On Curing the CPI's Substitution and New Goods Bias

Bert M. Balk*
Department of Statistical Methods
Statistics Netherlands
P. O. Box 4000, 2270 JM Voorburg, The Netherlands
E-mail `bblk@cbs.nl`

Draft, June 15, 1999

1 Introduction and executive summary

The Boskin *et al.* (1996) report criticized the U. S. Consumer Price Index for its bias, the wellknown components being (upper and lower level) substitution bias, new goods bias, quality change bias, and new outlets bias. Since this criticism easily can be extended to any other country's CPI, the report at the same time served to set the research agenda for the years to come.

In a recent article, Shapiro and Wilcox (1997) proposed a method for "picking the low-hanging fruit" of the CPI bias. To cure for the (upper level) substitution bias, they suggested to use instead of the Laspeyres price index a base-period-expenditure-share-weighted generalized mean price index. Using obvious but later on to be explained notation, this index can be written as

$$\left[\sum_{n\in I^0} s_n^0 (p_n^1/p_n^0)^{1-\sigma}\right]^{1/(1-\sigma)} . \qquad (1)$$

* The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands.

Unlike superlative indices such as those of Fisher and Törnqvist, this index can be calculated in real time, provided that the value of the parameter σ is known.

A generalized mean price index can be conceived as a cost of living index corresponding to an underlying (homothetic) preference structure which is characterized by the fact that for any pair of commodities the elasticity of substitution is the same (hence the name 'CES': constant elasticity of substitution). The parameter σ is then simply equal to this elasticity. Thus, employing (1) with $\sigma > 0$ is a first step in the direction of accounting for substitution behavior.

Also from another perspective our attention was drawn to the CES preference structure. Feenstra (1994) and Feenstra and Markusen (1994) proposed, albeit in a different context (import demand and growth accounting respectively), a modified version of the CES cost function. This version allows for varying ranges of commodities through time. See also Feenstra and Shiells (1997). So perhaps we have here a cure for what is called the new goods bias.

In this paper I review the (properties of the) cost of living index associated with a modified CES preference structure. The paper can be summarized as follows.

Section 2 is concerned with the necessary definitions and notation. It is assumed that the consumer's preferences are represented by a CES type unit cost function, defined on commodity sets that are variable but overlapping through time. In section 3 it is then shown that the cost of living index can be expressed in at least five different ways as the product of a conventional price index, defined with respect to the set of ongoing commodities, and a factor depending on the change of the range of commodities.

Each of these expressions, however, contains the parameter σ. In section 4 I propose some simple methods for estimating the elasticity of substitution. These methods do not require the estimation of a complete demand system.

A simple theoretical consideration shows that the value of σ must be larger than 1, which seems to conflict with empirical evidence. This evidence, however, pertains to commodity groups that are assumed to be stable through time. Thus, the potential conflict between theory and empirics can be resolved by assuming a two-level structure in the consumer's preferences, consisting of (at the upper level) unchanging commodity groups and (at the lower level) changing ranges of commodities within these groups. This is the topic explored in section 5. In particular it is shown that the cost of living index can be represented in at least $5 \times 5 \times G$ ways, where G is the number of commodity groups. There are now $G+1$ parameters involved, but these can be estimated in a relatively easy way.

2 The CES preference structure with variable sets of commodities

Let the sets of commodities be variable (but overlapping) through time, and define $I^t \subset \{1, ..., N\}$ as the set of commodities which are available in period t. We will denote by x_n^t and p_n^t respectively the non-negative quantity consumed and the corresponding positive price of commodity $n \in I^t$ at period t; x^t will denote the vector of x_n^t and p^t will denote the vector of p_n^t; x and p will denote generic quantity and price vectors of appropriate size.

It is assumed that the preference structure of the representative consumer exhibits homotheticity, and that the unit cost (expenditure) function is of the CES type. Thus, it is assumed that the period t cost function is

$$C(p, u \mid I^t) = u(\sum_{n \in I^t} b_n p_n^{1-\sigma})^{1/(1-\sigma)}, \tag{2}$$

where $u > 0$ is a utility level, $\sigma \geq 0, \sigma \neq 1$ is the (demand) elasticity of substitution and $b_n > 0$ $(n \in I^t)$ are quality or taste parameters.[1] Notice that it is assumed that σ as well as b_n $(n \in I^t)$ be time-invariant. By Shephard's Lemma, the optimal expenditure shares are

$$s_n(p \mid I^t) = \frac{b_n p_n^{1-\sigma}}{\sum_{n \in I^t} b_n p_n^{1-\sigma}} \; (n \in I^t). \tag{3}$$

Due to homotheticity, these expenditure shares are independent of the utility level u.

