Jonker, Nicole

**Working Paper**

## Constructing quality-adjusted price indices: a comparison of hedonic and discrete choice models

ECB Working Paper, No. 172

**Provided in Cooperation with:**

European Central Bank (ECB)

*Suggested Citation:* Jonker, Nicole (2002) : Constructing quality-adjusted price indices: a comparison of hedonic and discrete choice models, ECB Working Paper, No. 172, European Central Bank (ECB), Frankfurt a. M.

This Version is available at:

<https://hdl.handle.net/10419/152606>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

EUROPEAN CENTRAL BANK  
**WORKING PAPER SERIES**



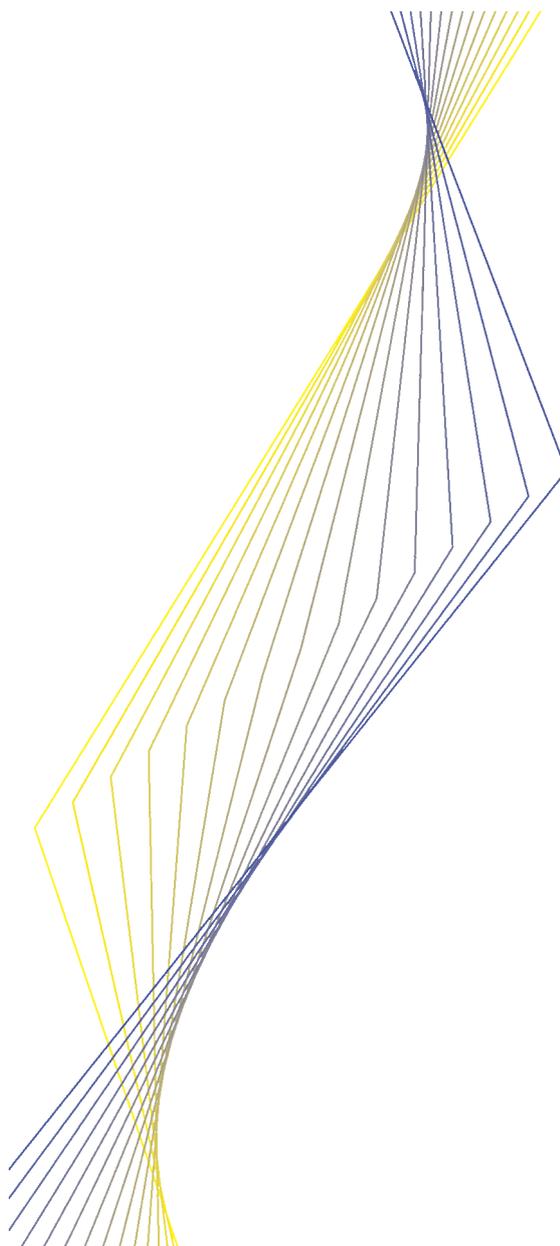
**WORKING PAPER NO. 172**

**CONSTRUCTING  
QUALITY-ADJUSTED PRICE  
INDICES: A COMPARISON OF  
HEDONIC AND DISCRETE  
CHOICE MODELS**

**BY NICOLE JONKER**

**September 2002**

**EUROPEAN CENTRAL BANK**  
**WORKING PAPER SERIES**



**WORKING PAPER NO. 172**

**CONSTRUCTING  
QUALITY-ADJUSTED PRICE  
INDICES: A COMPARISON OF  
HEDONIC AND DISCRETE  
CHOICE MODELS**

**BY NICOLE JONKER\***

**September 2002**

*\* I would like to thank Maria Demertzis, Erwin Diewert, Peter van Els, Carsten Folkertma, Lex Hoogduin and Mick Silver for their helpful comments.*

*This paper was presented at the CEPR/ECB Workshop in "Issues in the measurement of price indices" held at the European Central Bank on 16/17 November 2001.*

© European Central Bank, 2002

<b>Address</b>	<b>Kaiserstrasse 29 D-60311 Frankfurt am Main Germany</b>
<b>Postal address</b>	<b>Postfach 16 03 19 D-60066 Frankfurt am Main Germany</b>
<b>Telephone</b>	<b>+49 69 1344 0</b>
<b>Internet</b>	<b><a href="http://www.ecb.int">http://www.ecb.int</a></b>
<b>Fax</b>	<b>+49 69 1344 6000</b>
<b>Telex</b>	<b>411 144 ecb d</b>

*All rights reserved.*

*Reproduction for educational and non-commercial purposes is permitted provided that the source is acknowledged.*

*The views expressed in this paper are those of the authors and do not necessarily reflect those of the European Central Bank.*

ISSN 1561-0810

# Contents

Abstract	4
Non-technical summary	5
1 Introduction	6
2 “Historical” overview	7
2.1 Other methods	8
2.2 Hedonic models	8
2.3 Discrete choice models	9
3 Theoretical models	10
3.1 Hedonic method	11
3.2 Conditional logit model	13
3.3 Random coefficients discrete choice model	14
4 Empirical implementation	18
4.1 The hedonic method	18
4.2 The discrete choice method	20
4.3 A comparison of the two methods	24
5 Concluding remarks	25
6 References	27
European Central Bank working paper series	30

## Abstract

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. We compare two methods in this paper which can be used to correct for quality change bias, namely the hedonic method and a method based on the use of discrete choice models. We compare the underlying micro-economic models of the two methods as well as their empirical implementation. Although the discrete choice model has not been used often to calculate quality-adjusted price indices, past research shows that it might be beneficial to do so.

