NOTES ON PRODUCING AN ANNUAL SUPERLATIVE INDEX USING MONTHLY PRICE DATA

by

W. Erwin Diewert

July 2000

Discussion Paper No.: 00-08

DEPARTMENT OF ECONOMICS
THE UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, CANADA V6T 1Z1

http://web.arts.ubc.ca/econ/

NOTES ON PRODUCING AN ANNUAL SUPERLATIVE INDEX USING MONTHLY PRICE DATA

Erwin Diewert,
Department of Economics,
University of British Columbia,
Vancouver, Canada, V6T 1Z1
Email: diewert@econ.ubc.ca

Abstract

The main purpose of the presentation is to outline some alternative approaches on how a superlative annual consumer price index could be constructed using the monthly price information that is presently collected by statistical agencies. The first issue that must be addressed is: how should the monthly price information at the lowest level of aggregation be aggregated up (over months) to form an annual price level or price relative at this lowest level of commodity aggregation? Having constructed appropriate annual elementary indexes, the presentation discusses how to complete the process to construct an annual (calendar year) superlative index.

Considering the problem of seasonal commodities, it is noted that the construction of year over year (superlative) indexes for each month of the year should be free of seasonal influences. Moreover, the business community is typically quite interested in this class of index numbers so the paper recommends that statistical agencies produce them. This approach leads to another two stage aggregation of the micro price information into an overall annual index: first aggregate across commodities holding the month constant and then aggregate across months. It is noted that the alternative two stage annual indexes can be quite close to each other.

The two annual indexes are on a calendar year basis; i.e., the price and quantity data pertaining to all 12 months in 1999 are compared to the corresponding prices and quantities for the base year, say 1995. However, we can also construct a moving year or rolling year annual index, using price and quantity information collected each month. Thus the statistical agency could produce a new rolling year index every month, which of course would lag behind its present very timely CPI index due to the lags involved in collecting the relevant quantity (or expenditure) information. The real advantage of these moving year superlative indexes is that they are both timely and do not require any seasonal adjustment.

A problem with the indexes discussed so far is that they do not give us any information on short term price movements; i.e., all of these indexes discussed above compare prices in a month in the current year with the same month in the base year. The paper also discusses some of the problems involved in constructing superlative month to month price indexes.

Finally, some of the problems that are associated with the aggregation over households or regions are discussed.

1. Introduction[1]

The main purpose of these notes is to outline some alternative approaches on how a superlative *annual* consumer price index could be constructed using the *monthly* price information that is presently collected by statistical agencies. The first issue that must be addressed is: how should the monthly price information at the lowest level of aggregation be aggregated up (over months) to form an annual price level or price relative at this lowest level of commodity aggregation? We address this issue in section 2 below. In section 3, having constructed appropriate annual elementary indexes in section 2, we discuss how to complete the process to construct an annual (calendar year) superlative index. We also note that this two stage approach to the aggregation of price information (first aggregate across months holding the commodity constant and then aggregate across commodities) will be quite close (under certain conditions) to the construction of an annual superlative index in a single stage, where we regard each commodity in each month as a distinct commodity.[2] The advantage of this Mudgett (1955)-Stone (1956) approach to the construction of annual indexes is that it deals satisfactorily with the problem of seasonal commodities.[3]

Having introduced the problem of seasonal commodities in section 3, in section 4, we note that the construction of year over year (superlative) indexes for each month of the year should be free of seasonal influences.[4] Moreover, the business community is typically quite interested in this class of index numbers so we recommend that statistical agencies produce them. Having constructed year over year monthly indexes in section 4, in section 5, we can look at the problems involved in aggregating all twelve of these monthly indexes into a single annual index. This leads to another two stage aggregation of the micro price information into an overall annual index: first aggregate across commodities holding the month constant and then aggregate across months. It turns out that under certain conditions, the resulting two stage annual index will be quite close to the construction of an annual superlative index in a single stage. We also note that the two stage annual indexes constructed in section 5 can be quite close to the two stage annual indexes constructed in section 3.

The annual indexes constructed in sections 3 and 5 are annual indexes on a calendar year basis; i.e., the price and quantity data pertaining to all 12 months in 1999 are compared to the corresponding prices and quantities that pertain to the base year, say 1995. However, there is no need to construct a fixed base annual index for 1999 after the December 1999 price and quantity data are available and then wait a whole year before constructing the year 2000 annual fixed base index. Price and quantity information can be collected *each month*, so that when the January 2000 price and quantity information has been collected, we can construct a *moving year* or

[1] Discussions with Ralph Bradley, Rob Cage, Alan Dorfman, John Greenlees and Pat Jackman were very helpful. They are not responsible for any opinions expressed in this paper.

[2] This approach to the construction of annual indexes is due to Mudgett (1955) and Stone (1956).

[3] Thus the utility provided by a cold beer on a hot summer day could be quite different from the utility provided on a cold winter day.

[4] Fluctuations in temperature and the shifting of holidays across months mean that this statement is not quite true but most seasonal influences should be eliminated using year over year monthly indexes.

rolling year annual index that compares the January 2000 information with the January 1995 information, the February 1999 information with the February 1995 information, the March 1999 information with the March 1995 information,…., and the December 1999 information with the December 1995 information. Thus to the 1999 monthly data, we added the data for January 2000 and dropped the January 1999 data, giving us 12 consecutive months of price and quantity data which is in turn compared to the 12 base months. Thus the statistical agency could produce a new rolling year index *every month*, which of course would lag behind its present very timely CPI index due to the lags involved in collecting the relevant quantity (or expenditure) information. In section 6, we discuss how these moving year indexes could be constructed and note how they could be approximated by using a two stage aggregation procedure on the year over year monthly indexes discussed in section 4. The real advantage of these moving year superlative indexes is that they are both *timely* and *do not require any seasonal adjustment*.[5]

A problem with the indexes discussed in sections 3 to 6 is that they do not give us any information on *short term* price movements; i.e., all of these indexes compare prices in a month in the current year with the same month in the base year. Hence, it is simply impossible for these indexes to provide much information on short term price movements.[6] Thus in section 7, we briefly discuss superlative month to month price indexes.

Up to this point, the discussion has focussed on the problems involved in aggregating monthly data into annual data and has ignored the problems involved in aggregating over households or aggregating over different regions within the same country. In sections 8 and 9, we address these household and regional aggregation problems. In section 8, we look at the relationships between regional Laspeyres, Paasche and Fisher ideal indexes and their national counterparts while section 9 looks at the problems involved in aggregating regional Törnqvist indexes into a national index.

Section 10 summarizes some of the issues and gives some tentative recommendations.

2. Alternative Concepts for Annual Average Prices at the Lowest Level of Aggregation

In this section, we address the issue of how should the monthly price information at the lowest level of aggregation be aggregated over months to form an annual price level or price relative.

We first need to introduce some notation. We suppose that we have N distinct commodity classifications at the lowest level of aggregation and for each month m of year t, we have a "representative" price quote, $p_n^{t,m}$, for each commodity class n for n = 1,…,N.[7] Since we are

---

[5] Except that seasonal adjustment may be required due to changes in the number of working days in a month and or shifts in holidays like Easter from one month to another.
[6] The moving year indexes provide some information on the longer run trends in prices. These moving year indexes are the index number counterparts to statistical methods used to smooth and seasonally adjust price and quantity series, such as X-11. However, using moving year indexes or X-11 smoothed series to estimate a month to month change in prices will not give a *pure measure* of price change that depends *only* on the price and quantity information that pertains to those two months.
[7] This "representative" price quote could be an aggregate of a sample of price quotations on the price of commodity n for month m of year t. However, this raises issues about how the individual price quotes should be aggregated into

going to discuss the construction of annual superlative indexes, we need information on the quantities of commodity n purchased by the households in the domain of definition of the price index during month m of year t, $q_n^{t,m}$, that match up with the price quotes, $p_n^{t,m}$.

A *bilateral index number formula* is a function of four sets of variables: base period prices; current period prices; base period quantities and current period quantities. The quantity variables have to match up with the price variables in each period. Moreover, the inner product of the price and quantity vector pertaining to the same period should equal *expenditures* on the N goods in the index domain of definition for the reference population for that period. An implicit assumption associated with the use of a bilateral index number formula is that the price variables that enter into the formula are "representative" for all of the transactions pertaining to that commodity in the period under consideration. Thus implicitly, when we use a bilateral index number formula, we are assuming that the length of the period is short enough so that all of the transactions involving commodity n within each period can be adequately summarized by a single price.[8]

With the above limitations of bilateral index number theory in mind, we now turn our attention to the problem of aggregating our monthly price and quantity information on commodity n into annual price and quantity information. One very reasonable method for doing this aggregation is to define the annual consumption of commodity n in year t, Q_n^t say, to be the sum of the monthly consumption of commodity n in year t; i.e., define Q_n^t as follows:

$$(1)\quad Q_n^t \equiv \sum_{m=1}^{12} q_n^{t,m} ; \qquad n = 1,\ldots,N ; t = 0,1,\ldots,T.$$

If we want the annual value purchased by the reference population to equal the product of an annual price, say P_n^t, times the annual quantity Q_n^t, then P_n^t *must* be defined as follows:[9]

$$(2)\quad P_n^t \equiv \sum_{m=1}^{12} p_n^{t,m} q_n^{t,m} / \sum_{m=1}^{12} q_n^{t,m} = \sum_{m=1}^{12} p_n^{t,m} q_n^{t,m} / Q_n^t ; \qquad n = 1,\ldots,N ; t = 0,1,\ldots,T.$$

Definition (2) shows that the *annual unit value price for commodity n in year t*, P_n^t, is equal to a weighted average of the monthly prices, $p_n^{t,m}$, for m = 1,…,12. Define the *month m share of annual consumption of commodity n in year t* as

$$(3)\quad \sigma_n^{t,m} \equiv q_n^{t,m} / Q_n^t ; \qquad n = 1,\ldots,N ; t = 0,1,\ldots,T.$$

---

the overall "representative" price quote. In order to keep our exposition (and notation) relatively simple, we will not discuss these issues here.

[8] Fisher (1922; 318) and Hicks (1946; 122) both defined the length of period to be a time period short enough so that variations in price within the period can be ignored.

[9] The use of unit values and total quantities transacted during the period as the appropriate prices and quantities that should be inserted into a bilateral index number formula dates back to Walsh (1901; 96) (1921; 88) and Davies (1924); see Diewert (1995; 20-24). Hill (1993; 399) recommended the use of unit values and total quantities as the appropriate prices and quantities that should be used in the national accounts under normal conditions. However, when there is high inflation within the year, Hill (1996) did not recommend the use of annual unit values as prices. Diewert (1996) (1998) also noted that the annual unit value concept was not appropriate under conditions of high inflation but he also noted that even with low inflation, an annual unit value may not be the appropriate price concept if consumers have seasonal preferences over commodities.

Then using the quantity shares defined by (3), we can rewrite the annual unit value price for commodity n in year t as a (quantity) share weighted average of the monthly prices for commodity n in year t:

(4) $P_n^t \equiv \sum_{m=1}^{12} \sigma_n^{t,m} p_n^{t,m}$; n = 1,…,N ; t = 0,1,…,T.

Using (4), the long term annual price relative for commodity n in year t can be written as:

(5) $P_n^t / P_n^0 \equiv \sum_{m=1}^{12} \sigma_n^{t,m} p_n^{t,m} / \sum_{m=1}^{12} \sigma_n^{0,m} p_n^{0,m}$; n = 1,…,N ; t = 0,1,…,T.

