Notes on Unit Value Index Bias

Erwin Diewert and Peter von der Lippe,[1] Revised October 24, 2010
Discussion Paper 10-08,
Department of Economics,
University of British Columbia,
Vancouver, B.C.,
Canada, V6T 1Z1.
Email addresses: diewert@econ.ubc.ca and peter@von-der-lippe.org

Abstract

It is often the case that the value of a number of somewhat similar units (e.g., automobiles of a certain general type) is divided by the number of units in order to form a unit value price and these unit value prices are compared over two periods in order to form a unit value price index. This unit value price index or Drobisch price index can then be compared with other standard index number formulae and the bias in the index can be determined. The present paper presents most of the known results on this bias (and derives some new ones) in a coherent framework using a simple identity from the statistics literature. A related question first considered by Párniczky (1974) is also considered: does disaggregation of a unit value into more homogeneous subgroups reduce the unit value bias? The answer seems to be: probably yes.

Key words: Price indexes, unit value indexes, unit values, bias, Bortkiewicz, Drobisch, Paasche, Laspeyres, Fisher, Párniczky, Balk, von der Lippe.

JEL: C43, C81, E01, E31

---

[1] University of British Columbia, Canada and Universität Duisburg-Essen, Germany respectively. Our thanks to Bert Balk and Jens Mehrhoff for helpful comments. The financial support of the SSHRC of Canada is gratefully acknowledged.

1. Setting the Stage

It is sometimes the case that detailed price and quantity data for a group of closely related commodities (that have the same unit of measurement, such as a class of vehicles, types of grain, containers of similar products, etc.) are not available but information on the number of *units* is available in each period along with the *value of the products* in the shipment. In this case, the value of the products can be divided by the number of units and a *unit value price* is obtained for the period under consideration. If unit values for the product group can be calculated for two periods, then the ratio of the two unit values can be regarded as an (approximate) price index. This price index is known in the literature as a *unit value price index* or a *Drobisch* (1871) *index* in honour of the German measurement economist who first introduced this type of index. It is important to recognize that a Drobisch price index cannot be used over very heterogeneous items since the resulting index is not invariant to changes in the units of measurement. Thus a unit value price index can only be used over products that are measured in the same units and are “reasonably” homogeneous. The question that we are going to address in this paper is the following one: how bad is the bias in a unit value index that is constructed over “reasonably” homogeneous items that are not completely homogeneous?

A major problem with the Drobisch price index is that its axiomatic properties are not entirely satisfactory. In addition to not satisfying the invariance to changes in units test if the aggregation is over heterogeneous items, this index does not satisfy the *identity test*, which asks the index number to equal unity if the price vectors for the two periods under consideration remain the same.[2] However, as mentioned above, the focus of the present paper is not on the axiomatic properties of the Drobisch index; our focus is on determining the *bias* of the Drobisch price index when aggregating over “reasonably” homogeneous products as compared to standard indexes used in index number practice such as the Laspeyres (1871), Paasche (1874) and Fisher (1922) indexes.

A major problem with the Drobisch price index is that its axiomatic properties are not entirely satisfactory. In particular, this index does not satisfy the *identity test*, which asks the index number to equal unity if the price vectors for the two periods under consideration remain the same.[3] A related line of research has been to determine the *bias* of the Drobisch price index compared to standard indexes used in index number practice such as the Laspeyres (1871), Paasche (1874) and Fisher (1922) indexes. This research was initiated by Párniczky (1974) and significant contributions to this bias literature have been made by Timmer (1996), Balk (1998), (2008; 72-74), Silver (2009a) (2009b) (2009c) (2010) and von der Lippe (2007a) (2007b). Our goal in this paper is to present

[2] Balk (1998) and von der Lippe (2007a) looked at the axiomatic properties of the Drobisch index in a systematic way. For additional material on the axiomatic approach to index number theory, see Diewert (1992) (1995), Balk (1995) (2008) and the ILO (2004).

[3] Balk (1998) and von der Lippe (2007a) looked at the axiomatic properties of the Drobisch index in a systematic way. For additional material on the axiomatic approach to index number theory, see Diewert (1992) (1995), Balk (1995) (2008) and the ILO (2004).

the main results in this bias literature (and some new results) in a unified framework using a simple result from the statistics literature.

Some care should be used in interpreting this unit value bias literature. Economic agents often purchase and sell the same commodity at different prices over the accounting period under consideration but a bilateral index number formula requires that these multiple transactions in a single commodity be summarized in terms of a single price and quantity for the period. If the quantity is taken to be the total number of units purchased or sold during the period and it is desired to have the product of the single price and the total quantity transacted equal to the value of the transactions during the period, then the single price must be a unit value; i.e., total value transacted divided by total quantity transacted. Thus at this very first stage of aggregation, the "correct" price to insert into a bilateral index number formula is in fact the unit value for the narrowly defined commodity.[4] This unit value price should not be regarded as having a "bias". However, if there is further aggregation over "similar" commodities using unit value prices, then there can be unit value bias.

The simple result from the statistics literature is the following one. Let $x \equiv [x_1,...,x_N]$ and $y \equiv [y_1,...,y_N]$ be two N dimensional vectors and let $s \equiv [s_1,...,s_N]$ be an N dimensional share vector; i.e., s has nonnegative components ($s \geq 0_N$) which sum up to unity ($1_N \cdot s \equiv \sum_{n=1}^{N} s_n = 1$ where 1_N is a vector of ones and 0_N is a vector of zeros of dimension N). Define the *share weighted covariance* between x and y using the share vector s as follows:

$$(1)\ \mathrm{Cov}(x,y,s) \equiv \sum_{n=1}^{N} s_n(x_n - x^*)(y_n - y^*)$$

where $x^* \equiv s \cdot x = \sum_{n=1}^{N} s_n x_n$ and $y^* \equiv s \cdot y = \sum_{n=1}^{N} s_n y_n$ are the share weighted means of the components of the x and y vectors respectively. A straightforward computation shows that the following *covariance identity* holds:[5]

$$(2)\ \sum_{n=1}^{N} s_n x_n y_n = \mathrm{Cov}(x,y,s) + x^* y^* .$$

We will apply the covariance identity (2) in subsequent sections of this paper.

We conclude this section by formally defining the various indexes mentioned above.

The basic data are *two price vectors*, $p^t \equiv [p_1^t,...,p_N^t]$ and *two quantity vectors*, $q^t \equiv [q_1^t,...,q_N^t]$, for periods t = 0,1. For now, we assume that these price and quantity vectors have positive components and hence, the *period t value aggregates*, $V^t \equiv p^t \cdot q^t > 0$ are positive for each period t.

---

[4] This point was made many years ago by Walsh (1901; 96) (1921; 88) and Davies (1924) (1932) and more recently by Diewert (1995).

[5] This identity was used by von Bortkiewicz (1923) in order to establish his identity relating the Paasche and Laspeyres price indexes. Silver (2009c; 8) uses a correlation coefficient version of this identity to derive his unit value bias results. We will not cover the bias results of Silver in this paper since they generally involve two covariance effects and hence are more complex than our simpler results.

The *Laspeyres* (1871) and *Paasche* (1874) *price indexes* are defined as follows:[6]

(3) $P_L \equiv p^1 \cdot q^0/p^0 \cdot q^0 = \sum_{n=1}^N s_n^0(p_n^1/p_n^0)$;
(4) $P_P \equiv p^1 \cdot q^1/p^0 \cdot q^1 = [\sum_{n=1}^N s_n^1(p_n^1/p_n^0)^{-1}]^{-1}$

where the *period t expenditure share vector* is $s^t \equiv [s_1^t,...,s_N^t]$ for t = 0,1 and $s_n^t \equiv p_n^t q_n^t/p^t \cdot q^t$ for n = 1,...,N and t = 0,1. The *Fisher* (1922) *ideal price index* P_F is defined as the geometric mean of the Laspeyres and Paasche price indexes:

(5) $P_F \equiv [P_L P_P]^{1/2}$.

In order to define the Drobisch or unit value price index, it is necessary to restrict the N commodities under consideration to be measured in the same units. Thus it is *not* meaningful to add up units of crude oil with bushels of wheat in order to obtain an aggregate quantity for each period but it *is* meaningful to add up various grades of crude oil with differing chemical compositions or to add up bushels of wheat of varying quality. Thus in what follows, we assume that a meaningful *quantity aggregate* Q^t exists for each period, where Q^t is just the simple sum of the components of q^t:

(6) $Q^t \equiv 1_N \cdot q^t = \sum_{n=1}^N q_n^t$; t = 0,1.

