BLS WORKING PAPERS

U.S. DEPARTMENT OF LABOR
Bureau of Labor Statistics

OFFICE OF PRICES AND LIVING
CONDITIONS

Random Errors and Superlative Indexes

John S. Greenlees, U.S. Bureau of Labor Statistics

Working Paper 343
March 2001

I would like to thank Bert Balk, Ralph Bradley, Rob Cage, Jorgen Dalen, Alan Dorfman, Tim Erickson, Janice Lent, Rob McClelland, Travis Nesmith, March Reinsdorf, and David Swanson for helpful discussions and comments on earlier drafts. All responsibility for errors remains with me. The views expressed in this paper are solely those of the author and do not necessarily reflect the policies of the U.S. Bureau of Labor Statistics.

Abstract

This paper demonstrates that random sampling error in CPI price and expenditure data at the item-area level can have a distorting effect on empirical cost-of-living indexes computed using those data. In particular, the expected values of the Fisher Ideal and Tornqvist indexes can be distorted downward by random error in basic index relatives. This, in turn, can cause the estimated Laspeyres substitution bias in the CPI to be overestimated. The issue is illustrated empirically using CPI data for the period 1987 through 1995. Estimated substitution bias is sharply higher in each year when smaller CPI "replicate" samples are used to compute indexes than when the full samples—which are subject to less sampling error—are employed. To address this problem, the paper derives and applies a composite-estimation approach, in which CPI item-area indexes are replaced by a weighted average of those indexes and the U.S.-level item indexes. This approach causes the estimated superlative index values to be higher, and the estimated substitution bias consequently lower. For example, in a comparison of annual chain Laspeyres indexes to chain Tornqvist indexes the adjusted estimate of substitution bias is about 0.08 percentage point rather than the roughly 0.12 percentage point previously estimated.

I. Introduction

In recent years, considerable attention has focused on the question of whether, or to what extent, the U.S. Consumer Price Index (CPI) may overstate movements in the cost of living. Researchers have put forward several reasons for such potential overstatement, including inadequate recognition of quality change, failure to incorporate the gains to consumers from new products or merchandising techniques, and possible flaws in the formulas used to combine individual prices in producing the CPI.[1] Unfortunately, most of these problems are difficult to measure, and the size, direction, and even existence of the associated "biases" are a matter of some dispute.

There is more widespread agreement, however, on one CPI measurement issue, that of "substitution bias." This is the bias that arises because of the fixed-weight nature of the index. To construct the CPI, the Bureau of Labor Statistics (BLS) first computes indexes for approximately 8,000 "strata," defined by the combination of about 200 categories of items and about 40 geographic areas. These basic indexes are then aggregated using the fixed-weight Laspeyres formula and expenditure weights drawn from the Consumer Expenditure (CEX) Survey, conducted for the BLS by the Bureau of the Census.[2] Well-known economic theorems demonstrate that the use of fixed base-period weights should cause an overstatement of cost-of-living changes, because the ability of consumers to substitute in response to relative price changes is ignored.[3]

Substitution bias should exist also at the "lower level," if the basic indexes themselves are constructed as fixed-weight arithmetic averages.[4] At the "upper level" stage of combining the indexes, however, the availability of annual CEX data makes it possible to measure substitution bias much more rigorously. For this purpose, the "superlative index" results derived by Diewert (1976, 1992) have been crucial.

Aizcorbe and Jackman (1993), Aizcorbe, Cage, and Jackman (1996), and Shapiro and Wilcox (1997) have used CPI elementary indexes to construct Laspeyres indexes with varying base periods, as well as the superlative Fisher Ideal and Tornqvist counterparts. Based on this research, a rough consensus has emerged that the upper level substitution bias in the CPI is on the order of 0.15 percentage point per year.[5]

The purpose of this paper is to reconsider the above consensus. Although the existence of upper-level substitution bias is not in serious dispute among economists, the size of this bias is an empirical question. A well-known fact, but one little-examined in this context, is that the CPI basic indexes and associated expenditure weights used in the

---

[1] For the most recent summaries of these issues, see Shapiro and Wilcox (1996), Moulton (1996), U.S. Senate (1996), Bureau of Labor Statistics (1997), Boskin et al. (1998), and Abraham et al. (1998).

[2] The Laspeyres formula is used in the CPI in modified form, since the expenditure weights are linked into the index with a lag. See BLS (1997).

[3] See, for example, Pollak (1989).

[4] In January 1999, the BLS will begin to employ geometric means in the construction of CPI basic indexes, as a means of dealing with consumer substitution.

[5] See U.S. Senate (1996).

above-referenced studies are subject to sampling error. It will be demonstrated below that this error can have a considerable impact on the expected values of the index formulas being compared. It turns out that previous estimates of upper-level substitution bias are themselves "biased" in an upward direction.

Section II below presents several well-known formulas for price indexes and demonstrates how the expected values of those indexes can be affected by random error in the prices and expenditure weights. In particular, the expected values of the Fisher Ideal and Tornqvist indexes can be distorted downward by random error in basic index relatives. This, in turn, can cause the estimated Laspeyres substitution bias to be overestimated. The next three sections examine this issue empirically using CPI data for the period 1987 through 1995. Section III demonstrates that estimated substitution bias is sharply higher in each year when smaller "replicate" samples are used to compute indexes than when the full samples—which are subject to less sampling error—are employed. Section IV briefly notes that the problem of random error in price relatives cannot be solved simply by using U.S.-level CPI item indexes, rather than the basic item-area indexes to compute the aggregate superlative indexes. In Section V, I suggest instead a composite-estimation approach, in which CPI item-area indexes are replaced by a weighted average of those indexes and the U.S.-level item indexes. This should mitigate the problem of random sampling error and reduce or eliminate the downward bias in the estimated superlative indexes. Section VI addresses the question of "small sample bias" in CPI item-area indexes and how that possible problem relates to the issue addressed in this paper. Section VII concludes with a brief discussion of the implications of this research.