This is our first approach to defining a theoretical long term annual fixed base price relative for commodity n that could be inserted into an annual superlative index number formula. The annual quantity, Q_n^t defined by (1), would be the period t quantity for commodity n that would be used in the annual superlative index number formula. Let us call this approach to defining appropriate annual elementary prices the *annual unit value approach*.

The above approach is fine as far as it goes but it may not be appropriate for at least two reasons:

- The year may be too long a period to ignore within the period price variation. In particular, the approach breaks down if there is high inflation during the year; i.e., the implicit assumption that price variation within the year can be ignored breaks down.[10]
- The annual unit value approach may not be valid if consumers have *seasonal preferences* over some commodities.[11]

Thus it may be useful to take other points of view to the problem of aggregating monthly price information into annual prices. In particular, one could take a *stratified sampling* point of view. Thus we could regard each monthly long term price relative in year t, $p_n^{t,m} / p_n^{0,m}$ for m = 1,…,12, as a stratified sample from 12 monthly strata. Each long term price relative, $p_n^{t,m} / p_n^{0,m}$, compares a monthly price for commodity n in year t to the corresponding monthly price of commodity n in the base year, year 0. Thus if t = 1, then each of these price relatives, is expressing an *annual* rate of price change for that particular month m.

How should these 12 long term monthly price relatives, $p_n^{t,m} / p_n^{0,m}$, be combined into a single annual long term price relative? Obviously, using an unweighted arithmetic or geometric mean are possibilities. These two choices lead to the *Carli and Jevons long term annual price relatives for commodity n in year t*, P_C and P_J respectively:

(6) $P_C(p_n^0,p_n^t) \equiv \sum_{m=1}^{12} (1/12) (p_n^{t,m} / p_n^{0,m})$; n = 1,…,N ; t = 1,…,T;
(7) $P_J(p_n^0,p_n^t) \equiv \prod_{m=1}^{12} (p_n^{t,m} / p_n^{0,m})^{1/12}$; n = 1,…,N ; t = 1,…,T

where we define the *year t price vector for commodity n* as

(8) $p_n^t \equiv (p_n^{t,1}, p_n^{t,2},\ldots, p_n^{t,12})$; n = 1,…,N ; t = 0,1,…,T.

[10] See Hill (1996) and Diewert (1996) (1998).
[11] See Diewert (1996) (1998).

For completeness, we should note the following formula due to *Dutot*[12] that is used to aggregate price quotes at the elementary level:

(9) $P_D(p_n^0,p_n^t) \equiv [\sum_{m=1}^{12} (1/12)\, p_n^{t,m}] / [\sum_{m=1}^{12} (1/12)\, p_n^{0,m}]$; $n = 1,\ldots,N$; $t = 1,\ldots,T$.

Note that the ratio of averages form of the Dutot index (9) is similar in structure to our earlier annual unit value formula (5). In fact, if the quantity share weights in (5), the $\sigma_n^{t,m}$, are all equal, then (5) collapses down to (9).

The use of the Carli formula (6) is not recommended since it does not satisfy the time reversal test and in fact has an upward bias from the perspective of that test.[13] Diewert (1995) discussed the relative merits of the unweighted elementary aggregation formulae (6), (7) and (9) and generally expressed a preference for the use of the Jevons geometric formula, (7), both from the viewpoint of the economic approach as well as the axiomatic approach to elementary indexes. Hover, the lack of weighting in these indexes is a real drawback[14] and so we now turn our attention to weighted versions of the above formulae.

Let $\mu(a,b)$ be a weighted average of the two (positive) numbers a and b. Recall or earlier definition (3) of the month m share of year t consumption of commodity n, $\sigma_n^{t,m}$. We define the year t vector of these monthly quantity shares for commodity n by:

(10) $\sigma_n^t \equiv (\sigma_n^{t,1}, \sigma_n^{t,2},\ldots, \sigma_n^{t,12})$; $n = 1,\ldots,N$; $t = 0,1,\ldots,T$.

The mean function μ and the quantity weights defined by (10) above can be used to define the following *weighted counterparts* to the unweighted indexes (6), (7) and (9) above:

(11) $P_C(p_n^0,p_n^t,\sigma_n^0,\sigma_n^t) \equiv \sum_{m=1}^{12} \mu(\sigma_n^{0,m},\sigma_n^{t,m})\, (p_n^{t,m} / p_n^{0,m})$; $n = 1,\ldots,N$; $t = 1,\ldots,T$;
(12) $\ln P_J(p_n^0,p_n^t,\sigma_n^0,\sigma_n^t) \equiv \sum_{m=1}^{12} \mu(\sigma_n^{0,m},\sigma_n^{t,m}) \ln (p_n^{t,m} / p_n^{0,m})$; $n = 1,\ldots,N$; $t = 1,\ldots,T$;
(13) $P_D(p_n^0,p_n^t,\sigma_n^0,\sigma_n^t) \equiv [\sum_{m=1}^{12}\mu(\sigma_n^{0,m},\sigma_n^{t,m})\, p_n^{t,m}]/[\sum_{m=1}^{12}\mu(\sigma_n^{0,m},\sigma_n^{t,m})\, p_n^{0,m}]$;
$n = 1,\ldots,N$; $t = 1,\ldots,T$.

Note the difference between the weighted Dutot formula (13) and the annual unit value price relative defined earlier by (5): in (5), the quantity weights in the numerator and the denominator

[12] For references to the literature on elementary index number formulae, see Diewert (1995).
[13] Fisher (1922; 29-30 and 66) was the first to point this out.
[14] The price quote in a month where very little quantity is transacted gets the same weight as a price in a month where a large quantity is transacted. Walsh made this criticism of the unweighted stochastic approach to index number theory, as did Keynes (1930; 76-77) some years later: "It might seem at first sight as if simply every price quotation were a single item, and since every commodity (any kind of commodity) has one price-quotation attached to it, it would seem as if price-variations of every kind of commodity were the single item in question. This is the way the question struck the first inquirers into price-variations, wherefore they used simple averaging with even weighting. But a price-quotation is the quotation of the price of a generic name for many articles; and one such generic name covers a few articles, and another covers many. … A single price-quotation, therefore, may be the quotation of the price of a hundred, a thousand, or a million dollar's worths, of the articles that make up the commodity named. Its weight in the averaging, therefore, ought to be according to these money-unit's worth." Correa Moylan Walsh (1921; 82-83).

were in general *different* whereas in (13), the weights in the numerator and denominator are the *same*.

However, this is not the end of possible "reasonable" weighting schemes that could be applied (in theory) at the elementary level. Instead of using the monthly quantity shares ,$\sigma_n^{t,m}$, as weights, we could use monthly value shares, $s_n^{t,m}$, as weights. Thus define the year t monthly expenditure share for commodity n as:

(14) $s_n^{t,m} \equiv p_n^{t,m} q_n^{t,m} / \sum_{k=1}^{12} p_n^{t,k} q_n^{t,k}$; m = 1,…,12 ; n = 1,…,N ; t = 0,1,…,T.

Define the year t vector of these monthly expenditure shares for commodity n as

(15) $s_n^t \equiv (s_n^{t,1}, s_n^{t,2},\ldots, s_n^{t,12})$; n = 1,…,N ; t = 0,1,…,T.

Taking a Laspeyres or base period perspective, we can now define expenditure share weighted counterparts to the Carli, Jevons and Dutot annual price relatives; i.e., define *the Laspeyres type year t annual long term price relatives for commodity n* using the expenditure share weights (15) for these 3 formulae as follows:

(16) $P_{C,0}(p_n^0,p_n^t,s_n^0) \equiv \sum_{m=1}^{12} s_n^{0,m} (p_n^{t,m} / p_n^{0,m})$; n = 1,…,N ; t = 1,…,T;
(17) $\ln P_{J,0}(p_n^0,p_n^t,s_n^0) \equiv \sum_{m=1}^{12} s_n^{0,m} \ln (p_n^{t,m} / p_n^{0,m})$; n = 1,…,N ; t = 1,…,T;
(18) $P_{D,0}(p_n^0,p_n^t,s_n^0) \equiv [\sum_{m=1}^{12} s_n^{0,m} p_n^{t,m}]/[\sum_{m=1}^{12} s_n^{0,m} p_n^{0,m}]$; n = 1,…,N ; t = 1,…,T.

We now follow in the footsteps of Theil (1967; 136-137), who proposed a solution to the lack of weighting in the unweighted Jevons formula (7) in the context of a stochastic approach to index number theory. We argue as follows. Suppose we draw monthly long term price relatives for year t and commodity n at random in such a way that each dollar of expenditure on commodity n in the base period has an equal chance of being selected. Then the probability that we will draw the month m price relative is equal to $s_n^{0,m}$, the period 0 month m expenditure share for commodity n. Then the overall mean (period 0 weighted) price change for the year t long term price relative for commodity n is $\sum_{m=1}^{12} s_n^{0,m}(p_n^{t,m}/p_n^{0,m})$, which is just the right hand side of (16). Theil made just this argument, except he applied it to the logarithmic price ratios, $\ln (p_n^{t,m} / p_n^{0,m})$, instead of the price ratios themselves, $p_n^{t,m} / p_n^{0,m}$, and he ended up providing a stochastic interpretation for the index defined by (17).

In the context of producing superlative indexes, the lack of a symmetric treatment of the time periods 0 and t in the formulae (16)-(18) is troublesome: the base period plays an asymmetric role in these formulae.[15] In order to obtain symmetric formulae, we let $\mu(a,b)$ be a *symmetric mean* of the positive numbers a and b. Then we can use this mean function in order to define the following symmetric counterparts to the index number formulae (16)-(18):

[15] This lack of symmetry also troubled Theil (1967; 136-137) who went on to provide a symmetric version of formula (17). Thus in addition to his first stochastic index, $P_{T,0} \equiv \sum_{m=1}^{12} s_n^{0,m} \ln (p_n^{t,m}/p_n^{0,m})$, Theil argued that an equally valid stochastic index could be obtained using the period t shares, $s_n^{t,m}$, as probability weights, which leads to the index $P_{T,1} \equiv \sum_{m=1}^{12} s_n^{1,m} \ln (p_n^{t,m}/p_n^{0,m})$. Finally, Theil combined the two sets of probability weights by taking their arithmetic average, which leads to the final Theil type index, $P_T \equiv \sum_{m=1}^{12} (1/2)[s_n^{0,m} + s_n^{t,m}] \ln (p_n^{t,m}/p_n^{0,m})$. theil observed that the resulting index was not only symmetric in the data, but it also satisfied the time reversal test.

(19) $P_{C,\mu}(p_n^0,p_n^t,s_n^0,s_n^t) \equiv \sum_{m=1}^{12} \mu(s_n^{0,m},s_n^{t,m}) (p_n^{t,m} / p_n^{0,m})$; $n = 1,\ldots,N$; $t = 1,\ldots,T$;
(20) $\ln P_{J,\mu}(p_n^0,p_n^t,s_n^0,s_n^t) \equiv \sum_{m=1}^{12} \mu(s_n^{0,m},s_n^{t,m}) \ln (p_n^{t,m} / p_n^{0,m})$; $n = 1,\ldots,N$; $t = 1,\ldots,T$;
(21) $P_{D,\mu}(p_n^0,p_n^t,s_n^0,s_n^t) \equiv [\sum_{m=1}^{12} \mu(s_n^{0,m},s_n^{t,m}) p_n^{t,m}]/[\sum_{m=1}^{12} \mu(s_n^{0,m},s_n^{t,m}) p_n^{0,m}]$;
$n = 1,\ldots,N$; $t = 1,\ldots,T$.