Once the period t quantity aggregate Q^t is well defined, then we can divide the period t value aggregate, V^t, by Q^t in order to obtain the *period t unit value price* P^t:

(7) $P^t \equiv V^t/Q^t = p^t \cdot q^t/1_N \cdot q^t = p^t \cdot S^t$; t = 0,1

where the *period t quantity share vector* $S^t \equiv [S_1^t,...,S_N^t]$ is defined as follows:

(8) $S^t \equiv q^t/Q^t = q^t/1_N \cdot q^t$; t = 0,1.

Note that the period t quantity shares add up to one; i.e., we have:

(9) $1_N \cdot S^t = 1$; t = 0,1.

Thus the period t unit value P^t can be regarded as a *physical share weighted average* $\sum_{n=1}^N S_n^t p_n^t$ of the individual period t prices p_n^t where the period t physical share weights S_n^t must be distinguished from the period t expenditure shares s_n^t defined earlier.

With the above definitions in hand, we can define the *Drobisch* (1871) *price index* P_D and the corresponding *Drobisch* (or Dutot)[7] *quantity index* Q_D as follows:

[6] The corresponding Laspeyres and Paasche quantity indexes are defined as $Q_L \equiv p^0 \cdot q^1/p^0 \cdot q^0$ and $Q_P \equiv p^1 \cdot q^1/p^1 \cdot q^0$.

(10) $P_D \equiv P^1/P^0 = [p^1 \cdot q^1/p^0 \cdot q^0]/[1_N \cdot q^1/1_N \cdot q^0] = p^1 \cdot S^1/p^0 \cdot S^0$;
(11) $Q_D \equiv Q^1/Q^0 = 1_N \cdot q^1/1_N \cdot q^0$.

Note that $P_D Q_D$ equals the value ratio, V^1/V^0.

2. Comparisons of the Drobisch Index to the Paasche Index

Using (4) and (10) above, it is straightforward to compare the Drobisch index to the corresponding Paasche index:

(12) $P_D/P_P = \{[p^1 \cdot q^1/p^0 \cdot q^0]/[1_N \cdot q^1/1_N \cdot q^0]\}/[p^1 \cdot q^1/p^0 \cdot q^1] = p^0 \cdot S^1/p^0 \cdot S^0$

where we have used definitions (8) for the quantity shares S^t in order to derive the last equality in (12). Thus the *bias in the Drobisch index relative to the Paasche index* can be defined as follows:

(13) $[P_D/P_P] - 1 = [p^0 \cdot S^1/p^0 \cdot S^0] - 1$ using (12)
$= p^0 \cdot [S^1 - S^0]/P^0$ using (7) for t = 0
$= [p^0 - p^{0*} 1_N] \cdot [S^1 - S^0]/P^0$ since $1_N \cdot [S^1 - S^0] = 0$ using (9)
$= N \operatorname{cov}(p^0, S^1 - S^0, (1/N)1_N)/P^0$ using (1) and (2)

where $p^{0*} \equiv [\sum_{n=1}^N p_n^0]/N$ is the arithmetic average of the period 0 prices. Thus the Drobisch index will have an *upward bias* relative to the Paasche index if products n whose quantity shares are growing (so that S_n^1 is greater than S_n^0) are associated with period 0 prices p_n^0 which are above the arithmetic average of the period 0 prices p^{0*}.[8]

As is usual in the analysis of unit value index bias,[9] there are three cases where the bias will be zero:

- All prices in the base period are equal to the same price so that $p^0 = \alpha 1_N$ where $\alpha > 0$;
- The quantity shares remain constant over the two periods under consideration, which is equivalent to $q^1 = \lambda q^0$ where $\lambda > 0$; i.e., the two quantity vectors are proportional or
- The covariance (using share weights that are equal) between the base period prices p^0 and the difference in the quantity share vectors, $S^1 - S^0$, is zero.

[7] Balk (2008) refers to the quantity index defined by (11) as the Dutot quantity index. Dutot (1738) did not actually introduce the quantity index (11); instead he introduced the corresponding additive price index $P_{Dutot} \equiv 1_N \cdot p^1/1_N \cdot p^0$.
[8] Note that the first equation in (13) is particularly easy to interpret: we simply compare two weighted averages of the period 0 prices, p^0, using the quantity shares of period 1, S^1, in the numerator and the quantity shares of period 0, S^0, in the denominator.
[9] See Párniczky (1974; 234) or Balk (2008; 74).

The above application of the covariance identity (2) leads to a fairly simple bias formula for the Drobisch index. However, formula (13) is not equivalent to the bias formulae obtained by Párniczky (1974; 233) and Balk (1998) (2008; 74). These authors obtained bias formulae using share weighted covariance matrices rather than the above equally weighted covariance matrix. In order to obtain these alternative bias decompositions, it is necessary to make some additional definitions. Thus define the vector of *growth rates in quantity shares*, $G \equiv [G_1,...,G_N]$, where the components G_n are defined as follows:

$$(14)\ G_n \equiv [S_n^1/S_n^0] - 1\ ; \qquad n = 1,...,N.$$

Note that if $q^1 = \lambda q^0$ so that the period 1 quantity vector is proportional to the period 0 quantity vector, then $G = 0_N$ where 0_N is a vector of zeros of dimension N. Note also that the G_n satisfy the following equation:

$$\begin{aligned}(15)\ \textstyle\sum_{n=1}^N S_n^0 G_n &= \textstyle\sum_{n=1}^N S_n^0\{[S_n^1/S_n^0] - 1\} && \text{using (14)}\\ &= 0 && \text{using (9).}\end{aligned}$$

Using our earlier bias formula for the Drobisch index relative to the Paasche index (13), we have:

$$\begin{aligned}(16)\ [P_D/P_P] - 1 &= p^0\cdot[S^1 - S^0]/P^0\\ &= [\textstyle\sum_{n=1}^N p_n^0 S_n^0(S_n^1/S_n^0) - \sum_{n=1}^N p_n^0 S_n^0]/P^0\\ &= \{\textstyle\sum_{n=1}^N p_n^0 S_n^0[(S_n^1/S_n^0) - 1]\}/P^0\\ &= \textstyle\sum_{n=1}^N S_n^0 p_n^0 G_n/P^0 && \text{using definitions (14)}\\ &= \textstyle\sum_{n=1}^N S_n^0[p_n^0 - P^0]G_n/P^0 && \text{using (7) for t = 0 and (15)}\\ &= \mathrm{Cov}(p^0,G,S^0)/P^0 && \text{using (1), (2) and (15).}\end{aligned}$$

Comparing (16) with (13), it can be seen that in (16), the covariance uses the base period share weighted average P^0 of the period 0 prices p^0 in place of the arithmetic average of the period 0 prices p^{0*} and the covariance in (16) uses the base period share vector S^0 as the weighting vector as opposed to the equal weights vector $(1/N)1_N$.

As usual, there are three cases where the bias in the Drobisch index relative to the Paasche index will be zero:

- All prices in the base period are equal to the same price so that $p^0 = P^0 1_N$ where P^0 is the period 0 unit value price;
- The quantity vectors are proportional so that $G = 0_N$ or
- The covariance (using the base period share weights S^0) between the base period prices p^0 and the quantity share growth rate vector G is zero.

The bias formula is still not quite equal to the bias formula obtained by Párniczky. In order to obtain his formula, we need to undertake a bit more algebra. First define the vector of *growth rates in quantities* relative to overall growth in quantities, $g \equiv [g_1,...,g_N]$, where the components g_n are defined as follows:

(17) $g_n \equiv [q_n^1/q_n^0] - [Q^1/Q^0]$; $\qquad n = 1,...,N.$

From (15), we have the following equation:

$$\begin{aligned}(18)\ 0 &= \textstyle\sum_{n=1}^N S_n^0\{[S_n^1/S_n^0] - 1\} \\ &= \textstyle\sum_{n=1}^N S_n^0\{([q_n^1/Q^1]/[q_n^0/Q^0]) - 1\} && \text{using definitions (8)} \\ &= \textstyle\sum_{n=1}^N S_n^0\{[q_n^1/q_n^0]) - [Q^1/Q^0]\}/[Q^1/Q^0] \\ &= \textstyle\sum_{n=1}^N S_n^0 g_n/[Q^1/Q^0] && \text{using definitions (17).}\end{aligned}$$

Equation (18) shows that the base period quantity share weighted average of the relative quantity growth rates g is equal to zero. Put another way, the weighted mean of the quantity relatives q_n^1/q_n^0 using the base period quantity shares S_n^0 as weights is equal to the aggregate quantity relative, Q^1/Q^0, which is the Dutot quantity index; i.e., we have

$$(19)\ \textstyle\sum_{n=1}^N S_n^0[q_n^1/q_n^0] = Q^1/Q^0 = Q_D.$$

From the bias formula (16), we have the following expression:

$$\begin{aligned}(20)\ [P_D/P_P] - 1 &= \{\textstyle\sum_{n=1}^N p_n^0 S_n^0[(S_n^1/S_n^0) - 1]\}/P^0 \\ &= \textstyle\sum_{n=1}^N p_n^0 S_n^0\{[q_n^1/q_n^0]) - [Q^1/Q^0]\}/[Q^1/Q^0]P^0 && \text{using the algebra in (18)} \\ &= \textstyle\sum_{n=1}^N S_n^0 p_n^0 g_n/[Q^1/Q^0]P^0 && \text{using definitions (17)} \\ &= \textstyle\sum_{n=1}^N S_n^0[p_n^0 - P^0]g_n/[Q^1/Q^0]P^0 && \text{using (18)} \\ &= \mathrm{Cov}(p^0,g,S^0)/P^0 Q_D && \text{using (1), (2) and (18).}\end{aligned}$$

This is the bias formula for the Drobisch index relative to the Paasche index that was derived by Párniczky (1974; 233). We have the three usual cases where the bias will be zero; i.e., all prices in the base period are equal or the quantity vectors are proportional or the covariance $\mathrm{Cov}(p^0,g,S^0)$ is zero.