Key words: consumer price index, consumer behaviour, firm behaviour, discrete choice models

JEL Codes: C43, D11, D21, C25

## Summary

After the publication of the Boskin (1996) report on cost-of-living indices, a lot of interest in correctly estimating these indices was revived. One of the main results of the Boskin report was that conventional cost-of-living indices might overestimate the true cost-of-living index, which might be quite severe in empirical terms. For the USA, the CPI was overestimated by 1.1 percentage points on a yearly basis. One of the reasons for this overestimation is that the CPI suffers from several measurement errors, the most important of which being the quality change bias. With quality change bias we mean that for certain products the product specific price index is overstated because quality improvements of the products have not been accounted for in the price. In the academic world, the most popular method in constructing quality adjusted price indices is to apply hedonic methods. However, discrete choice models may also be useful in this context and have also been used for this purpose.

The two methods differ both theoretically as well as empirically and have both 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 consumers' utility maximising behaviour alone. A second main difference is that the hedonic approach assumes that a consumer has preferences over *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 indices 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 discontinuous changes in product characteristics relatively well compared to the conventional hedonic estimation methods. An example of a discontinuous product characteristic is the inclusion of an internal cd-writer in a personal computer compared with a personal computer without such a device. High-tech products often have a lot of this kind of discontinuous characteristics and the use of discrete choice models for the construction of high-tech product price indices is most likely to result in a better quality of the price index figure. A disadvantage related to the use of discrete choice models is that the data requirements are higher. One needs data on both product and consumer characteristics instead of only product characteristics. However, under some conditions, individual consumer data may not be necessary, as shown by Petrin (2001). Aggregated consumer data which can be matched with some of the product characteristics may also be sufficient.

## 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 categorized 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 indices 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 the same level of welfare (or utility) across two points in time. The Boskin committee concluded that the CPI had overestimated 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 government benefits 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, measuring the CPI correctly 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. With reference to the US, an overestimation of the cost-of-living index by one percentage point would result in a USD 1 trillion increase in national debt after a dozen of years.

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 indices. With quality change bias we mean that for certain products the product specific price index is overstated because quality improvements of the products have not been accounted for in the price.

Changes in products can be characterized as being *marginal* or *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 the case of computers, characteristics like the speed of the computer or the capacity of the hard disk are examples of continuous product characteristics whereas 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 that can be extended by extra features, like cars or computers. 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 that are unlikely to have extra features like fridges, washing machines, etc.

We focus on two different approaches to deal with quality change bias in price indices, namely the hedonic approach and the discrete choice modelling approach. The former approach is widely known and is already used by some statistical agencies, whereas interest in using the latter approach for constructing quality-adjusted price indices just revived in the 90s. However, we think that using this latter approach seems to be quite promising.

This paper proceeds as follows: section 2 describes several methods different statistical agencies use to correct for quality changes in the price index of products. It also gives an 'historical' overview of the development of hedonic price indices and the developments in discrete choice modelling with respect to price indices. Section 3 gives a description of the theoretical models underlying hedonic methods and discrete choice models. Section 4 elaborates on the empirical implementation of these two approaches and how to derive a quality-adjusted price index. Regarding discrete choice models, we also pay attention to evaluating the welfare effects of product changes in general. Section 5 summarises and concludes.

## **2 'Historical' overview**

Different statistical agencies use different techniques to adjust price indices for changes in product quality. For the sake of completeness, we describe three of them briefly at the beginning of this section. However, in the remainder of this section we focus on the developments of the hedonic approach and the discrete choice modelling approach.

## 2.1 Other methods

The 'matched model' approach provides a means to adjusting price indices for changes in product quality. With this method the price index is constructed using only 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. This technique is not suitable however for constructing price indices of product types involving rapid technical progress, like cars or computers, unless one uses chained price indices. Silver and Heravi (2001) give a thorough analysis of the shortcomings of the matched model approach. A second 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, a third possibility focuses on price changes for a basic product specification. It adjusts these price changes for alterations in the configuration of the basic model. For example, if a DVD player becomes part of the standard specification of a certain PC model from one period to the next, 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 from the current PC price<sup>1</sup>.

## 2.2 Hedonic models

Nowadays, statistical agencies use more and more hedonic price indices 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 vegetable prices. Court (1939) was the first to estimate a simple hedonic price equation for cars sold in 1925-1935. The Bureau of Labor Statistics official new car index, which was actually based on average list prices by brands without taking into account changing specifications, *rose* 45% over the 1925-1935 period, whereas Court's hedonic price index for new cars *decreased* 55% during the same period. Court included product characteristics as regressors in order to correct for product changes and

---

<sup>1</sup>The associated price of the DVD player can be retrieved from the previous period accessories' price list of the basic PC or by asking the manufacturer about the additional manufacturer's costs associated with installing the DVD player standard.

year dummies to reflect price changes. 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 e.g. Griliches (1961, 1964) on car prices, Chow (1967) on mainframe computer rental prices and the IBM study by Cole et. al. (1986) on individual components of a computer system. These early studies showed that quality-adjusted price indices of cars and computer systems decreased over time. Griliches (1964) was the first to consider Laspeyres and Paasche price indices for cars. The advantage of these price indices is that the valuation/shadow prices of product characteristics are allowed to change over time, whereas this is not the case with Court's hedonic price equation. For the Netherlands, Cramer (1966) estimated the first quality-adjusted price indices of new passenger cars sold between 1950-1966. In his study, Cramer showed, among others, that the quality-adjusted price index for new cars was about 30% lower than the unadjusted price index.