The utility function that is dual to the cost function (2) is given by

$$U(x \mid I^t) = (\sum_{n \in I^t} b_n^{1/\sigma} x_n^{(\sigma-1)/\sigma})^{\sigma/(\sigma-1)}. \tag{4}$$

[1] For the optimal quantities $x_n^t(p, u \mid I^t)$ $(n \in I^t)$ we find that $d\ln(x_n(p, u \mid I^t)/x_{n'}(p, u \mid I^t))/d\ln(p_n/p_{n'}) = -\sigma$ $(n \neq n')$, which must be non-positive. The Allen elasticity of substitution between n and n' $(n \neq n')$ is equal to σ. The Cobb-Douglas unit cost function $\prod_{n \in I^t} p_n^{b_n}$ is obtained as limiting case for $\sigma \to 1$. In turn, (2) is a specific case of the more general (indirect addilog) cost function which is implicitly defined by the equation $u \sum_{n \in I^t} b_n (p_n / C(p, u \mid I^t))^{1-\sigma_n} = 1$.

This function is said to go all the way back to Bergson (1936). Against the backdrop of production theory, its properties were discussed by Uzawa (1962) and McFadden (1963). A generalization beyond homotheticity was proposed by Gamaletsos (1973), Hasenkamp (1978), (1980), and Hasenkamp and Koo (1983). This generalization consists in replacing the quantities x_n by excess quantities $x_n - \gamma_n$, where $\gamma_n > 0$.

We consider two periods, a base period (denoted by 0) and a comparison period (denoted by 1). Our basic assumption is that the actual expenditure shares in these periods are equal to the optimal shares, that is

$$s_n^t \equiv \frac{p_n^t x_n^t}{\sum_{n \in I^t} p_n^t x_n^t} = s_n(p^t \mid I^t) \; (n \in I^t; t = 0, 1). \tag{5}$$

We will use the following definitions. Let $I^{01} \equiv I^0 \cap I^1$ be the set of all commodities common to the base period and the comparison period. It is assumed that $I^{01} \neq \emptyset$. For $n \in I^{01}$ and $t = 0, 1$ we define

$$s_n^{t*} \equiv \frac{p_n^t x_n^t}{\sum_{n \in I^{01}} p_n^t x_n^t} = \frac{s_n^t}{\sum_{n \in I^{01}} s_n^t}, \tag{6}$$

which are the period t expenditure shares relative to the set of commodities common to both periods. We also define

$$\lambda^t \equiv \frac{\sum_{n \in I^{01}} p_n^t x_n^t}{\sum_{n \in I^t} p_n^t x_n^t} = \sum_{n \in I^{01}} s_n^t, \tag{7}$$

which is the fraction of the period t expenditure attributable to the commodities which are common to both periods relative to the total expenditure of period t. By combining the two definitions, we obtain the following relationship:

$$s_n^t = s_n^{t*} \lambda^t \; (n \in I^{01}; t = 0, 1). \tag{8}$$

For later use we add that our basic assumption (5) implies that

$$s_n^{t*} = \frac{b_n (p_n^t)^{1-\sigma}}{\sum_{n \in I^{01}} b_n (p_n^t)^{1-\sigma}} \; (n \in I^{01}; t = 0, 1). \tag{9}$$

These expenditure shares can therefore be conceived as the optimal shares corresponding to minimum expenditure $C(p^t, u \mid I^{01})$ $(t = 0, 1)$.

3 Various representations of the cost of living index

The cost of living index for period 1 relative to period 0 will be defined by

$$P(p^1, p^0 \mid I^1, I^0) \equiv \frac{C(p^1, u \mid I^1)}{C(p^0, u \mid I^0)}. \tag{10}$$

Notice that, due to homotheticity, this index is independent of the utility level u. It is the ratio of the minimum expenditure that is necessary for obtaining utility level u under comparison and base period prices respectively, thereby taking into account the changed availability of commodities. Combining (10) with (2) we see that

$$P(p^1, p^0 \mid I^1, I^0) = \left[\frac{\sum_{n \in I^1} b_n (p_n^1)^{1-\sigma}}{\sum_{n \in I^0} b_n (p_n^0)^{1-\sigma}}\right]^{1/(1-\sigma)}. \tag{11}$$

There appear to be a number of ways to transform this expression into an expression consisting of observables. The *first* way proceeds by splitting the numerator multiplicatively into two parts, as follows

$$P(p^1, p^0 \mid I^1, I^0) = \left[\frac{\sum_{n \in I^1} b_n (p_n^1)^{1-\sigma}}{\sum_{n \in I^{01}} b_n (p_n^1)^{1-\sigma}} \frac{\sum_{n \in I^{01}} b_n (p_n^1)^{1-\sigma}}{\sum_{n \in I^0} b_n (p_n^0)^{1-\sigma}}\right]^{1/(1-\sigma)}. \tag{12}$$

Using assumption (5), definition (7), and relation (8), it is straightforward to infer that

$$P(p^1, p^0 \mid I^1, I^0) = \left[\frac{\lambda^1}{\lambda^0}\right]^{1/(\sigma-1)} \left[\sum_{n \in I^{01}} s_n^{0*} (p_n^1/p_n^0)^{1-\sigma}\right]^{1/(1-\sigma)}. \tag{13}$$