In the following section, we simply adapt the above analysis in order to obtain bias formulae for the Drobisch price index relative to the Laspeyres and Fisher indexes.

3. Comparisons of the Drobisch Index to the Laspeyres and Fisher Indexes

Using (3) and (10) above, it is straightforward to compare the Drobisch index to the corresponding Laspeyres index:

$$(21)\ P_D/P_L = \{[p^1\cdot q^1/p^0\cdot q^0]/[1_N\cdot q^1/1_N\cdot q^0]\}/[p^1\cdot q^0/p^0\cdot q^0] = p^1\cdot S^1/p^1\cdot S^0$$

where we have used definitions (8) for the quantity shares S^t in order to derive the last equality in (21). If the weighted average of period 1 prices using period 1 quantity share weights, $p^1\cdot S^1$, is greater than the weighted average of period 1 prices using period 0 quantity share weights, $p^1\cdot S^0$, then P_D will be greater than P_L and vice versa. The *bias in the Drobisch index relative to the Laspeyres index* can be defined as follows:

$$(22)\ [P_D/P_L] - 1 = [p^1\cdot S^1/p^1\cdot S^0] - 1 \qquad \text{using (21)}$$

$$= p^1\cdot[S^1 - S^0]/p^1\cdot S^0$$
$$= [p^1 - p^{1*}1_N]\cdot[S^1 - S^0]/p^1\cdot S^0 \qquad \text{since } 1_N\cdot[S^1 - S^0] = 0 \text{ using (9)}$$
$$= N\,\mathrm{Cov}(p^1,S^1 - S^0,(1/N)1_N)/p^1\cdot S^0 \qquad \text{using (1) and (2)}$$

where $p^{1*} \equiv [\sum_{n=1}^N p_n^1]/N$ is the arithmetic average of the period 1 prices. Thus the Drobisch index will have an *upward bias* relative to the Laspeyres index if products n whose quantity shares are growing (so that S_n^1 is greater than S_n^0) are associated with period 1 prices p_n^1 which are above the arithmetic average of the period 1 prices p^{1*}.

There are three cases where the bias will be zero:

- All prices in the current period are equal to the same price so that $p^1 = \alpha 1_N$ where $\alpha > 0$;
- The quantity shares remain constant over the two periods under consideration, which is equivalent to $q^1 = \lambda q^0$ where $\lambda > 0$; i.e., the two quantity vectors are proportional or
- The covariance (using share weights that are equal) between the current period prices p^1 and the difference in the quantity share vectors, $S^1 - S^0$, is zero.

In order to obtain our second bias decomposition that is an analogue to formula (16), it will be necessary to develop a formula for the Laspeyres index P_L relative to the Drobisch index P_D rather than for P_D/P_L.[10] As usual, some additional definitions will be required. Thus define the vector of *reciprocal growth rates in quantity shares*, $\Gamma \equiv [\Gamma_1,...,\Gamma_N]$, where the components Γ_n are defined as follows:

$$(23)\ \Gamma_n \equiv [S_n^1/S_n^0]^{-1} - 1\ ; \qquad n = 1,...,N.$$

Note that if $q^1 = \lambda q^0$ so that the period 1 quantity vector is proportional to the period 0 quantity vector, then $\Gamma = 0_N$. Note also that the Γ_n satisfy the following equation:

$$(24)\ \sum_{n=1}^N S_n^1\Gamma_n = \sum_{n=1}^N S_n^1\{[S_n^0/S_n^1] - 1\} \qquad \text{using (23)}$$
$$= 0 \qquad \text{using (9).}$$

Thus the period 1 share weighted average of the Γ_n is equal to 0. Using (21), we have P_D/P_L equal to $p^1\cdot S^1/p^1\cdot S^0$. Taking reciprocals of this equation and subtracting unity leads to the following equation which defines the bias of the Laspeyres index relative to the Drobisch index:

$$(25)\ [P_L/P_D] - 1 = [p^1\cdot S^0/p^1\cdot S^1] - 1$$
$$= p^1\cdot[S^0 - S^1]/P^1 \qquad \text{using (7) for } t = 1$$
$$= [\sum_{n=1}^N p_n^1S_n^1(S_n^0/S_n^1) - \sum_{n=1}^N p_n^1S_n^1]/P^1$$
$$= \{\sum_{n=1}^N p_n^1S_n^1[(S_n^0/S_n^1) - 1]\}/P^1$$
$$= \sum_{n=1}^N S_n^1p_n^1\Gamma_n/P^1 \qquad \text{using definitions (23)}$$

[10] Balk (2008; 74) developed a formula for P_D/P_L but it is more complex than the formula that we develop.

$= \sum_{n=1}^{N} S_n^{1}[p_n^{1} - P^1]\Gamma_n/P^1$ using (7) for t = 1 and (24)
$= Cov(p^1,\Gamma,S^1)/P^1$ using (1), (2) and (24).

The bias formula (25) for the Laspeyres price index relative to the Drobisch price index is the symmetric counterpart to the bias formula (16), which compared the Drobisch and Paasche price indexes. The covariance in (25) will be positive if higher than average reciprocal growth rates for quantity shares are associated with higher than average period 1 prices and under these conditions, P_L will be greater than P_D.

As usual, there are three cases where the bias in the Drobisch index relative to the Laspeyres index will be zero:

- All prices in period 1 are equal to the same price so that $p^1 = P^1 1_N$ where P^1 is the period 1 unit value price;
- The quantity vectors are proportional so that the vector of reciprocal quantity growth rates $\Gamma = 0_N$ or
- The covariance (using period 1 share weights S^1) between the period 1 prices p^1 and the reciprocal quantity share growth rate vector Γ is zero.

We will now modify the above bias formula in order to obtain a Laspeyres counterpart to the Paasche bias formula (20) obtained by Párniczky. As usual, we need to undertake a bit more algebra. Define the vector of *reciprocal growth rates in quantities* relative to overall reciprocal growth in quantities, $r \equiv [r_1,...,r_N]$, where the components r_n are defined as follows:

(26) $r_n \equiv [q_n^0/q_n^1] - [Q^0/Q^1]$; $n = 1,...,N.$

It can be seen that the share weighted mean of the r_n is zero if we use period 1 quantity shares S_n^1 as weights:

(27) $\sum_{n=1}^{N} S_n^{1}r_n = \sum_{n=1}^{N} S_n^{1}\{[q_n^0/q_n^1] - [Q^0/Q^1]\}$ using (26)
$= \sum_{n=1}^{N} [q_n^1/1_N{\cdot}q^1][q_n^0/q_n^1] - \sum_{n=1}^{N} S_n^{1} [Q^0/Q^1]$ using (8)
$= [1_N{\cdot}q^0/1_N{\cdot}q^1] - [Q^0/Q^1]$ using (9)
$= 0$ using (6).

Equation (27) shows that the period 1 quantity share weighted average of the reciprocal quantity growth rates r is equal to zero. Put another way, the weighted mean of the reciprocal quantity relatives q_n^0/q_n^1 using the period 1 quantity shares S_n^1 as weights is equal to the aggregate reciprocal quantity relative, Q^0/Q^1, which is the reciprocal of the Dutot quantity index; i.e., we have

(28) $\sum_{n=1}^{N} S_n^{1}[q_n^0/q_n^1] = Q^0/Q^1 = Q_D^{-1}$.

