

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: NBER Macroeconomics Annual 1996, Volume 11

Volume Author/Editor: Ben S. Bernanke and Julio J. Rotemberg, Editors

Volume Publisher: MIT Press

Volume ISBN: 0-262-02414-4

Volume URL: <http://www.nber.org/books/bern96-1>

Conference Date: March 8-9, 1996

Publication Date: January 1996

Chapter Title: Mismeasurement in the Consumer Price Index: An Evaluation

Chapter Author: Matthew D. Shapiro, David W. Wilcox

Chapter URL: <http://www.nber.org/chapters/c11028>

Chapter pages in book: (p. 93 - 154)

*Matthew D. Shapiro and David W. Wilcox*

UNIVERSITY OF MICHIGAN AND NBER; AND FEDERAL RESERVE BOARD

# *Mismeasurement in the Consumer Price Index: An Evaluation*

## *1. Introduction*

A number of analysts have claimed recently that the consumer price index overstates the true rate of increase of the cost of living. Our main objective in this paper is to undertake a detailed examination of the evidence on this claim. Where possible, we also report evidence on the variability of the bias in the CPI. The variability of the bias is of independent interest because—especially from the point of view of the monetary authority—any given average bias is more important the more variable it is. We also emphasize that estimates of the size of the bias are subject to uncertainty. We describe the extent of our uncertainty about each of the several components of the overall bias with a probability distribution, and we derive the distribution of the overall bias by aggregating these individual distributions.

A second objective is to present a new index for the price of cataract treatment. This index might serve as a prototype for an alternative approach to the pricing of medical care. Our results for this one course of treatment are not representative of what would be found in any comprehensive examination of the medical area; nonetheless, they do suggest that the overstatement of medical care inflation may be considerable.

A third objective is to discuss some of the implications of any bias in the

The opinions expressed herein are those of the authors alone, and may not represent the views of the Board of Governors of the Federal Reserve, or of the other members of its staff. We are grateful to Roland Bénabou, John Greenlees, Zvi Griliches, David Lebow, Brent Moulton, Marshall Reinsdorf, and David Stockton for helpful comments on earlier drafts, to Frank Diebold for very useful conversations, to Irving Shapiro, M.D., for assistance on the section on cataract surgery, and to Dwight Bibbs for excellent research assistance. Shapiro gratefully acknowledges support of the Alfred P. Sloan Foundation. Shapiro is Professor of Economics at the University of Michigan, Ann Arbor, MI 48109, and Research Associate at the National Bureau of Economic Research; Wilcox is Senior Economist at the Federal Reserve Board, Washington, D.C. 20551.

CPI, including implications for fiscal policy, monetary policy, and other measures of economic performance including real GDP and productivity.

A final objective is to suggest steps that we believe the Bureau of Labor Statistics (BLS) should consider as part of its ongoing efforts to improve the CPI.

We summarize our main conclusions as follows:

1. The CPI is one of the most carefully researched and best executed statistical programs in the United States. Many of the difficulties that have been the focus of public discussion recently lie at the frontiers of economic knowledge. Moreover, a very large fraction of the primary research concerning imperfections in the consumer price index has been conducted at the BLS, the agency that publishes the index. BLS personnel have been at the forefront of the effort to identify and quantify the influences that cause the CPI to be less than an ideal index. Over the years, the BLS has instituted a number of important improvements in the index based on this research.
2. Improving the index from its current state will not be easy. None of the problems still affecting the CPI is simple. Several of the remaining difficulties will require additional research before they can be addressed adequately. Even those cases in which the economics profession collectively “knows” in principle what to do may be resolvable only with a substantial commitment of additional resources.
3. There is enough evidence at this juncture to develop an informed opinion about the magnitude of the overall bias in the CPI. But, despite the efforts of the BLS and others, available evidence on the magnitude of several of the imperfections in the CPI is far less than complete. For this reason, we couch our statements about the size of the bias in the vocabulary of probability.
4. Based on our review of available evidence, we place the midpoint (median) of our subjective probability distribution for the overall bias in the CPI at just under 1.0 percentage point per year. We also estimate that about 80 percent of the mass of the distribution lies between 0.6 percentage point per year and 1.5 percentage points per year. Put slightly differently, we estimate that there is a 10 percent probability that the bias is less than 0.6 percentage point per year, and a 10 percent probability that it is greater than 1.5 percentage points per year.

Why is it important to have an assessment of the magnitude of the bias in the CPI? First, the CPI is the most widely followed measure of inflation. Users of all types, including members of the general public, policymakers,

and participants in financial markets, should have the best information available concerning the size of the bias.

Second, knowledge about the sources and magnitude of the bias could be important in guiding efforts to improve the index. Among other things, this type of knowledge is essential for judging the likely costs and benefits of investing additional resources in the index.

Third, the CPI has a substantial effect on the Federal budget. The Congressional Budget Office (1995) estimates that a permanent one-half-percentage-point reduction in the annual rate of growth of the CPI beginning in 1996 (holding all other elements of the economic environment constant) would reduce the Federal deficit in 2000 by \$26 billion relative to baseline projection, and the savings would escalate from there. This link between the CPI and the Federal budget has generated considerable political interest in the magnitude of the bias in the CPI.