If $I^0 = I^1 = I^{01}$, that is the range of commodities does not change between the two periods, then $\lambda^0 = \lambda^1 = 1$ and $s_n^{0*} = s_n^0$ $(n \in I^0)$, and (13) reduces to

$$P(p^1, p^0 \mid I^0, I^0) = \left[\sum_{n \in I^0} s_n^0 (p_n^1/p_n^0)^{1-\sigma}\right]^{1/(1-\sigma)}, \tag{14}$$

that is, a base-period-expenditure-share-weighted generalized mean price index. This specific result was obtained by Lloyd (1975), rediscovered by Moulton (1996) and suggested by Shapiro and Wilcox (1997). Lau (1979) proved the converse: if the cost of living index is given by (14) and the preference structure is assumed to be homothetic, then the unit cost function as well as the dual utility function have the CES functional structure.

Notice that for $\sigma = 0$ (14) reduces to the Laspeyres price index, and that for $\sigma \rightarrow 1$ we obtain the base-period-expenditure-share-weighted geometric mean price index.

The *second* way proceeds similarly by splitting the denominator of (11) into two parts. One then obtains

$$P(p^1, p^0 \mid I^1, I^0) = \left[\frac{\lambda^1}{\lambda^0}\right]^{1/(\sigma-1)} \left[\sum_{n \in I^{01}} s_n^{1*} (p_n^1/p_n^0)^{-(1-\sigma)}\right]^{-1/(1-\sigma)}. \qquad (15)$$

If $I^0 = I^1$, then (15) reduces to a comparison-period-expenditure-share-weighted generalized mean price index, a result also obtained by Lloyd (1975). If, in addition, $\sigma = 0$ then (15) reduces to the Paasche price index, while for $\sigma \rightarrow 1$ we obtain the comparison-period-expenditure-share-weighted geometric mean price index.

The *third* way proceeds by taking the square root of (13) and (15). By multiplying them, one obtains

$$P(p^1, p^0 \mid I^1, I^0) = \left[\frac{\lambda^1}{\lambda^0}\right]^{1/(\sigma-1)} \left[\frac{\sum_{n \in I^{01}} s_n^{0*} (p_n^1/p_n^0)^{1-\sigma}}{\sum_{n \in I^{01}} s_n^{1*} (p_n^1/p_n^0)^{-(1-\sigma)}}\right]^{1/2(1-\sigma)}. \qquad (16)$$

The second factor at the righthand side of (16) is a quadratic mean of order $2(1-\sigma)$ price index, computed with respect to the set I^{01}. This type of index has been introduced and discussed by Diewert (1976).

Notice that when $\sigma = 0$ the rightmost factor of (16) reduces to a Fisher price index, and that for $\sigma \rightarrow 1$ we obtain a Törnqvist price index.

The *fourth* way, using assumption (5), rearranges (3) to obtain

$$\sum_{n \in I^t} b_n (p_n^t)^{1-\sigma} = b_n (p_n^t)^{1-\sigma} / s_n^t \; (n \in I^t; t = 0, 1). \qquad (17)$$

Substituting this into (11), we obtain the following expression for the cost of living index number:

$$P(p^1, p^0 \mid I^1, I^0) = \frac{p_n^1/p_n^0}{(s_n^1/s_n^0)^{1/(1-\sigma)}} \qquad (18)$$

for $n \in I^{01}$. We now substitute (8) into (18), to obtain

$$P(p^1, p^0 \mid I^1, I^0) \left[\frac{\lambda^1}{\lambda^0}\right]^{1/(1-\sigma)} = \frac{p_n^1/p_n^0}{(s_n^{1*}/s_n^{0*})^{1/(1-\sigma)}} \; (n \in I^{01}). \tag{19}$$

Let us, just for a moment, abbreviate the lefthand side of this equation by P, then we can rewrite this equation as

$$\frac{\ln P - \ln(p_n^1/p_n^0)}{\ln(s_n^{1*}/s_n^{0*})} = -\frac{1}{1-\sigma} \; (n \in I^{01}). \tag{20}$$

It is then trivially true that

$$\sum_{n \in I^{01}} (s_n^{1*} - s_n^{0*}) \frac{\ln P - \ln(p_n^1/p_n^0)}{\ln(s_n^{1*}/s_n^{0*})} = 0, \tag{21}$$

which can also be written as

$$\sum_{n \in I^{01}} L(s_n^{1*}, s_n^{0*})(\ln P - \ln(p_n^1/p_n^0)) = 0, \tag{22}$$

where $L(.)$ is the logarithmic average.[2] But (22) is nothing else than the implicit definition of the Sato (1976) - Vartia (1976) price index, computed on the set of commodities which are common to both periods. Thus, $P = P^{SV}(p^1, x^1, p^0, x^0 \mid I^{01})$.[3] Substituting this into (19) and rearranging somewhat, one finally obtains

$$P(p^1, p^0 \mid I^1, I^0) = \left[\frac{\lambda^1}{\lambda^0}\right]^{1/(\sigma-1)} P^{SV}(p^1, x^1, p^0, x^0 \mid I^{01}). \tag{23}$$

This is essentially the result obtained by Feenstra (1994) and Feenstra and Markusen (1994), albeit in different contexts. They assumed (2) with time-dependent parameters b_n^t, but also assumed that $b_n^0 = b_n^1$ for all $n \in I^{01}$.

