Journal of Official Statistics
Vol. 10. No. 1, 1994, pp. 103–108

Letter to the Editor

Letters to the Editor will be confined to discussion of papers which have appeared in the Journal of Official Statistics and of important issues facing the statistical community.

A New Functional Form for Price Indexes' Elementary Aggregates

Jörgen Dalén's paper on "Computing Elementary Aggregates in the Swedish Consumer Price Index" (Journal of Official Statistics, Vol. 8, no. 2, 1992, pp. 129–147) provides important theoretical and empirical results on the construction of the basic components of consumer price indexes. Two especially noteworthy contributions are a new index number test and a new index number functional form that is now used by Statistics Sweden for aggregating outlets' prices to form the basic components of its Consumer Price Index (CPI). The purpose of this note is to propose a modification of Dalén's new price index that incorporates a logarithmic mean in lieu of an arithmetic mean. This modification offers two advantages: (1) Dalén's new index number test is satisfied exactly rather than just approximately; and (2) the geometric mean index is approximated more closely than by the index introduced by Dalén. This second advantage is especially noteworthy because Dalén's index itself approximates the geometric mean index better than other alternatives, including one resembling a Fisher's ideal index.

Dalén's new index number test, called the permutation test, is motivated by the observed tendency of outlets' prices for a good to oscillate. The second version of this test, which Dalén (p. 138) suggests is "enough for practical purposes," concerns the behavior of the functional form for the basic component price index when every price receives the same weight. It requires the index to equal 1 whenever the outlets in the index collectively offer the same set of prices, even if the identity of the outlet offering each particular price changes. This test, as well as other tests discussed by Dalén, is satisfied by the geometric mean index

$$G_{0t} = \exp\left\{\sum w_i \log(r_i)\right\}$$

where the w_i, which sum to 1, are outlets' weights in the basic component index and $r_i \equiv p_{it}/p_{i0}$.

The disadvantage of the geometric mean index is its dissimilarity to the Edgeworth index that Statistics Sweden uses to form higher level CPI aggregates. Dalén's new index provides for a unified approach at all levels of aggregation yet behaves quite similarly to a geometric mean index. It is

$$RA_{0t} = \frac{\sum w_i p_{it}/a(p_{i0}, p_{it})}{\sum w_i p_{i0}/a(p_{i0}, p_{it})}$$

where $a(p_{it}, p_{i0})$ denotes the arithmetic mean $(p_{it} + p_{i0})/2$. Ideally, the w_i represent outlets' shares of consumers' expenditure on the good in question over the period from time 0 to time t. If so, for any arbitrary outlet i, $w_i/a(p_{i0}, p_{it})$ is proportional to i's estimated quantity during the period from 0 to t calculated as the ratio of i's sales to consumers during that period to an estimate of i's average price during that period.

Assuming that prices have constant growth rates between times 0 and t implies a different estimator for the average price over the interval $[0, t]$. Let $p_{is} = p_{i0}e^{\rho_i s}$ for any $s \in [0, t]$. Then (assuming that $\rho_i \neq 0$), the average price during the interval $[0, t]$ is $\int_0^t (1/t)p_{i0}e^{\rho_i s} ds = (p_{it} - p_{i0})/[\ln(p_{it}) - \ln(p_{i0})]$, which is the logarithmic or "Vartia" mean, denoted $v(p_{i0}, p_{it})$. This mean may be thought of as lying two-thirds of the way from an arithmetic mean to a geometric mean because $v(p_{i0}, p_{it}) \approx [a(p_{i0}{}^{1/3}, p_{it}{}^{1/3})]^3$, while the geometric mean $g(p_{i0}, p_{it}) = \lim_{\epsilon \to 0}[a(p_{i0}{}^{\epsilon}, p_{it}{}^{\epsilon})]^{1/\epsilon}$, or $\exp\{a[\log(p_{i0}), \log(p_{it})]\}$. In fact, before the role of the logarithmic mean in ideal log-change index numbers was discovered, both $\frac{2}{3}g(p_{i0}, p_{it}) + \frac{1}{3}a(p_{i0}, p_{it})$ and $[g(p_{i0}, p_{it})]^{2/3} \times [a(p_{i0}, p_{it})]^{1/3}$ were suggested as approximations to it.

