Working to Improve the Consumer Price Index

Katharine G. Abraham, John S. Greenlees, and Brent R. Moulton

The recent report of the Advisory Commission appointed by the U.S. Senate Finance Committee begins with one overarching recommendation: "The BLS should establish a cost-of-living index (COLI) as its objective in measuring consumer prices" (U.S. Senate, 1996, p. 2). This seems basically right to us. Indeed, the BLS has long said that the cost-of-living framework guides operational decisions about the construction of the index (Gillingham, 1974; U.S. Bureau of Labor Statistics, 1984, 1988, 1992, 1997a). Putting things slightly differently, if the BLS staff or other technical experts knew how to produce a true cost-of-living index on a monthly production schedule, that would be what we would produce. Although we have no fundamental disagreement with the Commission about what the objective of the CPI program ought to be, we may disagree to some extent about the approaches and methods that are appropriate and feasible in working toward that objective. This is a point to which we will return below.

Our first purpose in the present paper is to comment on the Advisory Commission’s report and the recommendations it contains.1 We then will describe some of the initiatives currently underway at the BLS—some of which were undertaken

---

1 U.S. Bureau of Labor Statistics (1997b) provides a more comprehensive review of the report and its recommendations. This and other documents pertaining to the Consumer Price Index (CPI) are available on the Bureau’s Web site at (http://stats.bls.gov/cpihome.htm).

Katharine G. Abraham is Commissioner and John S. Greenlees is Assistant Commissioner for Consumer Prices and Price Indexes at the Bureau of Labor Statistics, U.S. Department of Labor, Washington, D.C. At the time this paper was written, Brent Moulton was Chief, Division of Price and Index Number Research, Bureau of Labor Statistics. He is now Associate Director for National Income, Expenditures, and Wealth Accounts, Bureau of Economic Analysis, U.S. Department of Commerce, Washington, D.C.
before the appearance of the Advisory Commission's report, others of which are part of a CPI improvement initiative that was included as part of the President's 1998 budget proposal—which attempted to address the bias issues that were highlighted by the Advisory Commission.

Substitution Bias

There is to our knowledge no disagreement among economists that, at least assuming stability of consumer preferences, an index such as the CPI that tracks the cost of purchasing a fixed market basket of goods and services represents an upper bound on the change in the cost of living. The BLS has attempted to explain this point to the general population of CPI users (for example, U.S. Bureau of Labor Statistics, 1983).

This is, however, a case in which what seems obvious to the average economist is not at all obvious to the average citizen. Many of the letters that we receive about the CPI focus on the substitution effect, and many of those who write find the whole concept confusing, or even offensive. One individual offered the following comments: "... your suggestion that switching from [butter to margarine] reduces inflation is flatly wrong. Were price increases... to lead people to live in cardboard boxes and eat gravel, that logic would lead to the statistical conclusion that there is no inflation at all." This isn't, of course, what economists who talk about the substitution effect have in mind. All else the same, an increase in any item's price will reduce consumer welfare, but the reduction will be less if consumers are able to substitute between items than if they are constrained to maintain their original pattern of consumption. That is, when prices rise, consumers generally won't need as much additional money to maintain their standard of living as they would to maintain their original consumption bundle. Substitution bias in the CPI consists, in essence, of the difference between the increase in the cost of the original consumption bundle and the (smaller) increase in expenditures required to maintain the original level of utility.

Operationally, as the Commission suggests, substitution bias in the CPI may show up at two levels. Lower-level substitution bias arises in the process of aggregating the roughly 80,000 price quotations collected each month to form a series of subindexes for categories (strata) of items such as "Apples," "Televisions," and "Prescription Drugs and Medical Supplies." Upper-level substitution bias occurs when those subindexes are aggregated to form the overall CPI. Neither the formula used to aggregate the individual price quotations to form the subindexes nor the formula used to aggregate the subindexes to form the overall CPI accounts for consumers' ability to substitute across items when the relative prices of those items change. Rather, the CPI tracks the cost of purchasing a fixed basket of goods and

---

2 More precisely, the subindexes are constructed within each of a set of CPI areas, and the subindexes are aggregated across item categories and areas to form the overall CPI.
services. Were the possibility of substitution taken into account, the CPI would rise less rapidly. The Commission estimates the combined magnitude of upper-level and lower-level substitution bias to be 0.4 percent per year.