If we take the perspective of Theil's stochastic approach to index number theory, then neither of the Dutot formulae, (18) or (21) seem to be satisfactory. A more satisfactory Dutot index from the stochastic viewpoint is:

(22) $P_D(p_n^0,p_n^t,s_n^0,s_n^t) \equiv [\sum_{m=1}^{12} s_n^{t,m} p_n^{t,m}]/[\sum_{m=1}^{12} s_n^{0,m} p_n^{0,m}]$; $n = 1,\ldots,N$; $t = 1,\ldots,T$.

The numerator in (22) is $\sum_{m=1}^{12} s_n^{t,m} p_n^{t,m}$ and it can be interpreted as the *year t expected price level for commodity n*, while the denominator in (22) is $\sum_{m=1}^{12} s_n^{0,m} p_n^{0,m}$ and it can be interpreted as the *year 0 expected price level for commodity n*. This long term annual price relative for commodity n in year t has the same disadvantage as our earlier unit value annual price relative defined by (5) above: the price weights in the numerator and denominator of these formulae are generally different.

How are we to choose the mean function μ and how are we to choose between the alternative weighted formulae (19) and (20) above? In order to answer this question, we need ask what properties we want our index to satisfy. We will ask that the index number formula satisfy two "reasonable" properties.

For our first "reasonable" property, we ask that the index, $P(p_n^0,p_n^t,s_n^0,s_n^t)$ say, satisfy the following *time reversal test*:

(23) $P(p_n^0,p_n^t,s_n^0,s_n^t)\, P(p_n^t,p_n^0,s_n^t,s_n^0) = 1$.

It can be seen that the generalized Carli index, $P_{C,\mu}(p_n^0,p_n^t,s_n^0,s_n^t)$ defined by (19), does not satisfy the time reversal test whereas the generalized Jevons index, $P_{J,\mu}(p_n^0,p_n^t,s_n^0,s_n^t)$ defined by (20), does satisfy this test provided that the function $\mu(a,b)$ is symmetric; i.e., we have $\mu(a,b) = \mu(b,a)$ for all $a > 0$ and $b > 0$.

For our second "reasonable" property, we ask that the index $P(p_n^0,p_n^t,s_n^0,s_n^t)$ satisfy the following *linear homogeneity in period t prices* property:

(24) $P(p_n^0,\lambda p_n^t,s_n^0,s_n^t) = \lambda\, P(p_n^0,p_n^t,s_n^0,s_n^t)$ for all $\lambda > 0$.

Applying this test to index numbers of the form defined by (20) where μ is a symmetric mean leads to the following condition that the weights $\mu(s_n^{0,m},s_n^{t,m})$ must satisfy:

(25) $\sum_{m=1}^{12} \mu(s_n^{0,m},s_n^{t,m}) = 1$ for all weights $s_n^{0,m}$ and $s_n^{t,m}$ which separately sum to 1 over m.

Condition (25) is also reasonable from the viewpoint of Theil's stochastic approach; we want the overall probabilities $\mu(s_n^{0,m},s_n^{t,m})$ (which treat the probabilities, $s_n^{0,m}$ and $s_n^{t,m}$, in each year for month m in a symmetric manner) of selecting each long term monthly price relative for commodity n, $p_n^{t,m}/p_n^{0,m}$, to sum to unity. In the Appendix, we show that the only homogeneous symmetric mean μ, which satisfies the functional equation (25), is the simple arithmetic mean. Using this simple mean, formula (20) becomes the Törnqvist-Theil index number formula defined as follows:[16]

$$(26)\ \ln P_T(p_n^0,p_n^t,s_n^0,s_n^t) \equiv \sum_{m=1}^{12} (1/2)[s_n^{0,m} + s_n^{t,m}] \ln (p_n^{t,m} / p_n^{0,m}) ; \qquad n = 1,\ldots,N ; t = 1,\ldots,T.$$

This concludes our discussion of the problem of how should the monthly price information at the lowest level of aggregation be aggregated over months to form an annual price level or price relative. From the national accounts perspective, the unit value formula (5) for constructing annual long term price relatives seems "best". From the viewpoint of (weighted) stochastic approaches to index number theory, we could take either a price *levels* approach or a price *relatives* approach to the problem of constructing annual prices. The stochastic price levels approach leads to the generalized Dutot formula (22) as being "best" while the stochastic price relatives or price ratios approach leads to the Törnqvist-Theil index (26) as being "best". Thus we have tentatively narrowed down the bewildering array of possible methods for aggregating monthly price information into annual prices into just three alternatives: (5), (22) or (26). The interesting point to note is that all three of our "best" approaches *involved weighting the monthly prices by monthly shares that reflected either the relative importance of monthly consumption of the commodity* (formula (5)) *or the relative importance of monthly expenditures on the commodity* (formulae (22) and (26)). Thus each of our "best" approaches to the monthly aggregation of prices problem involved the use of *monthly* quantity or expenditure *weights*. This makes intuitive sense since months are not of equal length and hence are not of equal economic importance. Of course, weather and custom further intensify the economic inequality in the consumption of any commodity n over the months of the year; i.e., there are definite *seasonal fluctuations* in the movement of prices and quantities and our methods for aggregating prices over months should take these fluctuations into account.

In the following section, we assume that the Törnqvist-Theil index number formula, P_T defined by (26), is used in order to aggregate up the monthly price information into annual prices. We then assume that the annual index is also constructed by using the Törnqvist-Theil formula. Finally, we indicate how this two stage aggregation procedure can be justified from the viewpoint of the economic approach to index number theory.

3. Two Stage Aggregation: Approach 1

Suppose that the long term annual price relatives for each commodity n are constructed using the Törnqvist-Theil index P_T defined by (26) above. We now go through the mechanics for

[16] Note that we have provided a somewhat deeper justification for Theil's choice of the probability weighting function, $\mu(a,b) = (1/2)a + (1/2)b$.

constructing a final overall annual index using the Törnqvist-Theil formula at the final stage of aggregation as well.

It is useful to add a bit of new notation. Let V_n^t be the *year t annual expenditures on commodity n*; i.e., we have:

$$(27)\quad V_n^t \equiv \sum_{m=1}^{12} p_n^{t,m} q_n^{t,m} \; ; \qquad n = 1,\ldots,N \; ; \; t = 0,1,\ldots,T.$$

For the base year, we define *the annual price for commodity n*, P_n^0 say, to be unity for all commodities and we define the *corresponding annual quantities*, Q_n^0 say, to be equal to the corresponding base year expenditures on commodity n, V_n^0; i.e., we make the following normalizations for the base year annual prices and quantities:

$$(28)\quad P_n^0 \equiv 1 \; ; \qquad n = 1,\ldots,N \; ;$$
$$(29)\quad Q_n^0 \equiv V_n^0 \; ; \qquad n = 1,\ldots,N.$$

For the years t which follow the base year, we define the *annual year t price for commodity n* by using the Törnqvist-Theil formula, (26); i.e., define P_n^t by:

$$(30)\quad \ln P_n^t \equiv \sum_{m=1}^{12} (1/2)[s_n^{0,m} + s_n^{t,m}] \ln (p_n^{t,m} / p_n^{0,m}) \; ; \qquad n = 1,\ldots,N \; ; \; t = 1,\ldots,T.$$

At this point, we need to deal with *strongly seasonal commodities*; i.e., it is very likely that for some commodity classes n, there are some months m where consumer expenditures on this commodity class are *zero*. Assuming that these zero expenditure months always occur for the same months each year, then the required modification for formula (30) is straightforward: simply sum over months m for which there are positive expenditures. In what follows, we assume that this modification has been made if necessary.

The corresponding *annual quantity for commodity n in year t*, Q_n^t, is defined by deflating the year t expenditure on commodity n, V_n^t, by the year t price for commodity n, P_n^t defined by (30); i.e., we have:

$$(31)\quad Q_n^t \equiv V_n^t / P_n^t \; ; \qquad n = 1,\ldots,N \; ; \; t = 1,\ldots,T.$$

Now define the *annual expenditure share on commodity n for year t* by

$$(32)\quad S_n^t \equiv V_n^t / \sum_{k=1}^{N} V_k^t \; ; \qquad n = 1,\ldots,N \; ; \; t = 0,1,\ldots,T.$$

Using the above definitions, we can now use the Törnqvist-Theil formula, (26), in order to aggregate the annual commodity prices, P_n^t, into an *overall price level*, P^t, as follows:

$$(33)\quad P^t \equiv \sum_{n=1}^{N} (1/2)[S_n^0 + S_n^t] \ln (P_n^t / P_n^0)$$
$$= \sum_{n=1}^{N} (1/2)[S_n^0 + S_n^t] \ln (P_n^t) \qquad t = 1,\ldots,T$$

where the last equality follows from the normalizations (28).

The two stage aggregation procedure, which led to the final overall price level for period t, P^t defined by (33), can be given an economic justification. Suppose that the consumer maximizes an annual utility function U, which has the following form:

$$(34)\quad u = U(q_1^1,\ldots,q_N^1;q_1^2,\ldots,q_N^2;\ldots;q_1^{12},\ldots,q_N^{12})$$

where q_n^m is the consumer's consumption of commodity n in month m of the year; i.e., we are following the example of Mudgett (1955) and Stone (1956; 74-75) and distinguishing *each commodity n in each month m as a separate commodity* which enters the consumer's utility function. Suppose further that the annual utility function U has the following *separable* functional form:[17]

$$(35)\quad U(q_1^1,\ldots,q_N^1;q_1^2,\ldots,q_N^2;\ldots;q_1^{12},\ldots,q_N^{12}) \equiv F[f_1(q_1^1,\ldots,q_1^{12}),f_2(q_2^1,\ldots,q_2^{12}),\ldots,f_N(q_N^1,\ldots,q_N^{12})]$$

so that there is an annual subaggregator function for commodity n, $f_n(q_n^1,\ldots,q_n^{12})$, that aggregates the monthly consumption of commodity n into an annual aggregate for commodity n, for n = 1,…,N. For simplicity, we will assume that the micro aggregator functions, $f_1,\ldots,f_N$, and the macro aggregator function F are all linearly homogeneous, concave and nondecreasing functions of their respective variables[18]. Thus these aggregator functions have unit cost function duals, $c_1,\ldots,c_N$ and C respectively, and these dual unit cost functions satisfy the same regularity conditions.[19]

Recall definitions (8), $p_n^t \equiv (p_n^{t,1}, p_n^{t,2},\ldots, p_n^{t,12})$, which defined the year t price vectors for commodity n. Under the usual assumptions of cost minimizing behavior on the part of consumers and given that the utility function U satisfies the separability conditions (35) plus the assumption that the c_n are translog unit cost functions, then it can be shown that the Törnqvist-Theil annual index numbers defined by (26) are *exact* for these translog unit cost functions[20]; i.e, we have:

$$(36)\quad \ln [c_n(p_n^t)/c_n(p_n^0)] = \textstyle\sum_{m=1}^{12} (1/2)[s_n^{0,m} + s_n^{t,m}] \ln (p_n^{t,m} / p_n^{0,m}) ; \qquad n = 1,\ldots,N ; t = 1,\ldots,T.$$

If in addition the unit cost function C that is dual to the macro aggregator function F is also has the translog functional form, then it can be shown that the annual price index P^t defined above by (33) is *exactly* equal to the following *economic cost of living index*:

$$(37)\quad P^t = C[c_1(p_1^t),c_2(p_2^t),\ldots,c_N(p_N^t)] / C[c_1(p_1^0),c_2(p_2^0),\ldots,c_N(p_N^0)] ; \qquad t = 1,\ldots,T.$$

Thus the use of the Törnqvist-Theil index number formula (26) in order to calculate annual prices from monthly prices not only has a strong justification from the stochastic approach to index number theory, it also has a very strong justification from the viewpoint of the economic approach to index number theory. Note that the two stage aggregation procedure that we have

---

[17] If some commodities n are not consumed in month m, then the corresponding quantities q_n^m are simply omitted in the list of arguments for the annual utility function U.
[18] The linear homogeneity assumption on F can be relaxed.
[19] See for example, Diewert (1993a; 120-122).
[20] See Diewert (1976; 121).

outlined in this section has the added benefit that it can deal with the problem of strongly seasonal commodities in a perfectly adequate manner. However, the separability assumptions that we have made in this section are not the most natural ones that can deal with the problems of seasonal commodities. In this section, we aggregated first over seasons, holding the commodity constant, and then aggregated over commodities. In the next two sections, we explore an alternative two stage aggregation procedure where we first aggregate over commodities holding the season constant and then aggregate over seasons.[21]

4. Year over Year Monthly Indexes

It seems clear that the separability assumptions on the annual utility function U that were made in (35) above are not the most natural. It is intuitively more appealing to think of the consumer as aggregating over commodities within a month in the first stage of aggregation rather than aggregating over months in the first stage. Thus in this section and the following one, we will assume that the consumer's annual utility function satisfies the following separability restrictions:

(38) $U(q_1^1,\ldots,q_N^1;q_1^2,\ldots,q_N^2;\ldots;q_1^{12},\ldots,q_N^{12}) \equiv G[g^1(q_1^1,\ldots,q_N^1);g^2(q_1^2,\ldots,q_N^2);\ldots;g^{12}(q_1^{12},\ldots,q_N^{12})]$.