Replacing $a(p_t, p_0)$ in the formula for RA_{0t} by the average price during $[0, t]$ under the constant growth rate assumption gives

$$RV_{0t} = \frac{\sum w_i p_{it}/v(p_{i0}, p_{it})}{\sum w_i p_{i0}/v(p_{i0}, p_{it})}$$

$$= 1 + \frac{\sum w_i \log(r_i)}{\sum w_i/v(r_i, 1)}.$$

Since RV_{0t} equals 1 whenever G_{0t} equals 1, it satisfies the permutation test exactly. Furthermore, Dalén notes that under the common condition of right skew in the distribution of the r_i, RA_{0t} has a slight downward bias compared with G_{0t}. In contrast, for the range of price changes that generally enter price indexes, RV_{0t} will be numerically indistinguishable from G_{0t}. For example, suppose that ten outlets furnish price quotes for a good, nine of which report no price change. If the tenth outlet reports an outlier of $r_{10} = 2$, then, rounding to the nearest thousandth, $G_{0t} = RV_{0t} = 1.072$, while $RA_{0t} = 1.069$. If $r_{10} = 10$ – an extreme outlier that would certainly be rejected for use in the U.S. CPI – then $G_{0t} = 1.259$, $RV_{0t} = 1.249$ and $RA_{0t} = 1.178$.

Letting $\mathbf{r}$ denote the vector of the r_i, $G(\mathbf{r})$, $RV(\mathbf{r})$, and $RA(\mathbf{r})$ have identical first and second derivatives at the point $\mathbf{r} = \mathbf{1}$. Dalén reports the third derivatives for $G(\mathbf{1})$ and $RA(\mathbf{1})$, though a minus sign appears to have been inadvertently inserted in front of his expression for $RA'''_{iij}(\mathbf{1})$. The third derivatives of $RV(\mathbf{1})$ are derived in the mathematical appendix. Table 1 displays all three sets of third derivatives.

Because $RV'''_{iii}(\mathbf{1})$ differs less from $G'''_{iii}(\mathbf{1})$ than does $RA'''_{iii}(\mathbf{1})$, right skew in $\mathbf{r}$ causes less discrepancy between $RV(\mathbf{r})$ and $G(\mathbf{r})$ than between $RA(\mathbf{r})$ and $G(\mathbf{r})$. Furthermore, unless $w_i \leq \frac{1}{6}$ (in which case both derivatives are near zero), $RV'''_{iij}(\mathbf{1})$ is closer to $G'''_{iij}(\mathbf{1})$ than is $RA'''_{iij}(\mathbf{1})$. The equality or near-equality of the first three derivatives of $RV(\cdot)$ and $G(\cdot)$ at $\mathbf{r} = \mathbf{1}$ accounts for the tendency for $G(\mathbf{r})$ and $RV(\mathbf{r})$ to agree precisely when expressed to three decimal places if the price changes in $\mathbf{r}$ are moderate.

A final noteworthy consequence of the similarity between $RV(\mathbf{r})$ and $G(\mathbf{r})$ is an expression in price-quantity form for

Table 1. Third derivatives of $G(\cdot)$, $RA(\cdot)$ and $RV(\cdot)$

	$G'''_{ijk}(\mathbf{1})$	$RA'''_{ijk}(\mathbf{1})$	$RV'''_{ijk}(\mathbf{1})$
$k = j = i$	$w_i(1 - w_i)(2 - w_i)$	$1.5w_i(1 - w_i)^2$	$w_i(1 - w_i)(2 - 1.5w_i)$
$k = i; j \neq i$	$-w_i w_j(1 - w_i)$	$-w_i w_j(1 - 1.5w_i)$	$-w_i w_j(\frac{7}{6} - 1.5w_i)$
$k \neq j \neq i$	$w_i w_j w_k$	$1.5w_i w_j w_k$	$1.5w_i w_j w_k$.

