ESTIMATING THE SAMPLING VARIANCE FOR ALTERNATIVE FORMS OF THE U. S. CONSUMER PRICE INDEX

Robert M. Baskin and Sylvia G. Leaver, U.S. Bureau of Labor Statistics
Robert M. Baskin ,USBLS, 2 Massachusetts Ave., N.E., Room 3655, Washington, D.C. 20212

Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.

KEY WORDS: Laspeyres, geometric mean, replicate, linearization, VPLX

The current functional form for the United States Consumer Price Index (CPI) is a Laspeyres index. The Bureau of Labor Statistics has been investigating alternative forms and is calculating a test series based on a geometric mean index. This report deals with estimating the sampling variance for this alternative form of the index.

This investigation is in two parts. First, a simulation is run to attempt to investigate the biases in different estimation methodologies. Second, variances for the test indices are estimated using available software.

In section one the estimators will be described. In section two the simulation is presented. The estimates for the test indices will be given in section three and conclusions and future directions will be presented in section four.

1. Introduction to the Form of the Index

For a full discussion of the CPI the reader is referred to Chapter 19 of the *BLS Handbook of Methods*, (1992). However, the following features of the CPI are important for the present discussion. According to the *Handbook*, p 176, "The CPI is a measure of the average change in the prices paid by urban consumers for a fixed market basket of goods and services." It is calculated monthly for the population of all urban families and also for the population of wage earners and clerical workers. The CPI is estimated for the total US urban population for all consumer items, but it is also estimated at other levels defined by geographic area and groups of items such as food, shelter, and transportation.

Pricing for the CPI is conducted in 88 PSUs in 85 geographic areas (New York area consists of 3 PSUs and the Los Angeles area consists of 2 PSUs). In the CPI area design there is random selection of PSUs according to a stratified design in which one PSU is selected from each stratum.

The CPI is a modified Laspeyres index, which is a ratio of the costs of purchasing a set of items of fixed quality and quantity in two different time periods. The index is estimated at the PSU level although not all PSUs are published. Let $IX_{it,s}$ denote the index at time t, in PSU i, relative to time period s. Then

$$IX_{it,s} = 100^{*}CW_{it} / CW_{is}$$

where CW_{it} and CW_{is} denote the aggregated weighted prices in PSU i for times t and s respectively.

The alternative form of the index being considered is a geometric mean index. This index is a ratio of modified cost weights which are the geometric mean of the weighted prices at the given time periods. If $IG_{it,s}$ denotes the alternative index, nicknamed the *geo-index*, at time t, in PSU i, relative to time period s, then

$$IG_{it,s} = 100^{*}GW_{it} / GW_{is}$$

where GW_{it} and GW_{is} denote the geometric means of weighted prices in PSU i for times t and s respectively. The weighted geometric mean of the time *t* prices is

$$GW_{it} = \prod_{j \in I_i} P_{ij,t}^{1/w_{ij}}$$

where P represents the price and w represents the sampling weight of item or unit *j*. It should be pointed out that the given form is algebraically equivalent to

$$GW_{it} = e^{\sum_{j \in I_i} 1/w_{ij} \log(P_{ij,t})}$$

This form is convenient for linearization and also shows that the geo-index is a function of price change at the item or unit level.

$$IG_{it,s} = \frac{GW_{it}}{GW_{is}} = e^{\sum_{j \in I_i} 1/w_{ij} \log(P_{ij,t}/P_{ij,s})}$$

For the properties of the different index forms the reader is referred to Diewert (1995). This paper will not deal with the relative merits of the different index forms but only attempt to investigate estimates of the sampling variance of the indices.

Previously, there has been much work on estimating sampling variance of the Laspeyres form of the CPI. Dippo and Wolter (1983) compared Taylor Series approximations to jackknifing. In a series of papers the Taylor series approach was used to estimate the sampling variance of the CPI in Leaver (1990), Leaver et. al. (1991), and Leaver and Swanson (1992). A large study of CPI variance was presented in Leaver and Valliant (1995). The current work builds on the work of the previous studies and is a beginning on estimating variances for the geo-index.

2. The Simulation

In order to investigate the bias of different forms of variance estimators, a computer simulation was run. An artificial population of values, considered either prices or rents, was generated for an initial period plus twelve consecutive time periods. The population index values were calculated for both the Laspeyres and geometric mean form.

Samples were generated from each of 100 populations 1000 times in order to simulate the sampling variance of the indices. The size of each population was 2000 items or units and a 10% sample was generated which would roughly correspond to monthly sample sizes for housing units in small to medium size PSUs.

