Journal of Official Statistics, Vol. 17, No. 3, 2001, pp. 369–390

Finite Sample Effects in the Estimation of Substitution Bias in the Consumer Price Index

Ralph Bradley[1]

Since it does not allow for substitution effects, the U.S. Consumer Price Index (CPI) has been criticized for overstating the Cost of Living increase. Previous studies estimate substitution bias by differencing a superlative index and the Laspeyres. However, both the Laspeyres and the superlative indexes are nonlinear in prices, and are subject to finite sample bias. This study shows that these previous estimates of the CPI's substitution bias are overstated because of the finite sample effects.

Key words: Substitution bias; superlative indexes.

1. Introduction

The U.S. Consumer Price Index (CPI) has been criticized for overstating the true rate of increase in the cost of living. The best-understood source of this bias is the CPI's failure to account for the substitution effect when relative prices change. Several studies have estimated the magnitude of this bias at the ''across-strata'' level.[2] Shapiro and Wilcox (1997) estimate the annual bias at .3 percent while Aizcorbe, Cage and Jackman (1996) have a .2 percent annual bias. Typically, this bias is estimated by taking the difference between the CPI and a superlative index such as the Fisher and the Törnqvist.[3]

However, these previous studies do not account for the statistical bias that comes from the effect of finite samples. When constructing the CPI, the United States Bureau of Labor Statistics (BLS) selects random samples within each stratum and computes a ''lower level'' price index for that stratum. These lower level indexes are then used as price proxies for the ''upper level'' or ''All-Items'' Index. The lower level indexes have a Laspeyres form for some strata and a Jevons (geometric mean) for other strata. This is called the hybrid approach. Kish, Namboordi, and Pillai (1962) and McClelland

[1] U.S. Bureau of Labor Statistics, 2 Massachusetts Ave. NE, Washington, DC 20212, U.S.A. Email: bradley_r@bls.gov

[2] A ''stratum'' in the CPI is a broad commodity item within an area. Examples are cereal in Boston, electrical utilities in San Diego, etc. As in previous studies, it is only possible to investigate ''across strata'' substitution bias since it is not possible to collect the expenditure data for each product within a stratum for each period.

[3] A superlative index allows for substitution effects. For a detailed description of superlative indexes see Diewert (1976). Boskin et al. (1996) in their report to the U.S. Senate Finance Committee urged that U.S. Bureau of Labor Statistics (BLS) compute a superlative type price index.

Acknowledgments: The author wishes to thank Rob Cage, Ken Stewart, Rob McClelland, Alan Dorfman, Dennis Fixler, Janice Lent, John Greenlees, Patricia Rozaklis, Michael Hoke and the participants at the 1999 NBER seminar on Index and Productivity Measurement for advice and assistance. The views represented in this article are solely those of the author, and do not reflect policies or procedures of the BLS.

and Reinsdorf (1997) showed that both these lower level indexes exhibit finite sample bias because of their non-linear nature and small sample size.[4] (The average stratum sample size is 13.) This small size is inadequate to purge the sampling bias and the sampling variance of the lower level index, and this induces sampling bias in the upper level index.[5]

Since the Törnqvist Index is a concave function of the lower level indexes, both sampling bias and variance in the lower level indexes will induce sampling bias in the upper level Törnqvist, while only the sampling bias of the lower level index will affect the sampling bias in the upper level Laspeyres. The reason that the sampling bias of the upper level Laspeyres is only affected by the sampling bias for the lower level index is shown in Greenlees (1998). The upper level Laspeyres is a linear combination of the lower level indexes.[6] Although the sampling bias and the sample variance of the lower level indexes are sources of sampling bias for the Törnqvist, I will show that these effects on the Törnqvist converge to zero as the number of strata grows. The same cannot be said for the effects of sampling bias for the upper level Laspeyres.

This note shows that when using the prices collected by BLS, the finite sample bias for the Laspeyres is positive and higher than it is for the Törnqvist. Therefore, the difference between a Laspeyres and Törnqvist cannot be attributed entirely to ‘‘across strata’’ substitution effects, and it gives a statistically biased estimate of the commodity substitution bias across strata.[7] When the U.S. Boskin Committee submitted their final critique of the CPI to the U.S. Congress, there was no mention of sampling bias, and they attributed the entire difference between the estimates of an upper level Laspeyres and Törnqvist to commodity substitution bias across strata. It is now apparent that this conclusion is not correct.

The basic reasoning behind the sampling bias of the lower level index is that it is a nonlinear transformation of averages that still exhibits sampling variance. The Laspeyres is basically a ratio of averages, and the source of its sampling bias comes from the result that for any two random variables x_i and y_i, and finite sample size n, the following occurs:

$$E\left(\frac{1/n\sum y_i}{1/n\sum x_i}\right) \neq \frac{E(y)}{E(x)} \tag{1}$$

And the source of sampling bias for the lower level Jevons Index comes from Jensen’s Inequality:

$$E\left(\exp 1/n\sum \ln(y_i/x_i)\right) > \exp(E(\ln(y/x)) \tag{2}$$

[4] The finite sample bias from ratio estimation is well covered in Cochran (1963). Kish, Nambroodri, and Pillai (1962) also derive the finite sample bias from ratio estimation.

[5] The upper level Törnqvist and Fisher Index are nonlinear and concave transformations of the lower level index. As a result of Jensen’s Inequality, any nonzero variance in the lower level index will induce a negative bias in the upper level index if the lower level relative is itself free from bias. However, the lower level index has positive bias, and therefore, these two effects have opposing effects on these superlatives.

[6] Although the upper level Laspeyres is linear in the lower level index, the lower level index is always a nonlinear transformation of the originally sampled prices. Therefore, the Laspeyres is nonlinear in the originally sample prices.

[7] The expenditure weights used in the upper level indexes also come from finite random samples. However, I show in the Appendix that the sampling bias in these weights converges very rapidly to zero. Dorfman, Leaver, and Lent (1999) also investigate the effects of random expenditure weights and also show that the sampling error has ‘‘relatively little impact on the bias of the estimator.’’ Therefore, this study does not analyze the effects of the expenditure weights.

Greenlees (1998) discusses the problem of constructing superlative indexes when random error is present in the lower level indexes. His analysis is based on the assumption that the lower level indexes are not subject to sampling bias. However, as previously mentioned, this assumption is not correct. Therefore, while both this study and Greenlees show that the sampling bias of the upper level Laspeyres index is only affected by the sampling bias of the lower level index and not its variance, our final results differ dramatically. Since he assumes no sampling bias for the lower level index, he concludes that the upper level Laspeyres does not suffer from sampling bias while this study concludes that it is the sampling bias in the upper level Laspeyres that is the major factor behind the sampling bias of the estimates of commodity substitution bias across strata. Greenlees concludes that it is the sampling variance in the lower level index that induces a downward sampling bias in the upper level Törnqvist whereas I show that this sampling variance has almost no effect.

Other studies have analyzed the precision of the CPI. Biggeri and Giommi (1987) acknowledge that the CPI could contain sampling bias, but their study focuses on evaluating variance estimation methods when there is sampling error of household expenditures as well as sampling error of prices. While the sampling error of household expenditure weights does add variance and bias to a price index, I show in the appendix of this article that these effects converge to zero as the number of strata increases. Since there are over 9,000 strata, there is little if any sampling bias coming from the expenditure weights.

The main finding of this study is that the previous range of estimates of .2 percent to .3 percent for across strata commodity substitution bias should be revised to a range of .1 percent to .2 percent.[8] The rest should be attributed to sampling bias. These percentages might not seem large, but a .1 percent change in the CPI has a direct 13.5 billion USD annual effect on the U.S. Federal Budget alone. If one includes the effects on pension systems, state and local government budgets, and private contracts, this could easily have over a 100 USD billion effect.