The *fifth* way starts by rewriting (20) as

$$\ln P = \ln(p_n^1/p_n^0) - \frac{1}{1-\sigma} \ln(s_n^{1*}/s_n^{0*}) \; (n \in I^{01}). \tag{24}$$

[2]The logarithmic average is, for $a, b > 0$, defined as $L(a, b) \equiv (a - b)/\ln(a/b)$ for $a \neq b$ and $L(a, a) \equiv a$.

[3]Viewed as a function of independent variables, the Sato-Vartia price index violates the monotonicity axiom, as demonstrated by Reinsdorf and Dorfman (1999).

Let α_n ($n \in I^{01}$) be positive constants, adding up to 1, such that

$$\sum_{n\in I^{01}} \alpha_n \ln(s_n^{1*}/s_n^{0*}) = 0. \tag{25}$$

One sees then immediately that

$$\ln P = \sum_{n\in I^{01}} \alpha_n \ln(p_n^1/p_n^0). \tag{26}$$

By substituting this into (19) one obtains

$$P(p^1, p^0 \mid I^1, I^0) = \left[\frac{\lambda^1}{\lambda^0}\right]^{1/(\sigma-1)} \prod_{n\in I^{01}} (p_n^1/p_n^0)^{\alpha_n}. \tag{27}$$

For $I^0 = I^1$ this result has been obtained by Banerjee (1983).

Summarizing the results obtained so far, we see that the cost of living index $P(p^1, p^0 \mid I^1, I^0)$, which accounts for the changed availability of commodities, can be decomposed in five ways[4] into the product of a conventional price index, computed on the set of commodities which are common to both periods, and a factor that depends on the magnitude of the change in the range of available commodities – as measured by the ratio of λ's.[5] Not all of these ways are equally interesting from the computational point of view. Most interesting are (13), (15) and (23). With respect to the conventional price index figuring in these expressions there appears to be a trade-off between using the expenditure shares of both periods or using the expenditure shares of but one period combined with knowledge of the value of the elasticity of substitution σ.

In an interesting contribution, Nahm (1997) showed that the cost of living index (and its dual Malmquist quantity index) can be decomposed into three parts. Recast in our notation, the decomposition reads

$$P(p^1, p^0 \mid I^1, I^0) = \frac{C(p^1, u \mid I^1)}{C(p^0, u \mid I^{01})} \frac{C(p^1, u \mid I^{01})}{C(p^0, u \mid I^0)} \left[\frac{C(p^1, u \mid I^{01})}{C(p^0, u \mid I^{01})}\right]^{-1}. \tag{28}$$

[4]Actually, the fifth way comprises an infinite number of representations.

[5]In retrospect, this ratio bears some resemblance to "The R Test for Homogeneity" proposed by Mudgett (1951, 55).

The first factor is a cost of living index based on the commodities which are common to both periods and the new commodities, the second factor is a cost of living index based on the common and the discontinued commodities, while the third factor, actually the denominator, is a cost of living index based on the common commodities only. Application of (13) and (15), thereby using (8), yields the following result:

$$P(p^1, p^0 \mid I^1, I^0) = \frac{\left[\sum_{n \in I^{01}} s_n^1 (p_n^1/p_n^0)^{-(1-\sigma)}\right]^{-1/(1-\sigma)} \left[\sum_{n \in I^{01}} s_n^0 (p_n^1/p_n^0)^{1-\sigma}\right]^{1/(1-\sigma)}}{\left[\sum_{n \in I^{01}} s_n^{0*} (p_n^1/p_n^0)^{1-\sigma}\right]^{1/(1-\sigma)}}. \quad (29)$$

The expression in the denominator follows from the interpretation of the shares s_n^{0*} $(n \in I^{01})$ as being the optimal expenditure shares corresponding to $C(p^0, u \mid I^{01})$. This expression can of course be replaced by equivalent expressions such as the rightmost factor in (15), (16), (23), and (27). It is noteworthy that in (29) the factor $(\lambda^1/\lambda^0)^{1/(\sigma-1)}$ has disappeared.

With respect to this factor we notice that λ^1 is equal to one minus the share of the new commodities in the period 1 expenditure, and λ^0 is equal to one minus the share of the discontinued commodities in the period 0 expenditure. One sees immediately that if these shares happen to be equal, then the value of the factor equals unity, irrespective of the magnitude of the elasticity of substitution.