To address the lower-level substitution problem, the Commission has suggested adoption of an alternative formula for aggregating price quotations that has been under investigation by the BLS over the past several years (Moulton, 1993; Moulton and Smedley, 1995; Reinsdorf and Moulton, 1997). The proposed alternative, termed the geometric mean formula, yields an exact measure of the change in the cost of living under the assumption of a unitary elasticity of substitution among items within a category, which implies that the share of consumer expenditures devoted to each item remains constant when relative prices change. Neither the assumption of no substitution underlying our current practice nor the assumption of a unitary elasticity of substitution underlying the geometric mean formula is likely to provide a close approximation in all cases.

It may be plausible to assume that consumers substitute freely between types of apples or between brands of television sets when their relative prices change, but not to assume similar substitutability between types of prescription drugs. Moreover, if the prices of all items within a category rise at about the same pace, as one might expect if the items are close substitutes for one another (for example, different brands of beer), just about any aggregation formula will produce pretty much the same result. Only if prices of the different items within an item category have diverged is the choice of aggregation formula of practical significance. Eventually, scanner data from retail establishments on the prices and quantities of the various items that are sold may be helpful for understanding the ongoing changes in the pattern of consumption within item categories, and thus for refining our method of within-stratum aggregation. Although the BLS has an active program of research in this area, the detailed data needed to measure lower-level substitution within most CPI item categories are simply not yet available. The BLS has begun issuing a monthly experimental measure that is constructed using the geometric mean formula in all index components, and will make a decision by the end of 1997, based on the existing incomplete evidence, as to which components of the official CPI should be constructed in that way.

The situation with respect to correcting for upper-level substitution is a bit different than that faced at the lower level of item aggregation. Because the BLS collects information on consumer expenditures for CPI item categories (through its Consumer Expenditure Survey), it is possible to construct a measure that accounts for substitution across those item categories in response to relative price changes, though not on the same schedule as the current CPI. Measures that ac-

---

8 The geometric mean formula would be applied to the items within a subindex for a particular item and area. The formula is

$$I_t = \pi(P_t/P_0)^{W_i},$$

where $I_t$ is the subindex in period $t$, $P_t$ and $P_0$ are the prices of good $i$ in periods $t$ and 0, and $W_i$ is the share of base period expenditures within the stratum represented by item $i$. 

count for substitution across items require both beginning-of-period and end-of-period information on both prices and expenditures. These measures, called superlative indexes, are based on the idea that observed changes in the pattern of consumers’ purchases provide information on their willingness to substitute among different items (Diewert, 1976, 1981). The BLS currently produces several superlative indexes on an experimental basis, and we have developed plans to produce them on a regular schedule, to a higher standard of precision and reliability.

Survey data on consumer expenditures, however, become available more slowly than data on prices. As a result, a true superlative measure can be produced only with a lag—at present, such a measure could not appear until the fall following the year to which it applies. Recognizing the unavoidable time lag in producing a true superlative measure, the Commission recommends that the BLS explore steps that might make the monthly CPI a better approximation to such an index. The BLS is open to this sort of approach, as can be seen in the variety of experimental measures we have published for some time, and we will continue such efforts. At this time, it is unclear what approach to approximating a superlative might be most satisfactory not only for the “all items” index, but also for the subaggregate indexes such as “food” or “transportation.” A price index based on statistical approximations would be more difficult to interpret and explain to users. Moreover, using an approximation would raise the issue of whether the index would need to be revised once the data to calculate the superlative measure became available. (The CPI historically has not been subject to revision, and this has been viewed as an important advantage by those adopting it for use in indexation provisions.) Were we to make such a fundamental change in the methods used to calculate the CPI without thorough testing and a substantial period of public consultation and education, we would fear damage to the credibility of our statistical system.

4 The specific recommendation contained in the Commission’s report was that the BLS adopt what the report refers to as a “trailing Tornqvist” measure, in which the stratum weights are updated annually and the geometric mean is used for cross-strata aggregation as well as within-stratum aggregation. As an empirical matter, this particular formula is likely to produce price changes that systematically understate the increases in the cost of living, as compared to a superlative index calculated using contemporaneous data (Aizcorbe, Cage and Jackman, 1996; Shapiro and Wilcox, 1997). More recently, other approximation strategies have been proposed, including a method based on the constant elasticity of substitution formula explored in Shapiro and Wilcox (1997).