Thus the month 1 micro aggregator function, $g^1(q_1^1,\ldots,q_N^1)$, aggregates up the consumption of each commodity q_n^1 in January into a January subutility, the month 2 micro aggregator function, $g^2(q_1^2,\ldots,q_N^2)$, aggregates up the consumption of each commodity q_n^2 in February into a February subutility and so on. The macro aggregator function G then aggregates up these monthly subutilities into an annual utility. Again, for simplicity, we will assume that the monthly micro aggregator functions, $g^1,\ldots,g^{12}$ and the macro aggregator function G are all linearly homogeneous, concave and nondecreasing functions of their respective variables[22]. Thus these aggregator functions have unit cost function duals, say $d^1,\ldots,d^{12}$ and D respectively, and these dual unit cost functions satisfy the same regularity conditions.

Define the vector of month m prices in year t as follows:

(39) $p^{t,m} \equiv (p_1^{t,m},\ldots,p_N^{t,m})$; $\quad m = 1,\ldots,12$; $t = 0,1,\ldots,T$.

We also need to define the *expenditure share of commodity n for month m of year t* as follows:

(40) $s^*{}_n^{t,m} \equiv p_n^{t,m} q_n^{t,m} / \sum_{k=1}^N p_k^{t,m} q_k^{t,m}$; $\quad n = 1,\ldots,N$; $m = 1,\ldots,12$; $t = 0,1,\ldots,T$.

Now define a vector of these shares for month m of year t as follows:

(41) $s^{*t,m} \equiv (s^*{}_1^{t,m},\ldots,s^*{}_N^{t,m})$ $\quad m = 1,\ldots,12$; $t = 0,1,\ldots,T$.

[21] These alternative two stage aggregation procedures for constructing an annual index in the context of seasonal monthly data were first discussed by Balk (1980a) (1980b) and Diewert (1980; 506-508).
[22] The linear homogeneity assumption on G can be relaxed.

The reader should note the difference between the earlier expenditure shares, $s_n^{t,m}$ defined by (14), and the new expenditure shares, $s*_n^{t,m}$, defined by (40): the $s_n^{t,m}$ sum to one over the months m for each commodity n and each year t whereas the $s*_n^{t,m}$ sum to one over the commodities n for each month m and each year t.

We can use the commodity shares of expenditure in years t and 0 and in month m to construct the following Törnqvist-Theil price index which compares the prices in month m of year t, $p^{t,m}$, with the prices of month m in the base year, $p^{t,0}$:

$$(42)\ \ln P_T(p^{0,m},p^{t,m},s*^{0,m},s*^{t,m}) \equiv \sum_{n=1}^{N}(1/2)[s*_n^{0,m} + s*_n^{t,m}]\ln(p_n^{t,m} / p_n^{0,m})\ ;\ m = 1,\ldots,12; t = 1,\ldots,T.$$

Under the usual assumptions of cost minimizing behavior on the part of consumers and given that the utility function U satisfies the separability conditions (38) plus the assumption that the d^m are translog unit cost functions, then it can be shown that the Törnqvist-Theil year over year monthly index numbers defined by (42) are *exact* for these translog unit cost functions; i.e, we have:

$$(43)\ \ln[d^m(p^{t,m})/d^m(p^{0,m})] = \sum_{n=1}^{N}(1/2)[s*_n^{0,m} + s*_n^{t,m}]\ln(p_n^{t,m} / p_n^{0,m})\ ;\ m = 1,\ldots,12\ ;\ t = 1,\ldots,T.$$

Thus the fixed base monthly Törnqvist-Theil indexes defined by (42), which compare the prices in month m of year t with the prices of the *same* month m in the base year, can be given a strong economic justification. Note that these 12 year over year monthly indexes should be free of seasonal influences. In the following section, we show how these monthly indexes can be aggregated up to give us an annual index.

5. Two Stage Aggregation: Approach 2

Suppose that the year over year fixed base monthly price indexes, $d^m(p^{t,m})/d^m(p^{0,m})$, are defined by equations (43), where the Törnqvist-Theil index number formula is used on the right hand side of (43). We now go through the mechanics for constructing a final overall annual index using the Törnqvist-Theil formula to aggregate over these year over year monthly indexes.

Again, we need a bit of new notation. Let $V^{t,m}$ be *the month m expenditures on all commodities in year t*; i.e., we have:

$$(44)\ V^{t,m} \equiv \sum_{n=1}^{N} p_n^{t,m} q_n^{t,m}\ ; \qquad m = 1,\ldots,12\ ;\ t = 0,1,\ldots,T.$$

For the base year, we define the *aggregate (over commodities) month m price level*, $P^{0,m}$ say, to be unity for all 12 months and we define the *corresponding aggregate monthly base period quantities*, $Q^{0,m}$ say, to be equal to the corresponding base year expenditures on all commodities during month m, $V^{0,m}$; i.e., we make the following normalizations for the base year monthly aggregated prices and quantities:

$$(45)\ P^{0,m} \equiv 1\ ; \qquad m = 1,\ldots,12\ ;$$
$$(46)\ Q^{0,m} \equiv V^{0,m}\ ; \qquad m = 1,\ldots,12.$$

For the years t which follow the base year, we define *the month m price levels for year t*, $P^{t,m}$, by using the Törnqvist-Theil formula, (42); i.e., define $P^{t,m}$ by:

(47) $\ln P^{t,m} \equiv \sum_{n=1}^{N}(1/2)[s^*{}_n^{0,m} + s^*{}_n^{t,m}] \ln(p_n^{t,m} / p_n^{0,m})$; m = 1,…,12 ; t = 1,…,T.

Again at this point, we need to deal with *strongly seasonal commodities*; i.e., it is very likely that for some months m, there are commodities n where consumer expenditures on this commodity class are *zero*. Assuming that these zero expenditure commodities in a given month are always zero expenditure commodities over all years for that month, then the required modification for formula (47) is straightforward: simply sum over the commodities n for which there are positive expenditures. In what follows, we assume that this modification has been made if necessary.

The corresponding *aggregate monthly quantity for month m in year t*, $Q^{t,m}$, is defined by deflating the month m expenditures on commodities in year t, $V^{t,m}$, by the month m aggregate price in year t, $P^{t,m}$ defined by (47); i.e., we have:

(48) $Q^{t,m} \equiv V^{t,m} / P^{t,m}$; m = 1,…,12 ; t = 1,…,T.

Now define the annual expenditure share on all commodities during month m in year t by

(49) $S^{t,m} \equiv V^{t,m} / \sum_{k=1}^{N} V^{t,k}$; m = 1,…,12 ; t = 0,1,…,T.

Using the above definitions, we can now use the Törnqvist-Theil formula in order to aggregate the monthly aggregate prices in year t, $P^{t,m}$, into an *overall price level,* P^{*t}, as follows:

(50) $$\begin{aligned} P^{*t} &\equiv \sum_{m=1}^{12} (1/2)[S^{0,m} + S^{t,m}] \ln (P^{t,m} / P^{0,m}) \\ &= \sum_{m=1}^{12} (1/2)[S^{0,m} + S^{t,m}] \ln P^{t,m} \end{aligned}$$

where the last equality follows from the normalizations (45).

The above two stage aggregation procedure, which led to the final overall price level for period t, P^{*t} defined by (50), can be given an economic justification. Under the usual assumptions of cost minimizing behavior on the part of consumers and given that the utility function U satisfies the separability conditions (38) plus the assumptions that the micro unit cost functions d^m are translog unit cost functions and that the macro unit cost function D is translog, then it can be shown that the two stage Törnqvist-Theil annual index number, P^{*t} defined by (50), is *exactly* equal to the following *economic cost of living index*:

(51) $P^{*t} = D[d^1(p^{t,1}),d^2(p^{t,2}),\ldots,d^{12}(p^{t,12})] / D[d^1(p^{0,1}),d^2(p^{0,2}),\ldots,d^{12}(p^{0,12})]$; t = 1,…,T;

where the vector of commodity prices in month m of year t is defined as

(52) $p^{t,m} \equiv (p_1^{t,m},\ldots,p_N^{t,m})$; m = 1,…,12 ; t = 0,1,…,T.

Thus the use of the Törnqvist-Theil index number formula in order to first calculate year over year monthly indexes and then aggregate these monthly indexes into an annual fixed base aggregate index also has a very strong justification from the viewpoint of the economic approach to index number theory. As was the case with the alternative two stage aggregation procedure outlined in section 3 above, the present two stage aggregation procedure can also deal with the problem of strongly seasonal commodities in a perfectly adequate manner. Furthermore, the separability assumptions that we have made in this section are perhaps the most natural ones that can deal with the problems caused by the existence of seasonal commodities.

The question now arises: how does the two stage price level P^{*t} constructed by (50) above compare to our earlier two stage estimate P^t defined by (33)? In order to provide an answer to this question, we will take a slight detour and outline yet another method for constructing an annual price index.

