The Econometric Foundations of Hedonic Elementary Price Indices[*]

Hans Wolfgang Brachinger[†]   Michael Beer

University of Fribourg Switzerland, Department of Quantitative Economics,
Bd de Pérolles 90, CH-1700 Fribourg, Switzerland.

Abstract

   Hedonic methods are currently considered state-of-the-art for handling quality changes when compiling consumer price indices. The present article proposes first a mathematical description of characteristics and of elementary aggregates. In a following step, a hedonic econometric model is formulated and hedonic elementary population indices are defined. These indices extend from simple indices based on some average quality to universal formulae that incorporate the full quality spectrum of the respective elementary aggregate. We emphasise that population indices are unobservable economic parameters that need to be estimated by suitable sample indices. It is shown that most of the hedonic elementary index formulae used in practice are sample versions of particular hedonic elementary population indices.

Key words: Consumer price index, hedonic regression, elementary aggregate, hedonic econometric model, hedonic elementary price index.

JEL classification: C43, E31.

1 Introduction

A consumer price index (CPI) measures the average price change of consumer goods in a market between two fixed time periods, assuming that their quality remains constant. In practice, however, the quality of the universe of products that households consume is continually changing. It is therefore necessary to estimate the contribution of the quality change to the observed price change in order to measure the quality-corrected 'pure' price change.

   The state-of-the-art manner of handling differences and changes in quality is the so-called *hedonic approach*. Its main idea is to identify the quality of a product – or, in other words, its 'potential contribution ... to the welfare and happiness of its purchasers and the community'

[†]Corresponding author. Tel.: +41 26 300 82 74; fax: +41 26 300 97 81.
Email addresses: hanswolfgang.brachinger@unifr.ch (H.W. Brachinger), research@michael.beer.name (M. Beer)
URL: http://www.unifr.ch/stat/

(Court, 1939, p. 107) – with a vector of product characteristics. In the hedonic approach, a regression equation is estimated relating the characteristics of the product to its price. Once such a relationship is established, the price of any similar item can be predicted by plugging its characteristics into the estimated *hedonic (regression) function*.

CPI concepts usually structure the basket of consumer goods in a hierarchical way. Individual price observations are transformed into a final index value through a sequence of aggregation steps. In the first stage, the price evolution is individually observed for restricted groups of homogeneous products, the so-called *elementary (expenditure) aggregates*. These aggregates usually serve as strata for data collection and form the building blocks of a CPI. For each of them, a so-called *elementary price index* is calculated. In further stages, these elementary price indices are 'averaged to obtain higher-level indices using the relative values of the elementary expenditure aggregates as weights' (ILO et al., 2009, para. 9.3). The need for adjusting price measurements for quality change appears at the level of elementary aggregates when individual prices are directly compared. Therefore, quality adjustment is purely an issue of elementary price indices. Elementary price indices where the quality adjustments are based on the hedonic approach are called *hedonic elementary price indices*.

The literature on hedonic methods in price statistics is steadily growing, with Triplett (2004) and ILO et al. (2004, Chap. 21) being two of the most recent comprehensive and fundamental overviews. The present paper contributes to this literature in proposing a formal framework for hedonic elementary price indices that incorporates and generalises these approaches. Our framework corroborates the existing theory by providing a novel conceptual approach from which most of the elementary hedonic price indices used in practice can be derived. Moreover, it defines the necessary concepts that allow, e.g., to examine state-of-the-art hedonic index estimators from an axiomatic viewpoint or to come up with new alternatives. We emphasise particularly the clear separation of elementary (population) indices as unobservable economic parameters from their estimators, the sample indices. In this aspect, the present piece of work abuts on the mindset of papers like, e.g., Dorfman et al. (1999), Brachinger (2002), Balk (2005), and Silver and Heravi (2007).

Section 2 provides a precise definition of characteristics and elementary aggregates. This definition permits a clear-cut and concise formulation of an econometric model underlying every hedonic price index. Section 3 discusses elementary price index concepts in general. These concepts are extended to universal formulae for hedonic elementary population indices in Section 4. Each of these indices is a well-defined economic parameter that eventually needs to be estimated from a random sample of observations. In Section 5, finally, sample versions of the universal formulae are presented and it is shown that from these sample indices most of the elementary indices used in practice can be derived. The paper closes with a short summary.

2 Elementary Aggregates and the Hedonic Econometric Model

2.1 Goods and characteristics

We begin by describing the basic entities of an elementary price index, namely some consumer goods offered in a market, the set of characteristics they exhibit, and the corresponding elementary aggregate. The formal language used for this purpose will allow us later to build an econometric model on top. From the outset, the characteristics of the goods play an important role as surrogates for the notion of quality. Their omnipresence in our framework will make the step from general (quality-unadjusted) to hedonic (quality-adjusted) elementary price indices

straightforward.

Let $\mathcal{O}$ denote the set of all consumer goods supplied in a market at a certain point in time.[1] Here, the notion of a *good* means physically tangible items like, e.g., used cars or personal computers as well as services and other immaterial entities to which a price can be assigned. Each of these goods exhibits a set of *characteristics*. Examples of such characteristics might be the volume or the physical mass of the good, the horsepower, mileage or colour of a used car, or the processor speed of a computer. Other non-physical characteristics comprise the location of sale or any after-sales service. Statistically speaking, a characteristic simply is a variable or an attribute. It may be scaled on different measurement levels from nominal to cardinal.

It is obvious that not every characteristic can be observed for a given good $o \in \mathcal{O}$, i.e. each characteristic m is generally only defined on a specific subset $\mathcal{O}_m$ of $\mathcal{O}$. Processor speed, for example, is a characteristic of computers but not (yet) of clothes or bicycles. So the domain $\mathcal{O}_m$ of the characteristic m : 'processor speed' contains the set of all computers, but also all other devices carrying a CPU including many modern household appliances, cars, and communication devices. m might not always be relevant to the purchaser of such goods, but technically, it is defined and measurable; we will come back to the economic relevance of certain characteristics later. Food, hospital services, and package holidays are examples of goods for which the characteristic 'processor speed' is not defined, hence they lie outside $\mathcal{O}_m$.

Although characteristics can be of any measurement scale, it is always possible to quantify their values such that they form a subset of the Euclidean real number space. This leads to the following definition:

Definition 2.1. *A **characteristic** m is a real-valued function $m : \mathcal{O}_m \longrightarrow \mathbb{R}$ defined on a non-empty subset $\mathcal{O}_m$ of $\mathcal{O}$. The set $\mathcal{O}_m$ is called the **domain** of m and, for each $o \in \mathcal{O}_m$, $m(o)$ will be called the m-**value** of o.*

For the sake of simplicity, the m-value of o may also be called its m-characteristic. The set of all characteristics will be denoted by $\mathcal{M} := \{m : \mathcal{O}_m \longrightarrow \mathbb{R} \,|\, \mathcal{O}_m \subset \mathcal{O}, \mathcal{O}_m \neq \emptyset\}$.

The reason why we put such emphasis on the domains of the characteristics is that they will now serve as building blocks for our definition of an elementary aggregate. Guidelines to practitioners on how elementary aggregates should be specified are traditionally rather vague and leave most decisions to the users' discretion. The authors of ILO et al. (2009, para. 9.7), e.g., confine themselves to requiring that elementary aggregates consist of goods that are 'as similar as possible', 'preferably fairly homogeneous', 'expected to have similar price movements', and 'appropriate to serve as strata for sampling purposes'. While such formulations may suffice in practice, they are much too cursory to serve as a building block of a hedonic econometric model. The following much more formal definition of an elementary aggregate contains all elements needed for the elaboration of our econometric framework.

Definition 2.2. *An **elementary aggregate** $\mathcal{G}$ is a set of goods in $\mathcal{O}$ having the following properties:*

1. The set $\mathcal{M}_\mathcal{G}$ of the characteristics defined for all elements of $\mathcal{G}$ is not empty, i.e.

$$\mathcal{M}_\mathcal{G} := \{m \in \mathcal{M} \,|\, \mathcal{O}_m \supset \mathcal{G}\} \neq \emptyset \,. \tag{1}$$

---

[1]We are going to raise the restriction to a single point in time in Section 4.