5 Under current procedures, the CPI is revised after first publication only in extraordinary circumstances, namely the discovery that data were incorrectly reported or processed. Such a correction has not taken place at the U.S. all-items level since 1974. Other published government statistics, including the Producer Price Index and the National Income and Product Accounts, routinely are revised as additional data and enhanced methods become available. For that reason, some have questioned why the CPI is not similarly revised. The CPI’s use in legislated or negotiated cost-of-living adjustments is the key factor justifying the policy that the index is final when first published. To take a recent example, the fact that the CPI is not revised retroactively was one of the principal reasons for its selection by the Department of the Treasury as the index for its inflation-protected bonds, first issued in January 1997. The comments received from individuals, academicians, investment management firms, dealers and institutional investors in response to Treasury’s 1996 Advance Notice of Proposed Rulemaking “indicated a clear consensus that the selected index should be: recognized widely, published frequently, accurate, easily obtainable, easily understood, and not revised retroactively” (Federal Register, September 27, 1996, italics added).
Quality and New Goods Problems

The Advisory Commission believes that the failure to make adequate adjustment for changes in the quality of the goods and services people buy and to account properly for the value to consumers of newly available goods, together with deficiencies in the way the CPI treats differences in the prices charged at different retail outlets, creates an upward bias in the CPI of 0.7 percentage points per year.

The BLS does have procedures in place that are designed to purge the effects of changes in item quality that occur when a new item must be substituted for an item that is no longer available for pricing in the CPI. In the case of certain item categories (most notably automobiles and apparel), these procedures involve direct adjustment for the observed price differential associated with the differences in the old and new items' characteristics. For example, since 1992, direct adjustments made in the CPI for the changing quality of autos include adjustments for improved corrosion protection, improved warranties, sealing improvements, stainless steel exhaust, longer-life spark plugs, improved steering gears, rust-resistant fuel injection, clearcoat paint, and more. In other item categories, the procedures most commonly applied, which are referred to as "linking" procedures, remove any difference in price between the new item and the disappearing item at the changeover point from measured price change. In essence, these procedures assume that the difference in price between the items at the time the substitution occurs reflects the difference in their value to consumers.6

It is clear that the linking procedures do not work perfectly. In the case of personal computers and other "high tech" consumer goods affected by rapid innovation, they fail to capture the full value of quality improvements (Berndt, Grilliches and Rappaport, 1995). Still, absent evidence for a particular index component to the contrary, there is no a priori basis for concluding that linking necessarily leads to overstatement of the rate of growth of prices. Indeed, there may be some presumption in the opposite direction. As already noted, the linking approach to quality adjustment treats the difference in price between the original and the replacement item as wholly attributable to a difference in quality. Price increases for many goods occur intermittently and often are timed to coincide with model replacements or other quality improvements. By attributing the full price difference between the old and new items to a difference in quality, the linking approach may lead to overadjustment for quality change and a resulting understatement of the rate of price growth (Triplett, 1971). Indeed, direct quality adjustments were introduced for apparel items as evidence emerged that linking procedures had led to a significant understatement of price change in that item category (Armnecht and Weyback, 1989; Liegy, 1993, 1994; Reinsdorf, Liegy and Stewart, 1996). Methods have been introduced to try to minimize the possibility of overadjusting for quality

6 It is not correct to say that, in the application of these procedures, "No judgment at all is made about the quality differential between the new and old item," as do Boskin and his co-authors in their accompanying paper in this issue. Rather, there is an implicit judgment that quality differences can be inferred from market price differences.
differences in the linking process more generally, but we cannot say it has been eliminated.7

In addition to the procedures put in place to account for changes in items' quality, the BLS also has established procedures for bringing new items and new outlets into the index. The expenditure share information used to aggregate the CPI subindexes historically has been updated only once every ten years or so, but the specific stores in which prices are collected and the specific items priced are reselected on a five-year cycle. It may be possible by using scanner data or other approaches to identify new product introductions sooner and rotate them into the index more quickly (Bradley, Cook, Leaver and Moulton, 1997). Although more frequent sample rotations undoubtedly would be desirable, it is already true that a considerable share of the resources available for producing the CPI are devoted to ensuring that the sample of items priced is representative of what consumers actually are purchasing.

The Advisory Commission does not argue, of course, that the BLS does not attempt to address quality/new goods biases, but rather that, in spite of the BLS efforts, residual bias remains. The report's approach to assessing this residual bias is to divide the index into 27 categories, and then to make a judgment about the magnitude of the bias in each case. Unfortunately, the evidence applicable to many of these categories is rather sparse. In a number of cases the Commission draws conclusions from the existing evidence with which we do not agree.