II. Random-Error Bias

Five of the standard index formulas used in empirical studies of substitution bias are shown below, where the individual price relatives R_{ij} and expenditure shares w_{ijt} refer to a basic stratum in the CPI corresponding to item i and area j.[6]

$$I^L = \sum_i \sum_j w_{ij0} R_{ij} \quad \text{(Laspeyres)}$$

$$I^P = \left[\sum_i \sum_j w_{ij1} R_{ij}^{-1} \right]^{-1} \quad \text{(Paasche)}$$

$$I^F = (I^L)^{1/2} (I^P)^{1/2} = \left[\sum_i \sum_j w_{ij0} R_{ij} \right]^{1/2} \left[\sum_i \sum_j w_{ij1} R_{ij}^{-1} \right]^{-1/2} \quad \text{(Fisher Ideal)}$$

[6] Throughout this paper, aggregate index formulas will be written as functions of basic index relatives rather than reference- and comparison-period basic index levels. This is done for convenience and to avoid confusion of the issue of interest here with that of small-sample bias in basic index relatives, as discussed by McClelland and Reinsdorf (1997). All analyses in this paper proceed from the assumption that the basic CPI index relatives are unbiased sample estimates of population relatives.

$$I^G = \prod_i \prod_j R_{ij}^{\;w_{ij0}} \quad \text{(Geometric Mean)}$$

$$I^T = \prod_i \prod_j R_{ij}^{\;w_{ij.}} \quad \text{(Tornqvist), where}$$

$$w_{ij.} = \frac{w_{ij0} + w_{ij1}}{2}$$

The Laspeyres and geometric mean formulas I^L and I^G use only expenditure data from the reference period *0*. These indexes will equal the cost-of-living index under the contrasting Leontief and Cobb-Douglas models of consumer behavior, respectively. That is, the Laspeyres index will be "correct" under the limiting assumption that, holding utility constant, relative prices do not affect quantities purchased. The geometric mean index will be "correct" if consumers respond to compensated relative price changes by maintaining constant expenditure shares *w*, unchanged from period *0* to the comparison period *1*.

The superlative Fisher Ideal and Tornqvist formulas I^F and I^T employ expenditure data from both the reference and comparison periods. As a result, they have the potential to reflect the extent to which consumers actually respond to relative price change. Under the assumption of consumer optimizing behavior, Diewert has shown that the superlative indexes will yield results close to the true cost-of-living index.[7] For example, if quantities remain constant, as under the Leontief behavioral model, the Laspeyres index and the Paasche index I^P will both equal the superlative Fisher Ideal index. By contrast, if expenditure shares are constant, as in the Cobb-Douglas model, the geometric mean and Tornqvist indexes will be equal.

In practice, the two superlative indexes, the Fisher Ideal and Tornqvist, yield extremely similar results. The extent to which they are below the Laspeyres results is the standard measure of upper-level substitution bias. For example, the Aizcorbe-Cage-Jackman results presented in Moulton and Stewart (1998) show that the annual increase in both chain Fisher and chain Tornqvist indexes each averaged 0.12 percentage points less than the average increase in a chain Laspeyres index over the 1987-95 period.[8]

The CPI's elementary indexes are computed from samples of individual priced items designed to be representative of the goods and services sold within the corresponding strata. In some cases, such as apples or motor fuel, the strata are relatively homogeneous, and price variation within the strata arises primarily from differences among stores, locations, and times of collection (CPI price data are collected throughout most of each month). Other strata, such as prescription drugs or women's dresses, are more broadly defined, and variance in observed price movements can be due to market conditions for particular products.

---

[7] The most familiar form of the result requires the assumptions of homotheticity and of a representative consumer.

[8] The average bias was 0.18 percentage point when fixed-base versions of Laspeyres and Tornqvist indexes were compared, and 0.13 percentage point when a fixed-base Laspeyres index was compared to a chain Tornqvist.

In the same way, the expenditure weights used in the CPI are designed as estimates of the aggregate spending distribution of urban U.S. consumers. Because of the finite sample sizes in the CEX, the expenditure shares w_{ijt} are subject to some sampling variance as well.

Each year, the BLS publishes estimates of sampling error for the CPI, broken down by region and major group. These estimates employ replicate data from both the CEX and CPI pricing samples, and thus reflect the variance in both expenditures and prices. According to the most recent published estimate, the standard error of 12-month price change for the all-items CPI is roughly 0.12 percentage point.[9] At lower levels, where sample sizes are smaller, and especially for the more volatile index components, the standard error is higher. As an extreme example, for apparel in the Northeast region the standard error is approximately 2.33 percentage points. Standard errors are not published for individual item-area strata, but would be higher still.

The intuition underlying this paper is that such sampling variation could distort empirical estimates of upper-level substitution bias. The theory underlying superlative indexes treats the expenditure weights and prices used in the formulas as known values corresponding to a specified population or a "representative household," and assumes that the response of expenditures to prices provides information on consumer preferences. By contrast, the observed data for item-area CPI components are sample estimates of population means, drawn from separate surveys of expenditures and prices. As a consequence, some of the observed variation in relative price is random sampling variation, which will be uncorrelated with the measured expenditure shares. As noted above, however, invariance of expenditure shares to relative price change is consistent with Cobb-Douglas consumer behavior, indicated by the closeness of a superlative aggregate index to its geometric mean counterpart. It is reasonable to hypothesize, then, that sampling variance in the basic CPI price indexes will cause the observed superlative indexes to diverge from the Laspeyres in the direction of the geometric mean, even if consumer behavior follows the Leontief model.

Of course, the fact that the basic indexes are stochastic raises questions about the "representative household" model, and about whether the expected values of the basic indexes are necessarily the most appropriate values to use in constructing aggregate indexes. The major point of this paper, however, is merely that prior estimates of superlative indexes and of substitution bias are systematically different from those that would be obtained if larger samples of prices were collected.

How this effect could work in practice can be seen in the context of the Tornqvist index formula shown above. If we assume that the measured relatives R_{ij} equal the product of true relatives P_{ij} and independent random errors θ_{ij} with mean one, then the

[9] See Swanson (1998).

expectation of the measured index $\hat{I}^T$ will be lower than the true I^T, a result that follows from Jensen's inequality and the concavity of the Tornqvist formula in price relatives. [10]

Relatively straightforward algebra can be used to show the approximate size of this bias. First, apply a second-order Taylor series expansion of the Tornqvist formula around the mean θ_{ij} of one:

$$\hat{I}^T = \prod_i \prod_j R_{ij}^{\;w_{ij.}} = \prod_i \prod_j P_{ij}^{\;w_{ij.}} q_{ij}^{\;w_{ij.}}$$

$$= I^T \prod_i \prod_j q_{ij}^{\;w_{ij.}}$$

$$\cong I^T \prod_i \prod_j \left[1 + w_{ij}\left(q_{ij}^{\;w_{ij.}} - 1\right) + .5 w_{ij}\,(w_{ij} - 1)\left(q_{ij}^{\;w_{ij.}} - 1\right)^2\right]$$

and

$$(1)\;\; E\left(\hat{I}^T\right) \cong I^T \prod_i \prod_j \left[1 + .5 w_{ij}\,(w_{ij} - 1) d_{ij}^{\,2}\right]$$

where

$$d_{ij}^{\,2} = \mathrm{var}\left(q_{ij}\right)$$

We can rewrite the bracketed term in (1) by employing the approximation to the exponential function. We then have:

$$(2)\;\; E\left(\hat{I}^T\right) \cong I^T \prod_i \prod_j e^{.5 w_{ij}(w_{ij}-1) d_{ij}^2} \leq I^T$$