Now, suppose that there are no discontinued commodities, so $\lambda^0 = 1$, and denote the share of the new commodities in the period 1 expenditure by s_N^1. Suppose further that between the base period and the comparison period the prices of the common commodities do not change, that is $p_n^1 = p_n^0$ $(n \in I^{01})$. One easily checks that in this case

$$P(p^1, p^0 \mid I^1, I^0) = (1 - s_N^1)^{1/(\sigma-1)}. \quad (30)$$

This expression must be smaller than 1, since otherwise the consumer, going from base period to comparison period, would be better off by disregarding the new commodities altogether and sticking to his base period consumption vector. But this is equivalent to requiring that the elasticity of substitution σ must be larger than 1.

Before closing this section we recall that the quality or taste parameters, occurring in the cost function (2) and the utility function (4), were assumed to be constant through time. One easily checks that replacing all b_n by b_n^t amounts to replacing p_n^t by $p_n^t/(b_n^t)^{1/(\sigma-1)}$ $(n \in I^{01}; t = 0,1)$ in all of the expressions for the cost of living index. Thus prices must be replaced by quality- or taste-corrected prices. One derives easily from (4) that

$$(b_n^t)^{1/(\sigma-1)} = \left[\frac{\partial U(x \mid I^t)^{(\sigma-1)/\sigma}}{\partial x_n^{(\sigma-1)/\sigma}}\right]^{\sigma/(\sigma-1)}, \quad (31)$$

which can be conceived as a kind of marginal utility.

4 Obtaining the elasticity of substitution

For obtaining the value of the elasticity of substitution σ we do not need to estimate a demand system as suspected by Shapiro and Wilcox (1997). There appear to be simpler routes, all prompted by the diverse representations of the cost of living index $P(p^1, p^0 \mid I^1, I^0)$.

To start with, it is obvious that the righthand side of (13) must be equal to the righthand side of (15), so that we can obtain σ as the solution of[6]

$$\left[\sum_{n\in I^{01}} s_n^{0*}(p_n^1/p_n^0)^{1-\sigma}\right]^{1/(1-\sigma)} = \left[\sum_{n\in I^{01}} s_n^{1*}(p_n^1/p_n^0)^{-(1-\sigma)}\right]^{-1/(1-\sigma)}, \quad (32)$$

given that we know the base and comparison period expenditure shares.

The solution can be obtained by a simple numerical procedure. For $\sigma = 0$ the lefthand side of (32) reduces to $\sum_{n\in I^{01}} s_n^{0*}(p_n^1/p_n^0)$, which is the Laspeyres price index with respect to the commodity set I^{01}. Similarly, the righthand side reduces to $[\sum_{n\in I^{01}} s_n^{1*}(p_n^1/p_n^0)^{-1}]^{-1}$, which is the Paasche price index with respect to the commodity set I^{01}. It is to be expected that the Laspeyres price index is greater than the Paasche price index. When we let $\sigma \to \infty$ the lefthand side of (32) approaches $\min_{n\in I^{01}}(p_n^1/p_n^0)$, while the righthand side approaches $\max_{n\in I^{01}}(p_n^1/p_n^0)$. Moreover, we know that any generalized mean $(\sum_n a_n(z_n)^k)^{1/k}$ is monotonously increasing in k (Hasenkamp 1978, Theorem 2).

Secondly, the equality of the righthand sides of (13) and (23) leads to

[6]Basically this procedure was suggested in Proposition 7 appended to the doctoral dissertation of Balk (1984).

$$\left[\sum_{n\in I^{01}} s_n^{0*}(p_n^1/p_n^0)^{1-\sigma}\right]^{1/(1-\sigma)} = P^{SV}(p^1, x^1, p^0, x^0 \mid I^{01}), \tag{33}$$

from which the value of σ can also be obtained with help of a simple numerical procedure. If fact, this equation lends support to the actual procedure used by Shapiro and Wilcox (1997). They had observations pertaining to time periods $t = 0, 1, 2, ..., T$ and determined the value of the elasticity of substitution by solving

$$\min_{\sigma} \frac{1}{T}\sum_{t=1}^{T} \mid \left[\sum_{n\in I^{t-1,t}} s_n^{t-1,*}(p_n^t/p_n^{t-1})^{1-\sigma}\right]^{1/(1-\sigma)} - P^{SV}(p^t, x^t, p^{t-1}, x^{t-1} \mid I^{t-1,t}) \mid, \tag{34}$$

except that the Sato-Vartia price index was replaced by the Törnqvist price index. These two indices, however, appear in practice to be close approximations to each other.

Thirdly, the equality of the righthand sides of (15) and (23) leads to

$$\left[\sum_{n\in I^{01}} s_n^{1*}(p_n^1/p_n^0)^{-(1-\sigma)}\right]^{-1/(1-\sigma)} = P^{SV}(p^1, x^1, p^0, x^0 \mid I^{01}) \tag{35}$$

as yet another way of obtaining the value of the elasticity of substitution.

