SOME OBSERVATIONS ON PRICE INDEX ESTIMATORS

Alan Dorfman, Sylvia Leaver, and Janice Lent
U.S. Bureau of Labor Statistics

1. Introduction

Few statistics produced by governmental agencies have greater impact than *price indexes.* Statisticians, however, seldom study these indexes. This is not because index construction is straightforward. Indeed, in terms of the controversies surrounding price indexes and the complexity of the data integration they require, they might make a claim to be the *most* complicated of statistics. A *price index* is a measure of change from one time period to another of the purchasing power of a given population's monetary unit. The term price index is also used to refer to the formula—chosen from among a wealth of rival formulas—used to calculate the measure of change. Economists have developed a body of theory which they use to compare price index formulas, applied at the population level, to a *cost of living index (COLI),* the ratio of minimal costs needed in the two time periods to achieve a given standard of living. In practice, government agencies apply a given price index formula to a sample from the target population, yielding an *index estimator* of the selected population index. The estimator itself, however, is also often referred to as a price index, and the resulting ambiguity of the term tends to blur the distinctions between the following questions: (1) What is the relation of the (population) price index, calculated by a given formula, to the ideal *COLI*? (2) What relations exist between various population indexes? (3) What relations exist between various index estimators? (4) What is the relation of a given index estimator to a particular target index? These questions form the substrate of this paper. We focus on the *consumer price index,* but the proposed answers apply to other sorts of price indexes as well.

2. Background

The purpose of a consumer price index is said to be "to measure changes over time in the general level of prices of goods and services that a reference population acquire, use or pay for consumption."[1] A measure of relative change in the price of a particular item j between time periods 1 and 2 is the price ratio p_{j2}/p_{j1}, where p_{jt} represents the price of item j at time $t \in \{1, 2\}$. One may therefore obtain a simple price index for a given population just by averaging these price ratios over the collection of N items purchased by the population:

$$A = \frac{1}{N} \sum_{j=1}^{N} \frac{p_{j2}}{p_{j1}}.$$

The resulting arithmetic mean index A is called the Carli index.[2] Assigning weights to the price ratios in A produces the Sauerbeck index, $A_{\underline{w}} = \sum_{j=1}^{N} w_j p_{j2}/p_{j1}$, where $\sum_{j=1}^{N} w_j = 1$. We shall see that the Carli and Sauerbeck indexes both possess properties that economists consider undesirable; for decades, in fact, these indexes have been explicitly rejected by economists such as Irving Fisher (1922). From a statistical standpoint, the deficiency of the

[1] Turvey (1989) p. 4. (This quote is from the "Resolution concerning consumer price indices" of the Fourteenth International Conference of Labour Statisticians.)

[2] Carli, G. (1764).

arithmetic mean indexes may be seen to stem from the actual probability distributions of the prices and hence of the price relatives p_{j2}/p_{j1} : empirical evidence suggests that they follow an approximate log-normal distribution, the skewness of which renders A upwardly biased relative to the median of the distribution. A more attractive method of aggregation may be to average the *logarithms* of the price relatives, which are more likely to be approximately normally distributed, and exponentiate the result. This leads to the geometric mean index

$$G_{\underline{w}} = \exp\left\{\sum_{j=1}^{N} w_j \ln\left(\frac{p_{j2}}{p_{j1}}\right)\right\} = \prod_{j=1}^{N}\left(\frac{p_{j2}}{p_{j1}}\right)^{w_j}.$$

Also known as the Jevons index, $G_{\underline{w}}$ (or G, in the case of equal weights) preserves the distribution of the price ratios, i.e., $G_{\underline{w}}$ is log-normally distributed when the price ratios are mutually independent, log-normal random variables.

Rather than focusing on the price ratios of individual items, we may wish to consider the percentage change in the total price of a bundle of goods and services. This concept leads to the Laspeyres index

$$L = \frac{\sum_{j=1}^{N} q_{j1}p_{j2}}{\sum_{j=1}^{N} q_{j1}p_{j1}} = \sum_{j=1}^{N} w_{j1}\left(\frac{p_{j2}}{p_{j1}}\right),$$

where q_{jt} denotes the quantity of item j in the bundle at time t, and $w_{jt} = p_{jt}q_{jt}/\sum_{k=1}^{N} p_{kt}q_{kt}$. The weight w_{jt} represents the *expenditure share* for item j in period $t \in \{1, 2\}$. For the Laspeyres index, the quantities considered are based on consumer buying habits during the initial time period. When buying habits remain unchanged across time ($q_{j2} = q_{j1}$ for all j), the Laspeyres index is considered a true COLI. That is, it reflects the change in the cost of achieving a given living standard. In order to achieve a given standard, however, some consumers may be able to shift their purchases toward equivalent lower-cost items as relative prices change, and the Laspeyres index may exaggerate the increase in their actual cost of living. Similarly, the Paasche index,