Section 2 briefly describes the two-level construction of the CPI and superlative indexes when using the BLS sample of price quotes and expenditures. It then divides into subsections. Since this article focuses on the effects of sampling bias on the upper level index, the first subsection derives the finite sample bias of the upper level indexes and the resulting sampling bias of the estimate for commodity substitution bias, while the second subsection describes the results from previous studies on the sampling bias for the lower level indexes. The only new result in this subsection deals with the sampling bias of the natural log of the lower level Laspeyres index. Finally, Section 3 describes the results of a bootstrap experiment that verifies the properties established in Section 2, and then gives a ''bias corrected'' estimate of substitution bias.

A note of caution to the reader. This article deals with two distinct sources of bias. One is the substitution effect bias of the Laspeyres index (hereafter referred to as ''substitution bias''), and the other is the statistical bias that comes from estimating nonlinear forms with finite samples (hereafter referred to as ''sampling bias''). It is the intent of this article to show how the sampling bias induces a biased estimate of the substitution bias.

[8] Obviously, commodity substitution bias will change each period since the underlying distribution of prices changes each period.

2. Effects of Sampling Bias on Index Estimation

BLS uses two levels of aggregation to construct the All-Items index. At the lower level, an index (from here on it is referred to as the relative) is calculated for each stratum. In each stratum, a sample of prices is drawn, and from this sample the month to month relative is computed either as a Laspeyres index or a Jevons index. Since the sample is finite and since the relative is a nonlinear function of the sampled prices, the relative will exhibit both a finite sample bias and a nonzero variance. At the higher level, the All-Items upper level index is constructed by using the relative as a proxy for the true ratio of the current month's price to the previous month's prices and by using strata expenditure shares that come from the U.S. Consumer Expenditures Survey. The set of strata that is used in the construction of the All-Items index is fixed.

2.1. Sampling bias of across strata substitution bias

In this subsection, I borrow from the notation of Greenlees (1998). Let the N strata be indexed by i; $w_{i,t}$, $Q_{i,t}$, and $P_{i,t}$ denote respectively the ith stratum's expenditure share, quantity, and price for period, t.[9] The current period is t and the comparison period is $t-1$. Although the textbook version of the Laspeyres is

$$\frac{\sum_{i=1}^{N} P_{i,t} Q_{i,t-1}}{\sum_{i=1}^{N} P_{i,t-1} Q_{i,t-1}} = \sum_{i=1}^{N} (P_{it}/P_{it-1}) w_{i,t-1} \tag{3}$$

since BLS cannot derive an estimate for $w_{i,t-1}$ in period t, it attempts to estimate the ''modified Laspeyres'' for a fixed basket of goods where the expenditure share $w_{i,t-1}$ in (3) is replaced with a fixed expenditure share from an earlier period. This expenditure share is denoted by $w_{i,0}$. This modified Laspeyres is

$$I_t^L = \sum_{i=1}^{N} (P_{it}/P_{it-1}) w_{i,0} \tag{4}$$

Letting $R_{i,t}$ denote the lower level sample relative from month $t-1$ to t, BLS estimates $P_{i,t}/P_{i,t-1}$ with $R_{i,t}$.[10] The resulting CPI estimate of the upper level index (4) is

$$\hat{I}_t^L = \sum_{i=1}^{N} R_{i,t} w_{i,0} \tag{5}$$

The expenditure share, $w_{i,0}$ also comes from a sample estimate but I show in the appendix that the sampling effect of the weight is of almost zero order, and therefore, I do not address its impact in this study. To estimate substitution bias, previous studies attempt

[9] I treat the stratum as one unit whereas Greenlees breaks the stratum into ''items'' and ''areas.''
[10] Hence the name relative – it approximates the relative price of the current period t to the base period $t-1$.

to use historical BLS data, estimate the following Törnqvist index[11]

$$I_t^T = \prod_{i=1}^{N} (P_{i,t}/P_{i,t-1})^{1/2(w_{i,t-1}+w_{i,t})} \quad (6)$$

with:

$$\hat{I}_t^T = \prod_{i=1}^{N} R_{i,t}^{1/2(w_{i,t-1}+w_{i,t})} \quad (7)$$

Then the estimated commodity substitution bias is

$$\hat{I}_t^L - \hat{I}_t^T \quad (8)$$

The problem with this approach is that previous studies (Cochran 1963; Kish, Namboordi, and Pillai 1962; McClelland and Reinsdorf 1997) have shown that $R_{i,t}$ exhibits both sampling bias and variance, since the sample size that is used for the estimation of $R_{i,t}$ is not adequately large to eliminate both its variance and sampling bias. I summarize these results in the next section. Because of sampling bias, these previous studies have shown that $E(R_{i,t}) > P_{i,t}/P_{i,t-1}$, for all i except for some highly restrictive cases. The resulting sampling bias, B_t^L, in the upper level Laspeyres is:

$$B_t^L = E(\hat{I}_t^L) - I_t^L = \sum_{i=1}^{N} E\{R_{i,t} - (P_{it}/P_{it-1})\} w_{i,0} > 0 \quad (9)$$

It is readily apparent that the sampling bias of the Laspeyres is merely linear in the sampling bias of the relative.[12] Likewise the bias, B_t^T, for the Törnqvist is represented as the ratio[13]:

$$B_t^T = E\left(\frac{\hat{I}_t^T}{I_t^T}\right) = E\left\{\prod_{i=1}^{N} \{R_{i,t}(P_{it-1}/P_{it})\}^{1/2(w_{i,t-1}+w_{i,t})}\right\} > 1 \quad (10)$$

and it is readily apparent that the sampling bias of the Törnqvist is not linear in the sampling bias of the relative. Therefore, (8) is not a consistent estimate of commodity substitution bias, $I_t^L - I_t^T$. We can decompose the estimated commodity substitution bias as

$$\hat{I}_t^L - \hat{I}_t^T = I_t^L - I_t^T + B_t^L - (B_t^T - 1)I_t^T \quad (11)$$

where the first difference on the right hand side of (11) is the true commodity substitution bias and the second difference is the sampling bias. Aizcorbe, Cage, and Jackman (1996) and Shapiro and Wilcox (1997) base their conclusions on the left hand side of (11), but have not accounted for the bias that is present on the right hand side of (11).

I illustrate this problem with the following example.[14]

[11] These Törnqvist indexes are at least two years old. Therefore, the index, $\hat{I}_{Jan,1996}^T$, is available in January 1998. As a result, estimates for commodity substitution bias are at least two years old.

[12] However, the relative is nonlinear in the originally sampled prices; therefore, the upper level Laspeyres is nonlinear in the originally sampled prices.

[13] Since the Laspeyres is a summation, the sampling bias is characterized by a difference, and since the Törnqvist is a product, the sampling bias is characterized by division.

[14] This example has been modified from the one in Greenlees (1998). His original example assumes zero bias in the relative.

Example. Suppose that

$$R_{i,t} = (P_{i,t}/P_{i,t-1})\theta_{i,t} \tag{12}$$

where the log of the sampling error, $\ln(\theta_{i,t})$, is distributed as an i.i.d. $N(0, \sigma^2)$ for all i. Then, the following holds:

$$E(R_{i,t}) = (P_{i,t}/P_{i,t-1})\exp\{\sigma^2/2\} > (P_{i,t}/P_{i,t-1}) \tag{13}$$

By substituting (13) into (9), it immediately follows that

$$B_t^L = \{\exp\{\sigma^2/2\} - 1\}\sum_{i=1}^{N}(P_{i,t}/P_{i,t-1})w_{i,0} = I_t^L\{\exp\{\sigma^2/2\} - 1\} > 0 \tag{14}$$

The sampling bias of the modified Laspeyres is strictly positive, linear in I_t^L, and the sampling bias of R_t. However, both the sampling bias of R_t and B_t^L are exponentially increasing in σ^2. After some algebraic manipulation, the sampling bias of the Törnqvist is

$$B_t^T = \exp\left\{1/8\sigma^2\sum_{i=1}^{N}(w_{i,t} + w_{i,t-1})^2\right\} > 1 \tag{15}$$

Since $(w_{i,t} + w_{i,t-1})^2$ is $O(N^{-2})$, $B_t^T - 1$ is $O(N^{-1})$, while (14) shows that B_t^L is $O(1)$ and is strictly greater than zero. Using this result and (11), the sampling bias for the commodity substitution bias has the asymptotic property:

$$\lim_{N\to\infty} \hat{I}_t^L - \hat{I}_t^T - I_t^L - I_t^T = \lim_{N\to\infty} B_t^L > 0$$

I now establish more general results. For the rest of this subsection, I continue to posit that $E(R_{i,t}) > P_{i,t}/P_{i,t-1}$. From (9), it is clear that the sampling bias of the Laspeyres index is strictly positive, and does not converge to zero as $N \to \infty$.