In each generated sample, replicates were assigned and sample variance estimates were calculated by two methods. First, the index form was linearized and a variance estimate was calculated based on the replicates. Second a jackknife estimate was calculated The number of replicates generated was set at either 2, 4, or 8. In the current sample design the largest three PSUs have four replicates each, the other self-representing PSUs have two replicates each, and the lowest level published indices for the non-self-representing PSUs have four to eight replicates.

The simulations for the geo-index and for the Laspeyres index variance estimates behaved in an unremarkable way. The only thing of note was that the jack-knife and linearized estimates were closer for the geo-index than for the Laspeyres index. Both geo-index and Laspeyres index variance estimates were of similar magnitude and both were upwardly biased. This upward bias for the jackknife variance estimator has been observed in other empirical studies as described in Fay (1984). Table 1 gives the bias in the variance estimate, estimate - true value, as well as standard deviations of the variance estimates for each of the four methods. Graphs in Appendix 1 show the distribution of bias of estimates versus true value for each of the four methods. Each point in each graph is the bias of a variance estimate based on 1000 sample simulations. The pattern in the residuals is explained by the magnitude of the estimate and the fact that the estimates are positive.

Table 1. Empirical Biases and Standard Deviations of Variance Estimates in 1000 Samples of Size 200 from 100 Artificial Populations of 2000 Units.

Index	Method	Bias	Std. Dev
Geomean	Jackknife	0.000721	0.00396
Geomean	Linearize	0.000629	0.003909
Laspeyres	Jackknife	0.000772	0.004193
Laspeyres	Linearize	0.00061	0.003994

3. Test Indices on Production Data

The Bureau is currently producing a test geo-index. The Commodities and Services portion of the index has begun production but at this point there are not a sufficient number of months to study. The Housing portion of the index has produced five years of back indices from January 1991 to December of 1995. There are two item stratum producing Housing geo-indices. One index is for rent and the other is Owner's Equivalent Rent (OER). These test indices were used to estimate the sampling variance of geo-indices for Housing.

Because the test geo-index is not being run in our production environment a test Laspeyres index is being produced simultaneously with the geo-index for comparison. Estimates of variance for both the test geo-index and the test Laspeyres index were produced using a stratified random groups estimator implemented in VPLX. Details of this estimator are given in Leaver and Valliant (1995). For a description of the VPLX software see Fay (1990). Variances for the production Laspeyres index have also been produced for the same period of time.

Price change was calculated for both one and twelve month periods. Variances were estimated at the All-U.S. level and for each of the four Census regions.

Graphs of 12-month price change with two-standard error bands are presented in Appendix 2. Six-year price change estimates and standard errors are given in Table 2. The graphs show that the test indices are smoother and have smaller variances than the corresponding production index. The estimates of sampling variance for both test Laspeyres and test geo-indices are very similar.

Differences in price change estimates between the two test series and the production Laspeyres series and their standard errors were also estimated using the same VPLX methodology. Graphs of these differences in 12-month price change with two-standard error bands are also presented in Appendix 2. Differences in average 1-month price change and six-year price change and their standard errors are given in Table 3.

Table 2. Six-Year Price Change and Standard Errors for Rent and Owner's Equivalent Rent, U. S. All Cities CPI for Production Laspeyres, Test Laspeyres, and Test Geo-Index Estimators

Index Series	6-Year Percentage Price Change, 8912-9512	Standard Error
Production Laspeyres		
SE2101 Rent	17.8366	.4497
SE2201 OER	23.6049	.6694
Test Laspeyres		
SE2101 Rent	18.4308	.4011
SE2201 OER	20.2106	.5405
Test Geo-index		
SE2101 Rent	19.3718	.3815
SE2201 OER	20.6326	.5251

Table 3: Differences between Estimator Series and Their Standard Errors for Rent and Owners Equivalent Rent, U.S. All Cities Average CPI, December 1989-December 1995

Series Comparison and Item	Difference, Average 1-Month Price Change, 9001-9512	Standard Error	Difference, 6-Year Price Change	Standard Error
Test Laspeyres - Test Geo-index				
SE2101 Rent	-.0110	.0014	-.9410	.1147
SE2201 OER	-.0049	.0020	-.4220	.1720
Production Laspeyres - Test Laspeyres				
SE2101 Rent	-.0069	.0034	-.5942	.2923
SE2201 OER	.0389	.0042	3.394	.3736
Production Laspeyres - Test Geo-index				
SE2101 Rent	-.0179	.0037	-1.5352	.3159
SE2201 OER	.0340	.0042	2.9724	.3781