The theoretical relation between hedonic models, utility and production theory was established by Rosen only just in 1974 (see section 3.1), several decades after the estimation of the first hedonic price equation. More recent studies on quality-adjusted price indices are e.g. Berndt, Griliches and Rappaport (1995), Raff and Trajtenberg (1995) and Blow and Crawford (1998).

### 2.3 Discrete choice models

In the mid-seventies another branch of econometrics started to develop, namely that of discrete choice modelling. McFadden (1974) derived the conditional logit model (see e.g. Maddala, 1983) from random utility theory. In such a set-up, the effect of choice characteristics on choice probabilities is estimated through the estimation of the so-called vector  $\beta$ . The elements of  $\beta$  characterize 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 as well as discrete variables. In that sense, the conditional logit model analyses a problem similar to the hedonic price index problem, although it was never used as such.

Discrete choice models have been used, albeit not often, primarily 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 simulations. 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 and finds that the former estimates are consistently lower.

Recently, some papers have appeared which pay 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). Furthermore, evaluating the welfare effects caused by economic changes can prove problematic (Petrin, 2001). Using a random coefficients discrete choice model overcomes these problems. Nevo (2001) uses this model 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 indices 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).

### 3 Theoretical models

In this section we describe the hedonic method and the discrete choice method on explaining consumer choices. 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.

In both approaches there is a set of consumers who have preferences regarding the  $n$  measurable characteristics of a composite good  $x$  and regarding  $m$  other goods  $z_1, \dots, z_m$ . These preferences can be represented by a

utility function  $U$ :

$$u = U(x_1, x_2, \dots, x_n, z_1, z_2, \dots, z_m) \quad (1)$$

It is assumed that the utility function  $U$  is concave with respect to the product characteristics of good  $x$  and of the other goods  $z_1, z_2, \dots, z_m$ . 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  $(x_1, x_2, \dots, x_n)$  and a composite commodity, the quantity of which is denoted by  $z$ , i.e.

$$u = U(x_1, x_2, \dots, x_n, z) \quad (2)$$

There exist  $J$  variants of good  $x$  and the  $j^{\text{th}}$  variant is denoted by  $x_j$ . This product variant  $x_j$  can be described by a vector of  $n$  measurable characteristics,  $x_j = (x_{1j}, x_{2j}, \dots, x_{nj})$ . 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  $z$ , then the utility he attaches to consuming  $x_j$  becomes

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

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  $y_i = p_j + z$ .

### 3.1 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 the 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 assumed that producers of a certain good operate in a competitive environment. Therefore single producers take product prices as given and cannot influence them. The class of goods can be characterized by  $n$  measured characteristics and any location in the plane represents a vector  $x = (x_1, x_2, \dots, x_n)$  with  $x_k$  equal to the level of the  $k^{\text{th}}$  product characteristic. A price  $p(x) = p(x_1, x_2, \dots, x_n)$  is defined at each point in the plane. It is assumed that a large number of

product varieties exist to choose from. Consumers base their decisions with regard to consumption on utility maximising behaviour and producers base their production on profit maximising behaviour. The observed prices  $p(x)$  are the market clearing prices matching consumers' and producers' choices perfectly and leading to market equilibrium.

It is assumed that consumers maximise their utility subject to the non-linear budget constraint. This requires that consumers choose  $z$  and  $(x_1, x_2, \dots, x_n)$  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:

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

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  $C(M; \gamma)$  where the vector  $\gamma$  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  $\pi = M(j)p(j) - C(M(j), x_{1j}, x_{2j}, \dots, x_{nj})$  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 costs evaluated at the optimum bundle of characteristics:

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

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.

### 3.2 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 variant 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 optimises 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  $\varepsilon_{ijt}$ . 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.

$$U_{ijt} = V_{ijt} + \varepsilon_{ijt} \quad (6)$$

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

Assume that

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

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

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

### 3.3 Random coefficients discrete choice model

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). They start from a somewhat different, more restrictive, specification of the logit model than the one discussed in the previous paragraph. Their specification is as follows:

$$\begin{aligned} V_{ij}(x_j, y_i - p_j) &= \beta' x_j - \alpha p_j + \xi_j + \varepsilon_{ij} \\ &= \delta_j + \varepsilon_{ij} \end{aligned} \quad (10)$$

Here, the subscript  $t$  has been dropped in  $V_{ij}$ . This is because they did not use the random coefficients model to construct price indices but for deriving demand and supply curves and for making welfare comparisons. Furthermore, in equation 10 there is no interaction between consumer and product characteristics. Heterogeneity among consumers is captured by the error term  $\varepsilon_{ij}$ . In this specification  $\xi_j$  is the (mean) utility of the unobserved (for the econometrician, not for the consumers) characteristics of good  $j$  and  $\delta_j$  is the mean utility of good  $j$ . Without the inclusion of the  $\xi_j$ 's the only source of variance in the model is actually  $\varepsilon_{ij}$ . This specification is computationally simple. However, the authors stress that the separability of the utility into the product characteristics  $\delta_j$  and the consumer characteristics  $\varepsilon_{ij}$  result in unrealistic aggregate substitution patterns, cross- and own-price elasticities. For each  $\delta_j$  a unique vector of market shares exists. Therefore, cross-price elasticities between any two variants do not depend on the variant characteristics, conditional on the market shares of these variants. This is in contradiction with the intuition that couples of products which are more alike with respect to their characteristics should have higher cross-price elasticities than products which are less alike. Furthermore, equation 10 implies that two products with the same market share will also have the same own-price elasticities. This also does not have to be the case.