Our third economic approach to the construction of an annual price index is also the one that makes the least restrictive assumptions. Recall the general annual utility function U, which had no separability restrictions placed on it; i.e., recall (34) above. We now assume that U is dual to a general translog unit cost function, c say.[23] We further assume that c has the translog functional form. There will be a Törnqvist-Theil index number formula that is exact for this set of assumptions. In order to define this index, it is necessary to introduce yet a third set of definitions for expenditure shares. Thus we define the *expenditure share of commodity n in month m of total annual expenditures,* $s^{**}{}_n^{t,m}$, as follows:

(53) $s^{**}{}_n^{t,m} \equiv p_n^{t,m} q_n^{t,m} / \sum_{j=1}^{N} \sum_{k=1}^{12} p_j^{t,k} q_j^{t,k}$; $\qquad t = 0,1,\ldots,T.$

Now we can use the Törnqvist-Theil index number formula, *treating each commodity in each month in a year as a separate commodity*, in order to define the following sequence of annual indexes:

(54) $\ln P^{**t} \equiv \sum_{n=1}^{N} \sum_{m=1}^{12} (1/2)[s^{**}{}_n^{0,m} + s^{**}{}_n^{t,m}] \ln(p_n^{t,m} / p_n^{0,m})$; $\qquad t = 0,1,\ldots,T.$

Under the usual assumptions about cost minimizing behavior on the part of consumers, it can be shown that the single stage Törnqvist-Theil annual index number, P^{**t} defined by (54), is *exactly* equal to the following *economic cost of living index*:

(55) $P^{**t} = c[p^{t,1},p^{t,2},\ldots,p^{t,12}] / c[p^{0,1},p^{0,2},\ldots,p^{0,12}]$; $\qquad t = 1,\ldots,T.$

It turns out that the annual indexes P^{**t} defined by (54) will normally be very close to the two stage annual indexes P^{*t} defined earlier by (50). If we regard each of these two index number formulae as functions of the price and quantity vectors that pertain to periods 0 and t, then it can be shown that these two index number formulae approximate each other to the second order around any point where the two price vectors are equal and where the two quantity vectors are

[23] If there are strongly seasonal commodities so that say q_n^m does not appear in the utility function U defined by (34), then the corresponding price p_n^m will not appear as an argument in the dual cost function c and when we calculate our overall index, terms involving the subscript n and the superscript m should be omitted in the formula.

equal.[24] Thus if period t is not too distant from the base period and price and quantity fluctuations are not too violent, then P^{**t} will typically differ from P^{*t} only in the third decimal place. Similarly[25], it can be shown that P^t defined by (33) will approximate P^{**t} also to the second order around any point where the two price vectors are equal and where the two quantity vectors are equal. Hence, for normal time series data, *the two stage Törnqvist-Theil indexes constructed in section 3 above will be very close to their counterpart two stage Törnqvist-Theil indexes constructed in this section*. They will also be very close to the Törnqvist-Theil index that constructs the annual indexes in a single stage. This latter index is exact for an economic index that makes the least restrictive assumptions of the three economic models that we have considered.

This completes our discussion on the problems involved in constructing a sequence of superlative annual indexes using monthly data for calendar years. However, if we have monthly data, there is no need to restrict ourselves to the construction of annual indexes for only calendar years. In the following section, we look at the problems involved in constructing annual indexes for a moving sequence of 12 consecutive months.

6. Moving Year Annual Superlative Indexes

As was mentioned in the introduction, it seems inefficient to patiently collect monthly price and quantity information for 12 consecutive months and then produce an annual price index once a year. How can we make better use of the incremental price and quantity information that flows into the statistical agency each month?

Recall our discussion of year over year monthly indexes in section 4 above and our subsequent discussion in section 5 where we aggregated these monthly indexes into calendar year annual indexes. However, there is no need to restrict ourselves to calendar years: the methodology explained in section 5 can be adapted to making a price comparison for any 12 consecutive months with the base 12 months.

We will not spell out the details of this moving year methodology. Additional information on this method can be found in Diewert (1983b) (1996) (1999) and in Alterman, Diewert and Feenstra (1999). It does seem to work well for the empirical examples for which it has been implemented: nice smooth series are obtained which are free from seasonal fluctuations.[26] After centering, the moving year sequence of monthly indexes is very comparable to a monthly price series that has been seasonally adjusted by X-11, except that the moving year indexes tend to be somewhat smoother. Of course, the price index for every twelfth month reduces to a calendar year comparison with the base year.

---

[24] See Diewert (1978; 895).

[25] Again apply the result on the approximate consistency in aggregation of superlative index number formulas in Diewert (1978; 895).

[26] Comparisons of this moving year index number theory method for seasonal adjustment and more traditional econometric and statistical methods for seasonal adjustment may be found in Diewert (1999; 61-64) and in Alterman, Diewert and Feenstra (1999).

7. Month to Month Superlative Indexes

The indexes suggested above in sections 3,5 and 6 are all annual type indexes and thus they cannot give accurate information on the short term month to month change in prices.[27] Hence, there is a need for a month to month superlative price index.[28]

Using the Törnqvist-Theil formula as usual and the notation introduced in section 4, the short term movement of prices going from month m to month m+1 of year t is given by:[29]

$$(56)\quad \ln P_T(p^{t,m},p^{t,m+1},s^{*t,m},s^{*t,m+1}) \equiv \sum_{n=1}^{N}(1/2)[s^*{}_n{}^{t,m} + s^*{}_n{}^{t,m+1}]\ln(p_n{}^{t,m+1} / p_n{}^{t,m})\ ; \quad m = 1,\ldots,11\ ;\ t = 0,1,\ldots,T.$$

It can be seen that the short term price indexes defined by (56) are quite different than the indexes that we have defined up to this point and they serve a different purpose; i.e., to indicate month to month movements in prices rather than year over year movements.

Existing statistical agency short term price indexes tend to be of the Laspeyres type and so there could be a need for a superlative index along the lines of (56). We also note that existing statistical agency consumer price indexes tend to use *annual weights* from a somewhat *distant base year*. However, the weights in a short term superlative index like that defined by (56) are not only *current*, but they are also definitely *not* annual weights; they are *monthly weights* that pertain to the two months being compared. *If annual weights are used in (56), then the resulting index is definitely not superlative and indeed, it cannot be justified from either the economic approach to index number theory or from the perspective of Theil's stochastic approach to index number theory.* I also do not know of any test approaches to index number theory that deal with monthly prices and annual expenditure weights.

8. Aggregation over Regions and Households: Paasche and Laspeyres Indexes.

Up to this point, we have implicitly assumed a representative consumer model. In this section and the next one, we consider some of the problems involved in the construction of a superlative index when there are many households or regions in the economy and our goal is the production of a national index.

---

[27] The year over year monthly indexes introduced in section 4 definitely do not give any information on the short term month to month movement in prices.

[28] If there is a hyperinflation, all of the annual indexes that we have discussed are not very relevant without modification. In the case of a hyperinflation, the consumer's budget constraint for 12 consecutive months must have interest rate discount factors in order to sensibly compare prices between different months. A month to month index number is what is needed for indexation purposes under conditions of high inflation.

[29] Formula (56) needs to be modified when going from December to January. The formula also needs to be modified when there are strongly seasonal commodities; i.e, commodities n such that n is present in one period but not the other. In this case, the shares in (56) have to be redefined and the summation in (56) must be restricted to commodities n that are present in both periods. This is the *maximum overlap method* for constructing short term indexes that was used extensively by Alterman, Diewert and Feenstra (1999).

In this section, we will consider an economic approach to the CPI that is based on the *plutocratic cost of living index* that was originally defined by Prais (1959). This concept was further refined by Pollak (1980; 276) (1981; 328) who defined his *Scitovsky-Laspeyres cost of living index* as the ratio of total expenditure required to enable each household in the economy under consideration to attain its base period indifference surface at period 1 prices to that required at period 0 prices. Diewert (1983a; 190-192) (2000) generalized Pollak's analysis and we repeat some of his analysis below.

Suppose that there are H households (or regions) in the economy and suppose further that there are N commodities in the economy in periods 0 and 1 that households consume *and* that we wish to include in our definition of the cost of living.[30] Denote an N dimensional vector of commodity consumption in a given period by $q \equiv (q_1,q_2,\ldots, q_N)$. Denote the vector of period t market prices faced by household h by $p_h^t \equiv (p_{h1}^t,p_{h2}^t,\ldots,p_{hN}^t)$ for t = 0,1.[31] In addition to the market commodities that are in the vector q, we assume that each household is affected by an M dimensional vector of *environmental*[32] *or demographic*[33] *variables or public goods*, $e \equiv (e_1,e_2,\ldots,e_M)$. We suppose that there are H households (or regions) in the economy during periods 0 and 1 and the preferences of household h over different combinations of market commodities q and environmental variables e can be represented by the continuous utility function $f^h(q,e)$ for $h = 1,2,\ldots,H$.[34] For periods t = 0,1 and for households $h = 1,2,\ldots,H$, it is assumed that the observed household h consumption vector $q_h^t \equiv (q_{h1}^t,\ldots,q_{hN}^t)$ is a solution to the following household h expenditure minimization problem:

$$(57)\quad \min_q \{ p_h^t \cdot q : f^h(q,e_h^t) \geq u_h^t \} \equiv C^h(u_h^t,e_h^t, p_h^t) \; ; \; t = 0,1; \; h = 1,2,\ldots H$$

where e_h^t is the environmental vector facing household h in period t, $u_h^t \equiv f^h(q_h^t,e_h^t)$ is the utility level achieved by household h during period t and C^h is the cost or expenditure function that is dual to the utility function f^h.[35] Basically, these assumptions mean that each household has stable preferences over the same list of commodities during the two periods under consideration, the same households appear in each period and each household chooses its consumption bundle in the most cost efficient way during each period, conditional on the environmental vector that it faces during each period. Note that the household (or regional) prices are in general different across households (or regions).

---

[30] The aggregation issues are the same whether we are attempting to construct a month to month superlative index as in the previous section or an annual Stone-Mudgett index as in the earlier sections.

[31] Using our previous notation, $p_n^{t,m}$ was the price of commodity n in month m of year t. We now replace t,m by a single period index t, which could denote a specific rolling year if our goal is to construct an annual superlative index or it could denote a specific month if our goal is to construct a superlative month to month index. Previously, N denoted the number of commodity classes that we are willing to consider in each month. If we are considering month to month superlative indexes, then our new N is equal to our old N; if we are considering annual moving year superlative indexes, then our new N is equal to 12 times our old N.

[32] This is the terminology used by Pollak (1989; 181) in his model of the conditional cost of living concept.

[33] Caves, Christensen and Diewert (1982; 1409) used the terms *demographic variables* or *public goods* to describe the vector of conditioning variables e in their generalized model of the Konüs price index or cost of living index.

[34] We assume that each $f^h(q,e)$ is continuous and increasing in the components of q and e and is quasiconcave in the components of q.

[35] Note that $p\cdot q \equiv \sum_{n=1}^N p_n q_n$ is the inner product between the vectors p and q.

With the above assumptions in mind, we generalize Pollak (1980) (1981) and Diewert (1983a; 190)[36] and define the class of *conditional plutocratic cost of living indexes*, $P^*(p^0,p^1,u,e_1,e_2,\ldots,e_H)$, pertaining to periods 0 and 1 for the arbitrary utility vector of household utilities $u \equiv (u_1,u_2,\ldots,u_H)$ and for the arbitrary vectors of household environmental variables e_h for h = 1,2,…,H as follows:

$$(58)\quad P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u,e_1,e_2,\ldots,e_H) \equiv \sum_{h=1}^H C^h(u_h,e_h,p_h^1) / \sum_{h=1}^H C^h(u_h,e_h,p_h^0) .$$

The numerator on the right hand side of (58) is the sum over households of the minimum cost, $C^h(u_h,e_h,p_h^1)$, for household h to achieve the arbitrary utility level u_h, given that the household h faces the arbitrary vector of household h environmental variables e_h and also faces the period 1 vector of prices p_h^1. The denominator on the right hand side of (58) is the sum over households of the minimum cost, $C^h(u_h,e_h,p_h^0)$, for household h to achieve the same arbitrary utility level u_h, given that the household faces the same arbitrary vector of household h environmental variables e_h and also faces the period 0 vector of prices p_h^0.