3

2. *The intersection of the domains of all characteristics contained in $\mathcal{M}_{\mathcal{G}}$ is a subset of $\mathcal{G}$, i.e.*

$$\bigcap_{m \in \mathcal{M}_{\mathcal{G}}} \mathcal{O}_m \subset \mathcal{G} \,. \tag{2}$$

*Each element $o \in \mathcal{G}$ will be called an **item** of the elementary aggregate $\mathcal{G}$. The elements of $\mathcal{M}_{\mathcal{G}}$ are called **distinguishing characteristics** of $\mathcal{G}$.*

Property (1) means that all goods belonging to an elementary aggregate $\mathcal{G}$ have at least one characteristic in common. Conversely, if two goods do *not* belong to the same elementary aggregate, there must be a characteristic that is defined for one of these goods but not for the other. Property (2) ensures that each good carrying all characteristics of $\mathcal{M}_{\mathcal{G}}$ is contained in the elementary aggregate. Note that every elementary aggregate in the sense of Def. 2.2 is defined relative to the set $\mathcal{O}$ of all goods supplied on the market.

The following proposition shows that each elementary aggregate has some kind of maximality property in the sense that its distinguishing characteristics fully determine the items of the aggregate. In other words, there is no item of an elementary aggregate that is not contained in the intersection of the domains of all distinguishing characteristics.

Proposition 2.1. *Each elementary aggregate $\mathcal{G}$ equals the intersection of the domains of its distinguishing characteristics, i.e.*

$$\bigcap_{m \in \mathcal{M}_{\mathcal{G}}} \mathcal{O}_m = \mathcal{G} \,. \tag{3}$$

Proof. We have $\mathcal{G} \subset \mathcal{O}_m$ for all $m \in \mathcal{M}_{\mathcal{G}}$. Therefore, $\mathcal{G} \subset \bigcap_{m \in \mathcal{M}_{\mathcal{G}}} \mathcal{O}_m$. The inclusion in the other direction is given by property 2 of Def. 2.2, hence equality holds. $\qquad\square$

It should be noted that an elementary aggregate in the sense of Def. 2.2 may still comprise many different items. In particular, there is no explicit requirement regarding the similarity or homogeneity of the items contained. So if, e.g., the physical mass of an object was taken as the only distinguishing characteristic, the respective elementary aggregate would embrace the whole universe of physically tangible goods, excluding just services and other intangible products like computer software. Thus we define the term 'elementary aggregate' in a much broader sense than it is usually applied in practice. However, it follows from Prop. 2.1 that supplementing the set $\mathcal{M}_{\mathcal{G}}$ of distinguishing characteristics of an elementary aggregate with additional characteristics leads to a diminution of $\mathcal{G}$. By selecting the appropriate list of distinguishing characteristics, we may thus in practice reduce a very general aggregate to one which satisfies the homogeneity or similarity requirements cited above.

As a consequence of Prop. 2.1, it is furthermore possible to induce elementary aggregates from samples of individual goods. Let $\mathcal{O}^* \subset \mathcal{O}$ be any set of goods. These might be, e.g., different models of personal computers. Let $\mathcal{M}_{\mathcal{O}^*} := \{m \in \mathcal{M} \,|\, \mathcal{O}_m \supset \mathcal{O}^*\}$ be the set of all characteristics whose domains contain these goods, i.e. all characteristics that are defined for all elements of $\mathcal{O}^*$. In the case of personal computers, these would contain typical attributes such as CPU speed, RAM size, hard drive size, brand, length of warranty period, etc., but also others such as the serial number, which may not be relevant to the consumers' purchase decision. Assume that $\mathcal{M}_{\mathcal{O}^*}$ is not empty. Then, it is possible to specify the elementary aggregate $\mathcal{G}(\mathcal{O}^*)$ induced by $\mathcal{O}^*$. The *induced elementary aggregate* is defined as the intersection

of the domains of all characteristics in $\mathcal{M}_{\mathcal{O}^*}$, i.e.

$$\mathcal{G}(\mathcal{O}^*) := \bigcap_{m \in \mathcal{M}_{\mathcal{O}^*}} \mathcal{O}_m \,. \tag{4}$$

The set $\mathcal{O}^*$ is thus extended by all goods on the market that exhibit at least the same characteristics as the goods fixed in $\mathcal{O}^*$. Obviously, by means of (4), any given set of characteristics $\mathcal{M}$ induces an elementary aggregate $\mathcal{G}(\mathcal{M}) := \bigcap_{m \in \mathcal{M}} \mathcal{O}_m$.

Def. 2.2 of an elementary aggregate is admittedly guided by theoretical elegance rather than practical pertinence. Nobody will be able to provide a comprehensive list of the distinguishing characteristics of even the simplest elementary aggregate being used in practice. Nonetheless is such an abstract definition inevitable to make the vague notion of an elementary aggregate manageable from an econometric viewpoint. Moreover, we presume that the idea of distinguishing characteristics can serve as an implicit guideline for practitioners needing to decide on which items should belong to a certain elementary aggregate. The authors of ILO et al. (2004, para. 3.147 ff.) identify three main approaches to the classification of consumer goods, namely the classification by product type, by purpose, and by economic environment. Recommended practice is 'to use a purpose classification at the highest level, with product breakdowns below'. Inevitably, the characteristics of goods play a certain role when elementary aggregates are defined by product type at the lowest level. The main merit of the concept introduced in Def. 2.2 is thus the duality between the elementary aggregate and its distinguishing characteristics. This duality will be exploited below.

As a final note to this first section, it is worth highlighting that the distinguishing characteristics of an elementary aggregate provide some very useful means of identifying items that are 'equivalent' in a certain sense. We will use this property later to partition an elementary aggregate into classes of equivalent quality. Let $\{m_1, \ldots, m_K\} \subset \mathcal{M}_{\mathcal{G}}$ denote any finite subset of the distinguishing characteristics of an elementary aggregate $\mathcal{G}$. By assembling them to a vector function

$$\boldsymbol{m} : \mathcal{G} \longrightarrow \mathbb{R}^K, \quad o \longmapsto \boldsymbol{m}(o) := (m_1(o), \ldots, m_K(o))' \tag{5}$$

all the items of $\mathcal{G}$ are mapped to a K-dimensional vector of characteristics. This identification of goods with a characteristics vector leads to an equivalence relation on $\mathcal{G}$ defined by

$$o_1 \sim_{\boldsymbol{m}} o_2 \quad :\Longleftrightarrow \quad \boldsymbol{m}(o_1) = \boldsymbol{m}(o_2) \,. \tag{6}$$

Two items of an elementary aggregate are thus identified if and only if their $\boldsymbol{m}$-values, i.e. their m_1- to m_K-values coincide. The equivalence classes respective to the relation $\sim_{\boldsymbol{m}}$ will be called $\boldsymbol{m}$-equivalence classes. They partition $\mathcal{G}$ into subsets containing items with equal $\boldsymbol{m}$-values. The quotient set induced by this equivalence relation will be denoted by $\mathcal{G}/\sim_{\boldsymbol{m}}$.

2.2 Characteristics and prices

In the last section, we identified a good with a list of characteristics and we showed how goods can be grouped into elementary aggregates. The economic foundation of this approach is the consumer theory developed by Lancaster (1966, 1971). This theory assumes that 'one demands not just physical objects, but the qualities with which they are endowed' (Milgate, 1987, p. 546). Consumers' preferences are therefore originally directed towards the characteristics of a good, and the latter determine eventually the consumers' preference ordering between individual items of an elementary aggregate.

Lancaster (1971, p. 140 ff.) himself emphasised that some characteristics of a good are usually irrelevant for a consumer's purchase decision (such as the serial number of a personal computer). Irrelevant characteristics are especially those that are invariant for all items of an elementary aggregate. Inversely, Lancaster defines a characteristic as *relevant* when ignoring it would change the preference ordering between two items.