In equation 11 the random coefficients discrete choice models is shown. This model allows for the interaction between consumer and product characteristics. Individual consumers may have different preferences for the observable characteristics. This interaction is captured by the term  $\sum_k \sigma_k v_{ik} x_{jk}$ . Each consumer  $i$  has idiosyncratic tastes for the  $n$  observed characteristics  $v_i = (v_{i1}, v_{i2}, \dots, v_{in})$ . The parameter  $\sigma_k$  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  $\beta^\mu$  reflect the mean relative valuation of the characteristics. The contribution of  $x_k$  units of characteristic  $k$  to consumer  $i$ 's utility is  $(\beta_k^\mu + \sigma_k v_{ik})x_k$ .

$$V_{ij}(x_j, p_j) = \beta^\mu x_j - \alpha p_j + \xi_j + \sum_{k=1}^n \sigma_k v_{ik} x_{jk} + \varepsilon_{ij} \quad (11)$$

With the specification of the above utility, utility can still be divided in two components, i.e.  $\delta_j = x_j \beta^\mu - \alpha p_j + \xi_j$  (mean utility derived from product  $j$ ) and  $\mu_{ij} = \sum_{k=1}^n \sigma_k v_{ik} x_{jk} + \varepsilon_{ij}$  (consumer specific deviation from this mean). However, this second component now depends on both consumer preferences and product characteristics, unlike the second component  $\varepsilon_{ij}$  in equation 10. This will induce more realistic cross-price elasticities than the ones obtained by using the standard logit approach. The idea behind this is that consumers who attach a high utility to a specific product characteristic will tend to attach a high utility to all product variants which score high at that particular characteristic, inducing large substitution effects between these product variants.

Berry, Levinsohn and Pakes have extended equation 11 in order to implement prior information on the distribution of certain consumer characteristics. They also had specific ideas about the functional forms regarding the interaction between consumer and product characteristics and the interaction between income and product variant prices. They have had information on the distribution of income across households at their disposal. The random coefficients discrete choice model was nested into a Cobb-Douglas utility function in product characteristics and expenditures on other goods. Consumer  $i$ 's household income is denoted by  $y_i$ .

$$V(x_j, y_i - p_j) = (y_i - p_j)^\alpha G(x_j, \xi_j, v_i) e^{\varepsilon_{ij}} \quad (12)$$

This model is linear in logs. The utility  $V_{i0}$  is associated with the utility derived from consuming the outside good. An additional unobserved term  $v_{i0}$  has been added to  $V_{i0}$  in order to account for the possibility that there is more unobserved variance in the idiosyncratic component of the outside good than for the 'inside product variants'. With this model, interactions between consumer and product characteristics are allowed and it is possible to incorporate exogenous data on the income distribution.

$$\begin{aligned}
V_{i0}(x_0, y_i - p_0) &= \alpha \log(y_i) + \xi_0 + \sigma_0 v_{i0} + \varepsilon_{i0} & (13) \\
V_{ij}(x_j, y_i - p_j) &= \beta^\mu x_j + \alpha \log(y_i - p_j) + \xi_j + \sum_{k=1}^n \sigma_k v_{ik} x_{jk} + \varepsilon_{ij}, \quad (j=1, \dots, J)
\end{aligned}$$

What remains is the problem of endogenous prices. It is likely that there is some correlation between the product prices  $p_j$  and the unobserved product characteristics  $\xi_j$  which may actually be observed by the consumers. If this endogeneity problem is not properly dealt with, it induces biased estimates of the elasticities; price elasticities turn out to become unreasonably small in absolute value. The simultaneity problem becomes more complicated through the interaction between consumer and product characteristics and the discrete character of the choice set. Berry, Levinsohn and Pakes assume that  $\xi_j$  is mean independent of a set of exogenous instruments<sup>2</sup>. They construct a set of estimators which behave according to the assumptions imposed by the orthogonality condition. The empirical results show that to get sensible own-price elasticities it is indeed necessary to use a model and estimators which correct for price endogeneity. The aggregated demand function is derived by integrating out the choice function over the distribution of consumer characteristics (the  $v_{ik}$ 's).

In a later paper, Berry, Levinsohn and Pakes (2001) not only use product data but they use individual consumer data as well, matching consumer characteristics with observed consumers' choices. They use an extremely rich micro data set, containing both the actual and the likely second choice of consumers who bought a car in the US in 1993. Using micro consumer data changes the price endogeneity problem but does not make it disappear. A subset of the parameters is identified without the mean independence assumption made in Berry, Levinsohn and Pakes (1995). However, depending on the application, they need to estimate some additional parameters. Using individual consumer data reduces the unobserved heterogeneity in preferences of product characteristics.