The results for the Törnqvist differ from the results for the Laspeyres. I rewrite the Törnqvist formulae as

$$\begin{aligned} I_t^T &= \exp\left\{\sum_{i=1}^{N} 1/2(w_{i,t-1} + w_{i,t})\ln(P_{i,t}/P_{i,t-1})\right\} \\ \hat{I}_t^T &= \exp\left\{\sum_{i=1}^{N} 1/2(w_{i,t-1} + w_{i,t})\ln(R_{i,t})\right\} \end{aligned} \tag{16}$$

Using these formulae, I can show the following result:

Proposition. Suppose that for all i, $E(\ln R_{i,t}) = (\ln P_{i,t}/P_{i,t-1})$, $0 < var(\ln R_{i,t}) < \infty$, and $R_{i,t}$ is independently distributed across i. Then, (i) $p\lim_{N\to\infty} B_t^L > 0$, and (ii) $p\lim_{N\to\infty} B_t^T = 0$, implying that the asymptotic bias of $\hat{I}_t^L - \hat{I}_t^T$ is equal to the asymptotic bias of the Laspeyres index.

Proof. (i) Since $E(\ln R_{i,t}) = (\ln P_{i,t}/P_{i,t-1})$, and $var(\ln R_{i,t}) > 0$, Jensen's Inequality implies that $E(R_{i,t}) > (P_{i,t}/P_{i,t-1})$, and (i) follows from (9). (ii) Since $E(\ln R_{i,t})$ equals $(\ln P_{i,t}/P_{i,t-1})$ and $\ln R_{i,t}$ is independently distributed across all i, then $E\{\sum_{i=1}^{N} 1/2(w_{i,t-1} + w_{i,t}) \ln(R_{i,t})\} = \sum_{i=1}^{N} 1/2(w_{i,t-1} + w_{i,t}) \ln(P_{i,t}/P_{i,t-1})$. Using this last result and using the assumption that $\ln R_{i,t}$ is independently distributed with bounded variance across all i, it follows that

$$var\left\{\sum_{i=1}^{N} 1/2(w_{i,t-1} + w_{i,t})(\ln(R_{i,t}))\right\} = \sum_{i=1}^{N} 1/4(w_{i,t-1} + w_{i,t})^2 var(\ln(R_{i,t}))$$

$$\leq \max_j\{var((\ln(R_{j,t}))\} \sum_{i=1}^{N} 1/4(w_{i,t-1} + w_{i,t})^2 = O(N^{-1})$$

Therefore, using the Weak Law of Large Numbers, I get that $p\lim_{N\to\infty} \sum_{i=1}^{N} 1/2(w_{i,t-1} + w_{i,t}) \ln(R_{i,t}) = p\lim_{N\to\infty} \sum_{i=1}^{N} 1/2(w_{i,t-1} + w_{i,t}) \ln(P_{i,t}/P_{i,t-1})$. Since I_t^T and $\hat{I}_t^T$ are respectively continuous functions of $1/2(w_{i,t-1} + w_{i,t}) \ln(P_{i,t}/P_{i,t-1})$, and $\sum_{i=1}^{N} 1/2(w_{i,t-1} + w_{i,t}) \ln(R_{i,t})$, by the Slutsky Theorem[15], we get (ii) since $p\lim_{N\to\infty} B_t^T = p\lim_{N\to\infty} \hat{I}_t^T - p\lim_{N\to\infty} I_t^T = 0$.

Under the conditions of the Proposition, the sampling bias of the estimate of commodity substitution bias will be positive, and will not diminish as the number of strata increases. This proposition gives some keen insight into the problem of sampling bias. As I will show in the next section, if the lower level relative is a Jevons, then the log of this relative is an unbiased estimator of the natural log of the true stratum price (i.e., the conditions for the Proposition hold); however, by Jensen's Inequality, this will automatically induce an upward sampling bias in the Jevons relative, and induce a systematic sampling bias in the upper level Laspeyres. The Laspeyres sampling bias will not diminish as the number of strata increases while the sampling bias of the Törnqvist will diminish as the number of strata increases. Additionally, the Proposition demonstrates how my results differ from Greenlees (1998). Greenlees starts his analysis with the assumption that the relative is an unbiased estimate with sampling variance. If the relative is not biased then using (9), one concludes, as Greenlees does, that the Laspeyres index is not biased. However, if the relative is unbiased but has sampling variance, then by Jensen's Inequality, the Törnqvist will have negative sampling bias because it is a concave function of the relative. Therefore, Greenlees concludes that the Laspeyres has no sampling bias, but the Törnqvist has negative sampling bias. I conclude that the Laspeyres has positive sampling bias while the Törnqvist is consistent.

Since the natural log of a Jevons relative has no sampling bias, I get the following corollary from the Proposition.

Corollary. If $R_{i,t}$ is a Jevons relative, then $p\lim_{N\to\infty}(\hat{I}_t^L - \hat{I}_t^T) - I_t^L - I_t^T = \lim_{N\to\infty} B_t^L > 0$.

In the case where the condition in the Proposition, $E(\ln R_{i,t}) = (\ln P_{i,t}/P_{i,t-1})$, does not hold, we get weaker results. To establish these results, I pass the expectations operator

[15] See Billingsley (1986) Corollary 2 page 344.

through the second order expansion of $\hat{I}_t^T$ around $\ln(P_{i,t}/P_{i,t-1})$ to get:

$$B_t^T - 1 = \frac{\hat{I}_t^T}{I_t^T} - 1 \approx \sum_{i=1}^{N} 1/2(w_{i,t-1} + w_{i,t})E(\ln(R_{i,t}) - \ln(P_{i,t}/P_{i,t-1}))$$
$$+ \sum_{i=1}^{N}\sum_{k=1}^{N}\{1/8(w_{k,t-1} + w_{k,t})(w_{i,t-1} + w_{i,t})\,E(\ln(R_{i,t})$$
$$- \ln(P_{i,t}/P_{i,t-1})(\ln(R_{k,t}) - \ln(P_{k,t}/P_{k,t-1})\} \quad (17)$$

Denote the sampling bias as $bias(\ln(R_{i,t})) = E(\ln(R_{i,t}) - \ln(P_{i,t}/P_{i,t-1})$. Then (17) is rewritten as

$$B_t^T - 1 \approx \sum_{i=1}^{N} 1/2(w_{i,t-1} + w_{i,t})bias(\ln(R_{i,t})) + \sum_{i=1}^{N} 1/8(w_{i,t-1} + w_{i,t})^2\{var(\ln(R_{i,t})$$
$$+ bias(\ln(R_{i,t}))^2\} + \sum_{i=1}^{N}\sum_{k=1,k\neq i}^{N} 1/8(w_{k,t-1} + w_{k,t})(w_{i,t-1} + w_{i,t})bias(\ln(R_{i,t}))$$
$$\times bias(\ln(R_{k,t})) = A_{1,N} + A_{2,N} + A_{3,N} \quad (18)$$