Driven by the consumers' individual preferences, Lancaster's approach suggests that a good's price observed on the market is essentially determined by the relevant characteristics of that good. This assumption is called *hedonic hypothesis* in the literature (see e.g. Dickie et al., 1997; Triplett, 1987; United Nations, 1993). The hedonic hypothesis serves as the general basis for all hedonic price indices. In order to build up a solid theory of hedonic price indices, we propose formulating it as an econometric model in the following form:

Hedonic econometric model. *Let $\mathcal{G}$ be an elementary aggregate with distinguishing characteristics $\mathcal{M}_{\mathcal{G}}$. There exists a finite set of characteristics*

$$\mathcal{M}_{\mathcal{G}}^{\mathrm{pr}} = \{m_1, \ldots, m_{K_{\mathcal{G}}}\} \subset \mathcal{M}_{\mathcal{G}} \tag{7}$$

and a function $h_{\mathcal{G}} : \mathbb{R}^{K_{\mathcal{G}}} \longrightarrow \mathbb{R}_{\geq 0}$, such that the price $p(o)$ of any item $o \in \mathcal{G}$ can be written as

$$p(o) = h_{\mathcal{G}}\big(\boldsymbol{m}_{\mathcal{G}}^{\mathrm{pr}}(o)\big) + \epsilon(o) \tag{8}$$

with $\boldsymbol{m}_{\mathcal{G}}^{\mathrm{pr}}(o) = (m_1(o), \ldots, m_{K_{\mathcal{G}}}(o))'$. The residual term $\epsilon(o)$ is assumed to be stochastic with conditional expectation

$$\mathrm{E}\big(\epsilon(o) \mid \boldsymbol{m}_{\mathcal{G}}^{\mathrm{pr}}(o)\big) = 0\,. \tag{9}$$

*The set $\mathcal{M}_{\mathcal{G}}^{\mathrm{pr}}$ will be called the **set of price-relevant characteristics**, and $h_{\mathcal{G}}$ is the **hedonic function** of $\mathcal{G}$.*

This model exploits the idea that for an elementary aggregate for which the hedonic hypothesis holds, the set of distinguishing characteristics contains a finite subset of price-relevant characteristics. They determine the price up to a residual term that covers any quality-independent price component. Assumption (9) implies that the hedonic price of an item with a certain quality is given by the average price over all items of the same quality.

One of the central points here is that the vector of price-relevant characteristics is seen as a surrogate for a good's *quality*. Much in the spirit of the outline provided in the *System of National Accounts 1993* (see United Nations, 1993, para. 16.105 ff.), the term 'quality' subsumes all characteristics of an item which make it distinguishable from other items from an economic point of view. The hedonic function $h_{\mathcal{G}}$, being defined on the quotient set $\mathcal{G}/\sim_{\boldsymbol{m}_{\mathcal{G}}^{\mathrm{pr}}}$, maps each class of items of equivalent quality to a constant price.

Assumption (9) appears reasonable if all the equivalence classes $[o]$ are sufficiently homogeneous, which is the case when the number of price-relevant characteristics is large enough. Similarly to what was already mentioned above, the size and thus the homogeneity of the individual classes is a non-increasing function of the number of price-relevant characteristics, since adding additional characteristics generally leads to more and thus smaller equivalence classes.

We deliberately do not impose any restrictions on the functional form of the hedonic function since, at this stage, we see no reason to do so. Finding an appropriate candidate of a hedonic function that links the vector of price-relevant characteristics to the average price a consumer needs to pay for an item of equivalent quality is a purely statistical issue. Triplett (2004)

convincingly argues that 'imposing some rule for what the hedonic function "should" look like destroys part of the information that market prices convey'. Referring to Rosen (1974), he emphasises that 'the form of the hedonic function is entirely an empirical matter that is determined by the distributions of buyers around the hedonic surface, and not by the form of their utility functions.' Therefore, the hedonic function can neither provide an economic explanation for the behaviour of economic agents nor identify demand or supply. It just describes the statistical relationship between the market price and the quality of a good, no matter how the price and thus the purchasers' valuation of characteristics emerge.

In the framework developed so far, we described the universe of consumer goods available in a market at a certain point in time. By means of distinguishing and price-relevant characteristics, we provided a formal definition of the notion of an elementary aggregate and established a link between the price of a good and its quality. The following section now introduces the time dimension and defines the basic forms of elementary price indices.

3 Elementary price indices

3.1 Elementary aggregates over time

Elementary price indices measure the average price evolution of an elementary aggregate between two time periods. There is a base period 0, serving as reference period, and a current period 1 for which the prices are compared. As time passes, we may observe a certain variation of the items contained in a given elementary aggregate: new items appear on the market and are purchased by consumers, others disappear. This effect is particulary pronounced for products where there is a rapid turnover of differentiated models, such as computers, communication, and multimedia devices.

The duality between elementary aggregates and distinguishing characteristics introduced in the previous section allows for a constant understanding of the nature of an elementary aggregate over time. Instead of fixing the exact content of an aggregate, we fix the distinguishing characteristics and allow new objects to become part of the aggregate if and only if they carry at least all these fixed characteristics. More formally, this leads to the following definition of a current (elementary) aggregate.

Definition 3.1. *Let $\mathcal{T}$ be the set of all time periods considered and let, for any time $t \in \mathcal{T}$, denote $\mathcal{O}^t$ the set of all goods supplied on the market at time t. Let $\mathcal{G} = \mathcal{G}^0$ be any elementary aggregate defined relative to $\mathcal{O}^0$ and let $\mathcal{M}_{\mathcal{G}}$ be its set of distinguishing characteristics.*

*Then, for any time $t \in \mathcal{T}$, the **current aggregate** $\mathcal{G}^t = \mathcal{G}^t(\mathcal{M}_{\mathcal{G}})$ is defined as the elementary aggregate induced by $\mathcal{M}_{\mathcal{G}}$ on $\mathcal{O}^t$, i.e.*

$$\mathcal{G}^t(\mathcal{M}_{\mathcal{G}}) := \bigcap_{m \in \mathcal{M}_{\mathcal{G}}} \mathcal{O}_m^t, \tag{10}$$

where $\mathcal{O}_m^t \subset \mathcal{O}^t$ denotes the domain of characteristic m in period t.

*The **composite elementary aggregate** $\mathcal{G}^{\mathcal{T}}$ **induced by** $\mathcal{G}^0$ is defined by*

$$\mathcal{G}^{\mathcal{T}} := \bigcup_{t \in \mathcal{T}} \mathcal{G}^t(\mathcal{M}_{\mathcal{G}}). \tag{11}$$

7

Obviously, by means of (11), any elementary aggregate $\mathcal{G}$ defined relative to a base period induces a composite elementary aggregate $\mathcal{G}^{\mathcal{T}}$ for any set $\mathcal{T}$ of time periods. Technically, $\mathcal{G}^t(\mathcal{M}_{\mathcal{G}})$ can be empty for certain $t \in \mathcal{T}$.

If we focus on the bilateral comparison of a reference period 1 with a base period 0, one feature of our approach is that it yields with $\mathcal{G}^0 \cap \mathcal{G}^1 \subset \mathcal{G}^{\{0,1\}}$ a straightforward identification of the set of matched items for the two periods. Moreover, the disappearing items are assembled in the difference set $\mathcal{G}^0 \setminus \mathcal{G}^1$ while any new unmatched items are represented by $\mathcal{G}^1 \setminus \mathcal{G}^0$. The impossibility of matching price observations over time being *the* main motivation for applying quality adjustment techniques in price statistics, these difference sets are going to be of particular importance in our framework.