From the bias formula (25), we have the following expression:

(29) $[P_L/P_D] - 1 = \{\sum_{n=1}^{N} p_n^{1}S_n^{1}[(S_n^0/S_n^1) - 1]\}/P^1$

$= \sum_{n=1}^{N} p_n^1 S_n^1 \{[q_n^0/q_n^1]) - [Q^0/Q^1]\}/[Q^0/Q^1]P^1$ using (6) and (8)
$= \sum_{n=1}^{N} S_n^1 p_n^1 r_n/[Q^0/Q^1]P^1$ using definitions (26)
$= \sum_{n=1}^{N} S_n^1 [p_n^1 - P^1] r_n/[Q^0/Q^1]P^1$ using (27)
$= \mathrm{Cov}(p^1,r,S^1)/P^1 Q_D^{-1}$ using (1), (2) and (27).

This is the bias formula for the Laspeyres price index relative to the Drobisch price index that is the counterpart to the Párniczky (1974; 233) bias formula for the Drobisch index relative to the Paasche index. We have the three usual cases where the bias will be zero; i.e., all prices in the base period are equal or the quantity vectors are proportional or the covariance $\mathrm{Cov}(p^1,r,S^1)$ is zero.

We conclude this section by comparing the Drobisch price index P_D to the Fisher ideal price index P_F. The bias formulae (29) and (20) can be rewritten as follows:

(30) $[P_D/P_L] = 1 + [P^1/\mathrm{Cov}(p^1,r,S^1)Q_D]$;
(31) $[P_D/P_P] = 1 + [\mathrm{Cov}(p^0,g,S^0)/P^0 Q_D]$.

Using the above formula and definition (5) for the Fisher price index, it can be seen that the bias of the Drobisch index relative to the Fisher index is:[11]

(32) $[P_D/P_F] - 1 = \{1 + [P^1/\mathrm{Cov}(p^1,r,S^1)Q_D]\}^{1/2}\{1 + [\mathrm{Cov}(p^0,g,S^0)/P^0 Q_D]\}^{1/2} - 1.$

In a similar fashion, we can rewrite the bias formulae (13) and (22) as follows:

(33) $[P_D/P_P] = [p^0 \cdot S^1/p^0 \cdot S^0]$;
(34) $[P_D/P_L] = [p^1 \cdot S^1/p^1 \cdot S^0]$.

Thus using definition (5), we have

(35) $[P_D/P_F] = [p^0 \cdot S^1/p^0 \cdot S^0]^{1/2}[p^1 \cdot S^1/p^1 \cdot S^0]^{1/2}$ using (33) and (34)
$= [p^0 \cdot S^1\, p^1 \cdot S^1]^{1/2}/[p^0 \cdot S^0\, p^1 \cdot S^0]^{1/2}$
$\approx [(1/2)\, p^0 \cdot S^1 + (1/2)\, p^1 \cdot S^1]/[(1/2)\, p^0 \cdot S^0 + (1/2)\, p^1 \cdot S^0]$
where we have approximated the geometric means by arithmetic means
$= p^* \cdot S^1/p^* \cdot S^0$

where the vector of *arithmetic average prices* p^* is defined as $(1/2)p^0 + (1/2)p^1$. Thus the approximate bias of the Drobisch price index relative to the Fisher price index is

(36) $[P_D/P_F] - 1 \approx [p^* \cdot S^1/p^* \cdot S^0] - 1$
$= p^* \cdot [S^1 - S^0]/p^* \cdot S^0$
$= N\, \mathrm{cov}(p^*,S^1 - S^0,(1/N)1_N)/p^* \cdot S^0$ using (1) and (2).

[11] This is similar to the bias formula developed by Balk (2008; 74) but our components are a bit different due to our use of reciprocal rates of growth which Balk did not use.

The above approximate bias formula is very similar to the earlier bias formula comparing P_D to P_P and P_L, formulae (13) and (22). The main difference is that in (13), the base period price vector p^0 appeared in the covariance and in (22) p^1 appeared in the covariance whereas in (36), the reference price vector p^* is the arithmetic average of p^0 and p^1. Thus the Drobisch index will (likely) have an *upward bias* relative to the Fisher index if products n whose quantity shares are growing (so that S_n^1 is greater than S_n^0) are associated with prices $(1/2)p_n^0 + (1/2)p_n^1$ which are above the arithmetic average of the period 0 and period 1 prices $(1/2)p^{0*} + (1/2)p^{1*}$.

The approximate bias on the right hand side of (36) will be zero if any one of the following three conditions is satisfied (the third condition implies the first two conditions):

- All average prices (averaged over the two periods) are equal to the same price so that $(1/2)p^0 + (1/2)p^1 = \alpha 1_N$ where $\alpha > 0$;
- The quantity shares remain constant over the two periods under consideration, which is equivalent to $q^1 = \lambda q^0$ where $\lambda > 0$; i.e., the two quantity vectors are proportional or
- The covariance (using share weights that are equal) between the average prices $(1/2)p^0 + (1/2)p^1$ and the difference in the quantity share vectors, $S^1 - S^0$, is zero.

In the next section, we derive some additional bias formulae for the Drobisch index using a technique due to von der Lippe (2007a; 415-428) (2007b).

4. Alternative Bias Decompositions for the Drobisch Index

The analysis in this section starts with an identity relating the Paasche and Laspeyres price indexes that was first derived by von Bortkiewicz (1923). As usual, we will require a few new definitions. Define the vector of *price relatives* $\rho \equiv [\rho_1,...,\rho_N]$ where $\rho_n \equiv p_n^1/p_n^0$ for n = 1,...,N and the vector of *quantity relatives* $\tau \equiv [\tau_1,...,\tau_N]$ where $\tau_n \equiv q_n^1/q_n^0$ for n = 1,...,N. Recalling definitions (3) and (4), it can be verified that a share weighted average of the price relatives is equal to the Laspeyres price index and a share weighted average of the quantity relatives is equal to the Laspeyres quantity index if we use the base period shares s_n^0 as weights; i.e., we have

(37) $\sum_{n=1}^N s_n^0 \rho_n = P_L$;
(38) $\sum_{n=1}^N s_n^0 \tau_n = Q_L$.

The von Bortkiewicz (1923) identity is the following one[12]:

(39) $P_P - P_L = \mathrm{Cov}(\rho,\tau,s^0)/Q_L$

[12] The covariance identity (2) can be used to prove (39).

where the covariance in (39) is defined by (1). Empirically, for most value aggregates, it is found that the covariance in (39) is negative; i.e., above average growth in a price is associated with a below average growth in the corresponding quantity.

Multiply both sides of (39) by Q_L and note that P_PQ_L equals the value ratio, V^1/V^0. Thus we obtain the following identity:

(40) $V^1/V^0 = \mathrm{Cov}(\rho,\tau,s^0) + P_LQ_L$.

Now divide both sides of (40) by P_LQ^1/Q^0 and noting that $[V^1/V^0]/[Q^1/Q^0]$ equals the Drobisch price index P_D, we obtain the following expression for P_D relative to the Laspeyres price index P_L:

(41) $P_D/P_L = [\mathrm{Cov}(\rho,\tau,s^0)/P_LQ_D] + [Q_L/Q_D]$

where we have also used Q_D equals Q^1/Q^0. Now subtract 1 from both sides of (41) and we obtain von der Lippe's (2007b) formula for the bias of the Drobisch index relative to the Laspeyres price index:[13]

(42) $[P_D/P_L] - 1 = [\mathrm{Cov}(\rho,\tau,s^0)/P_LQ_D] + [Q_L/Q_D] - 1$.

The first term on the right hand side of (42) can generally assumed to be negative but what can be said about the last term, $[Q_L/Q_D] - 1$? Using our definitions for the Laspeyres and Drobisch quantity indexes, we have:

(43) $Q_L/Q_D = [p^0\cdot q^1/p^0\cdot q^0]/[1_N\cdot q^1/1_N\cdot q^0]$
$= p^0\cdot S^1/p^0\cdot S^0$ using definitions (8)
$= [P_D/P_P]$ using (16)
$= [\sum_{n=1}^{N} S_n^0 p_n^0 G_n/P^0] + 1$ using (16)
$= [\mathrm{Cov}(p^0,G,S^0)/P^0] + 1$ using (16) again

where the $G_n \equiv [S_n^1/S_n^0] - 1$ are the growth rates for the quantity shares. Substituting (43) into (42) leads to the following bias formula:

(44) $[P_D/P_L] - 1 = [\mathrm{Cov}(\rho,\tau,s^0)/P_LQ_D] + [\mathrm{Cov}(p^0,G,S^0)/P^0]$.

