POTENTIAL BENEFITS FROM THE USE OF SCANNER DATA IN THE CONSTRUCTION OF THE CPI

Ralph Bradley, US Bureau of Labor Statistics
2 Masschusetts Ave. Washington, DC 20212

Key Words: Formula Bias, Laspeyres Index, Törnqvist Index

Introduction

Currently, the US Bureau of Labor Statistics(BLS) sends field representatives to manually collect price quotes when constructing the Consumer Price Index(CPI). Many of the nation's retailing outlets now have scanner check-outs that record their prices and expenditures. There is an opportunity for BLS to use these scanner records in the construction of its CPI. Some of the potential benefits from the use of scanner data are - the ability to construct superlative indices such as the Törnqvist, to reduce formula bias, to reduce the mean squared error of the CPI, to incorporate the introduction of new products on a more timely biases, to provide a more accurate sampling procedure within stores and to account for expenditure shifts across outlets. I discuss each potential benefit separately in this article.

In order to determine the effects of using scanner data I constructed various types of indices using the A.C. Nielsen Data Base. This database contained expenditure and unit sales information on four items - ketchup, tuna, toilet tissue, and milk in 10 different market areas. The time period that the data covers is from January 1992 to December of 1993. Additional information on product characteristics such as size, brand name is included.

Superlative Indices

The CPI has been criticized on the grounds that it is not a true Cost of Living Index(COL). Its Laspeyres specification assumes that the consumer will not substitute across items when their relative prices change. This induces the CPI to be greater than the true COL.(See Fisher and Shell [1972] or Pollack [1989].) The Tornqvist Index is an alternative index that allows some flexibility in the functional form of the consumer's preferences and incorporates the ability of consumers to substitute among items when relative prices change. Under the current manual system of collecting price quotes, the BLS is not able to construct a Tornqvist Index because the expenditure information that accompanies the price quote is not available. However, it is available with the use of scanner data. Table 1 compares the difference between a modified Laspeyres CPI that is used by BLS and a Tornqvist. These indices were constructed using price and expenditure information from the A.C. Nielsen Academic Data Base. While this is not a comprehensive data base it does demonstrate that the Laspeyres formula produces a larger index growth than the Tornqvist.

Formula Bias

Formula bias was discovered by Reinsdorf [1994]. When constructing the CPI, BLS uses expenditure weights that come from an earlier period than the period when the sample of prices is initiated. The purpose for this is to reduce the statistical dependency between the weights and the prices in the sample. Currently, BLS must impute these weights because the prices in the period when these expenditure weights were observed are not available under current collection procedure. Bradley[1995] and McClelland [1996] found that these imputations most often produced a positive bias. This bias is referred to as formula bias.

The BLS found a method to reduce formula bias, and is known as seasoning. Under the seasoning method, the price that is used to impute the weights comes from a period that is different from the period of the prices that are used to construct the index. Bradley[1995] compares the nonseasoned CPI with the seasoned CPI, using the A.C. Nielsen Academic Database and found that seasoning reduced the formula bias from .5% to .1%. Thus, the use of seasoning has mostly eliminated formula bias, and there would not be much additional benefit with the use of exact weights from scanner data.

Reduction in the Mean Squared Error of the CPI

Scanner data will produce larger samples and this alone will reduce sample variance. Currently, some BLS samples sizes are as small as four quotes. The use of scanner data could easily increase these sample sizes to more than 500 quotes. Additional reductions in the mean squared error can be achieved by the reduction in potential sources of bias. One source of potential bias comes with item attrition. There are times when the price of an item is not observable, even though it is still sold in the area. When this occurs, the CPI does not include a price for that item in the first month that its price is not observable, and if the item continues to be unobservable, procedures to find a substitute with the same outlet have been established. The bias can occur if the substitution method does not represent the consumer's behavior when an item is discontinued in a particular outlet.

Another opportunity to reduce the mean squared error, is the ability to compute accurate variances for each product area so that the sample can be allocated in a way that the mean squared error can be reduced. The current method of calculating variances relies on sample splitting (or replicates in BLS terminology). This splitting creates inefficiencies.(See Leaver and Valiant [1995] for the use of sample replicates in deriving variance estimates.) Bradley[1996] derives a method to calculate population variances when using scanner data. This method does not use sample replicates. He then derives an optimal sample allocation scheme based on these sample variances. The shortcoming of this sample allocation scheme is that is assumes constant variances across time. I ran simulations where I computed the current BLS CPI component based on a porportional probability sampling scheme and seperately, I computed CPI components based on the optimal allocation algoithm as derived in Bradley[1996]. Table 2 compares these simulation results. The results are shown for three market areas and four product categories. In all but one market the mean squared error has been reduced. The one exception has variances that are not constant over time. If variances are seasonal then the optimal sample allocation will need to also be seasonal.

The Ability to Identify New Products

When the video cassette recorder (VCR) became a common home appliance, BLS did not have the appropriate expenditure information to determine that it had become an important home appliance. Therefore, there was a lag between the time that the VCR became an important household appliance and the time that it was incorporated into the CPI.

The expenditure information from the scanner data bases should better enable BLS to determine when a new product becomes an important component of the consumer's budget, and this new product can be rotated into the samples on a more timely basis.

Greater Sampling Accuracy

Currently, when the field representative selects an item for the sample of quotes, the desired selection propability is proportional to the items share of expenditures within the outlet. The field representative obtains estimates of these expenditure shares from the outlet's management or by some other proxy. It is very difficult under current procedures to get reliable expenditure weights.