It is easy to see that the bias is potentially large. For example, consider the unrealistic but illustrative case in which all index strata have equal share weights *w* and equal variances δ^2. Then, if the number of strata is large, the above expectation is approximately equal to $I^T\exp(-.5\delta^2)$. Finally, to provide a crude but suggestive numerical result, assume that there are 8,000 item-area cells and that the standard error of annual-average price change is on the order of 0.1 percentage point.[11] This aggregate standard error would be consistent with a value of δ^2=.008 for an individual stratum index relative, since the Laspeyres is an arithmetic mean of the stratum relatives (i.e., $.008/8000=.001^2$). Under these assumptions, then, the downward random-error bias in an annual-average Tornqvist index change would be approximately -0.4 percentage point. This certainly would constitute a large potential distortion—in particular, it is much greater than previous estimates of annual substitution bias in the CPI. Moreover, following this simple series of calculations, the

[10] This is demonstrated by Erickson (1998) under the assumption that expenditure shares are not subject to error. Below I will argue that the sampling error in expenditure shares is in fact much less serious than that of the price relatives.

[11] This is an estimate derived from computations underlying the analysis reported in Leaver and Cage (1997).

same distortion would apply to the observed geometric mean index, which has the same form as the Tornqvist (and the same expenditure weights if these are all assumed equal in both periods).

As would be expected given the similarity of estimated Tornqvist and Fisher Ideal indexes, one can demonstrate that the Fisher Ideal is subject to an equivalent distortion. Consider the Paasche component of the Fisher Ideal, the second bracketed factor of I^F above. Again using a second-order Taylor-series expansion for the θ_{ij} and again assuming that all stratum relatives have equal variances, the expectation of $1/\theta_{ij}$ is approximately $\exp(\delta^2)$. Consequently, it can be shown that each term in the summation is biased upward by the ratio $\exp(\delta^2)$. With this factor raised to a power of -.5 in the calculation of I^F, one can see that the distortion in the Fisher Ideal index is comparable to that of the Tornqvist in the presence of random error in price measurement.

Mean-one multiplicative errors of this sort will *not* affect the expectation of a Laspeyres index like the CPI, because of its linear form. Under normal conditions, the Laspeyres should also be expected to lie above the geometric mean and superlative indexes. Therefore, one can see that the hypothesized random error in observed price relatives would move both the superlative and the geometric mean indexes downward, away from the Laspeyres index. This, in turn, would lead to an overestimate of upper-level substitution bias and, potentially, an inappropriately favorable view of the performance of the geometric mean formula relative to the Laspeyres.

Although the emphasis here is on sampling error arising from the well-recognized sampling variance of elementary CPI indexes, it should be noted that similar problems could arise from non-sampling error. Random, uncorrelated errors in quality adjustment procedures, for example, could also lead to downward bias in superlative index estimates.

Random error applies also to the share weights, of course, which are derived from finite CEX survey samples. The BLS employs a three-year average of CEX data to construct the CPI expenditure weights, along with a composite-estimation technique to reduce variance in local area weights. The expenditure shares used in most superlative index research, however, apply to single years and are not composite-estimated. It was noted above that the Tornqvist formula is concave in price relatives. The same formula is convex in expenditure shares, so random error in expenditure measurement should likely lead to a bias in the opposite direction from error in the price relatives. In the likely presence of uncorrelated errors in both shares and relatives, the net effect is indeterminate in sign.[12]

III. Analysis using Replicate Samples

We shall use the term "random-error bias" to designate the situation caused by random sampling error in price relatives or expenditure weights, in which the expected value of estimated aggregate indexes differs from the values that would be obtained using actual price and expenditure totals. In this section, we present simulations of annual indexes to demonstrate whether random-error bias is significant enough to have qualitative

[12] See Erickson (1998).

implications. For this purpose, the period of study will cover the years 1987 through 1995, as in the previous results reported in Moulton and Stewart (1998), Shapiro and Wilcox (1997), and Greenlees (1998), and the underlying CPI data are the same as used in those studies.[13] Both the Laspeyres and superlative series can also be constructed on either a fixed or chain basis. In the latter approach, which is used here, the estimates of price change between, for example, 1989 and 1990 employ expenditure data from those two years (1989 only in the case of the Laspeyres).[14] The stratum-level price data used are annual-average CPI price indexes.

Chain index series for this period are presented in Table 1. Over the entire period, the chain Laspeyres index overstates annual changes in the cost of living, as represented by the chain Tornqvist index, by an average of 0.118 percentage point. The geometric mean index averages 0.030 percentage point below the Tornqvist. (As noted above, the Fisher Ideal and Tornqvist indexes yield extremely similar results. For convenience of exposition we focus on the Laspeyres, the Tornqvist, and the geometric mean in the remaining analysis.) Thus, the average geometric mean substitution bias is in the opposite direction from, and less than one-quarter as large as, the estimated Laspeyres substitution bias.

The question at issue is whether random error in the annual-average prices and annual expenditure weights creates a serious distortion in the construction of these superlative indexes. Examination of this question is made possible by the existence of replicate values of both expenditures and prices at the CPI stratum level. Both the CEX and the CPI pricing samples are divided into subsamples, called replicates, for purposes of variance estimation. For the purposes of this paper, the replicate identifiers can be used to define data sets for index estimation that are subject to greater sampling error than the full sample, and thus also subject to greater "random error bias." If the superlative indexes constructed at the replicate level are markedly lower, and the divergence between estimated Laspeyres and superlative indexes is markedly wider, than at the full-sample level, this would indicate that sample-size limitations could be distorting the superlative indexes.

The results of such an analysis are displayed in Figures 1 and 2. Figure 1 displays the substitution bias in chain Laspeyres indexes, computed using the full sample and using replicate samples 1 and 2.[15] The results are strongly consistent with the above

---

[13] This period is chosen for reasons of item category consistency. Prior to 1987, the CPI employed a different item structure. Updates in the item structure take place approximately every decade; in January 1998, CPI strata were again recategorized.