$$P = \frac{\sum_{j=1}^{N} q_{j2}p_{j2}}{\sum_{j=1}^{N} q_{j2}p_{j1}} = \frac{1}{\sum_{j=1}^{N} w_{j2}\left(p_{j2}/p_{j1}\right)^{-1}},$$

which is based on quantities for period 2, may *underestimate* the true change in the cost of living unless quantities remain constant across the two periods.

To achieve a compromise between the Laspeyres and Paasche indexes, we may take their geometric mean. This leads to Fisher's ideal index, advocated by Irving Fisher (1922) and defined as $F = \sqrt{LP}$. The Fisher index incorporates quantity information from both periods 1 and 2. In practice, data on quantities of items purchased are generally not available; estimates of quantity may be derived indirectly from estimated expenditure shares. Alternatively, expenditure shares may be used as weights in the index formula. When estimated expenditure shares for both periods 1 and 2 are available, we may average the shares across the two periods. The geometric mean index based on these average share weights is called the Törnqvist index:

$$T = \prod_{j=1}^{N}\left(\frac{p_{j2}}{p_{j1}}\right)^{w_{j,1,2}},$$

where $w_{j,1,2} = (w_{j1} + w_{j2})/2$. The Fisher and Törnqvist indexes are known as *superlative* indexes, and economic theory suggests that they approximate a COLI (see, for example, Diewert 1987). Intuitively, we can see that using information on buying patterns in both reference periods allows us to relax the assumption of constant quantities underlying the Laspeyres and Paasche indexes. Differences between values of the superlative indexes and the corresponding values of the Laspeyres, Paasche, and geometric mean indexes are referred to as *substitution effects;* i.e., these differences reflect the extent to which consumers alter their buying patterns in response to changes in relative prices.

The Test Approach to Comparing Price Indexes

One approach used by economists to compare price indexes is the *test approach,* in which several "tests" of reasonableness are defined. The *proportionality test,* for example, specifies that if, for every item j, $p_{j2} = cp_{j1}$ for some constant c, the index $I_{1,2}$, measuring price change between periods 1 and 2, must equal c. Indexes based on price ratios—including the Carli index—clearly pass this test. The *time reversal test* specifies that the "backward" price index $I_{2,1}$, calculated by interchanging the price information from periods 1 and 2, must equal the multiplicative inverse of $I_{1,2}$. For practical purposes, this means that chained indexes behave reasonably when prices change and then revert to their former values. If we have three time periods, 1, 2, and 3, and $p_{j1} = p_{j3}$ for all j, then $I_{1,2}I_{2,3} = I_{1,3} = 1$, regardless of the values of the p_{j2}. The superlative indexes F and T pass the time reversal test, as does the geometric mean index. The Laspeyres and Sauerbeck indexes, however, fail this test; the Sauerbeck, in particular, is upwardly biased in the presence of price/time reversal.[3] The test approach originated with Fisher (1922); for additional tests, see, for example, Diewert (1987) and Dalén (1992).

Composite and "Hybrid" Price Indexes

It is natural to separate the vast array of consumer goods into groups characterized by item category, the place where items are bought, or both, since distinctions of item-type and of location may give rise to different rates of price change. Correspondingly, we may write the basic indexes in "composite form," for example, the Laspeyres as $L = \sum_g w_{g1} L_g$, where $w_{gt} = \sum_{i=1}^{N_g} q_{git} p_{git} / \sum_g \sum_{i=1}^{N_g} q_{git} p_{git}$ is the expenditure share for the gth group, and $L_g = \left(\sum_i q_{gi1} p_{gi2} \right) / \left(\sum_i q_{gi1} p_{gi1} \right)$ is the "sub-index" for group g. The computation of sub-indexes is called "lower-level aggregation," while the process of combining sub-indexes into an overall index is called "upper-level aggregation." The formulae used for upper- and lower-level aggregation need not be the same; we can compute "hybrid" indexes, e.g., $L_{G_w} = \sum_g w_{g1} G_{gw}$—a Laspeyres aggregation of geometric mean sub-indexes.

