On the Construction of Price Indexes

Diane F. Primont, Southeast Missouri State University

Abstract. This paper discusses some important issues in constructing price indexes. The issues relate to the choice of an index number formula. Which is a "good" index number formula for aggregating prices? A variety of alternative index number formulae are explored based on the economic approach and the axiomatic (test) approach. In practice, use of the Laspeyres price index is common. For example, this is the formula used for the official US consumer price index (prior to 1999). Some of the problems that arise in adopting the Laspeyres price index are noted.

Keywords: Price indexes

1. INTRODUCTION

This paper discusses some important issues in constructing price indexes. The issues relate to the choice of an index number formula. Which is a "good" index number formula for aggregating prices? A variety of alternative index number formulae are explored based on the economic approach and the axiomatic (test) approach.

In practice, use of the Laspeyres price index is common. For example, this is the index number formula used for the official US consumer price index (prior to 1999). Some of the problems that arise in adopting the Laspeyres price index are noted.

2. INDEX NUMBER THEORY

Denote $\mathbf{p}^t=(p^t_1,\ldots,p^t_N)$ as the non-negative vector of prices of N commodities and $\mathbf{q}^t=(q^t_1,\ldots,q^t_N)$ as the corresponding vector of the quantities in period t=0,1. How do we construct a price index $\mathbf{P}$ that compares the aggregate of prices of the N commodities in period 1 with the aggregate of prices of the N commodities in period 0? This is the problem of index number theory. There are two main approaches to index number theory: the economic approach and the axiomatic (test) approach (Diewert, 1993a).

Each of these approaches is briefly discussed in the context of a consumer price index. The approaches can also be applied in the context of a producer price index. (See, for example, Alterman, *et. al.*, 1999, 10-33.) The objective of each is to determine which index number formulae for the price index are "good."

2.1 The Economic Approach

Under the economic approach, a consumer is assumed to choose $\mathbf{q}^t$ to minimize the cost of obtaining a particular level of well-being (say, u), given the price vector $\mathbf{p}^t$ and a vector of other variables that affect consumer well-being (Diewert, 1993c). Denote the vector of other variables as $\mathbf{v}=(v_1,\ldots,v_M)$. These variables would include such things as health, current and expected future income, publicly provides goods (such as national defense, police and fire protection, roads), and environmental quality.

Thus, the consumer's cost minimization problem is

$$C(\mathbf{p}^t,u)= \min_{q\geq 0}\{\mathbf{p}^t \bullet \mathbf{q}: U(\mathbf{q},\mathbf{vbar})\geq u\} \qquad (1)$$

where U is the consumer's utility function, $u = U(\mathbf{qbar},\mathbf{vbar})$, and $\mathbf{qbar},\mathbf{vbar}$ are reference vectors of quantities. C is linearly homogenous.

Cost minimizing behavior in periods 0 and 1 leads to a theoretical cost of living index defined as

$$\mathbf{P}(\mathbf{p}^1, \mathbf{p}^0, \mathbf{qbar}, \mathbf{vbar}) = C(\mathbf{p}^1,\mathbf{qbar},\mathbf{vbar}) / C(\mathbf{p}^0,\mathbf{qbar},\mathbf{vbar}) \qquad (2)$$

If we ignore the vector $\mathbf{v}$, then (2) may be written as

$$\mathbf{P}(\mathbf{p}^1, \mathbf{p}^0, u) = C(\mathbf{p}^1,u) / C(\mathbf{p}^0,u). \qquad (3)$$

The theoretical cost of living index (consumer price index) compares the cost to a consumer at two different points of time of maintaining a constant level of well-being. It is the ratio of minimum costs of achieving a given level of well-being in a reference period (say, period 1) to a base period (say, period 0).

The theoretical consumer price index given by either (2) or (3) is unobservable, since we don't know the exact functional form for *C* or *U*. How can this index be constructed in practice using the observable data $\boldsymbol{p}^t$ and $\boldsymbol{q}^t$ for period t=0,1?