3.2 Concepts of elementary price indices

Relating to what was said in the last paragraph, an elementary price index is typically calculated from two sets of matched price observations: individual goods are sampled from an elementary aggregate and their prices are collected over a succession of time periods. For the bilateral comparison of two time periods 0 and 1 this implies that only those items of the elementary aggregate $\mathcal{G} = \mathcal{G}^0$ are considered which remain available in period 1. Moreover, items newly appearing in period 1 are ignored as their price cannot be matched with a price in the base period. In other words, bilateral comparisons are *a priori* restricted to $\mathcal{G}^0 \cap \mathcal{G}^1$ (and raising this restriction will be the central purpose of quality-adjusted price indices).

There are basically two competing approaches for the specification of an elementary price index. One approach relates the *average price* of the elementary aggregate $\mathcal{G}$ in the current period 1 to its average price in the base period 0, whereas the other takes the *average price ratio* of the individual items as a measure for the change in price level observed from 0 to 1.

If we denote by μ a measure of location defined for any univariate distribution of positive real numbers (i.e. what we called 'average' above), the two approaches just described can be written as

$$EPI^{0:1}(\mathcal{G}) = \frac{\mu\big(\tilde{p}^1(\mathcal{G})\big)}{\mu\big(\tilde{p}^0(\mathcal{G})\big)} \tag{12}$$

and

$$EPI^{0:1}(\mathcal{G}) = \mu\left(\widetilde{p^1/p^0}(\mathcal{G})\right) , \tag{13}$$

respectively.[2] Note that these indices are population indices since they are defined on the whole population of items of a given elementary aggregate. As such, they are latent economic parameters that cannot be observed in practice.

Several elementary price index formulae co-exist in statistical practice which must be considered as functions 'that transform sample survey data into an index number' (Balk, 2005, p. 676) or, in other words, as an estimator or the sample version of a population index.[3] They base upon a sample of objects $o_1, \ldots, o_N \in \mathcal{G}^0 \cap \mathcal{G}^1$ available in both periods for which prices $p_n^t := p^t(o_n)$ were collected. The most widely used formulae are summarised in Table 1. They

---

[2]In these formulae, $\tilde{p}^t(\mathcal{G})$ stands for the distribution of the prices $\{p^t(o)\}$ and $\widetilde{p^1/p^0}(\mathcal{G})$ for the distribution of the price ratios $\{p^1(o)/p^0(o)\}$ of all items $o \in \mathcal{G}^0 \cap \mathcal{G}^1$ with $p^t(o)$ being the observed price of an item o at time t ($t \in \{0, 1\}$).

[3]We further refer to the papers by Dorfman et al. (1999) as well as Silver and Heravi (2007) for some discussion on the fundamental distinction between sample and population indices.

Population index	Sample index	Index type
(12)	$\widehat{EPI}_{\mathrm{D}}^{0:1} = \dfrac{\sum_{n=1}^{N} p_n^1}{\sum_{n=1}^{N} p_n^0}$	Dutot
	$\widehat{EPI}_{\mathrm{J}}^{0:1} = \dfrac{\sqrt[N]{\prod_{n=1}^{N} p_n^1}}{\sqrt[N]{\prod_{n=1}^{N} p_n^0}}$	Jevons
	$\widehat{EPI}_{\mathrm{HD}}^{0:1} = \dfrac{\left(\sum_{n=1}^{N}\left(p_n^1\right)^{-1}\right)^{-1}}{\left(\sum_{n=1}^{N}\left(p_n^0\right)^{-1}\right)^{-1}}$	'Harmonic Dutot'
(13)	$\widehat{EPI}_{\mathrm{C}}^{0:1} = \dfrac{1}{N} \sum_{n=1}^{N} \dfrac{p_n^1}{p_n^0}$	Carli
	$\widehat{EPI}_{\mathrm{J}}^{0:1} = \sqrt[N]{\prod_{n=1}^{N} \dfrac{p_n^1}{p_n^0}}$	Jevons
	$\widehat{EPI}_{\mathrm{HC}}^{0:1} = \left(\dfrac{1}{N} \sum_{n=1}^{N} \left(\dfrac{p_n^1}{p_n^0}\right)^{-1}\right)^{-1}$	'Harmonic Carli'

Table 1: Elementary sample indices

differ in the population index they target and in the way they implement the measure of location μ, namely, e.g., as arithmetic, geometric, or harmonic mean. The Jevons elementary price index formula targets both population indices simultaneously, since (12) and (13) coincide if μ is implemented by the geometric mean.

There has been much debate in the literature on which of these and other alternative elementary sample indices was the most favourable. We do not intend to take this discussion any further but refer to Chapter 20 of ILO et al. (2004) for a detailed and comprehensive overview. It is just worth highlighting that the discussion on what index type to prefer should start at the level of population indices where no sampling issues arise. If there is no apparent economic reason to favour either (12) or (13) and any specific choice of μ, there may be axiomatic and empirical arguments that lead to a preferred definition. We are going to take up this point later when we look at hedonic elementary population indices.

The most important issue of the elementary price indices introduced so far is their inability to cope with a changing universe of items contained in the elementary aggregate. Limiting the set of items to those which are available in all time periods considered is, in general, a far too restrictive strategy. For many specific aggregates, especially for those subject to rapid technological progress, the set of items available in the base period *and* in all current periods will be too small to represent well enough the range of items of the aggregate.

Therefore, the set of items for which prices are available in all time periods considered must be artificially enlarged. This is usually done by assigning ('imputing') estimated prices to those items of the aggregate which are unobservable in certain time periods. Conventional methods for imputing unobserved prices are typically *ad hoc* solutions that attempt to deduce

the price of an item by 'quality-adjusting' the observed price of another item of similar quality (see e.g. ILO et al., 2004, Chap. 7 or Triplett, 2004, Chap. II for a comprehensive overview). However, they lack a sound methodological foundation and may not work consistently for all individual items of an elementary aggregate. A more satisfying solution to the problem of imputing unobservable prices is offered by the hedonic approach.

Based on the hedonic econometric model introduced in Section 2, we are now going to extend the two population indices outlined above such that they incorporate the entire population of a composite elementary aggregate.

4 Hedonic elementary price indices

4.1 The hedonic econometric model revisited

The hedonic econometric model establishes a relationship between characteristics and prices of the items of an elementary aggregate. This relationship is valid for a fixed point in time. In order to explicit its time dependency and to facilitate the comparison of two or more time periods, we propose to reformulate the model as follows:

Hedonic econometric model (in time). *Let $\mathcal{G} = \mathcal{G}^0$ be any elementary aggregate defined relative to the set $\mathcal{O}^0$ of all goods supplied on the market at a base period 0 and let $\mathcal{M}_\mathcal{G}$ be its set of distinguishing characteristics. Let $\mathcal{G}^T$ be the composite elementary aggregate induced by $\mathcal{G}^0$ for a given set of time periods $\mathcal{T}$. There exists a finite set of characteristics*

$$\mathcal{M}_\mathcal{G}^{\mathrm{pr}} = \{m_1, \ldots, m_{K_\mathcal{G}}\} \subset \mathcal{M}_\mathcal{G} \tag{14}$$

and, for each period $t \in \mathcal{T}$, a function $h_\mathcal{G}^t : \mathbb{R}^{K_\mathcal{G}} \longrightarrow \mathbb{R}_{\geq 0}$, such that the price $p^t(o)$ of any item $o \in \mathcal{G}^t$ available at time t can be written as

$$p^t(o) = h_\mathcal{G}^t(\boldsymbol{m}_\mathcal{G}^{\mathrm{pr}}(o)) + \epsilon^t(o) \tag{15}$$

with $\boldsymbol{m}_\mathcal{G}^{\mathrm{pr}}(o) = (m_1(o), \ldots, m_{K_\mathcal{G}}(o))'$. For all $t \in \mathcal{T}$, the residual term $\epsilon^t(o)$ is assumed to be stochastic with conditional expectation

$$\mathrm{E}\big(\epsilon^t(o) \mid \boldsymbol{m}_\mathcal{G}^{\mathrm{pr}}(o)\big) = 0\,. \tag{16}$$

Note that we assume the set $\mathcal{M}_\mathcal{G}^{\mathrm{pr}}$ of price-relevant characteristics to be time-invariant. This condition is less restrictive than it first appears since we only request that $\mathcal{M}_\mathcal{G}^{\mathrm{pr}}$ be finite. It may thus well assemble the whole set of characteristics which prove to be price-relevant in at least one of the time periods considered. If a characteristic is price-irrelevant in a certain period of time, the corresponding hedonic function will just neglect it.