the Törnqvist price index $T(\mathbf{w}_0, \mathbf{w}_t, \mathbf{r}) \equiv \exp[\sum a(w_{i0}, w_{it}) \log(r_i)]$, where $\mathbf{w}_0$ and $\mathbf{w}_t$ are vectors of *good's* expenditures shares at times 0 and t. Let $\bar{q}_i = a(w_{i0}, w_{it})/v(p_{i0}, p_{it})$. Although the $\bar{q}_i$ are deflated expenditure *shares,* rescaling them by a measure of average expenditures would not change the value of the approximation. It is

$$T(\mathbf{w}_0, \mathbf{w}_t, \mathbf{r}) = \frac{\sum p_{it}\bar{q}_i}{\sum p_{i0}\bar{q}_i}.$$

If consumers' expenditures on goods are constant, as would occur with constant budgets and Cobb-Douglas preferences, then $\bar{q}_i$ is proportional to $[v(1/q_{i0}, 1/q_{it})]^{-1}$. Assuming that prices (and, therefore, quantities) change at a constant rate in the interval $[0, t]$, this, in turn, equals a harmonic mean of the quantities of good i that consumers buy at various times between time 0 and time t.

Marshall Reinsdorf
U.S. Bureau of Labor Statistics
Washington, D.C. 20212
U.S.A.

Appendix

In order to find the third derivatives at $\mathbf{r} = \mathbf{1}$ of $RV(\mathbf{r}) = 1 + [\Sigma w_i \log r_i]/[\Sigma w_i(\log r_i)/(r_i - 1)]$, it is useful to know the derivatives of $(\log r_i)/(r_i - 1) \equiv u(r_i)$ at $r_i = 1$. To find these derivatives, define $x(r_i)$ as $(r_i - 1)/(r_i + 1)$. In evaluating the derivatives of $u(\cdot)$ it is convenient to suppress the i subscripts.

Since $r = (1 + x)/(1 - x)$

$$\log r = \log \frac{1 + x}{1 - x}.$$

A Taylor series expansion for log $[(1 + x)/(1 - x)$ is $2(x + x^3/3 + x^5/5 + \cdots)]$ where $x < 1$ since $r > 0$. Furthermore, $r - 1 = 2x/(1 - x)$. Hence

$$u(r) = f(x(r))$$

where

$$f(x) \equiv (1 - x)\left(1 + \frac{x^2}{3} + \frac{x^4}{5} + \cdots\right).$$

We then have

$$\begin{aligned}
u'(r) &= f'(x(r))x'(r) \\
u''(r) &= f''(x(r))[x'(r)]^2 + f'(x(r))x''(r) \\
u'''(r) &= f'''(x(r))[x'(r)]^3 \\
&\quad + 3f''(x(r))x''(r)x'(r) \\
&\quad + f'(x(r))x'''(r).
\end{aligned}$$

Evaluating the derivatives of $f(x)$ at $x = 0$ gives

$$\begin{aligned}
f'(x) &= -\left(1 + \frac{x^2}{3} + \frac{x^4}{5} + \cdots\right) \\
&\quad + (1 - x)(\tfrac{2}{3}x + \tfrac{4}{5}x^3 + \cdots) = -1 \\
f''(x) &= -2(\tfrac{2}{3}x + \tfrac{4}{5}x^3 + \cdots) \\
&\quad + (1 - x)(\tfrac{2}{3} + \tfrac{12}{5}x^2 + \cdots) = \tfrac{2}{3} \\
f'''(x) &= -3(\tfrac{2}{3} + \tfrac{12}{5}x^2 + \cdots) \\
&\quad + (1 - x)(\tfrac{24}{5}x + \tfrac{120}{7}x^3 + \cdots) = -2.
\end{aligned}$$

Evaluating the derivatives of $x(r)$ at $r = 1$ gives

$$x'(r) = 2(r+1)^{-2} = \tfrac{1}{2}$$
$$x''(r) = 4(r+1)^{-3} = -\tfrac{1}{2}$$
$$x'''(r) = 12(r+1)^{-4} = \tfrac{3}{4}.$$

Substituting these values into the expressions for the derivatives of $u(\cdot)$ gives

$$u'(1) = (-1)(\tfrac{1}{2}) = -\tfrac{1}{2}$$
$$u''(1) = (\tfrac{2}{3})(\tfrac{1}{2})^2 + (\tfrac{1}{2}) = \tfrac{2}{3}$$
$$u'''(1) = (-2)(\tfrac{1}{2})^3 + 3(\tfrac{2}{3})(-\tfrac{1}{2})(\tfrac{1}{2}) - (\tfrac{3}{4}) = -\tfrac{3}{2}.$$