4. Findings

Table 2 shows some small but noticeable differences in variance estimates between the three series. The differences in standard error estimates between the test Laspeyres and test geo-index series are very small. The production series for both rent and owners equivalent rent are more variable than either test series. For owners' equivalent rent, this is largely attributable to the differences between the estimators prior to changes described in Henderson and Smedley (1994) which were applied to the production series in January 1995. These changes were applied to the test series for the entire six-year period. In the case of rent, it appears that the deletion of certain very low rents in both test series produced larger estimates of measured price change as well as dampened price change variability.

In terms of long term change, it is quite clear that the two test series are estimating different measures. The test Laspeyres estimator produced a lower price change measure than the test geo-index over the six year study. This difference is most remarkable in the rent series. Large differences in long term price change between production and test Laspeyres estimators in owners' equivalent rent are attributable to the formula difference discussed above.

The current research indicates the variance estimates for the geo-index behave similarly to the estimates for the Laspeyres index. There is no indication that the distinction in functional forms is producing a different estimate of variance.

The major task is to decide which type of variance estimator to use for the geo-index in a production environment.

5. Acknowledgments

The authors would like to thank Janet Williams and Richard Valliant for their careful reading of this paper and for helpful comments. The authors would also like to thank Jim Branscome, Shawn Jacobson, Frank Ptacek and Dave Swanson for insight into the sample design. Special thanks go to Bob Fay for providing the VPLX software. The authors would like to thank Janet Williams for support on this project.

6. References

Bureau of Labor Statistics (1992) *BLS Handbook of Methods*, Washington. DC: U.S. Government Printing Office, pp. 176-235.

Diewert, W.E. (1995), "Axiomatic and Economic Approaches to Elementary Price Indexes," University of British Columbia Discussion Paper No. 95-01.

Dippo, C. S., and Wolter, K. M. (1983), "A Comparison of Variance Estimators Using the Taylor Series Approximation," *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 113-119.

Fay, R. E., (1984), "Some Properties of Estimates of Variance Based on Replication Methods," *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 495-500.

Fay, R. E., (1990), "VPLX: Variance Estimates for Complex Samples," *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 266-271.

Henderson, Steve, and Smedley, Karin, (1994), "Improvements in Estimating the Shelter Indexes in the CPI," *CPI Detailed Report: Data for October 1994*, U.S. Bureau of Labor Statistics, Washington. DC: U.S. Government Printing Office, pp. 5-6.

Leaver, S. G., (1990), "Estimating Variances for the U.S. Consumer Price Index for 1978-1986," *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 290-295.

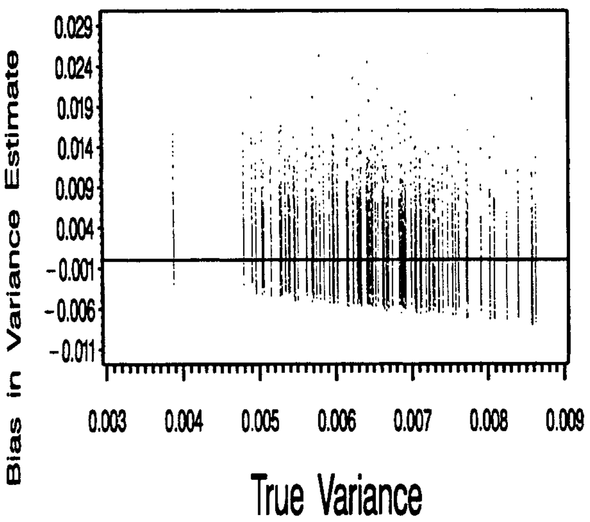
Leaver, S. G., Johnstone, J. E., and Archer, K. P. (1991), "Estimating Unconditional Variances for the U.S. Consumer Price Index for 1978-1986," *Proceedings of the Survey Research Methods Section*, American Statistical Association, pp. 614-619.

Leaver, S. G., and Swanson, D. C. (1992), "Estimating Variances for the U.S. Consumer Price Index for 1987-1991, "*Proceedings of the Survey Research Methods Section,* American Statistical Association, pp. 740-745.