While the bias formula (44) is an interesting one, it may not be as useful as our earlier bias formulae since only one covariance is involved in our earlier formulae.[14]

It is possible to develop Paasche counterparts to (41)-(44) by reversing the role of time. Define the vector of *reciprocal price relatives* $\rho^{-1} \equiv [\rho_1^{-1},...,\rho_N^{-1}]$ where $\rho_n^{-1} \equiv p_n^0/p_n^1$ for

[13] See also Balk (2008; 73-74).

[14] If we substitute the middle equation in (43) into (42), we obtain the identity $[P_D/P_L] - [P_D/P_P] = [\mathrm{Cov}(\rho,\tau,s^0)/P_LQ_D]$. Thus a negative covariance $\mathrm{Cov}(\rho,\tau,s^0)$ will just lead to the conclusion that P_L is greater than P_P, which is just the conclusion that we can draw from (39).

$n = 1,\ldots,N$ and the vector of *reciprocal quantity relatives* $\tau^{-1} \equiv [\tau_1^{-1},\ldots,\tau_N^{-1}]$ where $\tau_n^{-1} \equiv q_n^0/q_n^1$ for $n = 1,\ldots,N$. Recalling definitions (3) and (4), it can be verified that the following identities hold:

(45) $\sum_{n=1}^N s_n^1\rho_n^{-1} = P_P^{-1}$;
(46) $\sum_{n=1}^N s_n^1\tau_n^{-1} = Q_P^{-1}$.

Interchanging 0 and 1 in (40) leads to the following counterpart to (40):

(47) $V^0/V^1 = \mathrm{Cov}(\rho^{-1},\tau^{-1},s^1) + P_P^{-1}Q_P^{-1}$.

Again recalling that P_D is equal to $(V^1/V^0)/Q_D$, we can use (47) in order to obtain the following formula for the reciprocal of P_D relative to P_P:

(48) $(P_D/P_P)^{-1} = P_P(V^0/V^1)/Q_D$
$= [\mathrm{Cov}(\rho^{-1},\tau^{-1},s^1)P_P/Q_D] + [Q_D/Q_P]$ using (47).

The identity (48) is the counterpart to the earlier von der Lippe identity (41). We can expect $\mathrm{Cov}(\rho^{-1},\tau^{-1},s^1)$ to be negative but again, we need an analytical formula for the second term on the right hand side of (48), the ratio of the Drobisch quantity index Q_D to the Paasche quantity index Q_P. Using the definitions for Q_D and Q_P, we obtain the following decomposition:

(49) $Q_D/Q_P = [1_N\cdot q^1/1_N\cdot q^0]/[p^1\cdot q^1/p^1\cdot q^0]$
$= p^1\cdot S^0/p^1\cdot S^1$ using definitions (8)
$= P_L/P_D$ using (21)
$= [\sum_{n=1}^N S_n^1 p_n^1\Gamma_n/P^1] + 1$ using (25)
$= [\mathrm{Cov}(p^1,\Gamma,S^1)/P^1] + 1$ using (25) again

where the Γ_n are defined as the reciprocal share growth rates, $[S_n^1/S_n^0]^{-1} - 1$, for $n = 1,\ldots,N$. Substituting (49) into (48) leads to the following bias formula for the Paasche price index relative to the Drobisch price index:

(50) $[P_P/P_D] - 1 = [\mathrm{Cov}(\rho^{-1},\tau^{-1},s^1)P_P/Q_D] + [\mathrm{Cov}(p^1,\Gamma,S^1)/P^1]$.

Note that (48)-(50) imply the following identity:

(51) $[P_P/P_D] - [P_L/P_D] = \mathrm{Cov}(\rho^{-1},\tau^{-1},s^1)P_P/Q_D$.

Thus if $\mathrm{Cov}(\rho^{-1},\tau^{-1},s^1)$ is negative (the usual case), then (51) implies that P_P is less than P_L, which is also implied by the usual von Bortkiewicz identity, (39).

5. Does the Unit Value Bias Increase with Increased Aggregation?

It is generally thought that constructing broader unit value prices (i.e., aggregating over more specific products to form unit value prices) will lead to a greater degree of bias in a unit value price index as compared to the underlying "true" index. Párniczky (1974) addressed this issue and showed that this is not necessarily the case. However, his analysis was somewhat brief and it will be useful to address this issue in more detail.

Suppose we can decompose the N products in the aggregate under consideration into M subgroups where subgroup m has N_m distinct products for m = 1,...,M so that $N = N_1 + ... + N_M$. Let 1_m represent a vector of ones with dimension N_m for m = 1,...,M. Let the period t value of the products in subgroup m be V_m^t for t = 0,1 and m = 1,...,M. Unit value prices for subgroup m in period t, P_m^t, and group m total quantity in period t, Q_m^t, can be defined as follows:

$$(52)\ Q_m^t \equiv 1_m \cdot q_m^t\ ; \qquad t = 0,1;\ m = 1,...,M;$$
$$(53)\ P_m^t \equiv V_m^t/Q_m^t = p_m^t \cdot q_m^t/Q_m^t\ ; \qquad t = 0,1;\ m = 1,...,M$$

where the period t price and quantity vectors for subaggregate m are defined as p_m^t and q_m^t, $V_m^t \equiv p_m^t \cdot q_m^t$ and 1_m is a vector of ones of dimension N_m for m = 1,...,M and t = 0,1. The overall *period t unit value price* for the entire aggregate, P^t, is defined in the usual way as follows:

$$\begin{aligned}(54)\ P^t &\equiv \textstyle\sum_{m=1}^M p_m^t \cdot q_m^t / \sum_{m=1}^M 1_m \cdot q_m^t \qquad t = 0,1\\ &= \textstyle\sum_{m=1}^M p_m^t \cdot (q_m^t/1_m \cdot q_m^t)\sigma_m^t\\ &= \textstyle\sum_{m=1}^M p_m^t \cdot S_m^t\, \sigma_m^t\\ &= \textstyle\sum_{m=1}^M P_m^t\, \sigma_m^t\end{aligned}$$

where the *period t within group m quantity share vector* S_m^t is defined as follows

$$(55)\ S_m^t \equiv q_m^t/1_m \cdot q_m^t\ ; \qquad t = 0,1;\ m = 1,...,M;$$

and the *period t between subgroup m quantity share* σ_m^t is defined as follows:

$$(56)\ \sigma_m^t \equiv Q_m^t/\textstyle\sum_{i=1}^M Q_i^t = 1_m \cdot q_m^t/\sum_{i=1}^M 1_i \cdot q_i^t\ ; \qquad t = 0,1;\ m = 1,...,M.$$

Note that the various quantity shares sum to unity; i.e., we have

$$(57)\ 1_m \cdot S_m^t = 1\ ; \qquad t = 0,1;\ m = 1,...,M;$$
$$(58)\ \textstyle\sum_{m=1}^M \sigma_m^t = 1\ ; \qquad t = 0,1.$$

The final equation in (54) shows that the overall period t unit value price P^t is equal to a quantity share weighted average of the period t subgroup unit value prices, $\sum_{m=1}^M P_m^t\, \sigma_m^t$. Note also that (52), (53) and (55) imply that the period t, group m unit value price P_m^t can be expressed as the inner product of the period t subgroup m price vector p_m^t and the corresponding subgroup share vector S_m^t:

$$(59)\ P_m^t = p_m^t \cdot S_m^t\ ; \qquad t = 0,1;\ m = 1,...,M.$$

Using our new notation, the *Drobisch price index*, P_D, can be defined as follows:

$$(60)\ P_D \equiv P^1/P^0$$
$$= \sum_{m=1}^M p_m^1 \cdot S_m^1\ \sigma_m^1 / \sum_{m=1}^M p_m^0 \cdot S_m^0\ \sigma_m^0 \qquad \text{using (54)}$$
$$= \sum_{m=1}^M P_m^1 \sigma_m^1 / \sum_{m=1}^M P_m^0 \sigma_m^0 \qquad \text{using (59).}$$

Thus the Drobisch price index is equal to a quantity share weighted average of the period 1 subgroup unit values P_m^1 (using the period 1 between group shares σ_m^1) divided by a quantity share weighted average of the period 0 subgroup unit values P_m^0 (using the period 0 between group shares σ_m^0).

Again using our new notation, the *Paasche price index*, P_P, can be defined as follows:

$$(61)\ P_P \equiv \sum_{m=1}^M p_m^1 \cdot q_m^1 / \sum_{m=1}^M p_m^0 \cdot q_m^1$$
$$= \sum_{m=1}^M P_m^1 Q_m^1 / \sum_{m=1}^M p_m^0 \cdot S_m^1 Q_m^1 \qquad \text{using (52), (53) and (55)}$$
$$= \sum_{m=1}^M P_m^1 \sigma_m^1 / \sum_{m=1}^M p_m^0 \cdot S_m^1 \sigma_m^1 \qquad \text{using (56).}$$

Note that the numerators in the final equations in (60) and (61) are equal.