5 A two-level CES preference structure with variable sets of commodities

In section 3 we noticed that maintaining (2) with commodity sets being variable through time is only possible when the (overall) elasticity of substitution is larger than 1. For empirical economists such a value seems to be counterintuitive. Most empirical work points to a value between 0 and 1. However, all of this work has been executed on commodity groups as smallest units of measurement, thereby assuming that these groups as such are not changing through time.[7] This assumption is reasonable, given that it is mostly within groups that commodities appear and disappear.

[7]Shapiro and Wilcox's (1997) work was based on the 44 × 207 area-item strata of the U. S. CPI over the period 1987-1995. They obtained $\sigma = 0.7$.

This suggests as way out the assumption of a two-level structure in the preferences of the representative consumer. Thus, let there be G non-overlapping commodity groups. Let the sets of commodities belonging to each of these groups be variable (but overlapping) through time, and define $I_g^t \subset \{1, ..., N\}$ as the set of commodities belonging to group g $(g = 1, ..., G)$ which are available in period t. We will denote by x_{ng}^t and p_{ng}^t respectively the non-negative quantity consumed and the corresponding positive price of commodity $n \in I_g^t$ at period t; x_g^t will denote the vector of x_{ng}^t and p_g^t will denote the vector of p_{ng}^t; x_g and p_g will denote generic quantity and price vectors of appropriate size.

It is now assumed that the period t cost function is completely separable in the partition $I_1^t, ..., I_G^t$ and that the price aggregator functions of the groups are CES functions, that is

$$C(p, u \mid I_1^t, ..., I_G^t) = u \left[\sum_{g=1}^{G} \left(\sum_{n \in I_g^t} b_{ng} p_{ng}^{1-\sigma_g} \right)^{(1-\sigma)/(1-\sigma_g)} \right]^{1/(1-\sigma)} \tag{36}$$

where $u > 0$ is a utility level, $\sigma \geq 0, \sigma \neq 1$, $\sigma_g \geq 0, \sigma_g \neq 1$, $b_{ng} > 0$ $(n \in I_g^t; g = 1, ..., G)$. We see that the cost function is an (upper level) CES function of (lower level) CES functions.[8] The price aggregator functions of the groups are given by

$$\tilde{p_g} \equiv \left(\sum_{n \in I_g^t} b_{ng} p_{ng}^{1-\sigma_g} \right)^{1/(1-\sigma_g)} \quad (g = 1, ..., G). \tag{37}$$

By Shephard's Lemma the optimal expenditure shares can be obtained. Using the basic assumption that in both time periods $t = 0, 1$ these optimal shares are equal to the observed shares, one obtains the following expression for the observed expenditure shares of the commodities

$$s_{ng}^t = \frac{(\tilde{p}_g^t)^{1-\sigma}}{\sum_{g=1}^{G} (\tilde{p}_g^t)^{1-\sigma}} \frac{b_{ng} (p_{ng}^t)^{1-\sigma_g}}{\sum_{n \in I_g^t} b_{ng} (p_{ng}^t)^{1-\sigma_g}} \tag{38}$$

[8]The dual representations of (36) are discussed by Blackorby, Primont and Russell (1998). The two-level CES functional structure was introduced in production theory by Sato (1967) and in consumption theory (with excess quantities) by Brown and Heien (1972). The Allen elasticity of substitution between two different commodities n and n' is equal to $\sigma + (\sigma_g - \sigma)/s_g^t$ when $n, n' \in I_g^t$ and equal to σ when $n \in I_g^t$ and $n' \notin I_g^t$. Multi-level CES functions were considered by Keller (1976).

for $n \in I_g^t$ $(g = 1, ..., G)$. From this we obtain the following expression for the observed expenditure shares of the commodity groups

$$s_g^t \equiv \sum_{n \in I_g^t} s_{ng}^t = \frac{(\tilde{p}_g^t)^{1-\sigma}}{\sum_{g=1}^{G} (\tilde{p}_g^t)^{1-\sigma}} \; (g = 1, ..., G). \tag{39}$$

We further define for each commodity group

$$\lambda_g^t \equiv \frac{\sum_{n \in I_g^{01}} s_{ng}^t}{\sum_{n \in I_g^t} s_{ng}^t} = \frac{\sum_{n \in I_g^{01}} s_{ng}^t}{s_g^t} = \frac{\sum_{n \in I_g^{01}} b_{ng} (p_{ng}^t)^{1-\sigma_g}}{\sum_{n \in I_g^t} b_{ng} (p_{ng}^t)^{1-\sigma_g}} \; (g = 1, ..., G), \tag{40}$$

which is the fraction of the total expenditure on commodity group g at period t that is attributable to the commodities which are common to the base period and the comparison period. For each of these commodities the relative shares, that is, the shares with respect to the part of the commodity group that is common to both periods, are defined by

$$s_{ng}^{t*} \equiv \frac{s_{ng}^t}{\sum_{n \in I_g^{01}} s_{ng}^t} = \frac{b_{ng} (p_{ng}^t)^{1-\sigma_g}}{\sum_{n \in I_g^{01}} b_{ng} (p_{ng}^t)^{1-\sigma_g}} \; (n \in I_g^{01}; g = 1, ..., G). \tag{41}$$