Under the assumption of consumer cost-minimizing behavior, two obvious choices for the reference vector ***qbar*** are ***qbar***=$\boldsymbol{q}^1$ or ***qbar***=$\boldsymbol{q}^0$. If we choose ***qbar***=$\boldsymbol{q}^0$ then we get the Laspeyres price index ($\boldsymbol{P_L}$):

$$\boldsymbol{P_L}(\boldsymbol{p}^1,\boldsymbol{p}^0,\boldsymbol{q}^0) = \boldsymbol{p}^1\cdot \boldsymbol{q}^0 / \boldsymbol{p}^0\cdot \boldsymbol{q}^0 \qquad (4)$$

The Laspeyres price index is the ratio of the cost of the market basket of goods and services from period 0 at the prices in period 1 to the cost of the identical market basket at the prices in period 0. Because the market basket $\boldsymbol{q}^0$ is feasible in period 1, but not necessarily cost-minimizing,

$$C(\boldsymbol{p}^1,\boldsymbol{q}^1,\boldsymbol{v}^1) \leq \boldsymbol{p}^1\cdot \boldsymbol{q}^0 \ . \qquad (5)$$

The (observed) Laspeyres price index is an upper bound on the theoretical consumer price index.

Conversely, if we choose ***qbar***=$\boldsymbol{q}^1$, then we get the Paasche price index ($\boldsymbol{P_P}$):

$$\boldsymbol{P_P}(\boldsymbol{p}^1,\boldsymbol{p}^0,\boldsymbol{q}^1) = \boldsymbol{p}^1\cdot \boldsymbol{q}^1 / \boldsymbol{p}^0\cdot \boldsymbol{q}^1 \qquad (6)$$

The Paasche price index is the ratio of the cost of the market basket of goods and services from period 1 at the prices in period 1 to the cost of the identical market basket at the prices in period 0. Because the market basket $\boldsymbol{q}^1$ is feasible in period 0, but not necessarily cost-minimizing,

$$C(\boldsymbol{p}^0,\boldsymbol{q}^0,\boldsymbol{v}^0) \leq \boldsymbol{p}^0\cdot \boldsymbol{q}^1 \ . \qquad (7)$$

The (observed) Paasche price index is a lower bound on the theoretical consumer price index.

In addition, it can be shown that the theoretical consumer price index lies between the Laspeyres and Paasche price indexes, provided we choose a reference quantity vector ***qbar*** that is a weighted average of $\boldsymbol{q}^0$ and $\boldsymbol{q}^1$. Therefore, some weighted average of the Laspeyres and Paasche price indexes should be "close" to the unobserved theoretical consumer price index (Diewert, 1993b). For example, one such average is a *geometric mean* of the Laspeyres and Paasche price indexes, which is the Fisher ideal price index ($\boldsymbol{P_F}$):

$$\begin{aligned}\boldsymbol{P_F}(\boldsymbol{p}^1,\boldsymbol{p}^0,\boldsymbol{q}^1,\boldsymbol{q}^0) &= \boldsymbol{P_L}^{-}\boldsymbol{P_P}^{-} \\ &= (\boldsymbol{p}^1\cdot \boldsymbol{q}^0 / \boldsymbol{p}^0\cdot \boldsymbol{q}^0)^{-} (\boldsymbol{p}^1\cdot \boldsymbol{q}^1 / \boldsymbol{p}^0\cdot \boldsymbol{q}^1)^{-} \qquad (8)\end{aligned}$$

The (observed) Fisher ideal index provides a good approximation to the theoretical consumer price index (CPI).

Another method of constructing the theoretical price index from (2) or (3), according to the economic approach, is by choosing a specific functional form for the cost function *C*.