Recall that in section 2, we defined the bias in the Drobisch index relative to the Paasche index as $[P_D/P_P] - 1$. In this section, we will find it convenient to define the bias using a reciprocal measure. Thus define the *bias of the Paasche relative to the Drobisch index* as

$$(62)\ \text{Bias}(P_P/P_D) \equiv [P_P/P_D] - 1$$
$$= [\sum_{m=1}^M P_m^0 \sigma_m^0 / \sum_{m=1}^M p_m^0 \cdot S_m^1 \sigma_m^1] - 1 \qquad \text{using (61) and (62)}$$
$$= [\sum_{m=1}^M p_m^0 \cdot S_m^0 \sigma_m^0 / \sum_{m=1}^M p_m^0 \cdot S_m^1 \sigma_m^1] - 1 \qquad \text{using (59)}$$
$$= \{\sum_{m=1}^M p_m^0 \cdot [S_m^0 \sigma_m^0 - S_m^1 \sigma_m^1]\} / \sum_{m=1}^M p_m^0 \cdot S_m^1 \sigma_m^1.$$

The analysis in section 2 could be reworked at this point in order to analyze the reciprocal bias in terms of a covariance of the base period price vector p^0 and the vector of base period quantity shares less the vector of period 1 shares. Thus the reciprocal bias will be negative if products which have above average prices in the base period have growing quantity shares over the two periods under consideration.

Párniczky (1974; 235) introduced the idea of aggregating over all of the N individual product classes in two stages where unit value aggregation would be used in the first stage and normal index number theory would be used in the second stage. Thus if we use the Paasche formula in the second stage of aggregation, the basic price and quantity data, P_m^t and Q_m^t, that are used in the second stage formula are the unit values and subgroup total quantities that are defined by (52) and (53). Thus the second stage Paasche index or *hybrid Paasche price index* can be defined as follows:

$$(63)\ P_{HP} \equiv \sum_{m=1}^M P_m^1 Q_m^1 / \sum_{m=1}^M P_m^0 Q_m^1$$
$$= [\sum_{m=1}^M P_m^1 Q_m^1 / \sum_{m=1}^M Q_m^1] / [\sum_{m=1}^M P_m^0 Q_m^1 / \sum_{m=1}^M Q_m^1]$$

$= \sum_{m=1}^{M} P_m^1\sigma_m^1/\sum_{m=1}^{M} P_m^0\sigma_m^1$.

Note that the numerators in the final formulae for the true Paasche index P_P defined by (61) and for the hybrid Paasche index P_{HP} defined by (63) are equal. Define the *bias of the Paasche index relative to the hybrid Paasche index* as follows:

$$\text{(64) Bias}(P_P/P_{HP}) \equiv [P_P/P_{HP}] - 1$$
$$= [\sum_{m=1}^{M} P_m^0\sigma_m^1/\sum_{m=1}^{M} p_m^0\cdot S_m^1\sigma_m^1] - 1 \qquad \text{using (61) and (63)}$$
$$= \sum_{m=1}^{M} p_m^0\cdot[S_m^0 - S_m^1]\sigma_m^1/\sum_{m=1}^{M} p_m^0\cdot S_m^1\sigma_m^1 \qquad \text{using (59)}$$
$$= N\text{Cov}(p^0, S^{0*} - S^{1*},(1/N)1_N)/\sum_{m=1}^{M} p_m^0\cdot S_m^1\sigma_m^1 \text{ using (1), (2) and (57)}$$

where p^0 is the N dimensional vector of base period prices and the components of the vector $S^{0*} - S^{1*}$ are the vectors $[S_m^0 - S_m^1]\sigma_m^1$ stacked up into a single N dimensional vector. We have the following three sets of conditions which will imply that the hybrid Paasche is equal to the true Paasche index:

- All base period subaggregate price vectors p_m^0 are equal to the corresponding subaggregate unit values; i.e., $p_m^0 = P_m^0 1_m$ for m = 1,...,M;
- The subaggregate quantity shares remain constant going from period 0 to period 1; i.e., $S_m^0 = S_m^1$ for m = 1,...,M or
- The covariance between the base period price vector p^0 and the adjusted share difference vector $S^{0*} - S^{1*}$ defined above is zero; i.e., $\text{Cov}(p^0, S^{0*} - S^{1*},(1/N)1_N) = 0$.

Of course, the first two sets of conditions are special cases of the third condition. The first set of conditions is perhaps the most important for choosing how to construct the subaggregates: in order to minimize bias (relative to the Paasche price index), use unit value aggregation over products that sell for the same price in the base period.

We now address the question that was first considered by Párniczky (1974); i.e., if instead of having only one unit value over a large number of products, we disaggregate the data into M subgroups and calculate unit value prices and the corresponding quantities for these M subgroups and apply normal index number theory (in this case, we use the Paasche formula), do we reduce unit value bias? Our suspicion is that disaggregation will help reduce the bias; i.e., we expect that the $\text{Bias}(P_P/P_D)$ defined by (62) will be greater in magnitude than the magnitude of the $\text{Bias}(P_P/P_{HP})$ defined by (64). We can use the algebra developed above in order to obtain an exact relationship between these two bias measures. From (62), we have:

$$\text{(65) Bias}(P_P/P_D) = \{\sum_{m=1}^{M} p_m^0\cdot[S_m^0\sigma_m^0 - S_m^1\sigma_m^1]\}/\sum_{m=1}^{M} p_m^0\cdot S_m^1\sigma_m^1$$
$$=\{\sum_{m=1}^{M} [p_m^0\cdot S_m^0\sigma_m^1 - p_m^0\cdot S_m^1\sigma_m^1 + p_m^0\cdot S_m^0\sigma_m^0 - p_m^0\cdot S_m^0\sigma_m^1]\}/\sum_{m=1}^{M} p_m^0\cdot S_m^1\sigma_m^1$$
$$= \text{Bias}(P_P/P_{HP}) + \{\sum_{m=1}^{M} P_m^0[\sigma_m^0 - \sigma_m^1]/\sum_{m=1}^{M} p_m^0\cdot S_m^1\sigma_m^1\} \qquad \text{using (64) and (59)}$$
$$= \text{Bias}(P_P/P_{HP}) + M\text{Cov}(\mathbf{P^0}, \sigma^0-\sigma^1,(1/M)1_M)/\sum_{m=1}^{M} p_m^0\cdot S_m^1\sigma_m^1 \quad \text{using (2) and (58)}$$

where $\mathbf{P^0} \equiv [P_1^0,...,P_M^0]$ is the vector of period 0 unit value prices for the M subaggregates[15] and $\sigma^t \equiv [\sigma_1^t,...,\sigma_1^t]$ is the vector of period t subaggregate quantity shares where the σ_m^t were defined by (56).

Suppose that P_D is greater than P_P so that the reciprocal bias, $\text{Bias}(P_P/P_D)$ (equal to $[P_P/P_D] - 1$), is less than 1. Then from our analysis in section 2, we know that products which have above average prices in period 0 are positively correlated with growing quantity shares. It is *likely* (but not certain) that this positive correlation persists to the subaggregates so that subaggregates that have above average unit value prices P_m^0 in period 0 are associated with a positive $\sigma_m^1 - \sigma_m^0$ and hence the covariance in (65) will be negative under this hypothesis. It is also *likely* that the hybrid Paasche P_{HP} is equal to or greater than the true Paasche index P_P under these circumstances so that reciprocal bias, $\text{Bias}(P_P/P_{HP})$ (equal to $[P_P/P_{HP}] - 1$), is less than or equal to 1. Under these hypotheses, using (65), it can be seen that

(66) $\text{Bias}(P_P/P_D) < \text{Bias}(P_P/P_{HP}) \leq 1.$

Hence the magnitude of the bias of the hybrid Paasche index is less than the magnitude of the bias of the Drobisch index. Thus in this case, disaggregation of the unit value price index does reduce its bias.

A similar analysis can be made for the case where P_D is less than P_P so that the reciprocal bias, $\text{Bias}(P_P/P_D)$, is greater than 1. Then from our analysis in section 2, we know that products which have above average prices in period 0 are negatively correlated with growing quantity shares. It is *likely* (but not certain) that this negative correlation persists to the subaggregates so that subaggregates m that have above average unit value prices P_m^0 in period 0 are associated with a negative $\sigma_m^1 - \sigma_m^0$ and hence the covariance in (65) will be positive under this hypothesis. It is also *likely* that the hybrid Paasche P_{HP} is equal to or less than the true Paasche index P_P under these circumstances so that reciprocal bias, $\text{Bias}(P_P/P_{HP})$, is equal to or greater than 1. Under these hypotheses, using (65), it can be seen that

(67) $\text{Bias}(P_P/P_D) > \text{Bias}(P_P/P_{HP}) \geq 1.$

Hence the magnitude of the bias of the hybrid Paasche index is less than the magnitude of the bias of the Drobisch index. Thus in this case, disaggregation of the unit value price index again reduces its bias.