For general notation, we write $I = X_Y$ where X and Y are the formulas used for upper- and lower-level aggregation, respectively. Price index estimation then comprises three steps: (1) estimate sub-indexes Y_g, typically using establishment survey data (and perhaps some household data for weighting purposes); (2) estimate the stratum weights w_g, typically a

[3] Under most economic conditions, the Laspeyres index also suffers from an upward bias in the presence of price/time reversal. When quantities remain constant across time, however, the Laspeyres index passes the time reversal test.

function of expenditure shares estimated from household survey data; and (3) use the formula X, with raw ingredients $\widehat{Y}_g$ and $\widehat{w}_g$, to get $\widehat{I} = \widehat{X}_{\widehat{Y}}$. The result may be termed a "composite estimator."

3. Findings from a Simulation Study—"Laspeyresville"

We conducted a series of simulation experiments to assess (1) the behaviors of various population indexes and their sample estimators in relation to each other, and (2) the extent to which the sample estimators approximated their population counterparts. We first constructed an artificial population of households, items, and outlets. The economic behavior of our artificial population was of the Laspeyres type: quantities of items purchased by households were held constant from period to period, except for random variation, irrespective of changes in relative prices. The population of "Laspeyresville" consisted of 2000 households purchasing goods from 100 outlets. Commodities were divided into three item strata, each comprising 10 item types. All items within a stratum shared the same inflation rate, but their initial prices and quantities differed.

Next, in a manner similar to the sampling methods of the Consumer Expenditure and CPI price surveys—though considerably simplified—we drew repeated samples of households, items, and outlets from the Laspeyresville population. Though we did not attempt to replicate the complexities of CPI sampling (e.g., frame construction based on a Point of Purchase Survey), we constructed the samples in such a way that the *reliability* of the simulated sample estimators of expenditure shares and subindexes approximated that of the corresponding estimators used in the CPI. We ran 35 experiments, each consisting of one population construction from which 50 samples ("runs") were drawn. The variances of the population prices and quantities, as well as the sample sizes for households, outlets, and items, varied slightly across experiments but were the same for all samples within an experiment. For each run, samples of households, outlets, and items within outlets were drawn for three pricing periods: periods 1 and 2 (the periods between which price change was to be estimated) and a base period B preceding the two pricing periods. Estimates of expenditure weights for each item stratum were constructed from expenditures associated with the sampled households.

Period 1 to period 2 price indexes were computed for each stratum for the following indexes, using both population and sample data: Laspeyres, Paasche, Fisher, geometric mean indexes (with expenditure shares either constant or based in periods 1, 2, or B), Törnqvist, Carli, and Sauerbeck. Since outlet expenditure data were available, we constructed "pure" index estimators of these population indexes based on price and expenditure data from *only* the sample outlets. We also computed composite index estimators, based on both household and establishment data, as described in Section 2. We tabulated the distributional characteristics of the population indexes and their pure and composite estimators, and we compared the sample estimators with their population targets via means, standard deviations, root mean square error (RMSE) and predictive mean square error (PMSE) for the pure and composite Fisher index estimators. The RMSE was computed with reference to the population Fisher index. For pure indexes, the PSME was computed with respect to the pure Fisher estimate; for composite indexes, it was computed with respect to the composite Fisher estimate.

Table 1 shows the values of various population indexes (I), averaged over the experiments. We found that for all experiments, values of the population Laspeyres, Paasche, Fisher and Törnqvist indexes were essentially the same—as expected, since the population was constructed in a manner consistent with the "constant quantities" assumption underlying the Laspeyres index. The expenditure-weighted population geometric mean indexes also behaved as expected: G_1 was severely biased down, G_2 was similarly biased up, and G_B showed a moderate downward bias. The equally-weighted geometric mean G_c was surprisingly similar to the Laspeyres, while the Carli and Sauerbeck indexes were predictably biased up.

Table 1. Comparison of Average Laspeyresville Population Indexes

	F	L	P	T	$A_{\underline{w}}$	A	G_1	G_2	G_c
I	1.45599	1.45603	1.45595	1.45598	1.86251	1.94884	1.09669	1.91429	1.44750
$I-F$	0	0.00004	-0.00004	-0.00001	0.40652	0.49285	-0.35930	0.45830	-0.00849

With respect to composite estimators, $\widehat{L}_{GB}$, which follows a formula similar to that of the current CPI, tended to have larger RMSE than $\widehat{L}_L$, the formula it replaced in the CPI as of January 1999. $\widehat{L}_{GB}$ had a smaller RMSE only 9/35 times. The composite estimator $\widehat{G}_{L1}$ performed slightly better than $\widehat{L}_L$, having a lower RMSE 21/35 times, though these differences tended to be small. Among all the composite estimators, the unweighted geometric index $\widehat{G}_{L,c}$, which combined Laspeyres sub-indexes, had the lowest RMSE. It was generally lower than the composite Fisher, exceeding it in only three instances.