Berry, Levinsohn and Pakes present a random coefficients discrete choice model which allows preferences for product characteristics to vary as a linear function of observed and unobserved consumer characteristics. Subscript  $k$  refers to observed product characteristics (including price) and subscript  $r$  refers to observed household/consumer characteristics. The notations used

---

<sup>2</sup>Their mean independence assumption states that both supply (not discussed here!) and demand unobservables are mean independent of both product and cost characteristics. The pricing equation depends on the firm's supply and cost function.

sometimes differ slightly from those used in the Berry, Levinsohn and Pakes (1995) paper. Here,  $x_{jk}$  is the  $k^{th}$  observable product characteristic,  $\xi_j$  refers to the unobserved characteristics of product  $j$ ,  $z_{ir}$  is the  $r^{th}$  characteristic of consumer  $i$ ,  $v_{ik}$  denotes the idiosyncratic taste of consumer  $i$  for the  $k^{th}$  observed product characteristic and  $\varepsilon_{ij}$  is the remaining error term reflecting idiosyncratic individual preferences, independent of the product characteristics and of each other. The parameter  $\tilde{\beta}_{ik}$  measures the valuation of consumer  $i$  for product characteristic  $k$ . It can be decomposed into a mean valuation  $\beta_k^\mu$  by all consumers for this product characteristic, and two consumer-specific preference shifters  $\beta_{kr}^0$  and  $\beta_k^u$ .  $\beta_{kr}^0$  measures differences in the valuation of  $k^{th}$  product characteristic from the mean depending on observed consumer characteristics of consumer  $i$  whereas  $\beta_k^u$  measures this difference in valuation depending on the unobserved characteristics of consumer  $i$ .

$$V_{ij} = \sum_k x_{jk} \tilde{\beta}_{ik} + \xi_j + \varepsilon_{ij} \quad (14)$$

$$\tilde{\beta}_{ik} = \beta_k^\mu + \sum_r z_{ir} \beta_{kr}^0 + \beta_k^u v_{ik}$$

Substituting the specification of  $\tilde{\beta}_{ik}$  into  $V_{ij}$  yields

$$V_{ij} = \delta_j + \sum_k \sum_r x_{jk} z_{ir} \beta_{kr}^0 + \sum_k x_{jk} \beta_k^u v_{ik} + \varepsilon_{ij} \quad (15)$$

$$\delta_j = \sum_k x_{jk} \beta_k^\mu + \xi_j, \quad j = 0, 1, \dots, J$$

The specification of  $\delta_j$  enables the researchers to identify the parameters  $\delta$ ,  $\beta^0$  and  $\beta^u$  using micro data, even without making assumptions on the joint distribution function of  $(\xi, \varepsilon)$ . However, estimates of the parameter vector  $\beta^\mu$  are needed to determine the own- and cross-price elasticities. The number of observations of  $\delta$  which can be used to identify and estimate  $\beta^\mu$  equals the number of products and actually  $\beta^\mu$  needs to be estimated with the product data. Consequently, for the estimation of  $\beta^\mu$  assumptions are needed with respect to the joint distribution of  $(\xi, \varepsilon)$ . It is the same identification problem as the one in Berry, Levinsohn and Pakes (1995) and various assumptions can solve the identification problem. The authors make the assumption that the  $\xi_j$ 's are mean independent of the non-price characteristics of all of the products, but they also used other assumptions in order to identify  $\beta^\mu$ . A method of moments estimator has been used

to estimate the model. Market level data can be obtained by aggregating the consumer-specific choices implied by the individual utilities over the population's distribution of consumer characteristics.

The empirical results show that allowing for unobserved consumer characteristics greatly improves the reliability of the substitution patterns. The results also indicate that the use of the second choice data is also necessary to get reasonable estimates. Unfortunately, such data are generally unavailable or extremely expensive.

## 4 Empirical implementation

### 4.1 The hedonic method

There are various ways of estimating a hedonic price equation and consequently there are also a number of ways to construct price indices. Here, we present three related methods. In the first method equation 16 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  $\varepsilon_{jt}$ . 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 as to 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).

$$p_{jt} = f(c, x_{jt}) + \varepsilon_{jt} \quad (16)$$

With the second method one assumes that the implicit values of product characteristics do not change over the estimation period  $t_0 \dots T$ . Then one can pool the data from different periods and estimate equation 17 using period dummies  $D_t$ . Here, the implicit values of the continuous and discrete product characteristics are stored in the vector  $\beta$ . The parameter  $\alpha$  is an intercept term and  $\alpha_t$  ( $t \neq t_0$ ) acts as an intercept shift in log prices for period  $t$  compared to period  $t_0$ , once controlled for product characteristics.

$$\ln(p_{jt}) = \alpha + \alpha_{t_0+1}D_{t_0+1} + \alpha_{t_0+2}D_{t_0+2} + \dots + \beta x_{jt} + \varepsilon_{jt} \quad (17)$$

Analogously, the exponent of  $\alpha_t$  is an intercept shift in prices for period  $t$  compared to period  $t_0$ , once controlled for product characteristics. Equation 18 defines the quality-controlled price index  $I_t$  of prices at  $t$  relative to prices in the base period  $t_0$

$$I_t = \exp(\alpha_t) \quad (18)$$

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

$$\ln(p_{jt}) = \tilde{\alpha}_t + \beta_t x_{jt} + \varepsilon_{jt} \quad (19)$$