The inequalities (66) and (67) were established under hypotheses which were “reasonable” but they need not hold empirically.

[15] We need to distinguish the vector $\mathbf{P^0}$ from the scalar P^0, which is the overall unit value price for all N commodities in period 0.

It is possible to adapt the above analysis to the case where the Laspeyres formula is used as the target index rather than the Paasche index. Thus the "true" *Laspeyres price index*, P_L, is defined as follows:

$$(68)\ P_L \equiv \sum_{m=1}^{M} p_m^{1}\cdot q_m^{0}/\sum_{m=1}^{M} p_m^{0}\cdot q_m^{0}$$
$$= \sum_{m=1}^{M} p_m^{1}\cdot S_m^{0}Q_m^{0}/\sum_{m=1}^{M} p_m^{0}\cdot S_m^{0}Q_m^{0} \qquad \text{using (52), (53) and (55)}$$
$$= \sum_{m=1}^{M} p_m^{1}\cdot S_m^{0}\sigma_m^{0}/\sum_{m=1}^{M} p_m^{0}\cdot S_m^{0}\sigma_m^{0} \qquad \text{using (56)}$$
$$= \sum_{m=1}^{M} p_m^{1}\cdot S_m^{0}\sigma_m^{0}/\sum_{m=1}^{M} P_m^{0}\sigma_m^{0} \qquad \text{using (59).}$$

Recall formula (60) for the single stage unit value or Drobisch price index P_D and note that the denominators in the last lines of (60) and (68) are identical. The *bias of P_D relative to P_L* can be defined as follows:

$$(69)\ \text{Bias}(P_D/P_L) \equiv [P_D/P_L] - 1$$
$$= [\sum_{m=1}^{M} P_m^{1}\sigma_m^{1} - \sum_{m=1}^{M} p_m^{1}\cdot S_m^{0}\sigma_m^{0}]/\sum_{m=1}^{M} p_m^{1}\cdot S_m^{0}\sigma_m^{0} \qquad \text{using (60) and (68)}$$
$$= \sum_{m=1}^{M} p_m^{1}\cdot[S_m^{1}\sigma_m^{1} - S_m^{0}\sigma_m^{0}]/\sum_{m=1}^{M} p_m^{1}\cdot S_m^{0}\sigma_m^{0} \qquad \text{using (7)}$$
$$= N\,\text{Cov}(p^1,S^1 - S^0,(1/N)1_N)/\sum_{m=1}^{M} p_m^{1}\cdot S_m^{0}\sigma_m^{0} \qquad \text{using (1) and (2)}$$

where p^1 is the N dimensional vector of period 1 prices and S^t is the N dimensional vector of period t quantity shares for t = 0,1. Note that the last line in (69) is equivalent to our earlier bias formula (22). Thus P_D will be greater than P_L if the covariance in (69) is positive so that higher period 1 prices p_n^{1} are associated with growing quantity shares, $S_n^{1} - S_n^{0} > 0$.

Now consider a two stage aggregation procedure where the first stage consists of unit value aggregation but the second stage uses the Laspeyres formula applied to the first stage unit value prices and quantities. Thus define the *hybrid Laspeyres price index*, P_{HL}, as follows:

$$(70)\ P_{HL} \equiv \sum_{m=1}^{M} P_m^{1}Q_m^{0}/\sum_{m=1}^{M} P_m^{0}Q_m^{0}$$
$$= [\sum_{m=1}^{M} P_m^{1}Q_m^{0}/\sum_{m=1}^{M} Q_m^{0}]/[\sum_{m=1}^{M} P_m^{0}Q_m^{0}/\sum_{m=1}^{M} Q_m^{0}]$$
$$= \sum_{m=1}^{M} P_m^{1}\sigma_m^{0}/\sum_{m=1}^{M} P_m^{0}\sigma_m^{0}.$$

Note that the denominators in the final formulae for the true Laspeyres index P_L defined by (68) and for the hybrid Laspeyres index P_{HL} defined by (70) are equal. Define the *bias of the hybrid Laspeyres index relative to the Laspeyres index* as follows:

$$(71)\ \text{Bias}(P_{HL}/P_L) \equiv [P_{HL}/P_L] - 1$$
$$= [\sum_{m=1}^{M} P_m^{1}\sigma_m^{0}/\sum_{m=1}^{M} p_m^{1}\cdot S_m^{0}\sigma_m^{0}] - 1 \qquad \text{using (68) and (70)}$$
$$= \sum_{m=1}^{M} p_m^{1}\cdot[S_m^{1} - S_m^{0}]\sigma_m^{0}/\sum_{m=1}^{M} p_m^{1}\cdot S_m^{0}\sigma_m^{0} \qquad \text{using (59)}$$
$$= N\text{Cov}(p^1,S^{1**} - S^{0**},(1/N)1_N)/\sum_{m=1}^{M} p_m^{1}\cdot S_m^{0}\sigma_m^{0} \qquad \text{using (1), (2) and (57)}$$

where p^1 is the N dimensional vector of base period prices and the components of the vector $S^{0**} - S^{1**}$ are the vectors $[S_m^{1} - S_m^{0}]\sigma_m^{0}$ stacked up into a single N dimensional vector. As usual, we have the following three sets of conditions which will imply that the hybrid Paasche is equal to the true Paasche index:

- All period 1 subaggregate price vectors p_m^1 are equal to the corresponding subaggregate unit values; i.e., $p_m^1 = P_m^1 1_m$ for m = 1,...,M;
- The subaggregate quantity shares remain constant going from period 0 to period 1; i.e., $S_m^0 = S_m^1$ for m = 1,...,M or
- The covariance between the current period price vector p^1 and the adjusted share difference vector $S^{1**} - S^{0**}$ defined above is zero; i.e., $\mathrm{Cov}(p^1,S^{1**} - S^{0**},(1/N)1_N) = 0$.

Of course, the first two sets of conditions are special cases of the third condition. The first set of conditions important for choosing how to construct the subaggregates: in order to minimize bias (relative to the Laspeyres price index), use unit value aggregation over products that sell for the same price in the current period.

Looking at the bias formulae (69) and (71), it can be seen that if P_D is greater (less) than P_L so that $\mathrm{Cov}(p^1,S^1 - S^0,(1/N)1_N)$ is positive (negative), then it is *likely* that P_{HL} is also greater (less) than P_L so that $\mathrm{Cov}(p^1,S^{1**} - S^{0**},(1/N)1_N)$ is also positive (negative).

We can use the algebra developed above in order to obtain an exact relationship between the two bias measures (69) and (71). From (69), we have:

$$(72)\ \mathrm{Bias}(P_D/P_L) = [\textstyle\sum_{m=1}^M P_m^1\sigma_m^1 - \sum_{m=1}^M p_m^1\cdot S_m^0\sigma_m^0]/\sum_{m=1}^M p_m^1\cdot S_m^0\sigma_m^0$$
$$= \mathrm{Bias}(P_{HL}/P_L) - \{\textstyle\sum_{m=1}^M P_m^1[\sigma_m^1 - \sigma_m^0]/\sum_{m=1}^M p_m^1\cdot S_m^0\sigma_m^0\} \quad \text{using (71) and (59)}$$
$$= \mathrm{Bias}(P_{HL}/P_L) + M\mathrm{Cov}(\mathbf{P^1}, \sigma^1-\sigma^0,(1/M)1_M)/\textstyle\sum_{m=1}^M p_m^1\cdot S_m^0\sigma_m^0 \quad \text{using (2) and (58)}$$

where $\mathbf{P^1} \equiv [P_1^1,...,P_M^1]$ is the vector of period 1 unit value prices for the M subaggregates and $\sigma^t \equiv [\sigma_1^t,...,\sigma_1^t]$ is the vector of period t subaggregate quantity shares where the σ_m^t were defined by (56).