With scanner data, the desired expenditure weights can be derived with no estimation error. This will lead to a far superior sampling results, and will greatly reduce estimation error in the CPI

Ability to Account for Expenditure Shifts Across Outlets

Reinsdorf [1993] shows evidence that consumers substitute across outlets. While the Laspeyres specification assumes no substitution across products, it does not make such an assumption about outlet substitution. Scanner data will allow the BLS to monitor expenditure shifts across outlets on a more timely basis, and incorporate the price effects of these outlet shifts.

Bradley[1996] used the A.C. Nielsen Academic Database to simulate the construction of the components of the CPI using BLS current method of allowing no outlet substitution shift and a method where the representative individual pays the unit average price across all outlets for each item. Both simulations retained the BLS modified Laspeyres form. A summary of the results are outlined in Table 3.

It is evident that allowing for outlet substitution makes significant changes to the components of the CPI. However, the method that does not allow for outlet substitution will not automatically produce a higher index. Discount outlets do not necessarily guarantee lower prices on all items. It appears that they target their discounts to selected items.

Conclusion

While there are many potential benefits in the use of scanner data, there are also many challenges. The first challenge is that for the first time BLS will rely on private firms to supply its quotes. Thus, there is an additional delivery risk that these firms might not be able to deliver the necessary data on a timely basis. A second challenge is that a cross referencing system of product characteristics must be established so that items can be sampled from vendor data bases using BLS's current set of characteristics. A third challenge is that scanner data does not cover the universe of all items in the CPI. Therefore, the use of scanner data will need to be combined with the current manual methods of collecting price quotes. It is hoped that over time that the set of products that are recorded from scanner registers will increase.

TABLE 1.
A Comparison of thePopulation Laspeyres Indices with the Tornqvist Using 4 Items in 10 Areas

Period	Modified Laspeyres	Tornqvist
6/92-6/93	1.031	1.028
7/92-7/93	1.037	1.032
8/92-8/93	1.035	1.025
9/92-9/93	1.026	1.014
10/92-10/93	1.033	1.025
11/92-11/93	1.034	1.021
12/92-12/93	1.027	1.021

TABLE 2.
The Variances of the CPI Components using Current Sampling Methods and Optimal Sampling

Product-Market	Current Variance	Optimal Variance
Ketchup-01	.00289	.00235
Ketchup-02	.00222	.00205
Ketchup-03	.00412	.00396
Tuna-01	.00145	.00135
Tuna-02	.00106	.00098
Tuna-03	.00070	.00036
Toi. Tiss.-01	.00409	.00297
Toi.-Tiss-02	.00706	.00587
Toi.-Tiss-03	.00456	.00254
Milk-01	.00011	.00010
Milk-02	.00014	.00013
Milk-03	.00023	.00034

TABLE 3.
A Comparison of Index Components using the Fixed Outlet Weights and Unit Values Averaged Across Stores

Market Area-Product	Index 6/92-12/93
Market 01- Ketchup	
Fixed Weights	1.0475
Unit Values	1.0847
Market 02- Ketchup	
Fixed Weights	1.048
Unit Values	1.0212
Market 03- Ketchup	
Fixed Weights	0.9829
Unit Values	0.9797
Market 01- Milk	
Fixed Weights	1.0486
Unit Values	1.0494
Market 02- Milk	
Fixed Weights	0.9929
Unit Values	0.9806
Market 03- Milk	
Fixed Weights	1.1376
Unit Values	1.1345
Market 01- Toilet Tiss.	
Fixed Weights	1.0317
Unit Values	1.0495
Market 02- Toilet Tiss.	
Fixed Weights	1.0362
Unit Values	1.0427
Market 03- Toilet Tiss.	
Fixed Weights	0.9571
Unit Values	0.9813
Market 01- Tuna	
Fixed Weights	1.0339
Unit Values	1.0728
Market 02- Tuna	
Fixed Weights	1.0361
Unit Values	1.031
Market 03- Tuna	
Fixed Weights	0.9183
Unit Values	0.9166

The opinions expressed in this paper are my own and do not constitute policy of the U.S. Bureau of Labor Statistics. I would like to thank Brent Moulton. Marshall Reinsdorf and Bill Hawkes for useful comments.

References

Bradley R, “The Use of Scanner Data as a Means to Reduce the Mean Squared Error in the CPI” US BLS Mimeo, December 1995.

Bradley R. “Finite Sample Issues in the CPI” US BLS Mimeo, July 1996.

Fisher F.M. and Shell K., The Economic Theory of Price Indices, Academic Press, New York 1972

Leaver S. and Valiant R., “Statistical Problems in Estimating the U.S. Consumer Price Index”, Chapter 28 in Business Survey Methods, ed. Cox et. all, John Wiley and Sons, pages 543-566, 1995.

McClelland R., “Evaluating Formula Bias Indexes Using Simulations,” Draft Mimeo, U.S. Bureau of Labor Statistics, 1996.

Pollack R. A. The Theory of the Cost of Living Index, Oxford University Press, Oxford, 1989.

Reinsdorf M, “Price Dispersion, Seller Subsitution and the US CPI,” BLS Working Paper No. 252, 1994.

Reinsdorf, M., “The Effects of Outlet price Differentials on the U.S. Consumer Price Index.” In Foss, M.F. et al. ed. Price Measurement and the Uses Studies in Income and Wealth, Volume 57. Chicago: University of Chicago Press for the National Bureau of Economic Research, 1993, Pages 227-254.