Table 2 displays some pure and composite index estimates, along with their RMSE's and PMSE's. Among all sample based indexes, the pure unweighted geometric index $\widehat{G}_c$, though mildly biased low for the target $\widehat{F}$, consistently exhibited the lowest RMSE, usually at least 33% lower than the composite Fisher $\widehat{F}_L$. No other index was as consistently close to the target population index. The composite Fisher $\widehat{F}_L$ was mildly biased high for the target F. For the composite geometric mean indexes, the expenditure weights used at the item stratum level of aggregation had a more pronounced effect on the aggregate index values than did the upper level aggregation weights. This is most clearly seen in the differences between $\widehat{G}_{G2,1}$, which is biased up, and $\widehat{G}_{G1,2}$, which is biased down. Overall, the method of lower level aggregation dominated the effect of the method of higher level aggregation. In Table 2, we see only small differences, for example, between the values of $\widehat{G}_{L,1}$ and $\widehat{G}_{L,2}$; by contrast, the differences between $\widehat{L}_{G1}$ and $\widehat{L}_{G2}$ appear quite sharp.

For each of the 50 runs in each experiment, we counted the times certain indexes fell above or below some others. The pure Laspeyres estimate $\widehat{L}$ exceeded the pure Fisher estimate about half the time (24.7/50, on average). Similarly, the composite Laspeyres $\widehat{L}_L$ exceeded the population Fisher index an average of 25.9 out of 50 times. In sharp contrast, $\widehat{L}_L$ lay above the composite Fisher $\widehat{F}_L$ roughly 80% of the time (40.0/50), while the hybrid index $\widehat{G}_{L,1}$ exceeded $\widehat{F}_L$ only about one third of the time (16.7/50). The difference $\widehat{L}_L - \widehat{F}_L$ was generally small—resulting in a low PMSE for $\widehat{L}_L$—but consistently positive. (In the next section, we will present a theoretical explanation for this phenomenon.) In general, we found that the distance to the composite Fisher bore little resemblance to the RMSE. For example, the PMSE for $\widehat{G}_{L,c}$ was not particularly low. And though the RMSE's for $\widehat{G}_{L,1}$ and $\widehat{L}_L$ tended to be close, the PMSE for $\widehat{G}_{L,1}$ was consistently and sharply lower. These findings call into question the use of distance to $\widehat{F}_L$ as an appropriate basis for judging substitution behavior, as has been done in several recent papers on price index estimation.

Table 2. Comparison of Average Laspeyresville Pure and Composite Indexes

	$\widehat{F}$	$\widehat{L}$	$\widehat{G}_c$	$\widehat{F}_L$	$\widehat{L}_L$	$\widehat{G}_{G2,1}$	$\widehat{G}_{G1,2}$	$\widehat{G}_{L,1}$	$\widehat{G}_{L,2}$	$\widehat{L}_{G1}$	$\widehat{L}_{G2}$
I	1.468	1.467	1.452	1.467	1.482	1.926	1.136	1.464	1.471	1.146	1.963
$I-F$	0.012	0.011	-0.004	0.011	0.026	0.470	-0.320	0.008	0.015	-0.310	0.507
$RMSE$	0.181	0.176	0.108	0.176	0.183	0.566	0.355	0.178	0.177	0.346	0.612
$PMSE$	0	0.031	0.149	0	0.027	0.499	0.343	0.013	0.013	0.333	0.547

4. Order Relations of Index Estimators

In this section, we establish certain order relations between some index estimators. Following Dalén (1992), we write the Taylor series expansions[4] of the index formulas around the point $\underline{r}=\mathbf{1}$ in terms of the standardized moments of a set of price ratios $\underline{r}=\{r_j\}_{j=1}^N=\left\{\widehat{Y}_g\right\}_{g=1}^G$ (or, depending on the level of aggregation considered, $r_j=p_{j2}/p_{j1}$). Let the weights w_{jt} represent expenditure shares for time period $t\in\{1,2\}$, and define the moments

$$\mu_t=\sum_{j=1}^N w_{jt}r_j;\qquad \sigma_t^2=\sum_{j=1}^N w_{jt}\left(r_j-\mu_t\right)^2;\qquad \text{and}\qquad \gamma_t=\sum_{j=1}^N w_{jt}\left(r_j-\mu_t\right)^3.$$