The cost of living index for period 1 relative to period 0 is, analogous to (10), defined by

$$P(p^1, p^0 \mid I^1, I^0) \equiv \frac{C(p^1, u \mid I_1^1, ..., I_G^1)}{C(p^0, u \mid I_1^0, ..., I_G^0)}, \tag{42}$$

where I^1, I^0 is used as an obvious abbreviation. Substituting now (36) into (42), and employing (39), for the upper level each of the five ways discussed in section 3 could be followed. For instance, the first way results in

$$\begin{aligned} P(p^1, p^0 \mid I^1, I^0) &= \left[\sum_{g=1}^{G} s_g^0 \left(\frac{\sum_{n \in I_g^1} b_{ng} (p_{ng}^1)^{1-\sigma_g}}{\sum_{n \in I_g^0} b_{ng} (p_{ng}^0)^{1-\sigma_g}} \right)^{(1-\sigma)/(1-\sigma_g)} \right]^{1/(1-\sigma)} \\ &= \left[\sum_{g=1}^{G} s_g^0 \left(P(p_g^1, p_g^0 \mid I_g^1, I_g^0) \right)^{1-\sigma} \right]^{1/(1-\sigma)} . \end{aligned} \tag{43}$$

Next, for each commodity group g there are again five ways to follow. Again using the first way, one arrives finally at the following expression for the cost of living index

$$P(p^1, p^0 \mid I^1, I^0) = \left[\sum_{g=1}^{G} s_g^0 \left(\frac{\lambda_g^0}{\lambda_g^1} \sum_{n \in I_g^{01}} s_{ng}^{0*} (p_{ng}^1 / p_{ng}^0)^{1-\sigma_g} \right)^{(1-\sigma)/(1-\sigma_g)} \right]^{1/(1-\sigma)} . \tag{44}$$

Moreover, there is no necessity to restrict to the same way for each commodity group. Thus, (44) is but one of at least $5 \times 5 \times G$ different representations of the cost of living index.

Let us now suppose that $I_g^0 = I_g^1$ for $g = 2, ..., G$ and $I_1^0 \subset I_1^1$, that is, only in commodity group 1 there appear new commodities at the comparison period. Then $\lambda_g^0 = \lambda_g^1 = 1$ for $g = 2, ..., G$, $\lambda_1^0 = 1$, and $\lambda_1^1 = 1 - s_{N1}^1 / s_1^1$, where s_{N1}^1 is the sum of shares of the new commodities belonging to group 1. We also suppose that between both periods the prices of the common commodities do not change, that is $p_{ng}^0 = p_{ng}^1$ for $n \in I_g^{01}$ $(g = 1, ..., G)$. Then the cost of living index reduces to

$$P(p^1, p^0 \mid I^1, I^0) = \left[s_1^0 \left(1 - \frac{s_{N1}^1}{s_1^1} \right)^{-(1-\sigma)/(1-\sigma_1)} + 1 - s_1^0 \right]^{1/(1-\sigma)} , \tag{45}$$

which must be smaller than 1. The following chain of expressions

$$\begin{aligned}
&\ln P(p^1, p^0 \mid I^1, I^0) \\
&= \frac{1}{1-\sigma} \ln \left[s_1^0 \left(1 - \frac{s_{N1}^1}{s_1^1} \right)^{-(1-\sigma)/(1-\sigma_1)} + 1 - s_1^0 \right] \\
&\approx \frac{1}{1-\sigma} \left[s_1^0 \left(1 - \frac{s_{N1}^1}{s_1^1} \right)^{-(1-\sigma)/(1-\sigma_1)} - s_1^0 \right] \\
&\approx \frac{1}{1-\sigma} \left[s_1^0 \left(1 + \frac{1-\sigma}{1-\sigma_1} \frac{s_{N1}^1}{s_1^1} \right) - s_1^0 \right] \\
&= s_1^0 \frac{1}{1-\sigma_1} \frac{s_{N1}^1}{s_1^1}
\end{aligned} \tag{46}$$

demonstrates that for the cost of living index to be smaller than 1 it is required that $\sigma_1 > 1$, irrespective the value of σ. Of course, commodity group 1 was here only taken to ease the exposition.

We thus conclude that requiring all intra-group σ_g's to be larger than 1 is consistent with having the inter-group substitution elasticity σ to be smaller than 1.

The empirical estimation of the parameters σ_g and σ should proceed bottom-up. For each commodity group g each of the methods discussed in section 4 can be used to obtain both the value of σ_g and the price index number $P(p_g^1, p_g^0 \mid I_g^1, I_g^0)$. Then one level higher up, again each of these methods can be used to obtain the value of σ.