There are different product price indices which can be used. Five common price indices 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  $t$  relative 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 indices (COLI). COLIs indicate, roughly speaking, how much money a consumer would need in period  $t$  relative 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 the 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 indices in which the period  $t_0 - T$  is divided into sub-periods and for each sub-period an index is estimated. This reduces the problem of substitution bias. The price index at time  $t$  is then calculated by multiplying the sub-period price indices covering the period from  $t_0$  to  $t$ . 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 indices based on fixed weights,

which do not meet these criteria). Here it is used as a product price index.

$$\begin{aligned}
LPI_t &= \frac{\exp(\hat{\alpha}_t + \hat{\beta}_t x_{jt0}^\mu)}{\exp(\hat{\alpha}_{t0} + \hat{\beta}_{t0} x_{jt0}^\mu)} & (20) \\
LCPI_t &= LCPI_{t-1} * \frac{\exp(\hat{\alpha}_t + \hat{\beta}_t x_{jt-1}^\mu)}{\exp(\hat{\alpha}_{t-1} + \hat{\beta}_{t-1} x_{jt-1}^\mu)} \\
PPI_t &= \frac{\exp(\hat{\alpha}_t + \hat{\beta}_t x_{jt}^\mu)}{\exp(\hat{\alpha}_{t0} + \hat{\beta}_{t0} x_{jt}^\mu)} \\
PCPI_t &= PCPI_{t-1} * \frac{\exp(\hat{\alpha}_t + \hat{\beta}_t x_{jt}^\mu)}{\exp(\hat{\alpha}_{t-1} + \hat{\beta}_{t-1} x_{jt}^\mu)} \\
FP_t &= \sqrt{LPI_t PPI_t}
\end{aligned}$$

Equation 21 presents the third method for calculating price indices. This method is more straightforward than the second method. The assumption of constant implicit values of product characteristics is somewhat relaxed. Two-year regressions are estimated in which the intercept is allowed to shift between two adjacent years by including a dummy  $D_{\tilde{t}}$ , equal to one in year  $\tilde{t}$  and equal to zero in year  $\tilde{t} - 1$ . Then one assumes constant implicit values  $\tilde{\beta}_{\tilde{t}-1\tilde{t}}$  only between two adjacent years  $\tilde{t} - 1$  and  $\tilde{t}$  and not for the whole estimation period  $t_0 \dots T$ .

$$\ln(p_{it}) = \alpha + \alpha_{\tilde{t}} D_{\tilde{t}} + \tilde{\beta}_{\tilde{t}-1\tilde{t}} x_{jt} + \varepsilon_{it} \quad (21)$$

## 4.2 The discrete choice method

In this section, we discuss the use of discrete choice models in the derivation of consumers' preference parameters and when making welfare comparisons. We describe an approach aimed at constructing quality-adjusted price indices. Furthermore, we discuss three papers 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 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

$$\min_{x_1, \dots, x_n, z} \sum_{k=1}^n p_k x_k + p(z)z \quad (22)$$

subject to

$$\begin{aligned} U(x_1, \dots, x_n, z) &> = \tilde{u} \\ x_k, z &> = 0 \quad (k = 1, \dots, n) \end{aligned} \quad (23)$$

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  $z$ .

$$\begin{aligned} x_k^* &= H_k(p_1, \dots, p_n, p(z), \tilde{u}) = H_j(p_x, p(z), \tilde{u}), \quad (k = 1, \dots, n) \\ z^* &= H_z(p_1, \dots, p_n, p(z), \tilde{u}) = H_z(p_x, p(z), \tilde{u}) \end{aligned} \quad (24)$$

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

$$\sum_{k=1, \dots, n} p_k x_k^* + p(z)z^* = \sum_{k=1, \dots, n} p_k H_k + p(z)H_z = m(p_x, p_z, \tilde{u}) \quad (25)$$

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:

$$DCPI_t = \frac{m_t(p_{tx}, p_{tz}, \tilde{u})}{m_{t_0}(p_{t_0x}, p_{t_0z}, \tilde{u})} \quad (26)$$

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 indices 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 indices or, more generally, to identify and estimate consumers'

preference parameters and to retrieve demand and cost functions 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). Recently, Bajari and Lenkard's (2001), Berry, Levinsohn and Pakes (1995, 2001), Nevo (2001) and Petrin (2001) have contributed a great deal to this branch of research.

Berry, Levinsohn and Pakes (1995, 2001) has already been discussed in section 2. Nevo (2001) uses a random effects discrete choice model to estimate the demand for ready-to-eat cereal. He continues 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 uses 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 to 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.  $\xi_j^t$  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  $\varepsilon_{ijt}$ . He wonders what is actually captured by  $\xi_j^t$  and whether  $\xi_j^t$  changes over time. More specifically he asks himself whether  $\xi_j^t$  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 consumers' tastes (in which case  $\xi_j^t$  should be kept constant over time). Nevo adopts several combinations of assumptions (with respect to the outside good utility and the specification of  $\xi_j^t$ ) 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  $\xi_j^t$  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. Petrin's approach requires less rich data in comparison with Berry, Levinsohn and Pakes (2001), whereas it still yields reasonable

estimates of demand and supply curves, price elasticities and measures of consumer welfare. Petrin uses market-level data and combined them with data from another source relating the *average* characteristics of consumers to the characteristics of the products they buy. This type of consumer data are much easier to get than the individual consumer data on both actual and second choices Berry, Levinsohn and Pakes have had at their disposal. Petrin uses the extra information on average consumer characteristics in order to get a better identification of the marginal utility of income and the substitution patterns between vehicles in the family vehicle segment.