[14] Aizcorbe, Cage, and Jackman present both fixed and chained estimates, whereas Shapiro and Wilcox compare fixed (1986) base Laspeyres indexes to chained superlative indexes. The advantages and disadvantages of chained indexes, and the specific question of whether those indexes are subject to upward "chain drift" are beyond the scope of this paper.

[15] The CEX has two expenditure replicates for each CPI item-area stratum. Pricing samples, however, are divided into four or even six replicates in a few CPI areas. The replicate estimates in this paper use only pricing replicates 1 and 2, and so effectively lower the pricing sample size by more than 50 percent on average.

speculation. The average substitution bias, which was approximately 0.12 percentage point for the full-sample estimates, is 0.23 percentage point using Replicate 1 and 0.26 using Replicate 2. (It should be noted that the latter average is strongly affected by the 0.73 percentage point bias in 1990, which is too large to be shown in the figure.) Surprisingly, this occurs as much because the Laspeyres indexes are higher in the replicates as because the Tornqvist indexes are lower. On average, across the eight years and the two replicates, the Laspeyres estimates are approximately 0.076 percentage point higher than obtained using the full sample, and the Tornqvist indexes 0.056 percentage point lower.[16]

Figure 2 shows the corresponding substitution bias estimates for the geometric mean index relative to the Tornqvist. In contrast to the results in Figure 1, the full-sample bias estimates tend to be intermediate between the replicate estimates. On average, the use of smaller sample sizes appears to reduce the geometric mean index levels by approximately the same amount as it reduced the Tornqvist indexes. The estimated downward substitution bias in the geometric mean index (i.e., bias relative to the Tornqvist) increases only slightly in the replicate results, by approximately 0.003 percentage point on average. This contrasts with the approximate 0.13 percentage point increase in the estimated upward Laspeyres bias in the replicates as compared to the full sample. In several cases, the replicates yield geometric mean indexes higher than the corresponding Tornqvist indexes.

These results are generally consistent with the predicted effect of random variation in the basic CPI price indexes: using the smaller replicate samples, both the Tornqvist and geometric indexes are lower, and the estimate of Laspeyres substitution bias is greater. Unfortunately, there is no apparent way to use the replicate index results to determine the remaining random-error bias in the full-sample indexes. Increasing the sample size by a factor of two (from the replicates to the full sample) appears to reduce the estimated Laspeyres bias by about half. The biases involved are highly nonlinear, however, and the extent to which further sample size increases would change the bias estimate cannot be determined using the replicate results.

IV. Indexes Constructed Using U.S.-Level CPI Series

One natural way to approach the problem of random-error bias is by computing the superlative indexes not from the item-area indexes, but from the U.S.-level CPI item indexes, which are constructed as Laspeyres aggregates. Averaging the 44 area indexes for each item (weighted by expenditures) should drastically reduce the variance. Similarly, the U.S.-level expenditure shares are based on much larger CEX samples than the shares for individual areas. Comparing Laspeyres and superlative indexes constructed from U.S.-level subindexes should, therefore, provide some information about the extent of the random-error problem. It readily can be demonstrated, unfortunately, that this approach is

[16] It is unclear why the Laspeyres estimates should be so affected, although small-sample bias (see fn. 6 and Section VI) may be accentuated in the replicate indexes. Note also that Replicate 1 in 1990 is an outlier result in which the Laspeyres index is much higher than in either Replicate 2 or the full sample.

not a solution to that problem; superlative indexes constructed from Laspeyres U.S.-level subindexes likely lead to an understatement of the amount of substitution bias.

First, define the U.S.-level Laspeyres subindex for each item, constructed as a period-0 weighted average of the item-area index values:

$$\overline{R_i} = \sum_j \frac{w_{ij0}}{w_{i0}} R_{ij}$$
$$w_{it} = \sum_j w_{ijt}$$

The overall Laspeyres index constructed directly from item-area indexes will equal the Laspeyres index constructed from the Laspeyres U.S.-level item subindexes $\overline{R_i}$. Use of those U.S.-level indexes does, however, affect the expectation of the other index formulas. Note that the geometric mean index using Laspeyres U.S.-level indexes can be decomposed into two factors:

$$I^G = \left(\prod_i \overline{R_i}^{\,w_{i0}} \right) \left(\prod_i \left[\frac{\prod_j \left(R_{ij} \right)^{w_{ij0}/w_{i0}}}{\overline{R_i}} \right]^{w_{i0}} \right)$$

A similar decomposition can be performed for the Tornqvist index:

$$I^T = \left(\prod_i \overline{R_i}^{\,w_{i.}} \right) \left(\prod_i \left[\frac{\prod_j \left(R_{ij} \right)^{w_{ij.}/w_{i.}}}{\overline{R_i}} \right]^{w_{i.}} \right)$$

In each of these two equations, the first parenthesized factor is the aggregate index, constructed from the U.S.-level Laspeyres item subindexes. The second factor is a geometric average across items of a ratio of two item subindexes: the geometric-mean or Tornqvist item subindex, respectively, in the numerator and the Laspeyres subindex in the denominator. Since the arithmetic means must be higher than their geometric mean counterparts, the second factor in each case must be less than one. Thus, the geometric mean (Tornqvist) index computed from U.S.-level subindexes is larger than the index I^G (I^T).

Computed using the U.S.-level data, both the Tornqvist and geometric mean indexes average about a tenth of a percentage point higher than when basic item-area indexes were used. Because the Laspeyres index is unaffected, the Laspeyres substitution bias computed using U.S.-level data averages only about 0.03 percentage point over eight years and the downward bias in the geometric mean index relative to the Tornqvist is about 0.02 percentage point. It is seen from the above decompositions, however, that these estimates understate the total upper-level substitution bias, by ignoring variations in relative prices across areas. It is reasonable to consider them a lower bound on the size of that substitution bias.

V. A Correction Using Composite Estimation

The previous section argued that the use of U.S.-level indexes mitigated the problem of "random-error bias," but at a cost: namely, the loss of true (as opposed to sampled) inter-area relative price variation. Thus, although sampling error in the detailed item-area price indexes makes the Tornqvist indexes "too low," those indexes are "too high" when computed with U.S.-level indexes. In this section I propose a composite-estimation approach designed to mitigate or eliminate random-error bias, by combining local price *variation* information with the national-level information on mean prices. Specifically, the local index values are replaced by values whose variance around the random-error-purged national mean reflects "true" geographic price variation.