References

[1] Balk, B. M., 1984, Studies on the Construction of Price Index Numbers for Seasonal Products, Doctoral Dissertation, University of Amsterdam.

[2] Banerjee, K. S., 1983, "On the Existence of Infinitely Many Ideal Log-Change Index Numbers Associated with the CES Preference Ordering," Statistische Hefte / Statistical Papers 24, 141-148.

[3] Bergson, A., 1936, "Real Income, Expenditure Proportionality, and Frisch's 'New Methods of Measuring Marginal Utility'," Review of Economic Studies 4, 33-52.

[4] Blackorby, C., D. Primont and R. R. Russell, 1998, "Separability: A Survey," in *Handbook of Utility Theory, Volume 1*, edited by S. Barberá, P. J. Hammond and C. Seidl (Kluwer Academic Publishers, Dordrecht / Boston / London).

[5] Boskin, M. J., E. R. Dulberger, R. J. Gordon, Z. Griliches and D. W. Jorgenson, 1996, Toward a More Accurate Measure of the Cost of Living, Final Report to the [U. S.] Senate Finance Committee from the Advisory Commission To Study The Consumer Price Index.

[6] Brown, M. and D. Heien, 1972, "The S-Branch Utility Tree: A Generalization of the Linear Expenditure System," Econometrica 40, 737-747.

[7] Diewert, W. E., 1976, "Exact and Superlative Index Numbers," Journal of Econometrics 4, 115-145.

[8] Feenstra, R. C., 1994, "New Product Varieties and the Measurement of International Prices," The American Economic Review 84, 157-177.

[9] Feenstra, R. C. and J. R. Markusen, 1994, "Accounting for Growth with New Inputs," International Economic Review 35, 429-447.

[10] Feenstra, R. C. and C. R. Shiells, 1997, "Bias in U. S. Import Prices and Demand," in *The Economics of New Goods*, edited by T. F. Bresnahan and R. J. Gordon (University of Chicago Press, Chicago).

[11] Gamaletsos, T., 1973, "Further Analysis of Cross-Country Comparison of Consumer Expenditure Patterns," European Economic Review 4, 1-20.

[12] Hasenkamp, G., 1978, "Economic and Atomistic Index Numbers: Contrasts and Similarities," in *Theory and Applications of Economic Indices*, edited by W. Eichhorn, R. Henn, O. Opitz and R. W. Shephard (Physica-Verlag, Würzburg).

[13] Hasenkamp, G., 1980, *A Demand System Analysis of Disaggregated Consumption* (Vandenhoeck & Ruprecht, Göttingen).

[14] Hasenkamp, G. and A. Y. C. Koo, 1983, "United States Demand for Material Imports: Distinguished by Regions of Supply," in *Quantitative Studies on Production and Prices*, edited by W. Eichhorn, R. Henn, K. Neumann and R. W. Shephard (Physica-Verlag, Würzburg).

[15] Keller, W. J., 1976, "A Nested CES-Type Utility Function and its Demand and Price-Index Functions," European Economic Review 7, 175-186.

[16] Lau, L. J., 1979, "On Exact Index Numbers," The Review of Economics and Statistics 61, 73-82.

[17] Lloyd, P. J., 1975, "Substitution Effects and Biases in Nontrue Price Indices," The American Economic Review 65, 301-313.

[18] McFadden, D., 1963, "Constant Elasticity of Substitution Production Functions," Review of Economic Studies 30, 73-83.

[19] Moulton, B. R., 1996, Constant Elasticity Cost-of-Living Index in Share-Relative Form, Mimeo.

[20] Mudgett, B. D., 1951, *Index Numbers* (John Wiley & Sons, New York; Chapman & Hall, London).

[21] Nahm, D., 1997, Incorporating the Effect of New and Disappeared Products into the Cost of Living: An Alternative Index Number Formula, Macquary Economics Research Papers Number 26/97 (Department of Economics, Macquary University, Sydney).

[22] Reinsdorf, M. B. and A. H. Dorfman, 1999, "The Sato-Vartia Index and the Monotonicity Axiom," Journal of Econometrics 90, 45-61.

[23] Sato, K., 1967, "A Two-Level Constant Elasticity of Substitution Production Function," Review of Economic Studies 34, 201-218.

[24] Sato, K., 1976, "The Ideal Log-Change Index Number," The Review of Economics and Statistics 58, 223-228.

[25] Shapiro, M. D. and D. W. Wilcox, 1997, "Alternative Strategies for Aggregating Prices in the CPI," Federal Reserve Bank of St. Louis Review 79, 3, 113-125.

[26] Uzawa, H., 1962, "Production Functions with Constant Elasticities of Substitution," Review of Economic Studies 29, 291-299.

[27] Vartia, Y. O., 1976, "Ideal Log-Change Index Numbers," The Scandinavian Journal of Statistics 3, 121-126.