Petrin's specification of the utility function can be found in equation 11. The only difference between the specification he uses and equation 11 is that  $\alpha$ , denoting the marginal utility from income, may vary according to three income groups (of equal size). So  $\alpha = (\alpha_0, \alpha_1, \alpha_2)'$ . His estimation procedure strongly resembles Berry, Levinsohn and Pakes' (2001) GMM approach. The difference is that Petrin supplements their moments with a set of 11 new moments. The idea behind this new set of moments is that aggregated data are actually aggregated individual data. Therefore, aggregated data contain information on the average of variables measured at the individual level. Petrin's new moments include the moments matching the average probability of new vehicle purchase conditional on the income category (he distinguishes three categories). Furthermore, Petrin also adds moments matching the model's predicted averages to those observed in the external data source concerning the average family size of car buyers of four different types of vehicles and the moments matching the probability that the head of household is between 30 and 60 years of age for the same four vehicle types.

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 do 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 non-parametric 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 char-

acteristics but differ in price 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 consumers' 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 consumers' preference parameters.

### 4.3 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 consumers' 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 consumers' 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 rather than the products themselves. 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 matches the desired 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 characteristics, 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 in 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 perform 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 error terms are medium-sized 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 regarding 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 a 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.

## 5 Concluding remarks

After the publication of the Boskin (1996) report on cost-of-living indices, interest in correctly estimating these indices renewed. The main result of the report was that conventional cost-of-living indices 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 are not accounted for. An often used approach in the academic world to construct quality-adjusted price indices 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 consumers' utility maximising behaviour alone. A second main difference is that in the hedonic approach it is assumed that a consumer has preferences regarding *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 a 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 to 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; one needs data on both product and consumer characteristics instead of only product characteristics. However, recent research on random coefficients discrete choice models shows that individual consumer data may not be necessary. Data at the product level combined with aggregated consumer data matched with some of the product characteristics seems to perform just as well.

## 6 References

- Bajari, P and C. L. Benkard**, 2001, Demand estimation with heterogeneous 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.
- Berndt, E.R., Z. Griliches and N. Rappaport**, 1995, Econometric estimates of price indexes for personal coputers in the 1990's, *Journal of Econometrics*, vol. 68(1), 243-268.
- Berry, S., J. Levinsohn and A. Pakes**, 1995, Automobile prices in market equilibrium, *Econometrica*, vol. 63, 841-890.
- Berry, S., J. Levinsohn and A. Pakes**, 2001, Differentiated products demand systems from a combination of micro and macro data: The new car market, working paper
- Berry, S., and A. Pakes**, 2001, The pure characteristics discrete choice model with application to price indices, working paper
- 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.
- 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.

**Cole, R. et. al.**, 1986, Quality-adjusted price indexes for computer processors and selected peripheral quipment, *Survey of Current Business*, vol. 66(1), 41-50.

**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.

**Griliches, Z.**, 1964, Notes on the measurement of price and quality changes in: *Models of income determination*, Studies in income and wealth, NBER report, Princeton University Press.

**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: the case of the minivan, *NBER working paper*, No. 8227, Cambridge.
- Silver, M. and S. Heravi**, 2001, Why the matched model method may fail us: results from an hedonic and matched experiment using scanner data, working paper, Cardiff University, Cardiff.
- 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.

## European Central Bank working paper series

For a complete list of Working Papers published by the ECB, please visit the ECB's website (<http://www.ecb.int>).

- I 13 “Financial frictions and the monetary transmission mechanism: theory, evidence and policy implications” by C. Bean, J. Larsen and K. Nikolov, January 2002.
- I 14 “Monetary transmission in the euro area: where do we stand?” by I. Angeloni, A. Kashyap, B. Mojon, D. Terlizzese, January 2002.
- I 15 “Monetary policy rules, macroeconomic stability and inflation: a view from the trenches” by A. Orphanides, December 2001.
- I 16 “Rent indices for housing in West Germany 1985 to 1998” by J. Hoffmann and C. Kurz., January 2002.
- I 17 “Hedonic house prices without characteristics: the case of new multiunit housing” by O. Bover and P. Velilla, January 2002.
- I 18 “Durable goods, price indexes and quality change: an application to automobile prices in Italy, 1988-1998” by G. M. Tomat, January 2002.
- I 19 “Monetary policy and the stock market in the euro area” by N. Cassola and C. Morana, January 2002.
- I 20 “Learning stability in economics with heterogenous agents” by S. Honkapohja and K. Mitra, January 2002.
- I 21 “Natural rate doubts” by A. Beyer and R. E. A. Farmer, February 2002.
- I 22 “New technologies and productivity growth in the euro area” by F. Vijselaar and R. Albers, February 2002.
- I 23 “Analysing and combining multiple credit assessments of financial institutions” by E. Tabakis and A. Vinci, February 2002.
- I 24 “Monetary policy, expectations and commitment” by G. W. Evans and S. Honkapohja, February 2002.
- I 25 “Duration, volume and volatility impact of trades” by S. Manganelli, February 2002.
- I 26 “Optimal contracts in a dynamic costly state verification model” by C. Monnet and E. Quintin, February 2002.
- I 27 “Performance of monetary policy with internal central bank forecasting” by S. Honkapohja and K. Mitra, February 2002.
- I 28 “Openness, imperfect exchange rate pass-through and monetary policy” by F. Smets and R. Wouters, February 2002.