The effects of the sampling variance of $R_{i,t}$ are contained in $A_{2,N}$, and clearly $p\lim_{N\to\infty} A_{2,N} = 0$. This leads me to conclude that the effects of sampling variance converge to zero. If $bias(\ln(R_{i,t})) > 0$, for all i then $p\lim_{N\to\infty}(B_t^T - 1) \gtrsim 0$. However, it is important to note that $bias(R_{i,t}) > 0$ does not imply that $bias(\ln(R_{i,t})) > 0$. In fact, since ln() is a concave function, from Jensen's Inequality, there is an upper bound on $E(\ln(R_{i,t})) - \ln(P_{i,t}/P_{i,t-1})$:

$$E(\ln(R_{i,t})) - \ln(P_{i,t}/P_{i,t-1}) \leq \ln E(R_{i,t}) - \ln(P_{i,t}/P_{i,t-1})$$

or

$$bias(\ln(R_{i,t})) \leq \ln(bias(R_{i,t}))$$

This inequality is strict if $var(R_{i,t}) > 0$. It is possible that $E(\ln(R_{i,t}) - \ln(P_{i,t}/P_{i,t-1}) < 0$, for all i even though $E(R_{i,t}) - (P_{i,t}/P_{i,t-1}) > 0$. For example, this would occur if $\ln(R_{i,t})\tilde{\ }N(\delta, \sigma^2)$, $\delta < 0$, and $|\delta| < \sigma^2/2$. Thus, it is possible that $B_t^L = \hat{I}_t^L - I_t^L > 0$, while $B_t^T = \hat{I}_t^T - I_t^T < 0$. From these results, it is clear that systematic positive sampling bias in the Laspeyres index does not imply systematic positive sampling bias in the Törnqvist. If the sampling bias of the Laspeyres index is positive and the sampling bias of the Törnqvist is negative, then the sampling bias of the commodity substitution bias, $\hat{I}_t^L - \hat{I}_t^T$, will be strictly positive.

In the most general case, the results show that bias in the estimate of commodity substitution bias depends on the difference between the magnitude of the bias of the relative and the magnitude of sampling bias of the log of the relative. In the section, that describes the bias of the relative, I show that the bias of the log of the relative is smaller in magnitude than the bias of the relative for both the Laspeyres relative and the Jevons relative.

In the section on the sampling bias of the relative, I show that the log of sample Jevons has no sampling bias, and that the necessary conditions for no bias in the log of the Laspeyres are indeed sufficient conditions for positive sampling bias in the Laspeyres relative.

2.2. *The sampling bias of the relative*

Since the sampling bias of the upper level indexes is affected by the sampling bias of the lower level relative, this subsection briefly reviews the results of past studies that derive the sampling bias of the lower level relative. As previously mentioned, BLS estimates a Laspeyres relative for some strata and a Jevons relative for other strata.[16] The only new result in this section is the characterization of the sampling bias of the log of the Laspeyres relative. When BLS conducts its sampling of prices for a particular stratum, the probability that a particular item is chosen is proportional to the expenditure share for that item. This sampling method is referred to as ''Proportional to Spending'' sampling or PPS. The three parts in this subsection respectively establish the sampling bias for the Laspeyres relative, the log of the Laspeyres relative, and the Jevons relative. In the subsection that establishes the sampling bias of the Jevons relative, I show that the log of the Jevons relative is unbiased and thus I verify the Corollary of the previous section.

2.2.1. The sampling bias of the Laspeyres relative

When BLS estimates its Laspeyres relative, there is no quantity information available. Therefore, it uses the PPS sampling method along with a ''link price'' to estimate its Laspeyres relatives. Let R_t, m respectively denote the lower level relative for the stratum, and the sample size. $p_{j,t}$ is the observed price of the jth item in the stratum in time period t. The small p is used to distinguish it from the capital P used in the previous subsection to define the upper level price. The probability that the jth item is chosen equals its expenditure share within its stratum. In month t the sample of price quotes are selected and the sample estimate of the Laspeyres relative is computed as:

$$R_t = \frac{\frac{1}{m}\sum_{j=1}^{m} p_{j,t}/p_{j,l}}{\frac{1}{m}\sum_{j=1}^{m} p_{j,t-1}/p_{j,l}} \tag{19}$$

$p_{j,l}$ is the previously mentioned ''link price'' from ''link'' period $l < t-1$.[17] To briefly describe the theory behind this estimate, I denote the stratum population size of the items as M, and $w_{j,l}$ as the jth item's expenditure share in the ''link'' period. I take expectations of both the numerator and denominator of (19), and account for the PPS sampling scheme.

$$E\left(\frac{1}{m}\sum_{j=1}^{m} p_{j,t}/p_{j,l}\right) = \sum_{j=1}^{M} (p_{j,t}/p_{j,l})w_{j,l} \tag{20}$$

[16] This policy of using Laspeyres for some strata and Jevons for others is based on economic theory. The Laspeyres is used for strata with low substitution elasticities such as home heating oil, while the Jevons is used for strata with higher substitution elasticities such as canned vegetables.
[17] l must differ from both t and $t-1$ in order to avoid ''formula bias.'' For a detailed discussion on formula bias, see McClelland and Reinsdorf (1997).

Although $w_{j,l}$ is an expenditure share it also serves as a probability weight for the expectation operator and satisfies, $\sum_{j=1}^{M} w_{j,l} = 1$. Likewise for the denominator of (19), I get

$$E\left(\frac{1}{m}\sum_{j=1}^{m} p_{j,t-1}/p_{j,l}\right) = \sum_{j=1}^{M}(p_{j,t-1}/p_{j,l})w_{j,l} \tag{21}$$

Let S_l and $q_{j,l}$ be respectively the total amount of spending in the stratum and the unobserved quantity purchases for link period l. Since $w_{j,l} = q_{j,l}p_{j,l}/S_l$ in (20) and (21), it follows that

$$\frac{E\left(\frac{1}{m}\sum_{j=1}^{m} p_{j,t}/p_{j,l}\right)}{E\left(\frac{1}{m}\sum_{j=1}^{m} p_{j,t-1}/p_{j,l}\right)} = \frac{\sum_{j=1}^{M} p_{j,t}q_{j,l}}{\sum_{j=1}^{M} p_{j,t-1}q_{j,l}} = \frac{P_t}{P_{t-1}} \tag{22}$$

which is a modified Laspeyres index using quantity weights from the link period l. P_t/P_{t-1} is the target index that would be used in the upper level index if BLS could sample all M prices and quantities.

However, for a finite sample size m, Reinsdorf and McClelland (1997), Cochran (1963), and Kish, Namboordi, and Pillai (1962) show that

$$E(R) = E\left(\frac{\frac{1}{m}\sum_{j=1}^{m} p_{j,t}/p_{j,l}}{\frac{1}{m}\sum_{j=1}^{m} p_{j,t-1}/p_{j,l}}\right) \neq \frac{E\left(\frac{1}{m}\sum_{j=1}^{m} p_{j,t}/p_{j,l}\right)}{E\left(\frac{1}{m}\sum_{j=1}^{m} p_{j,t-1}/p_{j,l}\right)} \tag{23}$$

because $1/m\sum_{j=1}^{m} p_{j,t-1}/p_{j,l}$ in the denominator is a random variable with positive variance. The coefficient of variation for $p_{i,j,t}/p_{i,j,l}$ is denoted $cv(p_{i,j,t}/p_{i,j,l})$, and the correlation between $p_{i,j,t-1}/p_{i,j,l}$, and $p_{i,j,t}/p_{i,j,l}$ is $\rho(p_{i,j,t-1}/p_{i,j,l}, p_{i,j,t}/p_{i,j,l})$. All the studies cited above have shown that the second order approximation of this sampling bias is:

$$E(R_t) - \frac{P_t}{P_{t-1}} \approx \frac{P_t}{P_{t-1}}\left\{cv\left(\frac{p_{j,t-1}}{p_{j,l}}\right)^2 - \rho\left(\frac{p_{j,t-1}}{p_{j,l}}, \frac{p_{j,t}}{p_{j,l}}\right)cv\left(\frac{p_{j,t-1}}{p_{j,l}}\right)cv\left(\frac{p_{j,t}}{p_{j,l}}\right)\right\}\Big/ m \tag{24}$$

While there are cases where this can be negative, it is more often the case that it is positive. For example, if the coefficient of variation of $p_{j,t}/p_{j,l}$ remains constant over time and the first order autocorrelation is less than one then (24) is strictly positive. The prices in the BLS sample typically have these properties and therefore, the sampling bias (24) is positive.

It is extremely important to note that this sampling bias is not independent of the underlying price growth P_t/P_{t-1}. In the usual case when the term inside the brackets of (24) is positive, higher rates of inflation will induce larger sampling bias. Therefore, sampling bias will vary period to period if underlying inflation rates change period to period.