This approach is somewhat similar to the composite estimation of local area expenditure weights in the published CPI.[17] The focus here, however, is on composite estimation of local price indexes. The replicate-level results in Section III above were more consistent with distortion resulting from sampling error in elementary price indexes (which should bias the superlative indexes down) than from sampling error in expenditure shares (which should bias the superlative indexes up). Moreover, examination of the Fisher Ideal formula in Section II above also suggests that the impact of sampling error in expenditure shares is likely to be less severe than error in price relatives. The expenditure shares enter linearly in the summations of both the Laspeyres and Paasche components of the Fisher. Those components are combined in a nonlinear formula, but random share variation, *unlike random variation in price indexes*, does not affect the expectation of either component. Thus, the random variation in observed expenditure shares affects the expectation of the Fisher formula only to the extent of the sampling error in the component summations—that is, in the estimates of mean price change across all strata. The impact on the expectation of the Tornqvist index must be similar. The analysis of this section, therefore, is confined to the distribution of individual index relatives, not expenditure shares.

I begin by assuming, as in the heuristic example in Section II, that the index *relatives* R_{ij} are the product of "true" population relatives P_{ij} and random sampling errors θ_{ij}. Under the further assumption that the θ_{ij} have mean one and variance d_i^2, we found that[18]

$$(3)\quad E\left(\hat{I}^T\right)=\left(\prod_i\prod_j P_{ij}^{\,w_{ij.}}\right)\left(\prod_i\prod_j e^{.5w_{ij}(w_{ij}-1)d_i^2}\right)$$

---

[17] It should be noted that the composite estimation of expenditure weights is applied to a Laspeyres index and does not affect the aggregate CPI's expected value.

[18] It would be more general to subscript δ by both i and j, that is, allowing for the variance to differ by area as well as item. The constraint of a single δ for each item stratum was made for convenience of empirical analysis, given the need to estimate these variances.

The goal is to identify a formula for the Tornqvist, based on observed data rather than the unobserved P_{ij}, that has an expectation equal to that of the first parenthesized term of the above expression. To accomplish this, assume that the true area relatives P_{ij} are distributed around a mean item relative $\overline{P}_i$ with variance s_i^2. That is, we assume

$$P_{ij} = \overline{P}_i h_{ij}$$

where h_{ij} has mean 1 and variance s_i^2. Then, following the same algebra used previously, one obtains

$$E\left(\prod_i \prod_j P_{ij}^{\,w_{ij.}}\right) = \left(\prod_i \prod_j \overline{P}_i^{\,w_{ij.}}\right)\left(\prod_i \prod_j e^{.5 w_{ij}(w_{ij}-1)s_i^2}\right) = \left(\prod_i \overline{P}_i^{\,w_{i.}}\right)\left(\prod_i \prod_j e^{.5 w_{ij}(w_{ij}-1)s_i^2}\right)$$

where, as earlier,

$$w_{it} = \sum_j w_{ijt}$$

Under the assumptions above, and combining this result with equation (3) above we have

$$E\left(\hat{I}^T\right) = \left(\prod_i \overline{P}_i^{\,w_{i.}}\right)\left(\prod_i \prod_j e^{.5 w_{ij}(w_{ij}-1)(s_i^2 + d_i^2)}\right)$$

Next, we form a replacement for R_{ij} such that the indexes of interest have the same expectations as they would have if we were able to employ the unobservable P_{ij}'s. For this purpose, define the relative Q_{ij} as a geometric average of the (unobserved) U.S. mean $\overline{P}_i$ and the local area index R_{ij}, with weights depending on the variances of the h_{ij} and θ_{ij}:

$$Q_{ij} = b_i \overline{P}_i^{\,1-a_i} R_{ij}^{a_i}$$

$$a_i^2 = \frac{s_i^2}{s_i^2 + d_i^2}$$

$$b_i = e^{-.5 a_i (a_i - 1)(s_i^2 + d_i^2)}$$

One can show that the expectation of Q_{ij} approximately equals $\overline{P}_i$:

$$\begin{aligned}
E(Q_{ij}) &= E\left(b_i \overline{P}_i^{\,1-a_i} R_{ij}^{a_i}\right) \\
&= E\left(b_i \overline{P}_i^{\,1-a_i} \overline{P}_i^{\,a_i} h_{ij}^{\,a_i} q_{ij}^{\,a_i}\right) \\
&= \overline{P}_i e^{-.5 a_i (a_i - 1)(s_i^2 + d_i^2)} E\left(h_{ij}^{\,a_i} q_{ij}^{\,a_i}\right) \\
&\cong \overline{P}_i
\end{aligned}$$

This guarantees that, under our assumptions, the Laspeyres index will not be biased by the substitution of Q_{ij} for R_{ij}.

$$E\left(\sum_i \sum_j w_{ij0} Q_{ij}\right) \cong \sum_i \sum_j w_{ij0} \overline{P}_i = E\left(\sum_i \sum_j w_{ij0} P_{ij}\right)$$

It can be further demonstrated that with Q_{ij} inserted for R_{ij}, the Tornqvist index has the desired expectation.

$$E\left(\prod_i \prod_j Q_{ij}^{\ w_{ij.}}\right) = \left(\prod_i \overline{P}_i^{\,w_{i.}}\right)\left(\prod_i \prod_j e^{.5 w_{ij}(w_{ij}-1) s_i^{\,2}}\right)$$

The values $\overline{P_i}$, $s_i^{\,2}$, and $d_i^{\,2}$ are, of course, unobserved. Fortunately, the existence of replicate indexes enables us to form estimates of these parameters. We form an estimator d_i of d_i by computing the average of the area standard errors:

$$d_i = \frac{1}{J}\sum_j^J \left[\frac{1}{N_j\left(N_j - 1\right)}\sum_n^{N_j}\left(\ln\left[R_{ij}^n\right] - \ln\left[R_{ij}^{\cdot}\right]\right)^2\right]^{1/2}$$

where n indexes an area replicate and $R_{ij}^{\cdot}$ indicates the mean of the replicate relatives (not the full-sample relative) for the stratum. Then we use the across-area variation in R_{ij} to estimate the variance in P_{ij}:

$$s_i^2 = \frac{\sum_j^J \left(\ln\left[R_{ij}\right] - \ln\left[\overline{R_i}\right]\right)^2}{J - 1} - d_i^2$$

Finally, using the U.S.-mean index relative $\overline{R_i}$ to estimate $\overline{P_i}$, we compute the composite estimate of the local index relative as

$$\hat{Q}_{ij} = b_i \overline{R}_i^{\,1-a_i} R_{ij}^{a_i}$$

$$a_i^2 = \frac{s_i^{\,2}}{s_i^{\,2} + d_i^{\,2}}$$

$$b_i = e^{-.5(s_i^{\,2} + d_i^{\,2}) a_i (a_i - 1)}$$

The composite approach should mitigate the impact of random error bias and thereby yield improved Tornqvist and geometric mean estimates, as well as improved estimates of the substitution bias in the Laspeyres index. (As demonstrated above, the Laspeyres indexes themselves will not be affected systematically by the composite estimation approach.) Table 2 and Figure 3 compare annual estimates of substitution bias based on indexes constructed using the $\hat{Q}_{ij}$ instead of the R_{ij}.