- 129 “Non-standard central bank loss functions, skewed risks, and certainty equivalence” by A. al-Nowaihi and L. Stracca, March 2002.
- 130 “Harmonized indexes of consumer prices: their conceptual foundations” by E. Diewert, March 2002.
- 131 “Measurement bias in the HICP: what do we know, and what do we need to know?” by M. A. Wynne and D. Rodríguez-Palenzuela, March 2002.
- 132 “Inflation dynamics and dual inflation in accession countries: a “new Keynesian” perspective” by O. Arratibel, D. Rodríguez-Palenzuela and C. Thimann, March 2002.
- 133 “Can confidence indicators be useful to predict short term real GDP growth?” by A. Mourougane and M. Roma, March 2002.
- 134 “The cost of private transportation in the Netherlands, 1992-1999” by B. Bode and J. Van Dalen, March 2002.
- 135 “The optimal mix of taxes on money, consumption and income” by F. De Fiore and P. Teles, April 2002.
- 136 “Retail bank interest rate pass-through: the new evidence at the euro area level” by G. de Bondt, April 2002.
- 137 “Equilibrium bidding in the eurosystem’s open market operations” by U. Bindseil, April 2002.
- 138 “New” views on the optimum currency area theory: what is EMU telling us?” by F. P. Mongelli, April 2002.
- 139 “On currency crises and contagion” by M. Fratzscher, April 2002.
- 140 “Price setting and the steady-state effects of inflation” by M. Casares, May 2002.
- 141 “Asset prices and fiscal balances” by F. Eschenbach and L. Schuknecht, May 2002.
- 142 “Modelling the daily banknotes in circulation in the context of the liquidity management of the European Central Bank”, by A. Cabrero, G. Camba-Mendez, A. Hirsch and F. Nieto, May 2002.
- 143 “A non-parametric method for valuing new goods”, by I. Crawford, May 2002.
- 144 “A failure in the measurement of inflation: results from a hedonic and matched experiment using scanner data”, by M. Silver and S. Heravi, May 2002.
- 145 “Towards a new early warning system of financial crises”, by M. Fratzscher and M. Bussiere, May 2002.
- 146 “Competition and stability – what’s special about banking?”, by E. Carletti and P. Hartmann, May 2002.

- 147 "Time-to-build approach in a sticky price, stricky wage optimizing monetary model, by M. Casares, May 2002.
- 148 "The functional form of yield curves" by V. Brousseau, May 2002.
- 149 "The Spanish block of the ESCB-multi-country model" by A. Estrada and A. Willman, May 2002.
- 150 "Equity and bond market signals as leading indicators of bank fragility" by R. Gropp, J. Vesala and G. Vulpes, June 2002.
- 151 "G-7 inflation forecasts" by F. Canova, June 2002.
- 152 "Short-term monitoring of fiscal policy discipline" by G. Camba-Mendez and A. Lamo, June 2002.
- 153 "Euro area production function and potential output: a supply side system approach" by A. Willman, June 2002.
- 154 "The euro bloc, the dollar bloc and the yen bloc: how much monetary policy independence can exchange rate flexibility buy in an interdependent world?" by M. Fratzscher, June 2002.
- 155 "Youth unemployment in the OECD: demographic shifts, labour market institutions, and macroeconomic shocks" by J. F. Jimeno and D. Rodriguez-Palenzuela, June 2002.
- 156 "Identifying endogenous fiscal policy rules for macroeconomic models" by J. J. Perez, and P. Hiebert, July 2002.
- 157 "Bidding and performance in repo auctions: evidence from ECB open market operations" by K. G. Nyborg, U. Bindseil and I. A. Strebulaev, July 2002.
- 158 "Quantifying Embodied Technological Change" by P. Sakellaris and D. J. Wilson, July 2002.
- 159 "Optimal public money" by C. Monnet, July 2002.
- 160 "Model uncertainty and the equilibrium value of the real effective euro exchange rate" by C. Detken, A. Dieppe, J. Henry, C. Marin and F. Smets, July 2002.
- 161 "The optimal allocation of risks under prospect theory" by L. Stracca, July 2002.
- 162 "Public debt asymmetries: the effect on taxes and spending in the European Union" by S. Krogstrup, August 2002.
- 163 "The rationality of consumers' inflation expectations: survey-based evidence for the euro area" by M. Forsells and G. Kenny, August 2002.
- 164 "Euro area corporate debt securities market: first empirical evidence" by G. de Bondt, August 2002.

- 165 “The industry effects of monetary policy in the euro area” by G. Peersman and F. Smets, August 2002.
- 166 “Monetary and fiscal policy interactions in a micro-founded model of a monetary union” by R. M.W.J. Beetsma and H. Jensen, August 2002.
- 167 “Identifying the effects of monetary policy shocks on exchange rates using high frequency data” by J. Faust, J.H. Rogers, E. Swanson and J.H. Wright, August 2002.
- 168 “Estimating the effects of fiscal policy in OECD countries” by R. Perotti, August 2002.
- 169 “Modeling model uncertainty” by A. Onatski and N. Williams, August 2002.
- 170 “What measure of inflation should a central bank target?” by G. Mankiw and R. Reis, August 2002.
- 171 “An estimated stochastic dynamic general equilibrium model of the euro area” by F. Smets and R. Wouters, August 2002.
- 172 “Constructing quality-adjusted price indices: a comparison of hedonic and discrete choice models” by N. Jonker, September 2002.