In the food and beverages categories, for example, the Commission's estimates of upward biases rest exclusively on unsubstantiated judgments regarding the value to consumers of increased variety on grocery and liquor store shelves, together with the value of greater choice in restaurants. Moulton and Moses (1997) have examined evidence pertaining specifically to the value of the increased seasonal availability of fruits and vegetables; based on the magnitude of observed changes in consumption, they conclude that the Commission's estimates of that value are implausibly large. For new and used cars, the Commission estimates that the growth in prices has been overstated by 0.6 percentage point per year in the recent past, based on data showing that the average age of cars on the road has risen, together with an assumption that current CPI procedures do not capture any of the increases in automobile durability that may have occurred. This latter assumption, however, is incorrect; as noted earlier, a large number of direct quality adjustments related to auto durability have been made in the CPI over the past few years. For apparel, the Commission estimates that the CPI has overstated the rate of growth of prices by 1.0 percentage point per year since 1985, but this estimate rests on a comparison of the official CPI data with price indexes constructed using Sears catalogue prices for items remaining unchanged from one year to the next. We have serious reser-

7 Recent research by Moulton and Moses (1997) provides some interesting evidence on the combined impact of the various steps taken to adjust for quality change on the overall rate of growth of the CPI. As they document, the rate of growth of the CPI would be significantly larger were these procedures not applied. As they also make clear, however, these findings carry no direct implications concerning any quality-adjustment bias in the CPI.
vations about drawing any general conclusions based upon the prices charged by a single catalogue merchant, and are very skeptical of any index based only on the prices of unchanging items, particularly in a market segment where changing fashion is as important as it is in apparel.

To move beyond the particular examples, our reading of the Advisory Commission report is that the efforts made to identify possible positive biases in the CPI were considerably more systematic than those made to identify possible negative biases. Other analysts have pointed to reduced convenience and comfort of air travel, deteriorating quality of higher education, increases in travel time and driver irritation resulting from growing traffic congestion, and widespread declines in the quality of customer service as examples of quality decreases that are not accounted for in the CPI.

All of this, however, leaves us a long way from solving the quality/new goods problems the Commission believes exist with the CPI. There is much of what the Commission discusses that we do not know how to measure—or, to put it another way, for which economists simply have not developed operational procedures to correct the problems cited—and the Commission offers little in the way of practical assistance. Let us try to illustrate what we mean.

Surely, the variety of goods and services available to consumers has grown, and, surely, this variety is of value to consumers. Unfortunately, the techniques available for measuring the gains in consumer welfare from those new products (and the losses from product disappearances) are in their infancy, and may never be adaptable for implementation in a large, ongoing price measurement program like the CPI. To take another example, we would readily acknowledge that there have been major improvements in the medical treatment available for many serious health problems—improvements that have been of indubitable value to those suffering from the afflictions in question. Unfortunately, as a general matter, the BLS has no good way to measure the value of these improvements (say, increased mobility from improved knee surgery techniques).

Close to half of the quality/new goods bias the Commission believes exists in the overall CPI is judged to occur in just two areas of the index: medical care and high-tech consumer goods. These clearly are components of the index in which the BLS faces particularly difficult measurement problems. The Commission does not view ongoing BLS efforts in these areas as sufficient, but offers few concrete suggestions toward solutions to the difficult problems identified.

The report also discusses the question of new outlet bias, namely, how changes in the mix of retail outlets at which consumers shop ought to be treated. Current CPI procedures treat purchases of a particular item at different retail outlets as distinct transactions; the prices at the different stores are never directly compared. This could impart an upward bias to the CPI if, for example, stores offering lower prices but comparable service gained in market share. It also could impart a downward bias to the CPI if, for example, entry by low-priced outlets offering reduced services caused incumbent establishments to reduce their services in turn. As a practical matter, however, measurement of any such bias is complicated by the fact that different types of outlets commonly offer quite different shopping environ-
ments. Research on the factors affecting consumers' choices about where to shop ultimately may be helpful in devising appropriate procedures for dealing with changes in outlet mix.

For the BLS, the primary task is not to evaluate the bias estimates set forward by the Advisory Commission or other groups, but rather to employ the most accurate methods available for dealing with quality change and with new goods and outlets. Those methods must be rigorous, objective and reproducible, minimizing the role of analyst judgment. We recognize that there may be changes in consumer well-being associated with particular product or service innovations that are difficult or even impossible to capture using methods that meet these standards. We also believe, however, that adhering to the use of such methods is critical to maintaining the credibility of our official statistics.