The central aspect of this reformulation of the hedonic econometric model is the postulated time-dependency of the hedonic function $h_\mathcal{G}^t$. Fixed items are sold on the market for different prices at different points in time (this is what price statistics is all about), and the hedonic function mirrors these movements in how the market evaluates the inherent quality of an item. For an elementary aggregate where the hedonic econometric model holds, the quality-adjusted price evolution is thus fully represented by the evolution of the hedonic function over time. An appropriate comparison of the hedonic functions in a base and a current period may thus be seen as an implementation of an elementary price index measuring pure price change.

Before we proceed to the formulation of hedonic elementary population indices based on the idea just described, we propose simplifying the notation by 'randomising' the hedonic econometric model introduced above. Imagine a random draw from all the items of an elementary aggregate $\mathcal{G}^t$ at time t and denote by $\boldsymbol{M}^t$ the random vector of price-relevant characteristics and P^t the random variable representing the price of the drawn item.[4] By the hedonic econometric model, the relationship between $\boldsymbol{M}^t$ and P^t is given by

$$P^t = h_{\mathcal{G}}^t(\boldsymbol{M}^t) + \epsilon^t \tag{17}$$

where the random error ϵ^t has $\mathrm{E}\epsilon^t = 0$ for all $t \in \mathcal{T}$ and is assumed to be independent of $\boldsymbol{M}^t$. Within this additive error model, the hedonic function $h_{\mathcal{G}}^t$ therefore is exactly the conditional mean

$$h_{\mathcal{G}}^t(\boldsymbol{m}) = \mathrm{E}(P^t \,|\, \boldsymbol{M}^t = \boldsymbol{m}), \tag{18}$$

and the conditional distribution $\mathbb{P}(P^t \,|\, \boldsymbol{M}^t)$ depends on $\boldsymbol{M}^t$ only through $h_{\mathcal{G}}^t$.

4.2 Simple hedonic elementary population indices

In Section 3 we identified two approaches for defining elementary population indices either as the ratio of some average prices (12) or as some average of the price ratios (13). Both of these population indices were defined on the restricted set $\mathcal{G}^0 \cap \mathcal{G}^1$ of items available in both the base and the current period. With this restriction, it was ensured that the compared prices belonged to identical goods and thus that the qualities of the items compared were equal. Consequently, the measured price evolution was not subject to any bias due to quality change.

Hedonic elementary price indices adhere to the paradigm of fixed reference qualities, but they do not rely on a fixed set of items for which prices need to be available in both time periods. Once the hedonic function for a certain time period is determined, it is able to deliver imputed prices for virtually any vector of price-relevant characteristics and as such for any item quality. The idea of hedonic elementary price indices is now to fix the reference quality of an elementary aggregate through vectors of price-relevant characteristics. With the help of the hedonic functions, the reference qualities are mapped to corresponding prices that can ultimately be compared using one of the two elementary price index approaches introduced above.

The simplest form of a hedonic elementary price index takes just one vector $\boldsymbol{\mu}^*$ of price-relevant characteristics as reference quality. Irrelevant of the type of elementary population index used, this yields the index formula

$$HEPI^{0:1}(\mathcal{G}) = \frac{h_{\mathcal{G}}^1(\boldsymbol{\mu}^*)}{h_{\mathcal{G}}^0(\boldsymbol{\mu}^*)} = \frac{\mathrm{E}(P^1 \,|\, \boldsymbol{M}^1 = \boldsymbol{\mu}^*)}{\mathrm{E}(P^0 \,|\, \boldsymbol{M}^0 = \boldsymbol{\mu}^*)} \tag{19}$$

which we call *simple hedonic elementary population index*. It relates the imputed price of the reference quality $\boldsymbol{\mu}^*$ at time 1 to its imputed price at time 0.[5]

---

[4]The distribution of the random variable P^t corresponds to the distribution of prices denoted by $\tilde{p}^t(\mathcal{G})$ in Section 3, and the expectation $\mathrm{E}(P^t)$ is one possible implementation of $\mu(\tilde{p}^t(\mathcal{G}))$.

[5]Technically, the index (19) is only well-defined if $\boldsymbol{\mu}^*$ lies in $\boldsymbol{m}_{\mathcal{G}}^{\mathrm{pr}}(\mathcal{G}^0) \cap \boldsymbol{m}_{\mathcal{G}}^{\mathrm{pr}}(\mathcal{G}^1)$, i.e. in the domains of both $h_{\mathcal{G}}^0$ and $h_{\mathcal{G}}^1$. If this is not the case, a minimal requirement is that both hedonic functions can be extended to a domain including $\boldsymbol{\mu}^*$. This is normally not a problem in practice if a regression approach is chosen that allows for reasonable out-of-sample prediction.

The open question here is how $\boldsymbol{\mu}^*$ should be defined. The most obvious approach is to take some mean vector of price-relevant characteristics of the items available at the base or the current period. Formally, this gives us either $\boldsymbol{\mu}^* = \mathrm{E}\boldsymbol{M}^0$ or $\boldsymbol{\mu}^* = \mathrm{E}\boldsymbol{M}^1$ and with (19) corresponding implementations of the simple hedonic elementary population index.

The disadvantage of both of these implementations is that they asymmetrically favour the quality spectrum of the elementary aggregate at either the base or the current period. We therefore propose to work with a generalised reference quality distribution, represented by a random vector $\boldsymbol{M}$. The most natural choice for this reference distribution would probably be a mixture of $\boldsymbol{M}^0$ and $\boldsymbol{M}^1$, i.e.

$$\mathbb{P}_{\boldsymbol{M}} = g\,\mathbb{P}_{\boldsymbol{M}^0} + (1-g)\,\mathbb{P}_{\boldsymbol{M}^1}\,, \tag{20}$$

with $\mathbb{P}_{\boldsymbol{M}}$, $\mathbb{P}_{\boldsymbol{M}^0}$ and $\mathbb{P}_{\boldsymbol{M}^1}$ being the probability measures of $\boldsymbol{M}$, $\boldsymbol{M}^0$ and $\boldsymbol{M}^1$, respectively, and $g \in [0,1]$. If we set $\boldsymbol{\mu}^* = \mathrm{E}\boldsymbol{M}$, we get with $g = 0$ or $g = 1$ the two implementations of simple hedonic elementary population indices already introduced above and with $g = 1/2$ a sensible candidate of an index that symmetrically incorporates the quality spectrum in both the base and the current period.

For the special case of parametric hedonic functions, Brachinger (2002) introduced simple hedonic elementary population indices of the type (19) under the name of 'true hedonic price indices'. He distinguished explicitly the implementations obtained when $\boldsymbol{\mu}^* = \mathrm{E}\boldsymbol{M}$ with $g = 0$, 1, or $1/2$. Referring to their orientation towards either the base, the current, or both periods simultaneously, these implementations were called 'true hedonic Laspeyres price index', 'true hedonic Paasche price index', and 'true hedonic adjacent periods price index', respectively.

4.3 Full hedonic elementary population indices

Simple hedonic elementary population indices evaluate the 'distance' of the two hedonic functions in the base and in the current period at just one single quality point $\boldsymbol{\mu}^*$. Although this is certainly a valid practice, there are ways of better exploiting the full spectrum of the reference quality distribution and to obtain a more representative index value.