2.2.2. The sampling bias of the log of the Laspeyres relative

In the section on the construction of the upper level index, the sampling bias of the Törnqvist is characterized in part by the bias of the natural log of the relative. In the appendix, I show that the first order approximation of the sampling bias of the log of the Laspeyres is:

$$E(\ln(R_t)) - \ln\frac{P_t}{P_{t-1}} \approx \left(1/2\left\{cv\left(\frac{p_{j,t-1}}{p_{j,l}}\right)\right\}^2 - 1/2\left\{cv\left(\frac{p_{j,t}}{p_{j,l}}\right)\right\}^2\right)\Big/ m \tag{25}$$

Notice that if the coefficient of variation stays constant over time, this second order approximation is zero while the bias of the second order approximation of the relative as characterized by (24) is strictly positive. When this happens, the conditions for the Proposition hold and from the results of the section on the bias of the upper level index, it follows that the sampling bias of commodity substitution bias is strictly positive. For most strata in the BLS sample, there is little movement in the coefficient of variation over time and therefore (25) is usually close to 0.

2.2.3. The sampling bias of the Jevons relative

The estimate for the Jevons relative is:

$$R_t = \exp\left\{\frac{1}{m}\sum_{j=1}^{m}\ln(p_{j,t}/p_{j,t-1})\right\} \tag{26}$$

Notice that the link price, $p_{j,l}$ is not used. The theory behind this estimate is that based on PPS sampling, the following holds:

$$E(\ln R) = E\left\{\frac{1}{m}\sum_{j=1}^{m}\ln(p_{j,t}/p_{j,t-1})\right\} = \sum_{j=1}^{M}\ln(p_{j,t}/p_{j,t-1})w_{j,l} = \ln(P_t/P_{t-1}) \tag{27}$$

Since the left side of (27) is a simple average from a PPS sample, there is no sampling bias of the natural log of the Jevons relative. This then proves the Corollary. Then I get

$$\exp\{E\ln R_t\} = \prod_{j=1}^{M}(p_{j,t}/p_{j,t-1})^{w_{j,l}} = P_t/P_{t-1} \tag{28}$$

However, exp() is a convex function and from Jensen's inequality, the following holds:

$$E(R_t) = E(\exp\ln R_t) > \exp E\{\ln(R_t)\} = P_t/P_{t-1} \tag{29}$$

Therefore, the bias of the Jevons relative is strictly positive. For the upper level index this means that the component that uses Jevons relatives, the sampling bias of the Laspeyres will be strictly positive and $O(1)$ while by the Proposition in this article, the sampling bias of the Törnqvist will approach zero as the number of Jevons strata increases.

3. The Bootstrap Experiment

The main principle of the bootstrap is that the relationship between a population and its sample can be inferred by drawing new samples from the original sample, treating it as a ''virtual population.''[18] I continue to use the notation of the previous section. The purpose of this bootstrap is twofold. First, it verifies the results established in the previous

[18] Hall (1992) emphasizes that bootstrapping does not need to involve resampling. It could involve the derivation of parameters from iterated empirical distributions.

*Table 1. Descriptive results of the bootstrap experiment**

Upper level	Lower level	Original estimate %	Bootstrap average %	Percent difference %
1993				
Laspeyres	Laspeyres	103.355	103.448	0.089
Laspeyres	Jevons	103.045	103.121	0.074
Laspeyres	Hybrid	103.053	103.128	0.073
Paasche	Laspeyres	102.883	102.819	−0.062
Paasche	Jevons	102.564	102.515	−0.048
Paasche	Hybrid	102.575	102.517	−0.056
Fishers	Laspeyres	103.119	103.133	0.013
Fishers	Jevons	102.804	102.817	0.013
Fishers	Hybrid	102.814	102.822	0.008
Törnqvist	Laspeyres	103.114	103.126	0.011
Törnqvist	Jevons	102.801	102.813	0.012
Törnqvist	Hybrid	102.810	102.818	0.007
Jevons	Laspeyres	103.066	103.078	0.012
Jevons	Jevons	102.743	102.755	0.011
Jevons	Hybrid	102.753	102.760	0.007
1994				
Upper level	Lower level	Original estimate %	Bootstrap average %	Percent difference %
Laspeyres	Laspeyres	102.092	102.127	0.034
Laspeyres	Jevons	101.832	101.882	0.050
Laspeyres	Hybrid	101.827	101.899	0.071
Paasche	Laspeyres	101.364	101.265	−0.098
Paasche	Jevons	101.125	101.069	−0.055
Paasche	Hybrid	101.116	101.072	−0.044
Fishers	Laspeyres	101.728	101.695	−0.032
Fishers	Jevons	101.478	101.475	−0.003
Fishers	Hybrid	101.471	101.485	0.014
Törnqvist	Laspeyres	101.709	101.676	−0.032
Törnqvist	Jevons	101.458	101.456	−0.002
Törnqvist	Hybrid	101.452	101.466	0.015
Jevons	Laspeyres	101.643	101.608	−0.035
Jevons	Jevons	101.381	101.378	−0.003
Jevons	Hybrid	101.375	101.390	0.015

section, and then it estimates a bias correction so that one can decompose the current differences between a superlative and a Laspeyres into the commodity substitution bias effect and the sampling bias effect that was characterized by (11) in the previous section.

This bootstrap experiment differs from the conventional bootstrap simulation, since I resample each stratum separately rather than resample the entire sample. Suppose that stratum i's sample size is n_i. Then, my bootstrap resample (with replacement) is also n_i. I resample each stratum 999 times. After each resampling, I compute both a Laspeyres and Jevons relative. From the 999 Laspeyres and Jevons relatives for each stratum, I generate 999 corresponding upper level Laspeyres, and Törnqvist indexes. I also generate

Table 1. Continued

1995

Upper level	Lower level	Original estimate %	Bootstrap average %	Percent difference %
Laspeyres	Laspeyres	102.510	102.941	0.420
Laspeyres	Jevons	102.407	102.589	0.178
Laspeyres	Hybrid	102.193	102.565	0.364
Paasche	Laspeyres	102.076	101.952	−0.121
Paasche	Jevons	101.781	101.628	−0.149
Paasche	Hybrid	101.613	101.526	−0.086
Fishers	Laspeyres	102.293	102.446	0.149
Fishers	Jevons	102.093	102.108	0.014
Fishers	Hybrid	101.902	102.044	0.139
Törnqvist	Laspeyres	102.293	102.446	0.149
Törnqvist	Jevons	102.095	102.103	0.008
Törnqvist	Hybrid	101.905	102.042	0.134
Jevons	Laspeyres	102.309	102.481	0.168
Jevons	Jevons	102.163	102.173	0.010
Jevons	Hybrid	101.958	102.111	0.150

1996

Upper level	Lower level	Original estimate %	Bootstrap average %	Percent difference %
Laspeyres	Laspeyres	102.951	103.128	0.172
Laspeyres	Jevons	102.723	102.859	0.132
Laspeyres	Hybrid	102.712	102.857	0.141
Paasche	Laspeyres	102.155	102.019	−0.133
Paasche	Jevons	102.003	101.860	−0.140
Paasche	Hybrid	101.989	101.849	−0.137
Fishers	Laspeyres	102.552	102.572	0.019
Fishers	Jevons	102.362	102.359	−0.004
Fishers	Hybrid	102.350	102.352	0.002
Törnqvist	Laspeyres	102.562	102.592	0.028
Törnqvist	Jevons	102.374	102.379	0.005
Törnqvist	Hybrid	102.362	102.372	0.010
Jevons	Laspeyres	102.565	102.583	0.017
Jevons	Jevons	102.365	102.367	0.002
Jevons	Hybrid	102.354	102.364	0.010

*The Hybrid Lower Level uses the Jevons for all food strata and Laspeyres for all utility strata.

these upper level indexes using a hybrid approach where the Laspeyres is used for the strata that continue to be Laspeyres under BLS policy, and Jevons for the strata that continue to be Jevons under current BLS policy.