**BLS Initiatives**

The Advisory Commission Report has performed a service by calling to the attention of policymakers the many and varied issues that the BLS faces in constructing the CPI. Indeed, the central argument of the report is that almost every assumption underlying the procedures used around the world for price index construction is called into question by the pace and form of market developments. The issues are not new to index number experts (many of the issues are discussed, for example, in the articles in the December 1993 Monthly Labor Review), but the quantitative and budgetary importance of price measurement problems and techniques have not always been appreciated by those who use the CPI.

The BLS has a vigorous program of research and development activities aimed at improving the CPI. Since 1994, for example, the BLS has revised its methods for making seasonal adjustments; introduced procedures to capture the effect of generic drugs in reducing prices; changed the ways in which shelter prices are estimated for renters and homeowners; identified and corrected a flaw in the weights attached to items entering the CPI sample; and reclassified the services provided in hospitals to capture patterns of treatment better.

The BLS currently is engaged in numerous CPI program enhancement efforts. The Bureau's evaluation of the geometric mean formula for use in constructing the CPI subindexes already has been mentioned; whatever decision is made about the geometric mean formula will be implemented with the data for January 1999. Some of the planned enhancements are part of the six-year CPI revision program now underway: introduction of a new CPI market basket with data for January 1998,

---

8 The Commission's report also contains several intermediate and longer run recommendations having to do with issues such as sampling, classification, and the conceptual issues underlying index construction. These recommendations have not received as much attention as the short run recommendations that deal directly with bias, but they represent thoughtful analysis by the Commission members of the measurement issues that the BLS faces in constructing the CPI. Space does not permit us to discuss these recommendations fully here; the interested reader is referred to U.S. Bureau of Labor Statistics (1997b).
based on 1993–95 consumer expenditure patterns; conversion to computer-assisted data collection and a telephone-based Point-of-Purchase Survey; improvements to the housing sample and estimator; and enhancements to the Consumer Expenditure Survey processing system. In addition, the BLS recently received the first year of funding for an expansion in the Consumer Expenditure Survey sample. This expansion will support a more timely CPI market basket, as well as the production of an official superlative index, more accurate than the current experimental indexes, to supplement the CPI.

The BLS also plans to make greater use of hedonic statistical techniques that account explicitly for changes in the characteristics of items being purchased. Starting in January 1998, the hedonic model developed and used in the Producer Price Index for adjusting personal computer prices will be employed in the Personal Computers and Peripheral Equipment category of the CPI. New research is underway on estimation of hedonic models using the CPI television sample (Moulton et al., 1997). New funding is expected to support collection of product price and characteristic data for samples of approximately 2,500 individual items each year, focused in the consumer durables categories. This supplementary information will enable the BLS to expand significantly the use of hedonic quality adjustment methods and to reduce its reliance on linking methods for categories in which new models are likely to show important quality change. New funding is also expected to support targeted collection of augmented item samples in dynamic market segments, with the goal of identifying and beginning to price new goods promptly once they appear in the marketplace.

Looking beyond the immediate future, the BLS expects to continue its ongoing research on the potential usefulness of price scanner data, which may be helpful for identification of new products and outlets, sampling of items, and ultimately in the computation of the CPI itself, as well as its research on the treatment of quality change, new goods and new outlets.

The BLS is intensely aware of the sensitive nature of the data it produces, and of the critical need for these data to be as accurate as possible. Although we believe that we can make important improvements in the CPI, we do not believe it to be possible to produce a perfect cost-of-living measure. It is, in fact, commonplace to observe that there is no single best measure of inflation. It is evident that the expanding number of users of the CPI have objectives and priorities that sometimes can come into conflict. The BLS response to this situation has been to develop a “family of indexes” approach, including experimental measures designed to provide information that furthers assessment of CPI measurement problems, or to focus on certain population subgroups, or to answer different questions from those answered by the CPI. All of these measures are carefully developed but have their own limitations. Those who use the data we produce should recognize these limitations and exercise judgment accordingly concerning whether and how the data ought to be used.

The authors would like to thank Timothy Taylor, Brad De Long, Alan Krueger, Kenneth Dalton, and Michael Boskin for their valuable suggestions on an earlier draft.