Table 2 and Figure 3 both show that, as expected, the Tornqvist indexes are higher relative to the Laspeyres than the previous, unadjusted estimates. The estimated Laspeyres substitution bias is thus reduced in each year. Over the eight years, the average substitution bias estimate is 0.080 percentage point, compared to the 0.118 percentage point average obtained previously. Whereas using the unadjusted item-area indexes yielded annual substitution bias estimates of at least one tenth of a percentage point in every year except 1993, the adjusted bias estimates fall below that level except for the 0.107 value in 1990.

Table 2 also shows the estimated downward biases in the geometric mean indexes relative to the Tornqvist. Note that correcting the random error bias raises both the Tornqvist and the geometric mean, so that it will not necessarily widen the gap between them. The eight-year average bias is about 0.03 percentage point whether or not the composite estimation approach is used. Thus, the adjusted geometric mean bias is still smaller than the Laspeyres substitution bias estimate, but the ratio of the two biases is smaller, about 2.7 using the adjusted estimates as opposed to about 3.9 using the unadjusted estimates.

Because the composite-estimation approach leads to a marked reduction in the estimated substitution bias in chain Laspeyres indexes, it is useful to apply the same procedure to fixed-base indexes as well. Aggregate Laspeyres, Tornqvist, and geometric mean indexes were estimated for each of the years 1988 through 1995 using composite estimates of basic index changes relative to the fixed base year of 1987. Unadjusted, the average Laspeyres substitution bias estimate is 0.174 percentage point per year; adjusted, the average is 0.130 percentage point per year. (The fixed-base geometric mean index averages 0.067 and 0.063 percentage point below the corresponding Tornqvist in the unadjusted and adjusted estimates, respectively.)[19]

VI. Geometric Mean Basic Indexes and Small Sample Bias

A fundamental assumption in the preceding analysis is that the basic item-area indexes in the CPI are unbiased estimates of their population counterparts—that is, that the multiplicative sampling errors θ_{ij} have mean one. Under this assumption, the aggregate Laspeyres indexes are unbiased, because they are linear functions of the basic indexes, but the aggregate superlative indexes are not. If it were instead assumed that the basic indexes were themselves biased estimates, the analysis of this paper would still hold, in the sense that the differences between Laspeyres and superlative indexes would still be exaggerated by the random variation in the basic indexes. The interpretation of this distortion, however, might be different.

During the 1987-1995 period studied here, the CPI basic indexes were computed using a modified Laspeyres or arithmetic mean formula. Research by the BLS in the early 1990’s identified an upward bias that arose from the way in which the formula was applied when item samples were replaced, or rotated. This problem, known as “formula bias” or “functional form bias,” was often cited as one of the biases in the CPI along with “upper-level substitution bias,” until it was eliminated in a series of steps during 1995 and 1996.[20]

---

[19] Unfortunately, the data set used here does not permit the analysis of indexes using a 1982-84 base period for expenditures, the period used in the CPI during the 1987-1995 period of study.

[20] For a description of this problem and its solution see, for example, Moulton (1996), pp. 165-168.

In January 1999, the BLS began using a geometric mean formula to compute most of the basic item-area indexes in the CPI.[21] The geometric mean indexes are not subject to formula bias, but, as shown by McClelland and Reinsdorf (1997), they suffer from a potential small-sample bias. A geometric mean basic index relative is computed as the antilog of the weighted mean of the logarithms of individual item price relatives. Under reasonable assumptions, then, the *logarithm* of the sample index relative will be an unbiased estimate of the *logarithm* of the population geometric mean index relative. The index relative itself, however, will be upwardly biased in finite samples.

The small-sample bias in geometric mean indexes is, in fact, the reverse of the problem discussed in this paper. It can be shown that the proportional upward small-sample bias in a geometric mean basic index relative is approximately half the variance of the relative. Thus, this upward bias is of the same order of magnitude as the downward "random-error bias" in the Tornqvist caused by basic index variance. (As noted in Section II above, if all index strata have equal share weights and equal variances δ^2, the expectation of the sample Tornqvist was $I^T\exp(-.5\delta^2)$). Another way of saying this is that if, instead of assuming unbiased basic indexes, we assume that the logarithms of basic indexes are unbiased, there would be no random-error bias in the Tornqvist index.

Having said this, it is important to emphasize that the basic indexes analyzed in this paper were not computed using the geometric mean formula and therefore were not subject to this small-sample bias. Moreover, and equally important, the issue of small-sample bias does not invalidate the conclusion that upper-level substitution bias has been overestimated. Whether or not one believes that basic CPI indexes are biased upward, the difference between Laspeyres and superlative aggregate indexes overstates the difference that would be observed if the underlying price samples were larger.

VII. Implications

The central result of this paper is that random error in CPI item-area price indexes causes estimated superlative indexes to be biased downward and Laspeyres substitution bias to be exaggerated. In particular, the composite-estimation analysis of Section V suggests that for chain Laspeyres indexes compared to chain Tornqvist indexes the estimated bias may be only about 0.08 percentage point rather than the roughly 0.12 percentage point previously estimated. Moreover, a similar comment applies to previous comparisons of fixed-base Laspeyres indexes to chain or fixed-base superlative indexes.

These are significant findings both qualitatively and quantitatively. Not only can random-error bias be illustrated in conceptual terms, it has been shown here to distort calculations using actual CPI data. To be sure, the analysis also shows that, even using the composite-estimation correction for random-error bias, estimated aggregate superlative indexes are closer to geometric mean indexes than to Laspeyres indexes. That is, Laspeyres substitution bias—or, equivalently, consumer substitution among CPI strata in response to relative price change—does *not* appear to be an illusion created by random variation in sample prices. This paper does, however, call into question much of the

[21] See Dalton et al (1998) for details on the introduction of the geometric mean formula.

conventional wisdom about superlative index calculation and previous empirical results. This, in turn, has several implications.