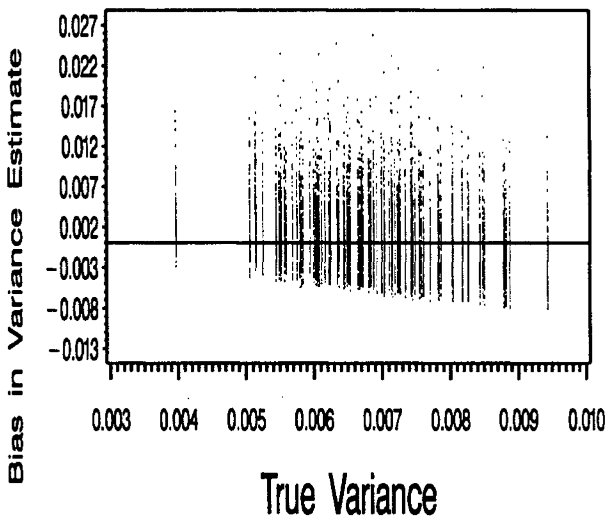
Leaver, S. G. and Valliant, R. L. (1995) "Chapter 28: Statistical Problems in Estimating the U.S. Consumer Price Index," *Business Survey Methods.* Wiley & Sons, Inc., 1995.

Appendix 1
Variance Estimates

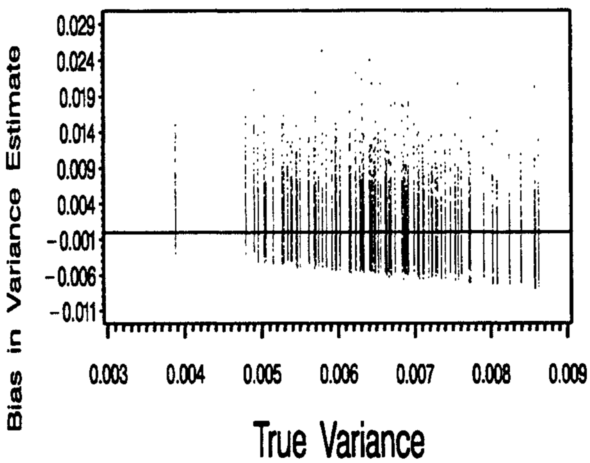
GeoMean Index
by Jackknife



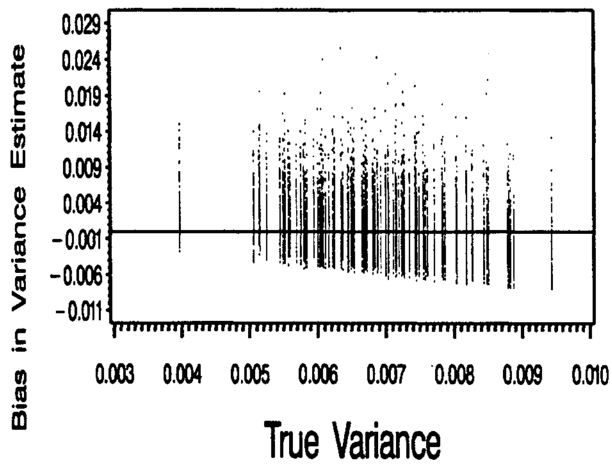
Laspeyres Index
by Jackknife



GeoMean Index
by Linearization

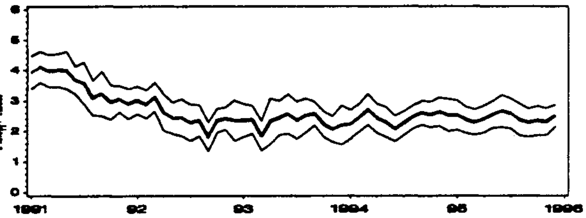


Laspeyres Index
by Linearization

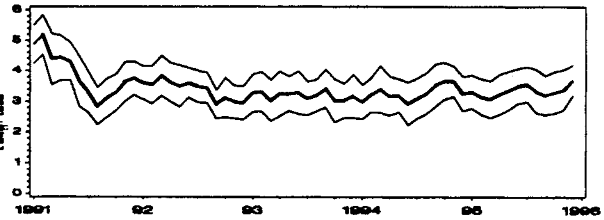


Appendix2: 12 – Month Price Change and Comparisons, 9101 – 9512
U.S. All Cities, Laspeyres and Geomeans Estimators, Rent and OER