We now specialize the general definition (58) by replacing the general utility vector u by either the period 0 vector of household utilities $u^0 \equiv (u_1^0,u_2^0,\ldots u_H^0)$ or the period 1 vector of household utilities $u^1 \equiv (u_1^1,u_2^1,\ldots u_H^1)$. We also specialize the general definition (58) by replacing the general household environmental vectors $(e_1,e_2,\ldots e_H) \equiv e$ by either the period 0 vector of household environmental variables $e^0 \equiv (e_1^0,e_2^0,\ldots e_H^0)$ or the period 1 vector of household environmental variables $e^1 \equiv (e_1^1,e_2^1,\ldots,e_H^1)$. The choice of the base period vector of utility levels and base period environmental variables leads to the *Laspeyres conditional plutocratic cost of living index*, $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^0,e^0)$[37], while the choice of the period 1 vector of utility levels and period 1 environmental variables leads to the *Paasche conditional plutocratic cost of living index*, $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^1,e^1)$. It turns out that these last two indexes satisfy some interesting inequalities, which we derive below.

Using definition (58), the *Laspeyres plutocratic conditional cost of living index*, $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^0,e^0)$, may be written as follows:

$$
\begin{aligned}
(59)\quad & P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^0,e_1^0,e_2^0,\ldots,e_H^0) \\
&\equiv \sum_{h=1}^H C^h(u_h^0,e_h^0,p_h^1) / \sum_{h=1}^H C^h(u_h^0,e_h^0,p_h^0) \\
&= \sum_{h=1}^H C^h(u_h^0,e_h^0,p_h^1) / \sum_{h=1}^H p_h^0\cdot q_h^0 \qquad \text{using (57) for } t = 0 \\
&\leq \sum_{h=1}^H p_h^1\cdot q_h^0 / \sum_{h=1}^H p_h^0\cdot q_h^0 \\
&\qquad \text{since } C^h(u_h^0,e_h^0,p_h^1) \equiv \min_q \{ p_h^1\cdot q : f^h(q,e_h^0) \geq u_h^0 \} \leq p^1\cdot q_h^0 \text{ and } q_h^0 \\
&\qquad \text{is feasible for the cost minimization problem for } h = 1,2,\ldots,H \\
&\equiv P_{DL}
\end{aligned}
$$

[36] These authors provided generalisations of the plutocratic cost of living index due to Prais (1959). Pollak and Diewert did not include the environmental variables in their definitions of a group cost of living index.

[37] This is the concept of a cost of living index that Triplett (1999; 27) finds most useful for measuring inflation: "One might want to produce a COL *conditional* on the base period's weather experience…. In this case, the unusually cold winter does not affect the *conditional* COL subindex that holds the environment constant. … the COL subindex that holds the environment constant is probably the COL concept that is most useful for an anti-inflation policy." Hill (1999; 4) endorses this point of view.

where P_{DL} is defined to be the *disaggregated (over households) Laspeyres price index*, $\sum_{h=1}^{H} p_h^1 \cdot q_h^0 / \sum_{h=1}^{H} p_h^0 \cdot q_h^0$, which uses the individual vectors of household or regional quantities for period 0, $(q_1^0,\ldots,q_H^0)$, as quantity weights.[38] If prices are equal across households (or regions), so that

(60) $p_h^t = p^t$ for $t = 0,1$ and $h = 1,2,\ldots,H$,

then the disaggregated Laspeyres price index P_{DL} collapses down to the usual aggregate Laspeyres index, P_L; i.e., if (60) holds, then P_{DL} in (59) becomes

(61)
$$\begin{aligned} P_{DL} &\equiv \sum_{h=1}^{H} p_h^1 \cdot q_h^0 / \sum_{h=1}^{H} p_h^0 \cdot q_h^0 \\ &= p^1 \cdot \sum_{h=1}^{H} q_h^0 / p^0 \cdot \sum_{h=1}^{H} q_h^0 \\ &= p^1 \cdot q^0 / p^0 \cdot q^0 \\ &\equiv P_L \end{aligned}$$

where the total quantity vector in period t is defined as

(62) $q^t \equiv \sum_{h=1}^{H} q_h^t$ for $t = 0,1$.

The inequality (59) says that the theoretical Laspeyres plutocratic conditional cost of living index, $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^0,e^0)$, is bounded from above by the observable (in principle) disaggregated Laspeyres price index P_{DL}. The special case of inequality (59) when the equal prices assumption (60) holds was first obtained by Pollak (1989; 182) for the case of one household with environmental variables and by Pollak (1980; 276) for the many household case but where the environmental variables are absent from the household utility and cost functions.

In a similar manner, using definition (58), *the Paasche conditional plutocratic cost of living index*, $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^1,e^1)$, may be written as follows:

(63)
$$\begin{aligned} &P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^1,e_1^1,e_2^1,\ldots,e_H^1) \\ &\quad\equiv \sum_{h=1}^{H} C^h(u_h^1,e_h^1,p_h^1) / \sum_{h=1}^{H} C^h(u_h^1,e_h^1,p_h^0) \\ &\quad= \sum_{h=1}^{H} p_h^1 \cdot q_h^1 / \sum_{h=1}^{H} C^h(u_h^1,e_h^1,p_h^0) && \text{using (57) for } t = 1 \\ &\quad\geq \sum_{h=1}^{H} p_h^1 \cdot q_h^1 / \sum_{h=1}^{H} p_h^0 \cdot q_h^1 && \text{using a feasibility argument} \\ &\quad\equiv P_{DP} \end{aligned}$$

where P_{DP} is defined to be the *disaggregated (over households) Paasche price index*, $\sum_{h=1}^{H} p_h^1 \cdot q_h^1 / \sum_{h=1}^{H} p_h^0 \cdot q_h^1$, which uses the individual vectors of household quantities for period 1, $(q_1^1,\ldots,q_H^1)$, as quantity weights.

If prices are equal across households (or regions), so that assumptions (60) hold, then the disaggregated Paasche price index P_{DP} collapses down to the usual aggregate Paasche index, P_P; i.e., if (60) holds, then P_{DP} in (63) becomes

[38] Thus the disaggregated Laspeyres index can be regarded as an ordinary Laspeyres index except that each commodity in each region is regarded as a separate commodity.

(64) $P_{DP} \equiv \sum_{h=1}^{H} p_h^1 \cdot q_h^1 / \sum_{h=1}^{H} p_h^0 \cdot q_h^1$
$= p^1 \cdot \sum_{h=1}^{H} q_h^1 / p^0 \cdot \sum_{h=1}^{H} q_h^1$
$= p^1 \cdot q^1 / p^0 \cdot q^1$
$\equiv P_P$.

Returning to the inequality (63), we see that the theoretical Paasche conditional plutocratic cost of living index, $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^1,e^1)$, is bounded from below by the observable disaggregated Paasche price index P_{DP}. Diewert (1983a; 191) first obtained the inequality (63) for the case where the environmental variables are absent from the household utility and cost functions and prices are equal across households.

Using the inequalities (59) and (63) and the continuity properties of the conditional plutocratic cost of living $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u,e)$ defined by (58), it is possible to modify the method of proof used by Konüs (1924) and Diewert (1983a; 191) and establish the following result:[39]

Proposition 1: Under our assumptions, there exists a reference utility vector $u^* \equiv (u_1^*,u_2^*,\ldots,u_H^*)$ such that the household h reference utility level u_h^* lies between the household h period 0 and 1 utility levels, u_h^0 and u_h^1 respectively for h = 1,…,H, and there exist household environmental vectors $e_h^* \equiv (e_{h1}^*,e_{h2}^*,\ldots,e_{hM}^*)$ such that the household h reference mth environmental variable e_{hm}^* lies between the household h period 0 and 1 levels for the mth environmental variable, e_{hm}^0 and e_{hm}^1 respectively for m = 1,2,…,M and h = 1,…,H, and the conditional plutocratic cost of living index $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^*,e^*)$ evaluated at this intermediate reference utility vector u^* and the intermediate reference vector of household environmental variables $e^* \equiv (e_1^*,e_2^*,\ldots,e_H^*)$ lies between the observable (in principle) disaggregated Laspeyres and Paasche price indexes, P_{DL} and P_{DP}, defined above by the last equalities in (59) and (63).

Thus the theoretical national consumer price index $P^*(p_1^0,\ldots,p_H^0,p_1^1,\ldots,p_H^1,u^*,e^*)$ is bounded from above by the disaggregated Laspeyres index P_{DL} and from below by the disaggregated Paasche index P_{DP}. Hence if P_{DL} and P_{DP} are not too different, a good point approximation to the theoretical national consumer price index will be the disaggregated Fisher index P_{DF} defined as:

(65) $P_{DF} \equiv [P_{DL}\, P_{DP}]^{1/2}$.

The disaggregated Fisher price index P_{DF} is computed just like the usual Fisher price index, except that each commodity in each region is regarded as a separate commodity.

Since statistical agencies do not calculate Laspeyres, Paasche and Fisher price indexes by taking inner products of price and quantity vectors as we have done in (59) and (63), it will be useful to obtain formulae for the Laspeyres and Paasche indexes that depend only on price relatives and expenditure shares. In order to do this, we need to introduce some notation. Define the *expenditure share of household h on commodity n in period t* as

[39] Note that the household cost functions must be continuous in the environmental variables which is a real restriction on the types of environmental variables which can be accommodated by the Proposition.

(66) $s_{hn}^t \equiv p_{hn}^t q_{hn}^t / \sum_{j=1}^N p_{hj}^t q_{hj}^t$; $t = 0,1$; $h = 1,2,\ldots,H$; $n = 1,2,\ldots,N$.

Define the *expenditure share of household h of total period t consumption* as:

(67) $S_h^t \equiv \sum_{n=1}^N p_{hn}^t q_{hn}^t / \sum_{k=1}^H \sum_{j=1}^N p_{hj}^t q_{hj}^t$
$= p_h^t \cdot q_h^t / \sum_{k=1}^H p_k^t \cdot q_k^t$ $t = 0,1$; $h = 1,2,\ldots,H$.

Finally, define the *national expenditure share of commodity n in period t* as:

(68) $\sigma_n^t \equiv \sum_{h=1}^H p_{hn}^t q_{hn}^t / \sum_{k=1}^H p_k^t \cdot q_k^t$ $t = 0,1$; $n = 1,2,\ldots,N$
$= \sum_{h=1}^H [p_{hn}^t q_{hn}^t / p_h^t \cdot q_h^t][p_h^t \cdot q_h^t / \sum_{k=1}^H p_k^t \cdot q_k^t]$
$= \sum_{h=1}^H s_{hn}^t \; p_h^t \cdot q_h^t / \sum_{k=1}^H p_k^t \cdot q_k^t]$ using definitions (66)
$= \sum_{h=1}^H s_{hn}^t S_h^t$ using definitions (67).

The *Laspeyres price index for region h* (or household h) is defined as:

(69) $P_{Lh} \equiv p_h^1 \cdot q_h^0 / \; p_h^0 \cdot q_h^0$ $h = 1,2,\ldots,H$
$= \sum_{n=1}^N (p_{hn}^1 / p_{hn}^0) p_{hn}^0 q_{hn}^0 / p_h^0 \cdot q_h^0$
$= \sum_{n=1}^N \; s_{hn}^0 (p_{hn}^1 / p_{hn}^0)$ using definitions (66).

Looking back at (59), the disaggregated national Laspeyres price index P_{DL} is defined as follows:

(69) $P_{DL} \equiv \sum_{h=1}^H \; p_h^1 \cdot q_h^0 / \sum_{h=1}^H p_h^0 \cdot q_h^0$
$= \sum_{h=1}^H \; [p_h^1 \cdot q_h^0 / p_h^0 \cdot q_h^0][p_h^0 \cdot q_h^0 / \sum_{h=1}^H p_h^0 \cdot q_h^0]$
$= \sum_{h=1}^H \; [p_h^1 \cdot q_h^0 / p_h^0 \cdot q_h^0] S_h^0$ using definitions (67) with $t = 0$
(70) $= \sum_{h=1}^H S_h^0 P_{Lh}$ using definitions (69)
$= \sum_{h=1}^H S_h^0 \sum_{n=1}^N \; s_{hn}^0 (p_{hn}^1 / p_{hn}^0)$ using the last line of (69)
(71) $= \sum_{h=1}^H \sum_{n=1}^N S_h^0 s_{hn}^0 (p_{hn}^1 / p_{hn}^0)$ rearranging terms.