First, the historical performance of the Laspeyres formula for use in aggregation of stratum indexes is not as unsatisfactory as previously may have been thought, relative to a cost-of-living index objective. The "correct" Tornqvist aggregate indexes are probably closer to the Laspeyres CPI than indicated by earlier studies.

By the same token, the results here cast doubt on the tentative estimate by Shapiro and Wilcox (1997) that a constant-elasticity of substitution (CES) index with a substitution parameter of 0.7 would closely track a superlative index. The "best" CES index likely has a lower substitution parameter, reflecting a smaller Laspeyres substitution bias.[22]

Finally, these results may provide lessons for the BLS superlative index planned for publication beginning in 2002. Numerous decisions must be made in preparation for that index, including: the role of composite estimation of local-area expenditures; the frequency of publication; the length of the expenditure base period; the advisability of chaining; and the possible use of a Laspeyres or CES "tail" for publication of preliminary index values. Among all those decisions, the impact of random error on superlative indexes should also be taken into consideration.[23]

---

[22] This analysis may also have implications for the particular price index relatives used in previous superlative index studies. The elementary price relatives used in this paper are computed from CPI annual averages of the basic monthly price data. Thus, they correspond to the annual expenditure data used to construct expenditure shares for aggregation in the indexes. This has not always been the practice in prior studies. Shapiro and Wilcox (1997), for example, use December-to-December price relatives, following the practice in Aizcorbe and Jackman (1993).

Following the logic of this paper, the December-to-December price relatives can be viewed as imperfect measures of the "correct" underlying price relatives. It is reasonable to suppose that consumers choose the annual expenditure shares of each item based on prices in effect throughout the year, not just in December. The December-to-December relatives should be correlated with the annual-average relatives, but with an even larger random "error" attached. It can be hypothesized, therefore, that substitution bias estimates will be larger when estimated using the December-to-December relatives. Partial confirmation of this hypothesis is provided by the fact that the average December-to-December bias estimates by Shapiro and Wilcox are somewhat higher (0.16 percentage point versus 0.14 percentage point, in comparisons of 1987-weighted Laspeyres to chain Fisher indexes) than corresponding annual-average estimates.

[23] It should be noted that the Fisher Ideal indexes produced for the National Income and Product Accounts should be relatively unaffected by this problem, since they employ U.S.-level CPI indexes rather than the item-area stratum indexes.

References

Abraham, Katharine G., John S. Greenlees, and Brent R. Moulton, “Working to Improve the Consumer Price Index,” *Journal of Economic Perspectives*, Winter 1998, pp. 27-36.

Aizcorbe, Ana M., and Patrick C. Jackman, “The Commodity Substitution Effect in CPI Data, 1982-1991,” *Monthly Labor Review*, 116, no. 12, December 1993, pp. 25-33.

Aizcorbe, Ana M., Robert A. Cage, and Patrick C. Jackman, “Commodity Substitution Bias in Laspeyres Indexes: Analysis Using CPI Source Data for 1982-1994,” paper presented at the Western Economic Association International Conference in San Francisco, July 1996 (Washington, DC, Bureau of Labor Statistics).

Boskin, Michael J., Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches, and Dale W. Jorgenson, “Consumer Prices, The Consumer Price Index, and The Cost of Living,” *Journal of Economic Perspectives*, Winter 1998, pp. 3-26.

Dalton, Kenneth V., John S. Greenlees, and Kenneth Stewart, “Incorporating a Geometric Mean Formula into the CPI,” *Monthly Labor Review*, 121, no. 10, October 1998, pp. 3-7.

Diewert, W. Erwin, “Exact and Superlative Index Numbers,” *Journal of Econometrics*, May 1976, 4, 115-45.

Diewert, W. Erwin, “Fisher Ideal Output, Input, and Productivity Indexes Revisited,” *Journal of Productivity Analysis*, September 1992, pp. 211-248.

Erickson, Tim. “Expected Values of Laspeyres and Geometric Mean Indexes When Price Relatives and Expenditure Weights Are Replaced by Well-Behaved Estimates,” DPINR Working Leaflet #1, Bureau of Labor Statistics, March 1998.

Greenlees, John S., “Expenditure Weight Updates and Measured Inflation,” in *Proceedings of the Third Meeting of the International Working Group on Price Indices* (B. M. Balk , ed.). Statistics Netherlands, 1998, pp. 57-64.

Leaver, Sylvia G., and Robert A. Cage, “Estimating the Sampling Variance for Alternative Estimators of the U.S. Consumer Price Index,” Bureau of Labor Statistics, 1997.

McClelland, Robert, “On the Use of the Geometric Mean and Laspeyres Formulas in the CPI,” Bureau of Labor Statistics, July 1998.

McClelland, Robert, and Marshall Reinsdorf, “Small Sample Bias in Geometric Mean and Seasoned CPI Component Indexes,” Bureau of Labor Statistics, July 1997.

Moulton, Brent R., “Bias in the Consumer Price Index: What is the Evidence?” *Journal of Economic Perspectives*, Fall 1996, pp. 159-177.

Moulton, Brent R., and Kenneth J. Stewart, “An Overview of Experimental U.S. Consumer Price Indexes,” in *Proceedings of the Third Meeting of the International Working Group on Price Indices* (B. M. Balk , ed.). Statistics Netherlands, 1998, pp. 41-56.

Pollak, Robert A. *The Theory of the Cost-of-Living Index.* New York: Oxford University Press, 1989).

Shapiro, Matthew D., and David W. Wilcox, "Mismeasurement in the Consumer Price Index: An Evaluation," in Ben S. Bernanke and Julio J. Rotemberg, eds., *NBER Macroeconomics Annual 1996*, (MIT Press, 1996)

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

Swanson, David, "Variance Estimates for Changes in the Consumer Price Index, January 1993-December 1997," *CPI Detailed Report,* February 1998, pp. 6-10.

U.S. Bureau of Labor Statistics, "Measurement Issues in the Consumer Price Index," paper prepared in response to a letter from Rep. Jim Saxton, Chairman of the Joint Economic Committee, June 1997.

U.S. Senate, Committee on Finance, *Final Report of the Advisory Commission to Study the Consumer Price Index.* Print 104-72, 104 Cong., 2 sess., (Washington, D.C., Government Printing Office, 1996).