Because of the large size of the entire CPI sample, (over 9,000 strata), I limit my bootstrap to food at home and utilities for the calendar years 1993, 1994, 1995, and 1996. Even with these restrictions, the entire size is 172,988 quotes for 279 strata for four years. Therefore, the average monthly sample is less than 13 quotes. For the hybrid indexes, the food strata relatives are Jevons, and the utility strata relatives are Laspeyres.

Table 2. Naïve confidence interval of the bootstrap index and its difference from the original index

1993					
Upper	Lower	Estimate	5%	Median	95%
Laspeyres	Laspeyres	Bootstrap index	103.0	103.4	103.9
		Bootstrap-original	−0.4	0.1	0.5
Laspeyres	Jevons	Bootstrap index	102.7	103.1	103.5
		Bootstrap-original	−0.3	0.1	0.5
Laspeyres	Hybrid	Bootstrap index	102.7	103.1	103.6
		Bootstrap-original	−0.3	0.1	0.5
Törnqvist	Laspeyres	Bootstrap index	102.7	103.1	103.5
		Bootstrap-original	−0.4	0.0	0.4
Törnqvist	Jevons	Bootstrap index	102.4	102.8	103.2
		Bootstrap-original	−0.4	0.0	0.4
Törnqvist	Hybrid	Bootstrap index	102.4	102.8	103.3
		Bootstrap-original	−0.4	0.0	0.5
1994					
Upper	Lower	Estimate	5%	Median	95%
Laspeyres	Laspeyres	Bootstrap index	101.7	102.1	102.5
		Bootstrap-original	−0.4	0.0	0.4
Laspeyres	Jevons	Bootstrap index	101.5	101.9	102.2
		Bootstrap-original	−0.3	0.1	0.4
Laspeyres	Hybrid	Bootstrap index	101.5	101.9	102.2
		Bootstrap-original	−0.3	0.1	0.4
Törnqvist	Laspeyres	Bootstrap index	101.3	101.7	102.0
		Bootstrap-original	−0.4	0.0	0.3
Törnqvist	Jevons	Bootstrap index	101.1	101.5	101.8
		Bootstrap-original	−0.3	0.0	0.3
Törnqvist	Hybrid	Bootstrap index	101.1	101.5	101.8
		Bootstrap-original	−0.3	0.0	0.4

In this section, I change the notation slightly. I denote the upper level index by $I^{u\text{–}l}$, where u denotes the upper level method, and l denotes the lower level index or the relative. For the upper level Törnqvist $u = T$ and for the upper level Laspeyres $u = L$. For the Laspeyres relative $l = L$, and for the Jevons (Geomean) relative $l = G$. Finally, for the hybrid approach $l = H$. For example, $I^{T\text{–}L}$ denotes an upper level Törnqvist using Laspeyres relatives, while $I^{L\text{–}H}$ is a Laspeyres upper level index using the hybrid approach for the relatives which is the current BLS method.

Table 1 lists the descriptive results from the experiment. The first two columns describe the index type. The column labelled ''Original Estimate'' is the All-Items index from the CPI data base. The column labelled ''Bootstrap Average'' lists the average of the 999 indexes computed from the resamplings. For each year, the first row lists the Laspeyres All-Items index computed with Laspeyres relatives. The second row for each year lists the Laspeyres All-Items index using a Jevons relative. From (9), the difference between the original Laspeyres estimates and the bootstrap average is a weighted average of the sampling bias in the relative. For all four years, this estimate of the sampling bias in the Laspeyres and Jevons relatives is positive, and thus, confirms the results of McClelland

Table 2. Continued

1995					
Upper	Lower	Estimate	5%	Median	95%
Laspeyres	Laspeyres	Bootstrap index	101.9	103.0	103.8
		Bootstrap-original	−0.6	0.5	1.3
Laspeyres	Jevons	Bootstrap index	101.8	102.6	103.3
		Bootstrap-original	−0.6	0.2	0.9
Laspeyres	Hybrid	Bootstrap index	101.5	102.6	103.3
		Bootstrap-original	−0.6	0.4	1.1
Törnqvist	Laspeyres	Bootstrap index	101.3	102.5	103.3
		Bootstrap-original	−0.9	0.2	1.0
Törnqvist	Jevons	Bootstrap index	101.4	102.1	102.7
		Bootstrap-original	−0.7	0.0	0.6
Törnqvist	Hybrid	Bootstrap index	101.0	102.1	102.8
		Bootstrap-original	−0.9	0.2	0.9

1996					
Upper	Lower	Estimate	5%	Median	95%
Laspeyres	Laspeyres	Bootstrap index	102.6	103.1	103.7
		Bootstrap-original	−0.4	0.2	0.8
Laspeyres	Jevons	Bootstrap index	102.4	102.9	103.4
		Bootstrap-original	−0.4	0.1	0.6
Laspeyres	Hybrid	Bootstrap index	102.3	102.9	103.4
		Bootstrap-original	−0.4	0.2	0.6
Törnqvist	Laspeyres	Bootstrap index	102.6	103.1	103.7
		Bootstrap-original	−0.4	0.2	0.8
Törnqvist	Jevons	Bootstrap index	101.9	102.4	102.9
		Bootstrap-original	−0.5	0.0	0.5
Törnqvist	Hybrid	Bootstrap index	101.9	102.4	102.9
		Bootstrap-original	−0.5	0.0	0.5

and Reinsdorf (1997) of the previous section. Notice that while the hybrid represents the combination of both Laspeyres and Jevons relatives, its results are not necessarily ‘‘in between’’ the Laspeyres and the Jevons. The reason is that it is possible for the Jevons sampling bias to be smaller than the Laspeyres bias for food, and the reverse to be true for utilities. This would make the hybrid sampling bias less than both the Laspeyres and the Geomeans.

The results for the All-Items Laspeyres computed with the Jevons relative and the Törnqvist All-Items computed with the Jevons relative confirms the Proposition. The sampling bias of the Törnqvist All-Items index is far smaller than the sampling bias of the Laspeyres All-Items when the relative is the Jevons. More importantly, the Corollary is verified where the sampling bias of the estimate of commodity substitution bias is ‘‘close to’’ the sampling bias of the upper level Laspeyres and this sampling bias is strictly positive.

Although not addressed in Section 2, the Paasche indexes all have negative sampling bias. The resulting Fisher index is then the geometric average of a Laspeyres that has upward sampling bias, and a Paasche that has downward sampling bias. Since the Fisher

Table 3. *A comparison of the original and bias corrected indexes and estimates of commodity substitution bias*

1993			
Indexes	Original %	Bias corrected %	Percent difference %
I^{L_L}	103.355	103.263	−0.089
I^{T_L}	103.114	103.103	−0.011
Substitution bias estimates	0.234	0.156	−33.462
I^{L_G}	103.045	102.970	−0.074
I^{T_G}	102.801	102.789	−0.012
Substitution bias estimates	0.238	0.176	−26.056
I^{L_H}	103.053	102.978	−0.073
I^{T_H}	102.810	102.802	−0.007
Substitution bias estimates	0.236	0.171	−27.704
1994			
Indexes	Original %	Bias corrected %	Percent difference %
I^{L_L}	102.092	102.057	−0.034
I^{T_L}	101.709	101.741	0.032
Substitution bias estimates	0.377	0.311	−17.605
I^{L_G}	101.832	101.781	−0.050
I^{T_G}	101.458	101.461	0.002
Substitution bias estimates	0.368	0.316	−14.113
I^{L_H}	101.827	101.755	−0.071
I^{T_H}	101.452	101.437	−0.015
Substitution bias estimates	0.370	0.313	−15.246

is ‘‘close’’ to the Törnqvist, the properties of the Törnqvist should approximate the properties of the Fisher. Again, this is verified in Table 1.