First consider the indexes based on a single set of fixed normalized weights $\underline{w}_1=\{w_{j1}\}_{j=1}^N$. The arithmetic mean index is $A_{\underline{w}_1}=\mu_1$, and, to the third order,

$$G_{\underline{w}_1}=\prod_{j=1}^N r_j^{w_{j1}}\approx\mu_1-\sigma_1^2\left(1-\frac{\mu_1}{2}\right)+\frac{\gamma_1}{3},$$

while

$$F_{\underline{w}_1}=\left\{\frac{\sum_{j=1}^N w_{j1}r_j}{\sum_{j=1}^N w_{j1}/r_j}\right\}^{1/2}\approx\mu_1-\sigma_1^2\left(1-\frac{\mu_1}{2}\right)+\frac{\gamma_1}{2}.$$

Clearly, $F_{\underline{w}_1}-G_{\underline{w}_1}\approx\gamma_1/6$ is generally a small positive difference when the price ratios r_j are right skewed (e.g., log-normally distributed). Thus, in practice, we should expect index estimates based on $F_{\underline{w}_1}$ and $G_{\underline{w}_1}$ to closely approximate one another. By contrast,

$$A_{\underline{w}_1}-F_{\underline{w}_1}\approx\sigma_1^2\left(1-\frac{\mu_1}{2}\right)-\frac{\gamma_1}{2}.$$

Differences in estimates based on $A_{\underline{w}_1}$ and $F_{\underline{w}_1}$ may therefore be substantial and should increase with σ_1^2. Now consider the corresponding indexes based on two sets of weights $\underline{w}_1=\{w_{j1}\}_{j=1}^N$ and $\underline{w}_2=\{w_{j2}\}_{j=1}^N$. To the second order, we have

$$F_{\underline{w}_1,\underline{w}_2}=\left\{\frac{\sum_{j=1}^N w_{j1}r_j}{\sum_{j=1}^N w_{j2}/r_j.}\right\}^{1/2}\approx\frac{1}{2}\left(\mu_1+\mu_2\right)-\frac{1}{2}\sigma_2^2-\frac{1}{8}\left(\mu_2-\mu_1\right)^2.$$

Similarly,

$$T_{\underline{w}_1,\underline{w}_2}=\prod_{j=1}^N r_j^{(w_{j1}+w_{j2})/2}\approx\frac{1}{2}\left(\mu_1+\mu_2\right)-\frac{1}{4}\left(\sigma_1^2+\sigma_2^2\right)-\frac{1}{8}\left(\mu_2-\mu_1\right)^2.$$

[4] The Constant Elasticity of Substitution (CES) index (see Shapiro and Wilcox 1997) may be expanded and analyzed in similar fashion.

Thus, to the second order, $F_{\underline{w}_1,\underline{w}_2} - T_{\underline{w}_1,\underline{w}_2} \approx (\sigma_1^2 - \sigma_2^2)/4$, while

$$F_{\underline{w}_1,\underline{w}_2} - G_{\underline{w}_1} \approx \frac{1}{2}(\mu_2 - \mu_1) + \frac{1}{2}\left(\sigma_1^2 - \sigma_2^2\right) - \frac{1}{8}(\mu_2 - \mu_1)^2,$$

and

$$T_{\underline{w}_1,\underline{w}_2} - G_{\underline{w}_1} \approx \frac{1}{2}(\mu_2 - \mu_1) + \frac{1}{4}\left(\sigma_1^2 - \sigma_2^2\right) - \frac{1}{8}(\mu_2 - \mu_1)^2.$$

These differences may be either positive or negative, depending on the weights and price ratios. The first term, $(\mu_2 - \mu_1)/2$, may be expected to dominate the differences. A positive value for this term indicates that consumers spent more in the second period on items whose prices had risen between the two periods, i.e., they did not adjust the quantities they purchased enough to entirely compensate for changes in relative prices. The second term is indeterminate in sign, while the third is clearly negative but of lower order. Examining the difference between $F_{\underline{w}_1,\underline{w}_2}$ and the arithmetic mean index, we have, to the second order,

$$A_{\underline{w}_1} - F_{\underline{w}_1,\underline{w}_2} \approx \frac{1}{2}\sigma_2^2 - \frac{1}{2}(\mu_2 - \mu_1) + \frac{1}{8}(\mu_2 - \mu_1)^2.$$

Since the first term $\sigma_2^2/2$ is positive and will likely dominate, we should expect $A_{\underline{w}_1}$ to exceed $F_{\underline{w}_1,\underline{w}_2}$; moreover, the two indexes will diverge as the variability in the r_j increases.