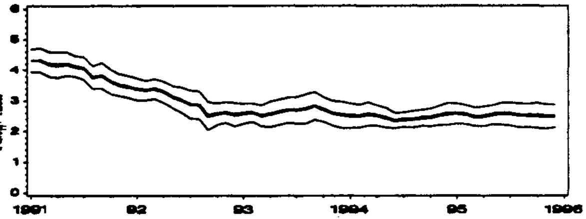
SE2101 Rent
Production Laspeyres



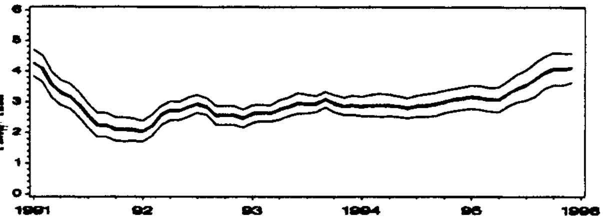
SE2201 Owners Equivalent Rent
Production Laspeyres



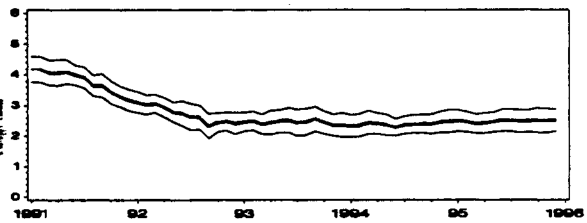
SE2101 Rent
Test Geomeans



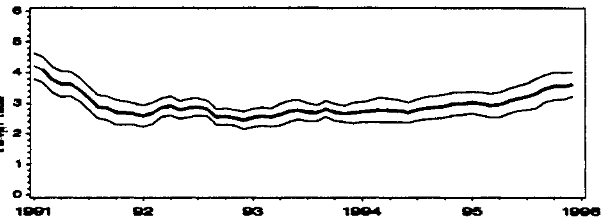
SE2201 Owners Equivalent Rent
Test Geomeans



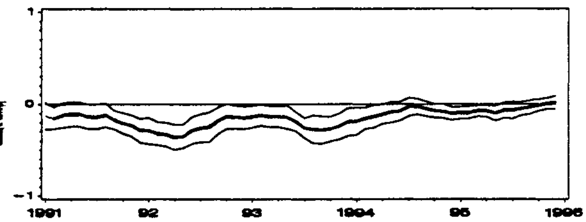
SE2101 Rent
Test Laspeyres



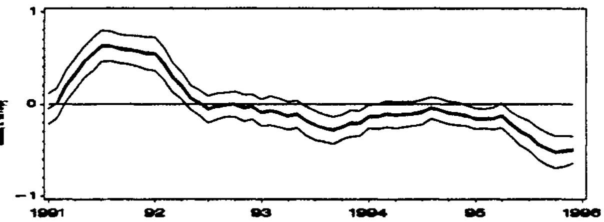
SE2201 Owners Equivalent Rent
Test Laspeyres



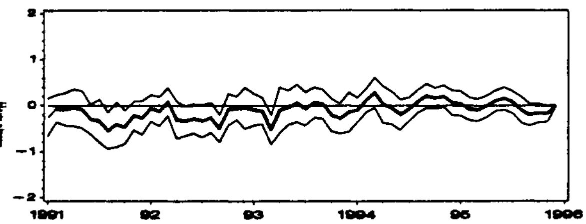
SE2101 Rent
Test Laspeyres – Test Geomeans



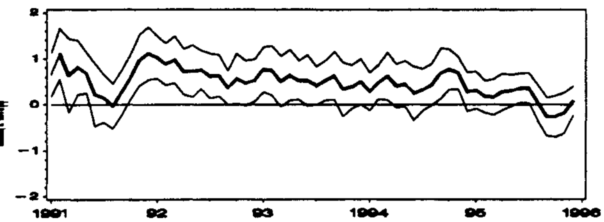
SE2201 Owners Equivalent Rent
Test Laspeyres – Test Geomeans



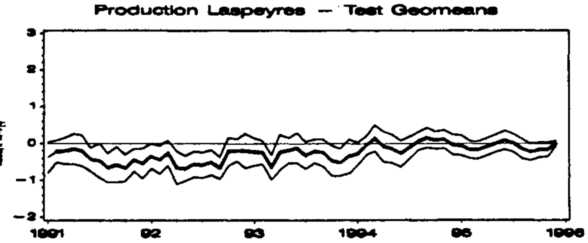
SE2101 Rent
Production Laspeyres – Test Laspeyres



SE2201 Owners Equivalent Rent
Production Laspeyres – Test Laspeyres



SE2101 Rent
Production Laspeyres – Test Geomeans



SE2201 Owners Equivalent Rent
Production Laspeyres – Test Geomeans