One such way is to transform the whole reference quality distribution with the help of the two hedonic functions and to compare the resulting price distributions using the approaches described in Section 3.2. If we take the expectation as measure of location μ, the population indices (12) and (13) translate into *full hedonic elementary population indices* defined by

$$HEPI^{0:1}(\mathcal{G}) = \frac{\mathrm{E}h_{\mathcal{G}}^1(\boldsymbol{M})}{\mathrm{E}h_{\mathcal{G}}^0(\boldsymbol{M})} \tag{21}$$

and

$$HEPI^{0:1}(\mathcal{G}) = \mathrm{E}\left[\frac{h_{\mathcal{G}}^1(\boldsymbol{M})}{h_{\mathcal{G}}^0(\boldsymbol{M})}\right]\,. \tag{22}$$

In both cases, the expectations are built over the whole range of $\boldsymbol{M}$ and cover thus the reference quality distribution as a whole.[6]

We see that the distribution of $\boldsymbol{M}$ in principle does not need to be related to either $\boldsymbol{M}^0$ or $\boldsymbol{M}^1$, although a mixture like (20) is probably still the most reasonable choice. The minimum

---

[6]Diewert et al. (2008) showed that for the widely used special case of log-linear hedonic functions and under certain assumptions for the reference quality distribution used (which are satisfied when $g = 0$ or $g = 1$), both the simple and the full hedonic elementary indices are equivalent.

assumption to be made is that the range of $\boldsymbol{M}$ is contained in the domain of both $h_{\mathcal{G}}^0$ and $h_{\mathcal{G}}^1$. Note that, following (18), we have

$$\mathrm{E}h_{\mathcal{G}}^t(\boldsymbol{M}) = \mathrm{E}_{\boldsymbol{M}}(\mathrm{E}_{P^t \mid \boldsymbol{M}^t}(P^t \mid \boldsymbol{M})) \tag{23}$$

$$= \int_{\mathbb{R}^{K_{\mathcal{G}}}} \left[\int_{\mathbb{R}} p \, \mathrm{d}\mathbb{P}_{P^t \mid \boldsymbol{M}^t}(p \mid \boldsymbol{m}) \right] \mathrm{d}\mathbb{P}_{\boldsymbol{M}}(\boldsymbol{m}) \, .$$

for $t \in \{0, 1\}$. Here, $\mathbb{P}_{P^t \mid \boldsymbol{M}^t}$ stands for the probability measure of the conditional distribution of P^t given $\boldsymbol{M}^t$, and $\mathrm{E}_{P^t \mid \boldsymbol{M}^t}$ is the expectation with respect to this probability measure. Moreover, $\mathbb{P}_{\boldsymbol{M}}$ is the probability measure respective to the distribution of $\boldsymbol{M}$, and $\mathrm{E}_{\boldsymbol{M}}$ is its expectation. If one considers continuous random variables and vectors, equation (23) can be rewritten as

$$\mathrm{E}h_{\mathcal{G}}^t(\boldsymbol{M}) = \int_{\mathbb{R}^{K_{\mathcal{G}}}} \left[\int_{\mathbb{R}} p \, \frac{f_{(P^t, \boldsymbol{M}^t)}(p, \boldsymbol{m})}{f_{\boldsymbol{M}^t}(\boldsymbol{m})} \, \mathrm{d}p \right] f_{\boldsymbol{M}}(\boldsymbol{m}) \, \mathrm{d}\boldsymbol{m} \tag{24}$$

with $f_{(P^t, \boldsymbol{M}^t)}$ being the common probability density of P^t and $\boldsymbol{M}^t$, $f_{\boldsymbol{M}^t}$ the marginal density of $\boldsymbol{M}^t$ and, finally, $f_{\boldsymbol{M}}$ the density of $\boldsymbol{M}$. It can be seen that for this equation to be well-defined, the support of $f_{\boldsymbol{M}}$ needs to be contained in the support of $f_{\boldsymbol{M}^t}$ for $t \in \{0, 1\}$. In other words, for each vector $\boldsymbol{m} \in \mathbb{R}^{K_{\mathcal{G}}}$ with $f_{\boldsymbol{M}^t}(\boldsymbol{m}) = 0$, it is necessary that $f_{\boldsymbol{M}}(\boldsymbol{m}) = 0$. This has to be taken into consideration when the reference quality $\boldsymbol{M}$ is chosen. In particular, $\mathbb{P}_{\boldsymbol{M}}$ must not attribute a positive probability to any set of characteristics vectors that does not have a positive probability with respect to $\mathbb{P}_{\boldsymbol{M}^0}$ and $\mathbb{P}_{\boldsymbol{M}^1}$ as well, i.e. within the populations available in both the base and current period.

In practice, therefore, it is even useful to assume that $\mathbb{P}_{\boldsymbol{M}^0}$ and $\mathbb{P}_{\boldsymbol{M}^1}$ attribute a positive probability to any non-discrete set of vectors in the characteristics space, i.e. the cartesian product of the ranges of all price-relevant characteristics. This ensures that out-of-sample-prediction is possible and that there are no formal restrictions on the distribution of reference characteristics.

4.4 Universal formulae for hedonic elementary price indices

In the last two sections, we introduced alternative definitions of a hedonic elementary population index. There the expectation operator was used as a special choice of a measure of location. There is, however, no a priori reason for this restriction. A natural generalisation of this approach results if we admit transformations of the price distributions. The expectation of the transformed price distribution characterises the location of this distribution. A measure of location of the original price distribution then results from backtransforming the expectation of the transformed price distribution.

Based on these reflections, the full hedonic elementary population indices (21) and (22) can be generalised to

$$HEPI^{0:1}(\mathcal{G}) = \frac{\varphi^{-1}(\mathrm{E}\varphi(h_{\mathcal{G}}^1(\boldsymbol{M})))}{\varphi^{-1}(\mathrm{E}\varphi(h_{\mathcal{G}}^0(\boldsymbol{M})))} \tag{25}$$

and

$$HEPI^{0:1}(\mathcal{G}) = \varphi^{-1}\left(\mathrm{E}\left[\varphi\left(\frac{h_{\mathcal{G}}^1(\boldsymbol{M})}{h_{\mathcal{G}}^0(\boldsymbol{M})} \right) \right] \right) \tag{26}$$

where φ is a continuous and injective function that maps a connected subset of $\mathbb{R}$ to $\mathbb{R}$ and φ^{-1} is its inverse.

With respect to the usual elementary price index formulae, three particular φ-functions play an important role. These are the identity, the hyperbolic transformation $\varphi(x) = x^{-1}$ as well as the natural logarithm $\varphi(x) = \ln x$. We will see below that depending on the choice of φ among these alternatives, the well-known hedonic elementary sample indices can be derived. Note that both definitions, (25) and (26), coincide if $\varphi(x) = \ln x$. This is due to the linearity of the expectation and the properties of the natural logarithm.[7]

We propose with (25) and (26) two universal prototypes of hedonic elementary population indices that leave, however, some degrees of freedom for the choice of φ and of the reference distribution of $\boldsymbol{M}$. We argued already that the latter is reasonably defined as a (symmetric) mixture of the base and the current period characteristics. However, there is no evident argumentation that favours either choice of φ except for the coincidence of both formulae if $\varphi(x) = \ln x$. Beer (2007b) discussed this question in the light of the well-known axiomatic approach to statistical price indices (Eichhorn, 1978; Eichhorn and Voeller, 1976) and proved that (25) is slightly preferable to (26) since it satisfies all proposed index axioms if $\varphi(\lambda x) = \varphi(\lambda) + \varphi(x)$ or $\varphi(\lambda x) = \varphi(\lambda)\varphi(x)$ for all $\lambda, x \in \mathbb{R}$. However, this latter condition holds for all three φ-functions proposed above, so the axiomatic approach does not seem to be sufficient for choosing one 'best' universal hedonic elementary population index. There are obviously further arguments that need to be considered.