Table 3 gives index estimates, computed from CPI data, using both the index formulae and the second-order Taylor series approximations (indicated by a superscript 0). In this case, the "price relatives" r_j are actually sub-indexes, and the weights are estimated expenditure shares for the corresponding item categories. Clearly the second-order approximations work well for these estimators of year-to-year change. The final column of the table shows the estimated values of $(\mu_2 - \mu_1)/2$, which, for these data, is the dominant term in the differences $F_{\underline{w}_1,\underline{w}_2} - G_{\underline{w}_1}$ and $T_{\underline{w}_1,\underline{w}_2} - G_{\underline{w}_1}$. The positive values of this term—for all but one year—are consistent with the superlative index estimates exceeding the geometric mean estimates. The terms $(\sigma_1^2 - \sigma_2^2)/2$ and $-(\mu_2 - \mu_1)^2/8$ contribute little to the total differences.

Table 3. Index Estimates and their Taylor Series Approximations

Year	$\widehat{A}_{\underline{w}_1}$	$\widehat{G}_{\underline{w}_1}$	$\widehat{G}^0_{\underline{w}_1}$	$\widehat{T}_{\underline{w}_1,\underline{w}_2}$	$\widehat{T}^0_{\underline{w}_1,\underline{w}_2}$	$\widehat{F}_{\underline{w}_1,\underline{w}_2}$	$\widehat{F}^0_{\underline{w}_1,\underline{w}_2}$	$(\widetilde{\mu}_2 - \widetilde{\mu}_1)/2$
87-88	1.03975	1.03806	1.03803	1.03838	1.03833	1.03839	1.03835	0.00027331
88-89	1.04555	1.04394	1.04385	1.04451	1.04441	1.04447	1.04438	0.00059465
89-90	1.05118	1.04942	1.04920	1.04962	1.04940	1.04967	1.04941	0.00019134
90-91	1.03919	1.03759	1.03747	1.03786	1.03774	1.03783	1.03773	0.00029062
91-92	1.02880	1.02726	1.02722	1.02744	1.02739	1.02743	1.02740	0.00017333
92-93	1.02764	1.02624	1.02615	1.02674	1.02665	1.02677	1.02667	0.00047575
93-94	1.02591	1.02461	1.02456	1.02486	1.02481	1.02483	1.02481	0.00024865
94-95	1.02737	1.02600	1.02596	1.02619	1.02615	1.02617	1.02614	0.00019976
95-96	1.02809	1.02685	1.02683	1.02665	1.02664	1.02656	1.02659	-.00014877
96-97	1.01976	1.01848	1.01850	1.01873	1.01876	1.01867	1.01878	0.00023259

5. Relations of Index Estimators to Population Indexes

Having described the relations of several forms of index estimator *to each other,* we now turn to the question of the relation of the estimators to the population indexes they target.[5] In the U.S. CPI, the *sub-indexes* targeted are now predominantly in the form of a geometric mean. Theory and experience show the sample geometric mean biased up for the population geometric mean. For suppose we aim at a population unweighted geometric mean $G_g = \prod_{j=1}^{N_g} r_{gj}^{1/N_g}$ and, by the usual additivity properties of estimators, $\log(G_{gs}) = \sum_s \pi_j^{-1} \log r_{gj}$ is unbiased for $\log(G_g)$ (where π_j^{-1} are the sample weights); then, exponentiating, we see that $\mathrm{E}(G_{gs}) > G_g$. Such biases decrease with increasing sample size, but the bias of the aggregated index will also depend on the method of aggregation. In recent years there has been considerable interest in employing alternate modes of aggregation, in the hope of achieving an index that is closer to a cost of living index. Since the Fisher and Törnqvist closely approximate each other, both as estimators (see above) and as population indexes (Diewert 1987), it is enough to study one of these. We focus primarily on the Törnqvist, but we first mention one result of interest regarding the Fisher.