Equation (70) shows that the national Laspeyres price index is equal to a (period 0) regional expenditure share weighted average of the regional Laspeyres price indexes. Equation (71) shows that the national Laspeyres price index is equal to a period 0 expenditure share weighted average of the regional price relatives, (p_{hn}^1 / p_{hn}^0), where the corresponding weight, $S_h^0 s_{hn}^0$, is the period 0 national expenditure share of commodity n in region h.

The *Paasche price index for region h* (or household h) is defined as:

(72) $P_{Ph} \equiv p_h^1 \cdot q_h^1 / \; p_h^0 \cdot q_h^1$ $h = 1,2,\ldots,H$
$= 1 / \sum_{n=1}^N (p_{hn}^0 / p_{hn}^1) p_{hn}^1 q_{hn}^1 / p_h^1 \cdot q_h^1$
$= 1 / \sum_{n=1}^N \; s_{hn}^1 (p_{hn}^1 / p_{hn}^0)^{-1}$ using definitions (66)
$= [\sum_{n=1}^N \; s_{hn}^1 (p_{hn}^1 / p_{hn}^0)^{-1}]^{-1}$.

From (63), the disaggregated national Paasche price index P_{DP} is defined as follows:

(73) $P_{DP} \equiv \sum_{k=1}^H \; p_k^1 \cdot q_k^1 / \sum_{h=1}^H p_h^0 \cdot q_h^1$

$= 1 / \{\sum_{h=1}^{H} [p_h^0 \cdot q_h^1 / p_h^1 \cdot q_h^1][p_h^1 \cdot q_h^1 / \sum_{k=1}^{H} p_k^1 \cdot q_k^1]\}$

$= 1 / \sum_{h=1}^{H} [p_h^1 \cdot q_h^0 / p_h^0 \cdot q_h^0]^{-1} S_h^1$ using definitions (67) with t = 1

(74) $= [\sum_{h=1}^{H} S_h^1 P_{Lh}^{-1}]^{-1}$ using definitions (72)

$= [\sum_{h=1}^{H} S_h^1 \sum_{n=1}^{N} s_{hn}^1 (p_{hn}^1 / p_{hn}^0)^{-1}]^{-1}$ using the last line of (72)

(75) $= [\sum_{h=1}^{H} \sum_{n=1}^{N} S_h^1 s_{hn}^1 (p_{hn}^1 / p_{hn}^0)^{-1}]^{-1}$ rearranging terms.

Equation (73) shows that the national Paasche price index is equal to a (period 1) regional expenditure share weighted harmonic mean of the regional Paasche price indexes. Equation (75) shows that the national Paasche price index is equal to a period 1 expenditure share weighted harmonic average of the regional price relatives, (p_{hn}^1 / p_{hn}^0), where the weight for this price relative, $S_h^1 s_{hn}^1$, is the period 1 national expenditure share of commodity n in region h.

Now if prices are equal across regions, the formulae (71) and (75) simplify. (71) becomes:

(76) $P_{DL} = \sum_{h=1}^{H} \sum_{n=1}^{N} S_h^0 s_{hn}^0 (p_{hn}^1 / p_{hn}^0)$

$= \sum_{h=1}^{H} \sum_{n=1}^{N} S_h^0 s_{hn}^0 (p_n^1 / p_n^0)$ using assumptions (60)

$= \sum_{n=1}^{N} \sigma_n^0 (p_n^1 / p_n^0)$ using (68) for t = 0

and (75) becomes

(77) $P_{DP} = [\sum_{h=1}^{H} \sum_{n=1}^{N} S_h^1 s_{hn}^1 (p_{hn}^1 / p_{hn}^0)^{-1}]^{-1}$

$= [\sum_{h=1}^{H} \sum_{n=1}^{N} S_h^1 s_{hn}^1 (p_n^1 / p_n^0)^{-1}]^{-1}$ using assumptions (60)

$= [\sum_{n=1}^{N} \sigma_n^1 (p_{hn}^1 / p_{hn}^0)^{-1}]^{-1}$ using (68) for t = 1.

Thus with the assumption that commodity prices are the same across regions, in order to calculate national Laspeyres and Paasche indexes, we require only "national" price relatives and national commodity expenditure shares for the two periods under consideration. However, if there is regional variation in prices, then the simplified formulae (76) and (77) are not valid and we must use our earlier formulae, (71) and (75).

It is now necessary to discuss a practical problem that faces statistical agencies: namely, that existing household consumer expenditure surveys, which are used in order to form estimates of household expenditure shares, *are not very accurate*. Thus the detailed commodity by region expenditure shares, $S_h^0 s_{hn}^0$ and $S_h^1 s_{hn}^1$, which appear in the formulae (71) and (75) are generally measured with very large errors. Hence, it may lead to less overall error if the regional commodity expenditure shares s_{hn}^t are replaced by the national commodity expenditure shares σ_n^t defined by (68). Whether this approximation is justified would depend on a detailed analysis of the situation facing the statistical agency.

In the following section, instead of using the conditional plutocratic aggregate price index P* defined by (58) as our target national index, we use Theil's stochastic approach as our national target index.

9. Aggregation over Regions and Households: Theil's Stochastic Approach

Recall Theil's (1967; 136-137) stochastic approach to the determination of the aggregate average price change between periods 0 and 1, explained in section 2 above. Using the notation introduced in the previous section, in order to apply Theil's approach in the present context, we simply weight the logarithm of each regional price relative, p_{hn}^1/ p_{hn}^0, by its average expenditure share of national expenditure over the two periods under consideration, $(1/2)S_h^0 s_{hn}^0 + (1/2)S_h^1 s_{hn}^1$, in order to obtain the logarithm of the average price change, P_T; i.e., we have

$$(78)\quad \ln P_T = \sum_{h=1}^H \sum_{n=1}^N (1/2)[S_h^0 s_{hn}^0 + S_h^1 s_{hn}^1] \ln (p_{hn}^1/ p_{hn}^0).$$

Using the usual second order approximation properties of the Fisher and Törnqvist-Theil indexes[40], it can be shown that the disaggregated Fisher index P_{DF} defined by (65) above approximates P_T defined by (78) to the second order around any equal (or proportional) price and quantity vectors for the two periods under consideration.

Suppose instead of using Theil's stochastic approach to index number theory, we wanted to use the economic approach to index number theory and aggregate up the regional or individual household Törnqvist indexes into a national price index. Thus, define the logarithm of the Törnqvist index P_{Th} for household (or region) h as follows:

$$(79)\quad \ln P_{Th} \equiv \sum_{n=1}^N (1/2)[s_{hn}^0 + s_{hn}^1] \ln(p_{hn}^1/ p_{hn}^0) ; \qquad h = 1,2,\ldots,H$$

where the household expenditure shares s_{hn}^t are defined by (66) in the previous section. Now suppose each household h has preferences that are dual to the general (nonhomothetic) translog cost function $C_h(u_h,p_h)$. Under the assumption of optimizing behavior for each household in each period, Diewert (1976; 122) shows that the Törnqvist index P_{Th} is exactly equal to the Könus true cost of living $C_h(u_h^*,p_h^1) / C_h(u_h^*,p_h^0)$ where u_h^* is the geometric mean of the period 0 and 1 utility levels for household h; i.e., we have[41]

$$(80)\quad P_{Th} = C_h(u_h^*,p_h^1) / C_h(u_h^*,p_h^0) ; \qquad h = 1,2,\ldots,H.$$

We now want to form some sort of average of the individual household true Könus cost of living indexes or true price indexes. In order to obtain an index that most closely resembles the stochastic index (78), we will take a weighted geometric average of the household h true indexes. Thus define the *social cost of living index P_S with geometric weights w_h* as follows:

$$(81)\quad \ln P_S(w_1,\ldots,w_H) \equiv \sum_{h=1}^H w_h \ln [C_h(u_h^*,p_h^1) / C_h(u_h^*,p_h^0)].$$

For a democratic index, we would choose the weights $w_h = 1/H$; i.e., we would choose equal weights.

[40] See Diewert (1978; 886-889).
[41] For analogous exact results in household translog models with environmental variables, see Caves, Christensen and Diewert (1982; 1409-1411).

For a plutocratic index, a reasonable choice for the weights would be the average expenditure share of region (or household) h in national expenditure over the two periods under consideration:

(82) $w_h = (1/2)S_h^0 + (1/2)S_h^1$; $h = 1,2,\ldots,H$

where the S_h^t are the shares of household h in period t's national consumption; see (67) above. Now substitute (79), (80) and (82) into (81) and we obtain the following expression for the plutocratic social cost of living index:

(83) $\ln P_S[(1/2)S_1^0+(1/2)S_1^1,\ldots, (1/2)S_H^0+(1/2)S_H^1]$
$= \sum_{h=1}^H \sum_{n=1}^N (1/2)[S_h^0 + S_h^1](1/2)[s_{hn}^0 + s_{hn}^1] \ln (p_{hn}^1/ p_{hn}^0)$.

Now compare (78) with (83); i.e., compare the stochastic approach measure of aggregate price change to the plutocratic economic approach. Somewhat surprisingly, we see that the two estimates of national price change between the two periods are not identical.[42] It can also be seen that the two measures of aggregate price change are equal if either: (i) the regional shares S_h^t are identical for the two periods or (ii) the household expenditure shares on commodities are identical for each household for the two periods; i.e., if either

(84) $S_h^0 = S_h^1$ for $h = 1,2,\ldots,H$ or
(85) $s_{hn}^0 = s_{hn}^1$ for $h = 1,2,\ldots,H$ and $n = 1,2,\ldots,N$.

However, in general, the stochastic approach estimate of aggregate price change defined by (78) will not equal the plutocratic social cost of living index defined by (83), which means that the statistical agency will have to make a choice between these alternative approaches.

As was mentioned in the previous section, the individual household (or regional) commodity expenditure shares, s_{hn}^0 and s_{hn}^1, which appear in the formulae (78) and (83) are generally measured with very large errors. Hence again, it may lead to less overall error if the regional commodity expenditure shares s_{hn}^t are replaced by the national commodity expenditure shares σ_n^t defined by (68).

10. Conclusions

In addition to the existing timely CPI that most statistical agencies currently produce, it would be useful for the agency to produce *3 alternative classes of superlative consumer price indexes*:

- A short term month to month superlative CPI should be produced using monthly weights; see section 7 above. This index would be the best measuring stick for very short term movements in prices. This index would *not* be free of seasonal influences.
- Twelve separate year over year superlative indexes for each month of the year should be produced; see section 4 above. These indexes are of interest to the business community

[42] It can be verified that both sets of share weights sum to unity.

because they are free of seasonal influences and give an accurate picture of price movements over the past year in each current month. These indexes are also the basic building blocks for the next class of index.

- A superlative index should be produced, which compares the prices of the last 12 months with the corresponding months in a base year; see section 6 above. These moving year indexes would be available on a monthly frequency and they would be free of seasonal influences. They would be counterparts to existing price series that are adjusted by various econometric and statistical methods such as X-11. However, I believe that these moving year superlative indexes have many advantages over series that are adjusted by existing seasonal adjustment methods.

A possible disadvantage of producing 4 separate CPI's is that the public might be confused. However, they all serve very different purposes. If I had to pick just a single one of the above 3 alternative superlative indexes, I would pick the moving year indexes. Next, I would pick the short term superlative index. It ranks behind my first choice because the existing CPI could serve much the same purposes.