Table 2 gives the 90 percent naive confidence intervals for the bootstrap indexes and their difference from the original indexes in the sample. It is apparent that zero always lies within the confidence interval of the difference between the bootstrapped and original index, and therefore, even for the All-Items Laspeyres, there is always a more than five percent chance of drawing a sample that actually will produce a Laspeyres that is less than the original Laspeyres even though the expected value of the difference is positive. Thus, there is still enough variance in the sampled priced quotes within the strata to produce a standard error in the All-Items index that is large relative to its bias. It should be noted that the variance for the entire sample should be smaller since there are only 279 strata in this experiment, but in the CPI data base there are over 9,000 so the confidence intervals using the full data set should be 17.6 percent $\approx \sqrt{279/9,000}$ the size of the bootstrapped confidence intervals.

In Table 3, I use the results of the bootstrap to do a correction for the sampling bias on the All-Items Laspeyres and Törnqvist for the Laspeyres, Jevons, and Hybrid relatives. Let $\hat{I}_t^{L\_L}$ denote the original sample estimate for the All-Items Laspeyres using

Table 3. Continued

1995			
Indexes	Original %	Bias corrected %	Percent difference %
I^{L_L}	102.510	102.081	−0.419
I^{T_L}	102.293	102.141	−0.149
Substitution bias estimates	0.212	−0.059	−127.776
I^{L_G}	102.407	102.226	−0.177
I^{T_G}	102.095	102.087	−0.008
Substitution bias estimates	0.306	0.135	−55.727
I^{L_H}	102.193	101.822	−0.363
I^{T_H}	101.905	101.768	−0.134
Substitution bias estimates	0.282	0.053	−81.205
1996			
Indexes	Original %	Bias corrected %	Percent difference %
I^{L_L}	102.951	102.775	−0.171
I^{T_L}	102.562	102.533	−0.028
Substitution bias estimates	0.379	0.236	−37.873
I^{L_G}	102.723	102.320	−0.393
I^{T_G}	102.374	102.158	−0.212
Substitution bias estimates	0.341	0.159	−53.395
I^{L_H}	102.712	102.568	−0.141
I^{T_H}	102.362	102.352	−0.010
Substitution bias estimates	0.342	0.211	−38.286

the Laspeyres relative, and let $\hat{\hat{I}}_t^{L\_L}$ denote its bootstrap average. Then the bias corrected index $\tilde{I}_t^{L\_L}$ is

$$\tilde{I}_t^{L\_L} = \frac{\left(\hat{I}_t^{L\_L}\right)^2}{\hat{\hat{I}}_t^{L\_L}} \tag{30}$$

The other indexes are bias corrected following the same procedure. The original estimate of the commodity substitution bias is estimated as:

$$\frac{\hat{I}_t^{L\_L}}{\hat{I}_t^{T\_L}} - 1 \tag{31}$$

The bias corrected estimate is

$$\frac{\tilde{I}_t^{L\_L}}{\tilde{I}_t^{T\_L}} - 1 \tag{32}$$

The column labelled ‘‘Original’’ in Table 3 gives the values for $\hat{I}_t^{L\_L}$, $\hat{I}_t^{T\_L}$, etc. and then the corresponding value for (31). The column labelled ‘‘Bias Corrected’’ lists $\tilde{I}_t^{L\_L}$, $\tilde{I}_t^{T\_L}$, etc. and then the corresponding value for (32).

The results here again confirm the results of the last section with the only exception in 1995 when the Laspeyres relative is used. Excluding that index, the original estimates of commodity substitution bias range from .2 percent to .4 percent. These are similar to the Shapiro and Wilcox (1997) results. However, the difference between the bias corrected estimates and the original estimates ranges from .06 percent to .2 percent. Therefore, there is not as much substitution bias as originally measured.

4. Conclusions

Sampling bias makes previous estimates of commodity substitution bias upwardly biased. From this study and other studies, there is on average a .25 percent difference between the CPI and a Törnqvist Index. However, the corrected bias averages .15 percent and the remaining .1 percent comes from sampling bias. Therefore, the bias of the CPI includes not only a commodity substitution bias, but also sampling bias that is greater in magnitude than the sampling bias in the corresponding superlative indexes.

It is therefore apparent that substitution elasticities across strata might not be as large as previously thought. The critique of the Boskin Report to the U.S. Senate Finance Committee (Boskin et al. 1996) that a main source of CPI bias is the unaccounted substitution effect perhaps needs to be revised to allow for the finite sample bias. One might conclude that this sampling bias can be reduced without changing the size of the BLS sample by changing the functional form of the CPI from a Laspeyres to a Törnqvist. However, at this point in time, the upper level expenditures weights cannot be generated on a timely basis to produce a current period Törnqvist. The Törnqvist indexes computed in this and other studies used historical data with a two-year lag. BLS could use an upper level Jevons and since the results for the Proposition would still hold for an upper level Jevons, changing the upper level from a Laspeyres to a Jevons could reduce the sampling bias, but perhaps at a cost of misspecified elasticity of substitution across strata.

Most other countries also have a multi-stage estimation of their CPI. If the initial samples of price quotes are small when calculating first stage price relatives, and if their upper level index is either a ‘‘fixed basket’’ or Laspeyres, then they too should be subject to the same sampling bias problem, and switching to a Törnqvist upper level index should also eliminate most of the sampling bias.

A. Appendix

A.1. Derivations of sampling bias for the log of the Laspeyres relative

I expand the sampling bias:

$$(\ln R_{i,t}) - (\ln P_t/P_{t-1}) = \ln\left(\frac{\sum_{j=1}^{m} p_{j,t}/p_{j,l})/m}{\sum_{j=1}^{m}(p_{j,t-1}/p_{j,l})m}\right) - \ln\left(\frac{\sum_{j=1}^{M}(p_{j,t}/p_{j,l})w_{j,l}}{\sum_{j=1}^{M}(p_{j,t-1}/p_{j,l})w_{j,l}}\right)$$

around $\sum_{j=1}^{m} p_{j,t}/p_{j,l})/m = \sum_{j=1}^{M}(p_{j,t}/p_{j,l})w_{j,l}$ and $\sum_{j=1}^{m}(p_{j,t-1}/p_{j,l})m = \sum_{j=1}^{M}(p_{j,t-1}/p_{j,l})w_{j,l}$ and then take expectations.

$$E(\ln R_{i,t}) - (\ln P_t/P_{t-1}) \approx \frac{1}{P_t} E\left(\sum_{j=1}^{m}(p_{j,t}/p_{j,l})/m - P_t\right) - \frac{1}{P_{t-1}} E\left(\sum_{j=1}^{m}(p_{j,t-1}/p_{j,l})/m - P_{t-1}\right) - 1/2\left(\frac{1}{P_t^2} E\left(\sum_{j=1}^{m}(p_{j,t}/p_{j,l})/m - P_t\right)^2 - \frac{1}{P_{t-1}^2} E\left(\sum_{j=1}^{m}(p_{j,t-1}/p_{j,l})/m - P_{t-1}\right)^2\right)$$

The first order terms are zero since by PPS, $E(\sum_{j=1}^{m}(p_{j,t}/p_{j,l})/m - P_t) = 0$ and $E(\sum_{j=1}^{m}(p_{j,t-1}/p_{j,l})/m - P_{t-1}) = 0$. Therefore, I get

$$E(\ln R_{i,t}) - (\ln P_t/P_{t-1}) \approx 1/2 \frac{var\left(\sum_{j=1}^{m}(p_{j,t-1}/p_{j,l})/m\right)}{\left(E\left[\sum_{j=1}^{m}(p_{j,t-1}/p_{j,l})/m\right]\right)^2} - \frac{var\left(\sum_{j=1}^{m}(p_{j,t}/p_{j,l})/m\right)}{\left(E\left[\sum_{j=1}^{m}(p_{j,t}/p_{j,l})/m\right]\right)^2}$$

$$\approx 1/2\left(\left\{cv\left(\frac{p_{j,t-1}}{p_{j,l}}\right)\right\}^2 - \left\{cv\left(\frac{p_{j,t}}{p_{j,l}}\right)\right\}^2\right/m.$$

A.2. Sampling effect on the expenditure weights

This section demonstrates that the sampling bias of the expenditure weights is of order 1/4,500,000.