Suppose that P_D is greater than P_L so that the bias, $\mathrm{Bias}(P_D/P_L)$, is greater than 1. Then from our analysis in section 3, we know that products which have above average prices in period 1 are positively correlated with growing quantity shares. It is *likely* (but not certain) that this positive correlation persists to the subaggregates so that subaggregates m that have above average unit value prices P_m^1 in period 1 are associated with a positive $\sigma_m^1 - \sigma_m^0$ and hence the covariance in (72) will be positive under this hypothesis. It is also *likely* that the hybrid Laspeyres P_{HL} is equal to or greater than the true Laspeyres index P_L under these circumstances so that the bias, $\mathrm{Bias}(P_{HL}/P_L)$, is equal to or greater than 1. Under these hypotheses, using (72), it can be seen that

$$(73)\ \mathrm{Bias}(P_D/P_L) > \mathrm{Bias}(P_{HL}/P_L) \geq 1.$$

Hence the magnitude of the bias of the hybrid Laspeyres index is less than the bias of the Drobisch index. Thus in this case, disaggregation of the single stage unit value price index into a two stage index where the second stage uses a Laspeyres formula does lead to a price index which is closer to the true Laspeyres index.

A similar analysis can be made for the case where P_D is less than P_L so that the bias, $Bias(P_D/P_L)$, is less than 1. In this case, the covariance in (69) is negative and it is *likely* that the covariances in (71) and (72) will also be negative. Under these conditions, using (72), we can deduce that:

(74) $Bias(P_D/P_L) < Bias(P_{HL}/P_L) < 1$.

Hence again, the magnitude of the bias of the hybrid Laspeyres index is less than the bias of the Drobisch index.

The techniques used in section 3 can be adapted to the present context in order to obtain bias formulae if the target index is a Fisher index.

6. Conclusion

In this paper, we have taken a systematic look at the existing literature on unit value biases and extended it somewhat. Our results indicate that it will *usually* be the case that the use of finer commodity classifications to generate unit value prices and quantities that are then inserted into a bilateral index number formula will generate closer approximations to an underlying preferred index.

It should be noted that some use of unit value aggregation is inevitable; i.e., it will always be necessary to aggregate household or establishment purchases or sales of individual products *over time* in order to obtain total purchases or sales of the unit under consideration and then these (within the period time aggregated) prices and quantities are used as inputs into a bilateral index number formula. Data limitations will generally lead to more unit value aggregation where there could be aggregation over households, establishments, geographical areas or products.[16] However, it is generally felt that as a theoretical target, more narrowly defined unit values will generally lead to more accurate price indexes and the present paper seems to reinforce this view.

But there is a problem with the analysis in this paper that needs to be addressed. The problem is that we have assumed that all N prices and quantities for the two periods under consideration are positive. But this condition can be far from being satisfied if we make the scope of our first stage unit values narrower and narrower. As the number of separate commodities N in the aggregate grows, it will generally be the case that more and more zero prices and quantities occur. This is due to the sporadic nature of shipments and purchases, particularly if the time period is relatively short. In limiting cases with a very large N, there can be an extreme lack of matching of products, leading to

[16] The scope of unit value aggregation was discussed in Diewert (1995) without any clear resolution. The ILO (2004) and IMF (2009) Manuals recommended “narrowly defined unit values” at the first stage of aggregation but the exact meaning of this advice is ambiguous. Silver (2010; S216-S220) discusses this issue in more detail.

nonsensical target indexes. For example, using the notation in the previous section, suppose that we have M = 2 so that there are two subaggregates. In period 0, suppose that the subaggregate 1 price and quantity vectors, p_1^0 and q_1^0 are positive so that the subaggregate 1 unit value price, $P_1^0 = p_1^0 \cdot q_1^0 / 1_1 \cdot q_1^0$, is positive and that the subaggregate 2 price and quantity vectors, p_2^0 and q_2^0 are vectors of zeros. Suppose that in period 1, the subaggregate 2 price and quantity vectors, p_2^1 and q_2^1 are positive so that the subaggregate 2 unit value price in period 1, $P_2^1 = p_2^1 \cdot q_2^1 / 1_2 \cdot q_2^1$, is positive and that the subaggregate 1 price and quantity vectors, p_1^1 and q_1^1 are vectors of zeros. Under these conditions, it can be seen that the true Laspeyres and hybrid Laspeyres price indexes, P_L and P_{HL}, are both equal to 0 and the true Paasche and hybrid Paasche price indexes, P_P and P_{HP}, are both equal to $+\infty$. Note that the Drobisch price index, P_D, is equal to P_2^1/P_1^0 and this will be a much more reasonable measure of price change, particularly if the products in the two subaggregates are all fairly "similar". Thus some caution is required in applying the results derived in this paper when there are zeros in the data.[17]

References

Balk, B.M. (1995), "Axiomatic Price Index Theory: A Survey", *International Statistical Review* 63, 69-93.

Balk, B.M. (1998), "On the Use of Unit Value Indices as Consumer Price Subindices", in: Lane, W. (Ed.) *Proceedings of the Fourth Meeting of the International Working Group on Price Indices*, Washington, DC: Bureau of Labour Statistics.

Balk, B.M. (2005), "Price Indexes for Elementary Aggregates: The Sampling Approach", *Journal of Official Statistics*, 21, 675-699.

Balk, B.M. (2008), "*Price and Quantity Index Numbers*, New York: Cambridge University Press.

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

Davies, G. R. (1932), "Index Numbers in Mathematical Economics", *Journal of the American Statistical Association* 27, 58-64.

Diewert, W.E. (1992), "Fisher Ideal Output, Input and Productivity Indexes Revisited", *Journal of Productivity Analysis* 3, 211-248.

Diewert, W. E. (1995), "Axiomatic and Economic Approaches to Elementary Price Indexes", Discussion Paper No. 95-01, Department of Economics, University of British Columbia, Vancouver, Canada. http://faculty.arts.ubc.ca/ediewert/axio.pdf

Drobisch, M.W. (1871), "Ǔber die Berechnung der Veränderungen der Waarenpreise und des Geldwerths", *Jahrbücher für Nationalökonomie und Statistik*, 16, 143 – 156.

[17] Silver (2010; S216) notes some other situations where it would be more appropriate to use a unit value price index rather than a superlative index. Basically, the use of a unit value index is appropriate when purchasers of a product buy essentially the same product from suppliers at different prices over the time period under consideration.

Dutot, C. (1738), *Réflections politiques sur les finances et le commerce*, Volume 1, La Haye: Les frères Vaillant et N. Prevost.

Fisher, I. (1922), The Making of Index Numbers, Boston: Houghton-Mifflin.

ILO, IMF, OECD, UNECE, Eurostat and World Bank (2004), *Consumer Price Index Manual: Theory and Practice*, ed. by P. Hill, ILO: Geneva.

IMF, ILO, OECD, UNECE and World Bank (2009), *Export and Import Price Index Manual*, ed. by M. Silver, IMF: Washington, D.C.

Laspeyres, E. (1871), "Die Berechnung einer mittleren Waarenpreissteigerung", Jahrbücher für Nationalökonomie und Statistik 16, 296-314.

Paasche, H. (1874), "Über die Preisentwicklung der letzten Jahre nach den Hamburger Borsennotirungen", *Jahrbücher für Nationalökonomie und Statistik* 12, 168-178.

Párniczky, G. (1974), "Some Problems of Price Measurement in External Trade Statistics", *Acta Oeconomica* 12:2, 229 – 240.

Silver, M. (2009a), "Do Unit Value Export, Import, and Terms of Trade Indices Represent or Misrepresent Price Indices?", *IMF Staff Papers* 56, 297-322. Washington D.C.: IMF.

Silver, M. (2009b), "Unit Value Indices" Chapter 2 in IMF, ILO, OECD, UNECE and World Bank (2008), *Export and Import Price Index Manual*, ed. by M. Silver, IMF: Washington, D.C.

Silver, M. (2009c), "An Index Number Formula Problem: The Aggregation of Broadly Comparable Items", *IMF Working Paper* WP/09/19, Washington D.C.: IMF.

Silver, M. (2010), "The Wrongs and Rights of Unit Value Indices", *Review of Income and Wealth* Series 56, Special Issue 1, S206-S223.

Timmer, M. (1996), "On the Reliability of Unit Value Ratios in International Comparisons", Research Memorandum GD-31, University of Groningen, Dec. 1996.

von Bortkiewicz, L. (1923), "Zweck und Struktur einer Preisindexzahl", *Nordisk Statistisk Tidsskrift* 2, 369-408.

von der Lippe, P.M. (2007a), *Index Theory and Price Statistics*, Frankfurt/Main: Peter Lang.

von der Lippe, P.M. (2007b), "Price Indices and Unit Value Indices in German Foreign Trade Statistics", unpublished Paper University of Duisburg- Essen. Internet: http://mpra.ub.uni-muenchen.de/5525/

Walsh, C. M. (1901), *The Measurement of General Exchange Value*, New York: Macmillan and Co.

Walsh, C. M. (1921), *The Problem of Estimation*, London: P.S. King & Son.