There are problems of aggregation over regions or households associated with producing any one of the above three classes of superlative index. These aggregation problems were briefly discussed in sections 8 and 9 above. The main problem is that monthly commodity expenditure shares by type of household or by region *are subject to large measurement errors*. This means that these expenditure shares will have to be approximated or estimated in some manner. We suggested the use of national expenditure shares in place of the regional expenditure shares but other estimation and smoothing methods might be appropriate. These regional aggregation problems mean that the construction of an annual superlative index number will not be a trivial exercise.

A final recommendation is that the BLS make every effort to obtain accurate expenditure data as well as accurate price data. The production of superlative indexes requires accuracy in both sets of variables.

Appendix: On Justifying the Arithmetic Average in the Törnqvist-Theil Index Formula.

The functional equation that we study is the following one:

(A1) $\sum_{m=1}^{12} \mu(s^m,S^m) = 1$ for all weights $s^m \geq 0$ and $S^m \geq 0$ such that

(A2) $\sum_{m=1}^{12} s^m = 1$ and $\sum_{m=1}^{12} S^m = 1$,

where μ is a homogeneous, symmetric mean. We follow Diewert (1993b; 361-364) and define a homogeneous, symmetric mean $\mu(x,y)$ as a function of two nonnegative variables that has the following properties:

(A3) $\mu(x,x) = x$; (mean property);

(A4) $\mu(x,y) = \mu(y,x)$; (symmetry property);
(A5) $\mu(x,y)$ is a continuous function over its domain of definition;
(A6) $\mu(x,y)$ is increasing in its components and
(A7) $\mu(\lambda x,\lambda y) = \lambda\mu(x,y)$ for all $\lambda > 0$ and $x \geq 0, y \geq 0$; (homogeneity property).

Let:

(A8) $s^m = (1/12)$ for $m = 1,2,\ldots,12$;
(A9) $S^m = (1/12) + (1/12)\, x_m$ for $m = 1,2,\ldots,11$ and
(A10) $S^{12} = (1/12) - (1/12) \sum_{m=1}^{11} x_m$.

Substitution of (A8)-(A10) into the functional equation (A1) leads to the following functional equation for $x_m \geq -1$ for each m and $\sum_{m=1}^{11} x_m \leq 1$:

(A11) $\sum_{m=1}^{11} \mu[1/12, (1/12)(1 + x_m)] + \mu[1/12, (1/12)(1 - \sum_{m=1}^{11} x_m)] = 1.$

Multiply both sides of (A11) by 12 and use the homogeneity property (A7) in order to transform (A11) into the following functional equation for $x_m \geq -1$ for each m and $\sum_{m=1}^{11} x_m \leq 1$:

(A12) $\sum_{m=1}^{11} \mu[1, 1 + x_m] = 12 - \mu[1, 1 - \sum_{m=1}^{11} x_m]$.

Now define the following functions f and g in terms of μ as follows:

(A13) $f(x) \equiv \mu(1,1 + x)$;
(A14) $g(y) \equiv 12 - f(-y) = 12 - \mu(1,1-y)$.

Substitution of (A13) and (A14) into (A12) leads to the following functional equation: for $x_m \geq -1$ for each m and $\sum_{m=1}^{11} x_m \leq 1$:

(A15) $\sum_{m=1}^{11} f(x_m) = g(\sum_{m=1}^{11} x_m)$.

From Theorem 2.6.3 in Eichhorn (1978; 39), the functional equation (A15) has the following solution:[43]

(A16) $f(x) \equiv \alpha x + \beta$;
(A17) $g(y) \equiv \alpha y + 11\beta$

where α and β are real numbers. Note that

(A18) $f(0) \equiv \mu(1,1)$ using (A13)
$= 1$ using (A3)
$= \alpha 0 + \beta$ using (A16)
$= \beta$.

[43] We require either property (A5) or (A6) on μ in order to apply Eichhorn's result.

Thus the parameter β in the solution (A16) and (A17) must equal 1. In our particular application, we need to check whether Eichhorn's general solution also satisfies the additional restriction in (A14); i.e., whether f and g defined by (A16) and (A17) also satisfy g(y) = 12 − f(−y). We have :

(A18) $g(y) \equiv \alpha y + 11\beta$ using (A17)
$= \alpha y + 11$ using (A18)
$= 12 - 1 - \alpha(-y)$
$= 12 - [\alpha(-y) + 1]$
$= 12 - f(-y)$ using definition (A16) with $\beta = 1$.

We now use the symmetry property (A4) in order to determine the constant α. Using (A13):

(A19) $f(x) \equiv \mu(1,1 + x)$
$= \mu(1 + x,1)$ using (A4)
$= (1 + x)\, \mu(1,[1 + x]^{-1})$ using (A7)
$= (1 + x)\, \mu(1,1 - x[1 + x]^{-1})$
$= (1 + x)\, f(-x[1 + x]^{-1})$ using definition (A13)
$= (1 + x)\, [\alpha(-x[1 + x]^{-1}) + 1]$ using (A16) and (A18).

Again using (A16) and (A18), we have:

(A20) $f(x) = \alpha x + 1$
$= (1 + x)\, [\alpha(-x[1 + x]^{-1}) + 1]$ using (A19).

Solving (A20) for α leads to the solution

(A21) α = _.

Thus

(A22) $f(x) = \mu(1,1 + x) = (1/2)\, x + 1$.

Now set 1 + x = c and multiply both sides of (A22) by a to obtain the following equation:

(A23) $a\, \mu(1,c) = a\, [(1/2)\{c - 1\} + 1]$ or
$\mu(a,ac) = (1/2)\, ac + (1/2)\, a$ using (A7) or
$\mu(a,b) = (1/2)\, b + (1/2)\, a$ letting $b \equiv ac$.

References

Alterman, W.F., W.E. Diewert and R.C. Feenstra, (1999*), International Trade Price Indexes and Seasonal Commodities*, Bureau of Labor Statistics, Washington D.C.

Balk, B.M. (1980a), *Seasonal Products in Agriculture and Horticulture and Methods for Computing Price Indices*, Statistical Studies No. 24, The Hague: Netherlands Central Bureau of Statistics.

Balk, B.M. (1980b), “Seasonal Commodities and the construction of Annual and Monthly Price Indexes”, *Statistische Hefte* 21, 110-116.

Caves, D.W., L.R. Christensen and W.E. Diewert (1982), “The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity”, *Econometrica* 50, 1393-14-14.

Davies, G.R. (1924), “The Problem of a Standard Index Number Formula”, *Journal of the American Statistical Association* 19, 180-188.

Diewert, W.E. (1976), “Exact and Superlative Index Numbers,” *Journal of Econometrics* 4, 115-146.

Diewert, W.E. (1978), Superlative Index Numbers and Consistency in Aggregation”, *Econometrica* 46, 883-900.

Diewert, W.E. (1980), “aggregation Problems in the Measurement of Capital”, pp. 433-528 in *the Measurement of Capital*, D. Usher (ed.), Chicago: The University of Chicago Press.

Diewert, W.E. (1983a), “The Theory of the Cost of Living Index and the Measurement of Welfare Change”, pp. 163-233 in *Price Level Measurement*, W.E. Diewert and C. Montmarquette (eds.), Ottawa: Statistics Canada, reprinted as pp. 79-147 in *Price Level Measurement*, W.E. Diewert (ed.), Amsterdam: North-Holland, 1990.

Diewert, W.E. (1983b), “The Treatment of Seasonality in a Cost of Living Index”, pp. 1019-1045 in *Price Level Measurement*, W.E. Diewert and C. Montmarquette (eds.), Ottawa: Statistics Canada.

Diewert, W.E. (1993a), “Duality Approaches to Microeconomic Theory”, pp. 105-175 in *Essays in Index Number Theory*, Volume 1, W.E. Diewert and A.O. Nakamura (eds.), Amsterdam: North-Holland.

Diewert, W.E. (1993b), “Symmetric Means and Choice Under Uncertainty”, pp. 355-433 in *Essays in Index Number Theory*, Volume 1, W.E. Diewert and A.O. Nakamura (eds.), Amsterdam: North-Holland.

Diewert, W.E. (1995), Axiomatic and Economic Approaches to Elementary Price Indexes”, Discussion Paper 95-01, Department of Economics, University of British Columbia, Vancouver, Canada, January.

Diewert, W.E. (1996), “Seasonal Commodities, High Inflation and Index Number Theory”, Discussion Paper 96-06, Department of Economics, University of British Columbia, Vancouver, Canada, V6T 1Z1.

Diewert, W.E. (1998), High Inflation, Seasonal Commodities and Annual Index Numbers”, *Macroeconomic Dynamics* 2, 456-471.

Diewert, W.E. (1999), “Index Number Approaches to Seasonal Adjustment”, *Macroeconomic Dynamics* 3, 1-21.

Diewert, W.E. (2000), “The Consumer Price Index and Index Number Purpose”, Discussion Paper 00-02, Department of Economics, University of British Columbia, Vancouver, Canada, V6T 1Z1.

Eichhorn, W. (1978), *Functional Equations in Economics*, Reading, MA: Addison-Wesley.
Fisher, I. (1922), *The Making of Index Numbers*, Boston: Houghton Mifflin Co.
Hicks, J.R. (1946), *Value and Capital*, Second Edition, Oxford: the Clarendon Press.
Hill, T.P. (1993), “Price and Volume Measures”, pp. 379-406 in *System of National Accounts 1993*, Eurostat, IMF, OECD, UN and World Bank, Luxembourg, Washington, D.C., Paris, New York, and Washington, D.C.
Hill, T. P. (1996), *Inflation Accounting: A Manual on National Accounting Under Conditions of High Inflation*, Paris: O.E.C.D.
Hill, T.P. (1999), “COL Indexes and Inflation Indexes”, paper tabled at the 5th Meeting of the Ottawa Group on Price Indices, Reykjavik, Iceland, August 25-27, 1999.
Keynes, J.M. (1930), *Treatise on Money*, Volume 1, London: MacMillan.
Konüs, A.A. (1924), “The Problem of the True Index of the Cost of Living”, translated in *Econometrica* 7, (1939), 10-29.
Mudgett, B.D. (1955), “The Measurement of Seasonal Movements in Price and Quantity Indexes”, *Journal of the American Statistical Association* 50, 93-98.
Pollak, R.A. (1980), “Group Cost-of-Living Indexes”, *American Economic Review*70, 273-278.
Pollak, R.A. (1981), “The Social Cost-of-Living Index”, *Journal of Public Economics* 15, 311-336.
Pollak, R.A. (1989), “The Treatment of the Environment in the Cost-of –Living Index”, pp. 181-185 in R.A. Pollak, *The Theory of the Cost-of-Living Index*, Oxford: Oxford University Press.
Prais, S.J. (1959), “Whose Cost of Living?”, *The Review of Economic Studies* 26, 126-134.
Stone, R. (1956), *Quantity and Price Indexes in National Accounts*, Paris: The Organization for European Cooperation.
Theil, H. (1967), *Economics and Information Theory*, Amsterdam: North-Holland.
Triplett, J.E. (1999), “Should the Cost-of-Living Index Provide the Conceptual Framework for a Consumer Price Index?”, paper presented at the 5th Meeting of the Ottawa Group on Price Indices, Reykjavik, Iceland, August 25-27, 1999.
Walsh, C.M. (1901), *The Measurement of General Exchange Value*, New York: MacMillan.
Walsh, C.M. (1921), *The Problem of Estimation*, London: P.S. King and Son.