With these results in hand, we now turn to the derivatives of $RV(\cdot)$. We have

$$RV(\mathbf{r}) = 1 + \frac{\Sigma w_i \log r_i}{\Sigma w_i u(r_i)}$$

$$\frac{\partial RV(\cdot)}{\partial r_i} = \frac{w_i/r_i}{\Sigma w_h u(r_h)} - \frac{[\Sigma w_h \log r_h] w_i u'(r_i)}{[\Sigma w_h u(r_h)]^2}$$

$$\frac{\partial^2 RV(\cdot)}{(\partial r_i)^2} = \frac{-w_i/r_i^2}{\Sigma w_h u(r_h)} - \frac{2(w_i^2/r_i)u'(r_i)}{[\Sigma w_h u(r_h)]^2} - \frac{[\Sigma w_h \log r_h] w_i u''(r_i)}{[\Sigma w_h u(r_h)]^2} + \frac{2[\Sigma w_h \log r_h][w_i u'(r_i)]^2}{[\Sigma w_h u(r_h)]^3}$$

$$\frac{\partial^2 RV(\cdot)}{\partial r_i \partial r_j} = -\frac{w_i w_j u'(r_j)/r_i}{[\Sigma w_h u(r_h)]^2} - \frac{w_i w_j u'(r_i)/r_j}{[\Sigma w_h u(r_h)]^2} + 2\frac{[\Sigma w_h \log r_h][w_i u'(r_i)][w_j u'(r_j)]}{[\Sigma w_h u(r_h)]^3}$$

$$\frac{\partial^3 RV(\cdot)}{(\partial r_i)^3} = \frac{2w_i/r_i^3}{\Sigma w_h u(r_h)} + \frac{(w_i/r_i)^2 u'(r_i)}{[\Sigma w_h u(r_h)]^2} + 2\frac{(w_i/r_i)^2 u'(r_i) - (w_i^2/r_i)u''(r_i)}{[\Sigma w_h u(r_h)]^2} + 4\frac{[w_i^3/r_i][u'(r_i)]^2}{[\Sigma w_h u(r_h)]^3} - \frac{w_i\{(w_i/r_i)u''(r_i) + [\Sigma w_h \log r_h]u'''(r_i)\}}{[\Sigma w_h u(r_h)]^2} + 2\frac{[\Sigma w_h \log r_h] w_i^2 u''(r_i)u'(r_i)}{[\sum w_h u(r_h)]^3} + 2\frac{w_i^3[u'(r_i)]^2/r_i + 2[\Sigma w_h \log r_h] w_i^2 u'(r_i)u''(r_i)]}{[\Sigma w_h u(r_h)]^3} - 6\frac{[\Sigma w_h \log r_h][w_i u'(r_i)]^3}{[\Sigma w_h u(r_h)]^4}.$$

Evaluating this expression at $\mathbf{r} = \mathbf{1}$ gives

$$2w_i + (-1/2)w_i^2 + 2w_i^2[(-1/2) - (2/3)] + 4w_i^3(1/4) - w_i^2(2/3) + 2w_i^3(1/4)$$
$$= (3/2)w_i^3 - (7/2)w_i^2 + 2w_i$$
$$= w_i(1 - w_i)(2 - \tfrac{3}{2}w_i).$$

We next evaluate $RV_{iij}(\mathbf{1})$. It is

$$\frac{\partial^3 RV(\cdot)}{(\partial r_i)^2 \partial r_j} = \frac{w_i w_j u'(r_j)/r_i^2}{[\Sigma w_h u(r_h)]^2} + 4\frac{(w_i^2/r_i) w_j u'(r_i)u'(r_j)}{[\Sigma w_h u(r_h)]^3} - \frac{w_i w_j u''(r_i)/r_j}{[\Sigma w_h u(r_h)]^2} + 2\frac{[\Sigma w_h \log r_h] w_i w_j u''(r_i) w_j u'(r_j)}{[\Sigma w_h u(r_h)]^3} - 6\frac{[\Sigma w_h \log r_h][w_i u'(r_i)]^2 w_j u'(r_j)}{[\Sigma w_h u(r_h)]^4} + 2\frac{[w_i u'(r_i)]^2 (w_j/r_j)}{[\Sigma w_h u(r_h)]^3}.$$