Let $w \in \Re^N$ denote the spending shares for the N strata, and $R_t \in \Re^N$ denote the relative for time period t. Then for the upper level indexes are denoted as

$$I^L(w, R_t) = \sum_{i=1}^{N} w_i R_{i,t} \tag{33}$$

and

$$I^T(w, R_t) = \exp\left(\sum_{i=1}^{N} w_i \ln(R_{i,t})\right) \tag{34}$$

Neither w nor R^t are directly estimated. Instead for each $i = 1, \ldots, N$. We estimate w from a fixed number of households, say H. Let $e_{i,j}$ denote household j's expenditures on goods in the ith strata. Then

$$\hat{w}_i = \frac{\sum_{j=1}^{H} e_{i,j}}{\sum_{i=1}^{N}\sum_{j=1}^{H} e_{i,j}} \tag{35}$$

Similarly, let $\hat{R}_t$ be a sample estimate of R_t that has a non zero positive bias and variance. Then the bias in the upper level index can be decomposed as follows:

$$\{I^L(\hat{w},\hat{R}_t) - I^L(w,\hat{R}_t)\} + \{I(w,\hat{R}_t) - I^L(w,R_t)\} \tag{36}$$

and

$$\left\{\frac{I^T(\hat{w},\hat{R}_t)}{I^T(w,\hat{R}_t)}\right\}\left\{\frac{I^T(w,\hat{R}_t)}{I^T(w,R_t)}\right\} \tag{37}$$

The first term in these decompositions is the effect of sample bias and variance in the estimation of the spending weight, and the second term reflects the sample bias and variance effect in the relative. The weight effect for the Laspeyres is

$$I^L(\hat{w},\hat{R}_t) - I^L(w,\hat{R}_t) = \sum_{i=1}^{N}(\hat{w}_i - w_i)\hat{R}_{i,t} \tag{38}$$

and for the Törnqvist, the following second order Taylor expansion holds:

$$\frac{I^T(\hat{w},\hat{R}_t)}{I^T(w,\hat{R}_t)} \approx 1 + \sum_{i=1}^{N}(\hat{w}_i - w_i)\ln\hat{R}_{i,t} + 1/2\sum_{i=1}^{N}\sum_{k=1}^{N}(\hat{w}_i - w_i)(\hat{w}_k - w_k)\ln\hat{R}_{i,t}\ln\hat{R}_{k,t} \tag{39}$$

Using the fact that since both weights add to 1, the following holds

$$E(\hat{w}_i - w_i) = \sum_{i=1}^{N}(\widehat{w}_i - w_i) = 0 \tag{40}$$

If $cov((\widehat{w}_i - w_i), R_{i,t}) = 0$, then

$$E(I^L(\hat{w},\hat{R}_t) - I^L(w,\hat{R}_t)) = \sum_{i=1}^{N} E(\hat{w}_i - w_i)E(\hat{R}_{i,t}) = 0 \tag{41}$$

and

$$E\left(\frac{I^T(\hat{w},\hat{R}_t)}{I^T(w,\hat{R}_t)}\right) \approx 1 + \sum_{i=1}^{N} E(\hat{w}_i - w_i)E(\ln\hat{R}_{i,t})$$

$$+ 1/2\sum_{i=1}^{N}\sum_{k=1}^{N} E[(\hat{w}_i - w_i)(\hat{w}_k - w_k)]E(\ln\hat{R}_{i,t}\ln\hat{R}_{k,t}) = 0 \tag{42}$$

If $cov((\widehat{w}_i - w_i), R_{i,t}) \neq 0$, then using

$$\hat{w}_i = \frac{\sum_{j=1}^{H} e_{i,j}/H}{\sum_{i=1}^{N}\sum_{j=1}^{H} e_{i,j}/H} = \frac{\bar{e}_i}{\sum_{i=1}^{N}\bar{e}_i} \tag{43}$$

and using Cochran (1963), Chapter 6, I get

$$\hat{w}_i - w_i \approx \frac{w_i}{H}\left(\left[cv\left(\sum_{k=1}^{N} e_{k,j}\right)\right]^2 - corr\left(\sum_{k=1}^{N} e_{ik}, e_{ij}\right) cv\left(\sum_{k=1}^{N} e_{k,j}\right) cv(e_{i,j})\right) \tag{44}$$

where $cv()$ denotes coefficient of variation and $corr$ denotes correlation. If $cov(\sum_{k=1}^{N} e_{k,j}, e_{ij}) = var(e_i)$, then (44) reduces to:

$$\hat{w}_i - w_i \approx \frac{w_i}{H}\left(\left[cv\left(\sum_{k=1}^{N} e_{k,j}\right)\right]^2 - \frac{var(e_{i,j})}{\sqrt{var(e_{i,j})}\sqrt{var\left(\sum_{k=1}^{N} e_{k,j}\right)}} cv\left(\sum_{k=1}^{N} e_{k,j}\right) cv(e_{i,j})\right)$$

$$\hat{w}_i - w_i \approx \frac{w_i}{H}\left(\left[cv\left(\sum_{k=1}^{N} e_{k,j}\right)\right]^2 - \frac{\sqrt{var(e_{i,j})}}{\sqrt{var\left(\sum_{k=1}^{N} e_{k,j}\right)}} cv\left(\sum_{k=1}^{N} e_{k,j}\right) cv(e_{i,j})\right) \tag{45}$$

$[cv(\sum_{k=1}^{N} e_{k,j})]^2$ is $O(1)$, $\sqrt{var(\sum_{k=1}^{N} e_{k,j})}$ is $O(N^{-1/2})$ and the other terms are $O(1)$. If $O(w_i) = 1/N$ for all i, then $O(\hat{w}_i - w_i) = O(1/N)O(1/H)$. In the BLS sample, $N \approx 9,000$ and $H \approx 5,000$ then $O(\hat{w}_i - w_i) = 1/(NH) = 1/4500000$, and using (38) and (39),

$$I^L(\hat{w}, \hat{R}_t) - I^L(w, \hat{R}_t) = O(1/H)$$

and

$$\frac{I^T(\hat{w}, \hat{R}_t)}{I^T(w, \hat{R}_t)} \approx 1 + O(1/H) = 1 + O(1/5000)$$

5. References

Aizcorbe, A., Cage, R., and Jackman, P. (1996). Commodity Substitution Bias in the Laspeyres Indexes: Analysis Using CPI Source Data 1982–1994, U.S. Bureau of Labor Statistics Manuscript.

Biggeri, L. and Giommi, A. (1987). On the Accuracy Precision of the CPI, Methods and Applications to Evaluate the Influence of Sampling Households. Proceedings of the 46th Session of the ISI, Tokyo, IP-12.1, 137–154.

Billingsley, P. (1986). Probability and Measure. New York: John Wiley and Sons, 2nd Edition.

Boskin, M.J., Dulberger, E.R., Gordon, R.J., Griliches, Z., and Jorgensen, D. (1996). Toward a More Accurate Measure of the Cost of Living: Final Report to the Senate Finance Committee for the Advisory Commission to Study the Consumer Price Index. U.S. Senate Finance Committee.

Diewert, E.W. (1976). Exact Superlative Index Numbers. Journal of Econometrics, 115–145.

Cochran, W.G. (1963). Sampling Techniques. New York: John Wiley and Sons. 2nd Edition.

Dorfman, A., Leaver, S., and Lent, J. (1999). A Note on Relations among Price Indices and Their Estimators. U.S. Bureau of Labor Statistics Manuscript.

Greenlees, J. (1998). Random Errors and Superlative Indexes. U.S. Bureau of Labor Statistics Manuscript.

Hall, P. (1992). The Bootstrap and Edgeworth Expansion. Springer-Verlag.

Kish, L., Namboordi, N.K., and Pillai, R.K. (1962). The Ratio Bias in Surveys. Journal of the American Statistical Association, 57, 92–115.

McClelland, R. and Reinsdorf, M. (1997). Small Sample Bias in the Geometric Mean Seasoned CPI Component Indexes. U.S. Bureau of Labor Statistics Manuscript.

Shapiro, M.G. and Wilcox, D.W. (1997). Alternative Strategies for Aggregating Prices in the CPI. Review. Federal Reserve Bank of St. Louis, 113–125.

Received January 2000
Revised April 2001