Result 1 *Fisher Estimator vis-a-vis Fisher.* Let

$$\widehat{F}_{\widehat{L}} = \left\{ \sum_g \tilde{s}_{g1} \widehat{L}_g \left(\sum_g \tilde{s}_{g2} \widehat{L}_g^{-1} \right)^{-1} \right\}^{1/2},$$

with $\tilde{s}_{gt}$ a consistent estimator of

$$s_{gt} = \frac{\sum\limits_{i \in S_g} q_{git} p_{lht}}{\sum\limits_g \sum\limits_{i \in S_g} q_{git} p_{lht}}$$

and $\widehat{L}_g$ a consistent estimator (based on prices survey) of L_g. Then

$$\begin{aligned} \widehat{F}_{\widehat{L}} \to F^+ = & \left\{ \sum_g s_{g1} L_g \left(\sum_g s_{g2} L_g^{-1} \right)^{-1} \right\}^{1/2} \\ \neq & \left\{ \sum_g s_{g1} L_g \left(\sum_g s_{g2} P_g^{-1} \right)^{-1} \right\}^{1/2} = F. \end{aligned}$$

Under typical economic conditions $L_g > P_g$, and F^+ will exceed F. ■

We turn to the Törnqvist, that is, to a geometric mean having the average of first and second period group expenditure shares for weights. More generally, we consider:

Result 2 *Weighted geometric mean, weights assumed known.* Let

$$T_I = \prod_g (I_g)^{w_g} \qquad \text{and} \qquad \widehat{T}_{\widehat{I}} = \prod_g \left(\widehat{I}_g \right)^{w_g}.$$

Consider the factors. Suppose

$$E\left(\widehat{I}_g\right) = I_g (1 + b_g), \quad var\left(\widehat{I}_g\right) = I_g^2 (1 + b_g)^2 v_g.$$

[5] Related work has been carried out by Greenlees (1998).

Thus $1 + b_g$ is the multiplicative bias, and v_g is the squared coefficient of variation, of $\widehat{I}_g$. Then $E\left(\widehat{I}_g^{w_g}\right) \approx I_g^{w_g} H\left(w_g, b_g, v_g\right)$, where

$$H\left(w_g, b_g, v_g\right) = 1 + w_g A_g + w_g^2 B_g$$

where $A_g = \left\{b_g - \frac{1}{2}\left[b_g^2 + v_g\right]\right\}$ and $B_g = \frac{1}{2}\left\{b_g + v_g\left[1 - \left(b_g - \frac{b_g^2}{2}\right)\right]\right\}$. Then $\widehat{I}_g^{w_g}$ is biased low for $I_g^{w_g}$ if and only if

$$w_g < \frac{\frac{1}{2}\left[b_g^2 + v_g\right] - b_g}{\frac{1}{2}\left\{b_g^2 + v_g\left[1 - \left(b_g - \frac{1}{2}b_g^2\right)\right]\right\}} \approx 1 - \frac{2b_g}{b_g^2 + v_g}. \quad \blacksquare$$

For $b_g = 0$, the bias of $\widehat{I}_g^{w_g}$ is negative (this generalizes a result of Greenlees (1998)). Thus the positive bias we noted above for Laspeyres sub-indexes may be advantageous if we use a Törnqvist aggregator. Of course, the above result does not *guarantee* that the upper- and lower-level aggregation biases will offset each other, since the relative sizes of the expenditure weights to the biases and coefficients of variation must be assayed, and this is a non-trivial prospect. Some bias can be added as a result of sampling error in constructing the weights, but the following result suggests it will not as a rule be significant.

Result 3 *Weighted geometric mean, weights estimated unbiasedly.* Let

$$T_I = \prod_g \left(I_g\right)^{w_g} \qquad \text{and} \qquad \widehat{T}_{\widehat{I}} = \prod_g \left(\widehat{I}_g\right)^{\widehat{w}_g}.$$

Assume $\widehat{w}_g = w_g + \epsilon_g + \overline{\epsilon}$, where $E\left(\epsilon_g\right) = 0$, $var\left(\epsilon_g\right) = \phi_g$, and $\overline{\epsilon} = \frac{1}{G}\sum_g \epsilon_g$. Also, assume

$$E\left(\widehat{I}_g\right) = I_g\left(1 + b_g\right), \quad var\left(\widehat{I}_g\right) = I_g^2\left(1 + b_g\right)^2 v_g,$$

as in Result 2. Then

$$E\left(\widehat{T}_{\widehat{I}}\right) \approx \prod_g I_g^{w_g} \prod_g H\left(w_g, b_g, v_g\right)\left(1 + \frac{1}{2}\sum_g \phi_g J_g\right) \approx T_I\left(1 + \sum_g w_g H_g^* + \frac{1}{2}\sum_g \phi_g J_g\right),$$

where $H_g^* = A_g + w_g B_g$, and

$$J_g = 2B_g - A_g^2 + \lambda_g^2 - 2\lambda_g\overline{\lambda} + \overline{\lambda}^2,$$

with $\lambda_g = \log\left(I_g\right) + A_g + 2w_g B_g$ and $\overline{\lambda} = \frac{1}{G}\sum_g \lambda_g$. $\blacksquare$