Evaluating this expression at $\mathbf{r} = \mathbf{1}$ gives

$$-\tfrac{1}{2}w_iw_j + w_i^2w_j - \tfrac{2}{3}w_iw_j + \tfrac{1}{2}w_i^2w_j$$
$$= \tfrac{3}{2}w_i^2w_j - \tfrac{7}{6}w_iw_j = w_iw_j(\tfrac{3}{2}w_i - \tfrac{7}{6}).$$

Finally, we evaluate the last derivative in Table 1. It is

$$\frac{\partial^3 RV(\cdot)}{\partial r_i \partial r_j \partial r_k} = 2\frac{w_iw_jw_ku'(r_j)u'(r_k)/r_i}{[\Sigma w_h u(r_h)]^3}$$
$$+ 2\frac{w_iw_jw_ku'(r_i)u'(r_k)/r_j}{[\Sigma w_h u(r_h)]^3}$$
$$+ 2\frac{[w_iu'(r_i)][w_ju'(r_j)][w_k/r_k]}{[\Sigma w_h u(r_h)]^3}$$
$$- 6\frac{[\Sigma w_h \log r_h][w_iu'(r_i)][w_ju'(r_j)][w_ku'(r_k)]}{[\Sigma w_h u(r_h)]^4}.$$

Evaluating this expression at $\mathbf{r} = \mathbf{1}$ gives $\frac{3}{2}w_iw_jw_k$.

References

Dalén, J. (1992). Computing Elementary Aggregates in the Swedish Consumer Price Index. Journal of Official Statistics, 8, 129–147.

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

Lorenzen, G. (1990). A Unified Approach to the Calculation of Growth Rates. The American Statistician, 44, 148–150.

Sato, K. (1974). Ideal Index Numbers that Almost Satisfy the Factor Reversal Test. Review of Economics and Statistics, 56, 549–552.

Sato, K. (1976). The Ideal Log-Change Index Number. Review of Economics and Statistics, 58, 223–228.

Theil, H. (1973). A New Index Number Formula. Review of Economics and Statistics, 55, 498–502.

Törnqvist, L., Vartia, P., and Vartia, Y.O. (1985). How Should Relative Changes Be Measured? The American Statistician, 39, 43–46.

Vartia, Y.O. (1976). Ideal Log-Change Index Numbers. Scandinavian Journal of Statistics, 3, 121–126.

Reply

I thank Marshall Reinsdorf for his many insightful and enlightening comments on my article. His letter is an indication of the increasing attention given by statistical agencies to the effects of different methods for computing elementary aggregates in consumer price indices.

I have a few detailed comments to make:

1. Reinsdorf's *RV* index satisfies the weak (VI-B in my article) but not the strong (VI) version of the permutation test.

2. Reinsdorf correctly points out an erroneous minus sign in one of my formulas. This error does not, however, influence my main result 10.2.3, where I have used the correct expression as given by Reinsdorf.

3. In 10.2.3 in my article I give the approximate expression

$$G - RA = \gamma/12 + (\mu - 1)\sigma^2/4.$$

A similar expression could be given for *RV*: $G - RV \approx (\mu - 1)\sigma^2/12$. Although this approximation is only valid for price ratios $\mathbf{r}$ sufficiently close to one, it confirms Reinsdorf's statement that *RV* is a closer approximation to *G* than is *RA*.

4. The *G* and *RA* formulas can be derived from well-known superlative index formulas (Törnqvist and Edgeworth, resp.) if weights are set equal. This seems not to be possible with the *RV* formula.

5. Reinsdorf's "fixed basket" approximation to the Törnqvist index is very

interesting, since the principal argument against using geometric means in price index computations seems to be that they could not be given a fixed basket interpretation. Reinsdorf's approximation shows the contrary. My own numerical experimentation indicates that this approximation is very good, but it still would be interesting to see a more elaborate mathematical explanation.

Jörgen Dalén
Statistics Sweden, Price Statistics
S-115 81 Stockholm
Sweden