Table 1
Full Sample Index Results, in Percent

	Index Percent Change			Substitution Bias	
Year	Laspeyres	Tornqvist	Geometric	Laspeyres	Geometric
1988	3.97%	3.84%	3.81%	0.131%	-0.031%
1989	4.56%	4.45%	4.39%	0.100%	-0.055%
1990	5.12%	4.96%	4.94%	0.149%	-0.019%
1991	3.92%	3.79%	3.76%	0.128%	-0.026%
1992	2.88%	2.74%	2.73%	0.132%	-0.018%
1993	2.76%	2.67%	2.62%	0.087%	-0.049%
1994	2.59%	2.49%	2.46%	0.103%	-0.024%
1995	2.74%	2.62%	2.60%	0.115%	-0.019%
Average	3.56%	3.44%	3.41%	0.118%	-0.030%

Table 2
Unadjusted and Adjusted Substitution Bias Estimates

	Laspeyres Index		Geometric Mean Index	
Year	Unadjusted	Adjusted	Unadjusted	Adjusted
1988	0.131%	0.094%	-0.031%	-0.028%
1989	0.100%	0.066%	-0.055%	-0.051%
1990	0.149%	0.107%	-0.019%	-0.023%
1991	0.128%	0.090%	-0.026%	-0.026%
1992	0.132%	0.090%	-0.018%	-0.021%
1993	0.087%	0.052%	-0.049%	-0.041%
1994	0.103%	0.058%	-0.024%	-0.025%
1995	0.115%	0.080%	-0.019%	-0.019%
Average	0.118%	0.080%	-0.030%	-0.029%

Figure 1. Annual Laspeyres Substitution Bias Estimates

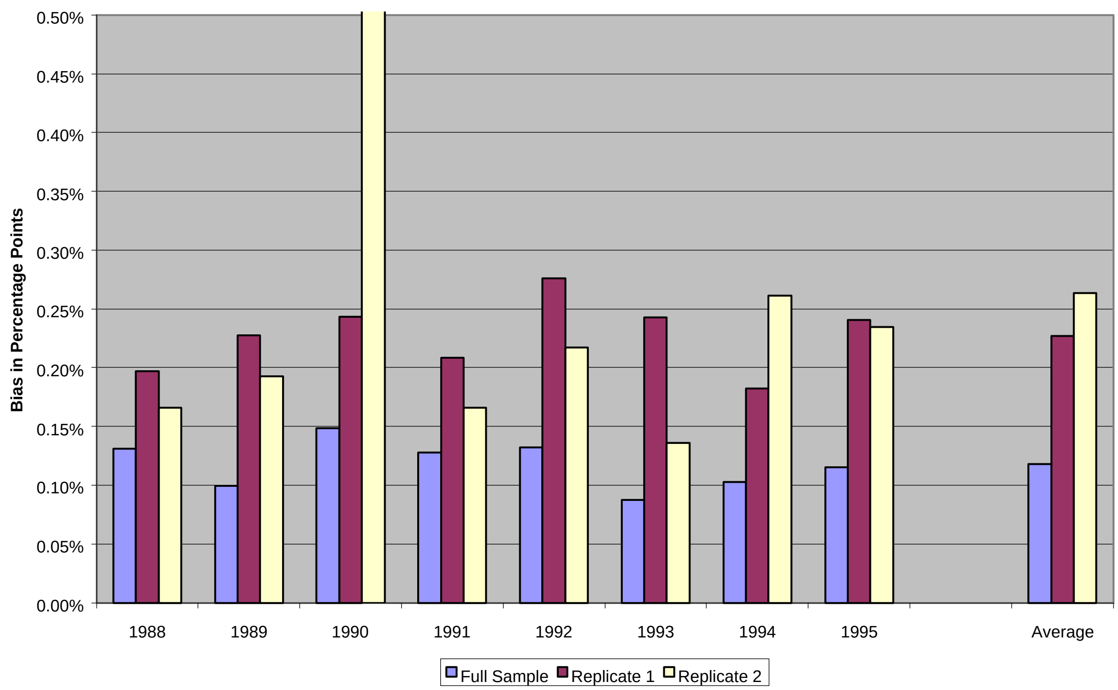

Figure 2. Annual Geometric Mean Substitution Bias Estimates

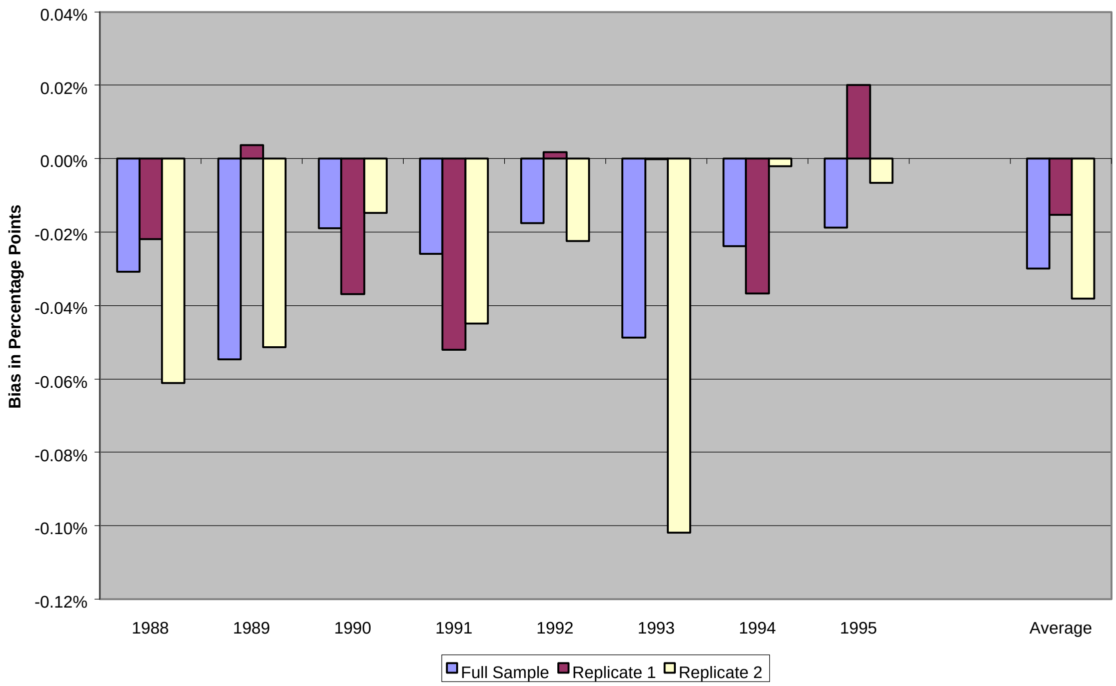

Figure 3. Unadjusted and Adjusted Laspeyres Substitution Bias Estimates