5 Estimators of hedonic elementary price indices

5.1 Hedonic imputation indices

So far, we have consistently navigated on the abstract level of index definitions and population indices which are, as we repeatedly stressed, economic parameters that cannot directly be observed and eventually need to be estimated. Assume that $\hat{h}_{\mathcal{G}}^t$ ($t \in \{0, 1\}$) are estimators of the hedonic functions $h_{\mathcal{G}}^t$ based on regressions of item characteristics to prices in both periods. Starting from an i.i.d. sample of reference characteristics vectors $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_N$ where $\boldsymbol{M}_n \overset{\text{i.i.d.}}{\sim} \boldsymbol{M}$ for all $n \in \{1, \ldots, N\}$, sample versions of the universal population indices defined by (25) and (26) are given by

$$\widehat{HEPI}^{0:1}(\mathcal{G}) = \frac{\varphi^{-1}\left(\frac{1}{N}\sum_{n=1}^{N}\varphi(\hat{h}_{\mathcal{G}}^1(\boldsymbol{M}_n))\right)}{\varphi^{-1}\left(\frac{1}{N}\sum_{n=1}^{N}\varphi(\hat{h}_{\mathcal{G}}^0(\boldsymbol{M}_n))\right)} \tag{27}$$

Formula	Transformation	Sample index	Index type
(27)	$\varphi(x) = x$	$\widehat{HEPI}_{\mathrm{D}}^{0:1} = \dfrac{\sum_{n=1}^{N} \hat{h}^1(\boldsymbol{M}_n)}{\sum_{n=1}^{N} \hat{h}^0(\boldsymbol{M}_n)}$	Dutot
	$\varphi(x) = \ln x$	$\widehat{HEPI}_{\mathrm{J}}^{0:1} = \dfrac{\sqrt[N]{\prod_{n=1}^{N} \hat{h}^1(\boldsymbol{M}_n)}}{\sqrt[N]{\prod_{n=1}^{N} \hat{h}^0(\boldsymbol{M}_n)}}$	Jevons
	$\varphi(x) = x^{-1}$	$\widehat{HEPI}_{\mathrm{HD}}^{0:1} = \dfrac{\left(\sum_{n=1}^{N}\left(\hat{h}^1(\boldsymbol{M}_n)\right)^{-1}\right)^{-1}}{\left(\sum_{n=1}^{N}\left(\hat{h}^0(\boldsymbol{M}_n)\right)^{-1}\right)^{-1}}$	'Harmonic Dutot'
(28)	$\varphi(x) = x$	$\widehat{HEPI}_{\mathrm{C}}^{0:1} = \dfrac{1}{N}\sum_{n=1}^{N} \dfrac{\hat{h}^1(\boldsymbol{M}_n)}{\hat{h}^0(\boldsymbol{M}_n)}$	Carli
	$\varphi(x) = \ln x$	$\widehat{HEPI}_{\mathrm{J}}^{0:1} = \sqrt[N]{\prod_{n=1}^{N} \dfrac{\hat{h}^1(\boldsymbol{M}_n)}{\hat{h}^0(\boldsymbol{M}_n)}}$	Jevons
	$\varphi(x) = x^{-1}$	$\widehat{HEPI}_{\mathrm{HC}}^{0:1} = \left(\dfrac{1}{N}\sum_{n=1}^{N}\left(\dfrac{\hat{h}^1(\boldsymbol{M}_n)}{\hat{h}^0(\boldsymbol{M}_n)}\right)^{-1}\right)^{-1}$	'Harmonic Carli'

Table 2: Hedonic elementary sample indices

and

$$\widehat{HEPI}^{0:1}(\mathcal{G}) = \varphi^{-1}\left(\frac{1}{N}\sum_{n=1}^{N}\varphi\left(\frac{\hat{h}_{\mathcal{G}}^1(\boldsymbol{M}_n)}{\hat{h}_{\mathcal{G}}^0(\boldsymbol{M}_n)}\right)\right) , \qquad (28)$$

respectively.

From these two formulae, by choosing φ among the alternatives mentioned above, we get the five hedonic elementary sample indices displayed in Table 2 which are hedonic counterparts to the elementary sample indices summarised in Table 1. We recognise that the elementary index formulae most widely used in practice (see e.g. ILO et al., 2004, paras. 20.38–45) prove to be estimators of the population indices (25) and (26). Among these are the indices attributed to Dutot, Jevons, and Carli, and the one that is called 'Harmonic Carli' here. Moreover, we find a 'Harmonic Dutot' sample index which to our knowledge does not appear in the literature. Note that when using $\varphi(x) = \ln x$ both general sample indices (27) and (28) lead to the Jevons elementary price index formula.

Once the φ-function is fixed and the distribution of $\boldsymbol{M}$ is defined, the single remaining influence factor that determines the nature and the performance of the hedonic elementary sample indices (27) and (28) and thus eventually the statistical quality of the actual index estimates is the regression approach used to estimate the hedonic functions. As we already discussed in Section 2, estimating the relationship between characteristics and price is a purely statistical issue with no *a priori* restriction on the functional form or regression approach to choose. As Triplett (2004, p. 186) stated, 'Any empirical form that fits the data is consistent with the

theory.' So the entire repertoire of regression analysis can be applied to find an approach that best fits the data and delivers price predictions with the highest possible precision. The only point to remember is that estimated hedonic functions are normally used to perform some out-of-sample predictions where they should still provide plausible estimates.

In practice, the prevalent regression approaches for estimating hedonic functions are linear, semi-log and double-log models which perform well for many data sets. Curry et al. (2001) were among the few authors who argued for a more flexible functional form, although the neural network approach they tested at the example of TVs did not show to be favourable to the linear or semi-log models.[8] Beer (2007b) investigated the use of conventional models compared to a partial least squares approach in an empirical study on used cars data. There, the winning model in terms of lowest bootstrap aggregate prediction error was an adaptive semi-log approach where individual regressions with automated variable selection and outlier detection were carried out and used for prediction for each of the car models in the sample. Interestingly, the study also showed that the Jevons hedonic elementary sample index was least sensitive to the (mis-)specification of the hedonic function. Further research is needed to consolidate this finding.

It seems ambitious to analyse the statistical qualities of HEPI estimators given the potential complexity of the hedonic functions and the generally unknown form of the reference characteristics distribution. However, appropriate bootstrap approaches have been proposed for estimating confidence intervals for hedonic elementary price indices (see Beer, 2007a,b for details). Conditioned on the functional form of the hedonic regression, they deliver an insight on how accurately hedonic elementary sample indices estimate the corresponding population indices. It was shown empirically in the cited reference that confidence intervals were consistently shorter for the Jevons than for the Dutot formula, suggesting that the index (25) could be estimated more accurately if $\varphi(x) = \ln x$ than if $\varphi(x) = x$ and thus providing even stronger evidence in favour of the Jevons formula. Again, the theoretical grounds of this finding remain a topic for further research.

5.2 Time dummy hedonic indices

We shall close this section by a comment on the *time dummy variable method*, which is a widely used alternative to the hedonic imputation indices discussed above (see e.g. Griliches, 1971, p. 59, Silver and Heravi, 2003, pp. 280–1, Triplett, 2004, p. 48–55, or Diewert et al., 2008). There, the price and characteristics data of both the base and the current period are pooled and the price-relevant characteristics $\boldsymbol{m} = (m_1, \ldots, m_K)'$ are supplemented by a time dummy variable t. Then a joint parametric hedonic function $h_{\mathcal{G}}^{\{0,1\}}$ is estimated on the basis of the pooled sample. From the estimated hedonic function $\hat{h}_{\mathcal{G}}^{\{0,1\}}$ two period-specific hedonic functions $\hat{h}_{\mathcal{G}}^t$ $(t = 0, 1)$ are easily recovered through

$$\hat{h}_{\mathcal{G}}^t(\boldsymbol{m}) := \hat{h}_{\mathcal{G}}^{\{0,1\}}(\boldsymbol{m}, t). \tag{29}$$

These can be plugged into all hedonic elementary sample index formulae presented above.