For example, suppose we assume consumer cost-minimizing behavior in both periods and choose a Cobb-Douglas cost function for (3)

$$C(\boldsymbol{p}, u) = (\prod_n p_n^{\,s_n})\, u \qquad (9)$$

where $s_n = (p_n\, q_n)/(\boldsymbol{p}\cdot \boldsymbol{q})$ and $\sum_n s_n = 1$. The Cobb-Douglas cost function is a first-order approximation to a continuously differentiable cost function. In the Cobb-Douglas cost function, the expenditure weights s_n are constant. All elasticities of substitution in consumption between commodities equal minus one. In this case, the theoretical CPI will exactly equal the observed price index:

$$\begin{aligned}\boldsymbol{P}(\boldsymbol{p}^1,\boldsymbol{p}^0,u) &= (\prod_n p^1_n{}^{s_n}) / (\prod_n p^0_n{}^{s_n}) \\ &= \prod_n (p^1_n / p^0_n)^{s_n} \qquad (10)\end{aligned}$$

which is a weighted geometric mean of the relative prices of the *N* commodities (Diewert, 1993c, 201-202). The weights s_n are commodity *n*'s share in total expenditures.

Instead, suppose we assume cost minimizing behavior, and choose a homogenous quadratic cost function in (3), then

$$C(\boldsymbol{p},u) = (\boldsymbol{p}\_B\boldsymbol{p})^{-}\, u \qquad (11)$$

where *B* is an *N_N* symmetric matrix of coefficients. This is a *flexible* cost function: it provides a second order approximation to a twice continuously differentiable cost function. In this case, the theoretical consumer price index is exactly equal to the observed Fisher ideal price index $\boldsymbol{P_F}$ (Diewert, 1993c, 205-206). Therefore, the Fisher ideal price index is "superlative." Diewert (1993d, 1993e) derives other superlative price index number formulae. Among the class of superlative index numbers, however, Diewert (1993d, 246) argues that the Fisher ideal price index is preferred.

2.2 The Axiomatic (Test) Approach

The axiomatic approach places no assumptions on consumer behavior: both $\boldsymbol{p}^t$ and $\boldsymbol{q}^t$ are considered exogenous variables for $t = 0,1$.

This approach starts with a number of "reasonable" axioms or tests that the price index $\boldsymbol{P}(\boldsymbol{p}^1,\boldsymbol{p}^0,\boldsymbol{q}^1,\boldsymbol{q}^0)$ should satisfy. These axioms, hopefully, determine a unique functional form or formula for $\boldsymbol{P}$. While there is no definitive list of axioms to be satisfied, Diewert (1993e) identifies 20 reasonable axioms that have been proposed by researchers over the years.

One of the most important of these axioms is the "time reversal test" which states that

$$\boldsymbol{P}(\boldsymbol{p}^1,\boldsymbol{p}^0,\boldsymbol{q}^1,\boldsymbol{q}^0) = 1 / \boldsymbol{P}(\boldsymbol{p}^0,\boldsymbol{p}^1,\boldsymbol{q}^0,\boldsymbol{q}^1) \quad . \tag{12}$$

If the price and quantity data for the periods 0 and 1 are interchanged, i.e., the order of time is reversed, the resulting price index will equal the reciprocal of the original price index.

Diewert (1993e) shows that the only index number formula that satisfies all 20 of the proposed axioms is the Fisher ideal price index given by (8). On the other hand, both the Laspeyres and Paasche price indexes, given by (4) and (5), respectively, satisfy 17 of the axioms, but, unfortunately, both fail the important time reversal test given by (12).

The axiomatic approach suggests that the Fisher ideal price index is the "best" index number formula.

2.3 Choice of an Index Number Formula

Both the economic and the axiomatic approaches to index number theory provide a strong justification for using the Fisher ideal price index (or other superlative index number formula) instead of the Laspeyres or Paasche price indexes.