If the w_g are estimated reasonably well, say with coefficient of variation less than 0.2, then the squared coefficient of variation ϕ_g / w_g^2 will be quite small. Then a comparison of components of $\frac{1}{2}\sum_g \phi_g J_g$ and $\sum_g w_g H_g^*$ suggests that the former will be relatively small, to the point of being negligible. In other words, estimating the weights, if done reasonably precisely and without bias, should have relatively little impact on the bias of the estimator. No work has been done on determining the variance of $\widehat{T}_{\widehat{I}}$ in the situation corresponding to the above two results. Intuitively, the less noise there is in estimating I_g and $\widehat{w}_g$, the smaller the variance is likely to be, but the relative impact of the various components awaits further investigation.

6. Discussion

In the Introduction we noted four questions, and this paper has indicated some answers, which we here summarize: (1) Economic theory suggests that a *COLI* is approximated by indexes calculated by the Fisher and the Törnqvist formulas. (2) The size and direction of the gap between these superlative indexes and the Laspeyres, Paasche, and geometric mean indexes reflect the substitution behavior of consumers, that is, how much they switch from one set of goods to another merely because of changes in relative prices. These population indexes are, however, never available in practice. Instead we can at best compare index *estimates,* which are almost invariably based on data arising from at least two distinct surveys, one yielding estimates of sub-indexes, the other yielding weights by which to aggregate the sub-indexes. (3) The relations between the index estimators are very largely influenced by formal relations between the estimating formulas—characterized by certain low order Taylor series expansions—and only indirectly by the substitution behavior of consumers.

Thus the inference of substitution behavior from the relations among index estimates, shown in various studies of data from the U.S. CPI,[6] is called into question. That the geometric mean usually falls below the superlative indexes *can* be taken as a sign that consumers do not substitute freely across major categories of goods. On the other hand, as indicated by the "Laspeyresville" study of Section 3 and implied by the Taylor expansions of Section 4, the fact that the Laspeyres estimator is typically well above the superlative estimator *cannot* in itself be taken as evidence that consumers are substituting goods in response to price change. (4) The relation between a sample weighted aggregated geometric index estimator and a target index of the same form depends on the bias and variance of each sub-index and their relation to the aggregating weights. Results 2 and 3 of Section 5 may offer some guidance for experimentation with various sub-indexes and weights. If nothing else, they serve as reminders of the distinction between an index and an index estimator.

References

Aizcorbe, A. and Jackman, P. 1993. "The Commodity Substitution Effect in CPI Data, 1982-1991: Anatomy of Price Change", *Monthly Labor Review*, December 1993, pp. 25-33. U.S. Government Printing Office.

Carli, G. 1764. "Del valore e della proporzione de 'metalli monetati," reprinted in pp. 297-336 in *Scrittori classici italiani di economia politica,* Volume 13, Milano: G.G. Destefanis, 1804. (Reference obtained from Diewert 1995.)

Dalén, J. 1992. "Computing Elementary Aggregates in the Swedish Consumer Price Index." *Journal of Official Statistics,* Vol.8. No. 2, 1992, pp. 129-147. Statistics Sweden.

Diewert, W. E. 1987. "Index numbers." In *The New Palgrave: A Dictionary of Economics,* (edited by J. Eatwell, M. Milate, and P. Newman), MacMillan, London.

Diewert, W. E. 1995. On the Stochastic Approach to Index Numbers. Discussion paper no. DP 95-31, Sep. 1995, Department of Economics, University of British Columbia, Vancouver, Canada V6T 1Z1.

Fisher, I. 1922. *The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability.* Sentry Press, New York, NY 10019.

Greenlees, J. 1998. "Random Errors and Superlative Indexes." August 9, 1998. Unpublished BLS manuscript.

Shapiro, M. and Wilcox, D. 1997 "Alternative Strategies for Aggregating Prices in the CPI," *Federal Reserve Bank of St. Louis Review*, May/Jun 1997, pp. 113- 125

Turvey, R. et al. 1989. *Consumer Price Indexes: An ILO manual.* International Labour Office, Geneva.

[6] For example, Aizcorbe and Jackman (1993), Shapiro and Wilcox (1997).