An interesting situation emerges if we adopt the semi-log functional form for estimating the

---

[8]The failure of their neural network approach was particularly due to its instability on out-of-sample predictions.

hedonic function. Then the relevant regression equation is given by

$$\ln P = \beta_0 + \delta\, t + \sum_{k=1}^{K} \beta_k M_k + \epsilon \tag{30}$$

and the estimated hedonic functions $\hat{h}_{\mathcal{G}}^t$ can be written as

$$\hat{h}_{\mathcal{G}}^t(\boldsymbol{m}) = \exp\left(\hat{\beta}_0 + \hat{\delta}\, t + \sum_{k=1}^{K} \hat{\beta}_k\, m_k\right). \tag{31}$$

with $\hat{\beta}_0, \ldots, \hat{\beta}_k$ and $\hat{\delta}$ being the OLS estimates of the corresponding coefficients in (30). Obviously,

$$\hat{h}_{\mathcal{G}}^1(\boldsymbol{m}) = \exp\ \hat{\delta} \times \hat{h}_{\mathcal{G}}^0(\boldsymbol{m}) \tag{32}$$

for all $\boldsymbol{m}$, and all of the sample index formulae listed in Table 2 reduce to $\widehat{HEPI}^{0:1} = \exp\ \hat{\delta}$. They are thus completely independent of the reference quality distribution used.

Although the property of independence just described sounds appealing, we agree with, e.g., Diewert et al. (2008) who argue in favour of hedonic imputation indices. In contrast to the time dummy approach, they have the advantage of not imposing any constraint on the functional form and eventually on the parameters of the hedonic functions. We are convinced that the flexibility of the functional form is important and therefore that any technical restrictions should be avoided.

6 Summary

We started the present piece of work with a formal definition of elementary aggregates. From the outset, much emphasis was put on the duality between elementary aggregates and their distinguishing characteristics. The latter played a central role when we translated the hedonic hypothesis known from the literature into a hedonic econometric model.

After discussing the two fundamental concepts of elementary price indices, we defined a list of hedonic elementary population indices reaching from simple indices where the entire quality range of an object was represented by a single vector of price-relevant characteristics to two universal index formulae showing much flexibility in how the price distributions in the base and current period are compared. As population indices are unobservable economic parameters, there is a need for sample indices acting as appropriate estimators of these parameters. We were able to show that most of the index formulae used in practice could be derived naturally within the proposed framework.

Neither of the universal formulae of hedonic elementary population indices proposed in this paper is completely determined. So the user needs make some further decisions in order to obtain a concrete target population index and eventually a corresponding sample index formula for practical applications. Based on axiomatic and empirical reflections, however, we argued that the most attractive candidate of a hedonic elementary price index for an elementary aggregate $\mathcal{G}$ and for any given reference quality distribution $\mathbb{P}_{\boldsymbol{M}}$ was

$$HEPI^{0:1}(\mathcal{G}) = \exp\left(\mathrm{E}\left[\ln\left(\frac{h_{\mathcal{G}}^1(\boldsymbol{M})}{h_{\mathcal{G}}^0(\boldsymbol{M})}\right)\right]\right). \tag{33}$$

An estimator of this index is given by the Jevons formula

$$\widehat{HEPI}_{\mathrm{J}}^{0:1} = \sqrt[N]{\prod_{n=1}^{N} \frac{\hat{h}^1(\boldsymbol{M}_n)}{\hat{h}^0(\boldsymbol{M}_n)}} \tag{34}$$

based on estimates $\hat{h}^0$ and $\hat{h}^t$ of the hedonic functions and with $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_N$ sampled symmetrically from the price-relevant characteristics in both the base and the current period. Moreover, candidate estimates of the hedonic functions should be evaluated based on their predicting abilities rather than on their conformity with any given functional form.

References

Balk, B. M., 2005. Price indexes for elementary aggregates: The sampling approach. Journal of Official Statistics 21 (4), 675–99.

Beer, M., 2007a. Bootstrapping a hedonic price index: Experience from used cars data. AStA Advances in Statistical Analysis 91 (1), 77–92.

Beer, M., 2007b. Hedonic elementary price indices: Axiomatic foundation and estimation techniques. Ph.D. thesis, University of Fribourg Switzerland.
URL http://www.michael.beer.name/phdthesis

Brachinger, H. W., 2002. Statistical theory of hedonic price indices. DQE Working Paper 1, Department of Quantitative Economics, University of Fribourg Switzerland.
URL http://ideas.repec.org/p/fri/dqewps/wp0001.html

Court, A. T., 1939. Hedonic price indexes with automotive examples. In: The Dynamics of Automobile Demand. General Motors Corporation, New York, pp. 99–117.

Curry, B., Morgan, P., Silver, M., 2001. Hedonic regressions: Mis-specification and neural networks. Applied Economics 33 (5), 659–671.

Dickie, M., Delorme, Jr, C. D., Humphreys, J. M., 1997. Hedonic prices, goods-specific effects and functional form: inferences from cross-section time series data. Applied Economics 29 (2), 239–49.

Diewert, E., Heravi, S., Silver, M., 2008. Hedonic imputation versus time dummy hedonic indexes. Working Paper 14018, National Bureau of Economic Research, Cambridge, MA.
URL http://www.nber.org/papers/w14018.pdf

Dorfman, A. H., Leaver, S., Lent, J., 1999. Some observations on price index estimators. Statistical Policy Working Paper 29, Bureau of Labor Statistics.
URL http://www.bls.gov/ore/pdf/st990080.pdf

Eatwell, J., Milgate, M., Newman, P. (Eds.), 1987. The New Palgrave: A Dictionary of Economics. The Macmillan Press Limited, London.

Eichhorn, W., 1978. What is an economic index? an attempt of an answer. In: Eichhorn, W., Henn, R., Opitz, O., Shephard, R. W. (Eds.), Theory and Applications of Economic Indices. Physica-Verlag, Würzburg, pp. 3–42.

Eichhorn, W., Voeller, J., 1976. Theory of the Price Index. Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, Berlin.

Griliches, Z., 1971. Hedonic price indexes for automobiles: An econometric analysis of quality change. In: Griliches, Z. (Ed.), Price Indexes and Quality Change. Harvard University Press, Cambridge, pp. 55–87.

ILO, IMF, OECD, UNECE, Eurostat, The World Bank (Eds.), 2004. Consumer Price Index Manual: Theory and Practice. International Labour Office, Geneva.

ILO, IMF, OECD, UNECE, Eurostat, The World Bank (Eds.), 2009. Consumer Price Index Manual: Theory and Practice. International Labour Office, Geneva, partially revised version published online [accessed 10 April 2009].
URL http://www.ilo.org/public/english/bureau/stat/guides/cpi/

Lancaster, K. J., 1966. A new approach to consumer theory. The Journal of Political Economy 74 (2), 132–57.

Lancaster, K. J., 1971. Consumer Demand: A New Approach. No. 5 in Columbia Studies in Economics. Columbia University Press, New York.

Milgate, M., 1987. Goods and commodities. In: Eatwell et al. (1987), pp. 546–9.

Rosen, S., 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. The Journal of Political Economy 82 (1), 34–55.

Silver, M., Heravi, S., 2003. The measurement of quality-adjusted price changes. In: Feenstra, R. C., Shapiro, M. D. (Eds.), Scanner Data and Price Indexes. Vol. 64 of Studies in Income and Wealth. The University of Chicago Press, Chicago, pp. 277–316.

Silver, M., Heravi, S., 2007. Why elementary price index number formulas differ: Evidence on price dispersion. Journal of Econometrics 140 (2), 874–83.

Triplett, J. E., 1987. Hedonic functions and hedonic indexes. In: Eatwell et al. (1987), pp. 630–4.

Triplett, J. E., 2004. Handbook on hedonic indexes and quality adjustments in price indexes: Special application to information technology products. Working Paper 2004/9, OECD Directorate for Science, Technology and Industry, Paris.
URL http://dx.doi.org/10.1787/643587187107

United Nations (Ed.), 1993. System of National Accounts 1993. United Nations, Ch. XVI. Price and Volume Measures.
URL http://unstats.un.org/unsd/sna1993/