However, in practice, use of the Laspeyres price index is common. For example, the official CPI (prior to January 1999) published by the U.S. Bureau of Labor Statistics (BLS) and the Exvessel Price Index published by the U.S. National Marine Fisheries Service are Laspeyres price indexes.

As discussed in section 2.1, one difficulty with using the Laspeyres price index $\boldsymbol{P}_L$ is that $\boldsymbol{P}_L$ is an upper bound on the theoretical consumer price index. The (observed) Laspeyres price index tends to *overstate* the theoretical CPI. In other words, $\boldsymbol{P}_L$ is biased upward relative to the theoretical CPI. The bias occurs because the Laspeyres price index given by (4) uses a *fixed* market basket of commodities $\boldsymbol{q}^0$. The quantities of the commodities are held fixed at $\boldsymbol{q}^0$ in subsequent periods even as prices change. If the price of Granny Smith apples rises and the price of red Delicious apples falls from period 0 to period 1, the consumer purchases the same quantities as before of each, according to (4). Similarly, if the price of peaches falls relative to the price of apples, the quantities of peaches and apples are held constant. In reality, however, we would expect consumers to substitute toward the relatively cheaper item, i.e., demand more red Delicious apples and fewer Granny Smith apples, or demand more peaches and fewer apples. Thus, this bias is referred to as the "substitution bias." (BLS, 1997b)

Note that an output (producer) price index calculated as a Laspeyres price index also suffers from the "substitution bias." However, the bias is in the reverse direction. The (observed) Laspeyres price index tends to *understate* the theoretical output price index. It is biased downward relative to the theoretical output price index. When an output price rises, the producer will supply more, *ceteris paribus*. (See, for example, Alterman, *et. al.,* 1999, 10-33 for a discussion of an output price index.)

Despite the substitution bias of the Laspeyres price index, the BLS has continued to use this formula to construct the official CPI rather than a superlative index number formula. A superlative price index formula, such as the Fisher ideal price index given by (8), uses both base period (period 0) and current (period 1) expenditures (or quantities). The BLS obtains expenditure data from the Consumer Expenditure household interview survey. However, these data are available only with a lag of about one year. Furthermore, the data are collected on a quarterly basis, whereas the official CPI is published monthly. Thus, data are not available to construct a superlative price index, at least on a monthly basis. The BLS is currently investigating the possibility of publishing an annual CPI using a superlative index number formula. (BLS, 1997b)

How important is the substitution bias? According to Alterman, *et. al.*, (1999, 22-23) "if the price change for each commodity *n* is "small" going from period 0 to period 1 (i.e., p_n^0 is close to p_n^1 for each *n*) and the quantity change for each commodity *n* is also "small" going from period 0 to period 1 (i.e., q_n^0 is close to q_n^1 for each *n*), then the difference between [the Laspeyres, Paasche, and Fisher ideal price indexes] will also be "small."" It is much less likely that price and quantity change will be "small" the greater the time difference between the reference period (period 0) and the current period (period 1). For example, suppose the current period is 1999. Price and quantity change for each commodity from a base period of 1982-84 is likely to be *much* larger that that from a base period of 1993-95 (the new reference period of the CPI introduced in January 1999). Updating the reference period should reduce the difference between $\boldsymbol{P_L}$ and $\boldsymbol{P_F}$, but the substitution bias will remain (Greenlees, *et. al.*, 1996, 4).

How large is the substitution bias in the official CPI? The BLS distinguishes between two levels of substitution bias, which arise from the way the official CPI is constructed. During the first step, called the "lower" level, individual prices for specific items falling in an item category (or stratum) are aggregated. So, for example, individual prices for all different kinds of apples (Granny Smith, red Delicious, etc.) are used to construct a Laspeyres price index for the item category *apples* for each of the geographic areas in the US. Similarly, the item categories *lettuce*, *milk*, *citrus fruits*, and *carbonated drinks*, use individual prices of all different kinds of lettuce, milk, citrus fruits, and carbonated drinks, respectively, to form $\boldsymbol{P_L}$ for each item category. In the second step, called the "upper" level, all of the various $\boldsymbol{P_L}$ for the item categories are used to construct a $\boldsymbol{P_L}$ for all item categories across all geographic regions. (BLS, 1997b)

BLS researchers have estimated the size of both the upper and lower level substitution biases. The substitution bias at the upper level is estimated to add 0.15 percentage point per year to the annual rate of increase of the official CPI (BLS, 1997b). The substitution bias at the lower level is estimated to add 0.2 to 0.24 percentage point per year to the annual rate of increase in the official CPI (BLS, 1998, 1997b). Thus, if the official consumer price index increases at a rate of 3.0 percentage points per year, the theoretical (true) increase in consumer prices is, instead, 2.6-2.75 percentage points per year.

In an effort to mitigate the lower level substitution bias in the official CPI, the BLS has revised the index number formula it uses for constructing item category price indexes, i.e., at the lower level, as of January 1999. Instead of the Laspeyres price index, the BLS has adopted a geometric mean price index, given by (10), for a large number of item categories (representing over 60 percent of total consumer spending) (BLS, 1998). From section 2.1, the (observed) geometric mean price index is exact for a Cobb-Douglas cost function, which has an elasticity of substitution in consumption of minus one, and uses constant expenditure share weights. Because expenditure share weights are constant, expenditure data for the current period are not needed. Suppose the price index for the item category *apples* is calculated using a geometric mean price index. Since the expenditure shares are constant, this means that the fraction of spending on *apples* attributed to Granny Smith apples, red Delicious apples, and each other type of apple is fixed. Consider the case of just two types of apples, Granny Smith apples and red Delicious apples. If the price of Granny Smith apples rises and the price of red Delicious apples falls, the geometric mean index formula implies that the quantity of Granny Smith apples bought falls and the quantity of red Delicious apples bought rises in such a way that the same original dollar amount is spent on each type of apple. (For more examples, see BLS, 1997a.) In this way, the geometric mean formula accounts for consumer substitution among items in an item category that would be expected to occur when relative prices change. However, even in the official CPI published since January 1999, the substitution bias at the upper level remains because the Laspeyres formula is employed at this level.

3. Summary and Conclusions

How do we construct a price index $\boldsymbol{P}$ that compares the aggregate of prices of the *N* commodities in period 1 with the aggregate of the prices of the *N* commodities in period 0? This is the problem of index number theory. There are two major approaches to index number theory that offer some guidance in choosing a price index formula: the economic approach and the axiomatic (test) approach.

Under the economic approach, prices are treated as exogenous and the consumer is assumed to choose the quantities of each of the *N* commodities to minimize the cost of achieving some given level of well-being in both periods. These assumptions allow the derivation of an (unobserved) theoretical price index.

Without placing additional restrictions on the cost function (or underlying utility function), we could evaluate several index number formulae. For a consumer price index, the theoretical price index lies between the (observed) Laspeyres price index and the (observed) Paasche price index. The (observed) Laspeyres price index is an upper bound on the (unobserved) theoretical price index, while the (observed) Paasche price index is a lower bound. The (observed) Laspeyres price index is biased upward, while the (observed) Paasche price index is biased downward relative to the (unobserved) theoretical consumer price index. The (observed) Fisher ideal price index, which is a geometric mean of the (observed) Laspeyres price index and the (observed) Paasche price index, was found to provide a good approximation to the (unobserved) theoretical price index.

For an output (producer) price index, the bias runs in the opposite direction. The (observed) Laspeyres price index is biased downward and the (observed) Paasche price index is biased upward relative to the (unobserved) theoretical output price index.

If we are willing to restrict the form of the cost function (or underlying utility function), then we can derive an (observed) price index that will be exactly equal to the (unobserved) theoretical price index. Two examples are the geometric mean price index and the Fisher ideal price index. These are called "exact" price index number formulae. If the specified cost function is also "flexible," then the corresponding (observed) price index formula is not only "exact," but is also a "superlative," such as the Fisher ideal price index.

No unique choice of an index number formula emerges from this approach. However, Diewert (1993d) argues that the (observed) Fisher ideal price index is the preferred approximation to the (unobserved) theoretical price index.

The axiomatic (or test) approach treats observed price and quantity data as exogenous, and, thus, makes no assumption regarding consumer behavior. The approach starts with a number of "reasonable" axioms or tests that the price index ***P*** should satisfy. Diewert (1993e) identifies 20 axioms. Only the Fisher ideal price index satisfies all 20. Of the two frequently used price index numbers, the Laspeyres price index and the Paasche price index, 17 axioms are satisfied, but both fail the important "time reversal test."

Both the economic and the axiomatic approaches to index number theory provide a strong justification for using the Fisher ideal price index (or other superlative index number formula) instead of the Laspeyres or Paasche price indexes. However, in practice, use of the Laspeyres price index is common. One example is the official US consumer price index (prior to 1999). The justification for the BLS's use of the Laspeyres price index is the unavailability of current quantity or expenditure data needed to construct a superlative price index on a monthly basis.

However, the fixed nature of the quantities of the commodities means that the Laspeyres price index exhibits a "substitution bias," which causes it to overstate the theoretical (true) increase in consumer prices. Nevertheless, if the Laspeyres price index must be used, this gap between the Laspeyres price index and the theoretical price index can be reduced by (*i*) frequent updating of the base period to reflect more recent expenditure patterns and (*ii*) using another index number formula -- such as the geometric mean price index -- to account for substitutability between similar commodities where applicable.

4. References

Alterman, William F., W. Erwin Diewert, and Robert C. Feenstra, *International Trade Price Indexes and Seasonal Commodities*, Washington, DC: US Bureau of Labor Statistics, 1999.

Bureau of Labor Statistics, The Experimental CPI using Geometric Means (CPI-U-

XG), [http://stats.bls.gov/cpigmrp.htm], 1997a.

Bureau of Labor Statistics, Measurement Issues in the Consumer Price Index, [http://stats.bls.gov/cpigm697.htm], 1997b.

Bureau of Labor Statistics, Planned Change in the Consumer Price Index Formula, [http://stats.bls.gov/pdf/cpipgm02.pdf], 16 April 1998.

Diewert, W. E., Overview of Volume 1, in *Essays in Index Number Theory Volume 1*, W. E. Diewert and A. O. Nakamura, eds., Amsterdam: North Holland Elsevier Science Publishers B.V., 1-31, 1993a.

Diewert, W. E., Index Numbers, in *Essays in Index Number Theory Volume 1*, W. E. Diewert and A. O. Nakamura, eds., Amsterdam: North Holland Elsevier Science Publishers B.V., 71-104, 1993b.

Diewert, W. E., The Economic Theory of Index Numbers: A Survey, in *Essays in Index Number Theory Volume 1*, W. E. Diewert and A. O. Nakamura, eds., Amsterdam: North Holland Elsevier Science Publishers B.V., 177-221, 1993c.

Diewert, W. E., Exact and Superlative Index Numbers, in *Essays in Index Number Theory Volume 1*, W. E. Diewert and A. O. Nakamura, eds., Amsterdam: North Holland Elsevier Science Publishers B.V., 223-252, 1993d.

Diewert, W. E., Fisher Ideal Output, Input and Productivity Indexes Revisited, in *Essays in Index Number Theory Volume 1*, W. E. Diewert and A. O. Nakamura, eds., Amsterdam: North Holland Elsevier Science Publishers B.V., 317-353, 1993e.

Greenlees, John S. and Charles C. Mason, Overview of the 1998 Revision of the Consumer Price Index, *Monthly Labor Review*, 3-9, December 1996.