The rest of the paper is organized as follows: Section 2 describes the methods used to construct the CPI in the United States. Section 3 provides a framework for the analysis of imperfections in the CPI. Section 4 reviews available evidence on the nature and magnitude of various imperfections in the CPI. Section 5 discusses our prototypical index for the price of cataract treatment. We both assess the shortcomings in the current official treatment of medical care prices and present a method for constructing a better index. Section 6 assesses some of the consequences of imperfections in the CPI as an index of the cost of living. Finally, Section 7 advances some suggestions about what might be done to improve the CPI.

## 2. *How the CPI is Constructed: A Brief Primer*

This section gives a thumbnail sketch of the methodology that the BLS uses to construct the CPI. Our goal is not to provide a comprehensive treatise, but rather to touch on the main features of the methodology that will be relevant for the discussion that follows. The primary source of information on this topic is Chapter 19 of the *Handbook of Methods* (Bureau of Labor Statistics, 1992).

### 2.1 PRICES

Each month, the BLS collects about 70,000 price quotations from roughly 21,000 outlets in 88 regions around the country known as primary sampling units (PSUs). In the five largest urban regions (comprising eight PSUs), prices are collected every month for all items; in the other regions, prices are collected monthly for food, fuels, and a few other items, and bimonthly for other items (Bureau of Labor Statistics, 1992, p. 178).

Separately, the BLS collects information from about 40,000 renters or landlords and about 20,000 homeowners for the housing components of the CPI (Abraham, 1995, p. 107). These individual price quotations are aggregated into the overall CPI in two stages.

In the first stage, individual price readings are aggregated into 9108 strata—one for each of 207 *items* in each of 44 *areas*. For example, prices at individual filling stations in the Chicago consolidated metropolitan statistical area are aggregated to form an index for the price of motor fuel in that area. Other examples of items at the stratum level include ground beef, women's dresses, new cars, physicians' services, and information-processing equipment. As these examples suggest, some strata (e.g., ground beef) are quite homogeneous, while others (e.g., physicians' services and information-processing equipment) decidedly are not.

Collectively, the 207 items are meant to provide exhaustive coverage of all consumer expenditures (treating owners' housing expenditures on a rental equivalence basis, and including only that portion of spending for medical care which is financed either out of pocket or by the portion of health insurance coverage paid for by individuals).<sup>1</sup> Of the 44 areas, 32 actually correspond to individual geographical locations in 29 cities, which are self-representing in the index on account of their size. The remaining 12 areas are composites constructed from the 56 primary sampling units which are not themselves areas. These 12 areas provide representation in the index for the smaller and mid-size cities in each of several regions of the country.

The modified Laspeyres formula for the first stage of aggregation is given by

$$\frac{P_{it}}{P_{il}} = \frac{\sum_j q_{jb} p_{jt}}{\sum_j q_{jb} p_{jl}}, \quad (1)$$

where  $P_{it}$  is the price index for item–area stratum  $i$  in period  $t$ ,  $p_{jt}$  is the price of individual item  $j$  in period  $t$ , and  $q_{jb}$  is an index of the quantity of item  $j$  purchased during a base period  $b$ . The time period  $l$  referenced in the denominator of both the left- and right-hand sides is the *link period*, the date when the weighting structure represented by the  $q$ 's is introduced into the index. In a true Laspeyres formula, the base period  $b$  would coincide with the link period  $l$ ; at this first stage of aggregation, the base period precedes the link period by about two years on average.

1. At present, only 184 of the item strata are actually priced; the other 23 strata, which collectively account for less than 2 percent of the weight of the overall index, are moved in line with the fluctuations of various priced strata.

We provide more information about the implementation of this formula in Section 4 below.

In the second stage, the item–area strata are aggregated into higher-level indexes (including the overall index) using another modified Laspeyres formula:

$$\frac{I_t}{I_L} = \frac{\sum_i Q_{iB} P_{it}}{\sum_i Q_{iB} P_{iL}}, \quad (2)$$

where  $I_t$  is a higher-level index,  $P_{it}$  is the price index for stratum  $i$  in period  $t$  from the left-hand side of equation (1), and  $Q_{iB}$  is the quantity of stratum  $i$  consumed in the base period  $B$ . Once again, the Laspeyres formula is modified rather than true because the second-stage base period  $B$  differs from the second-stage link period  $L$ .

## 2.2 SAMPLES AND WEIGHTS

An extensive array of sample-based information undergirds the calculation of the CPI. In brief, this information base can be described as follows: The quantities that are used in the second stage of aggregation are derived from the Consumer Expenditure Survey. This survey collects detailed information covering all out-of-pocket expenditures from a national sample of households.

Historically, the Bureau of Labor Statistics has updated these quantities (popularly known as “the marketbasket”) about once per decade. For current data, the weights used in the second stage of aggregation reflect an average of results derived from the surveys for 1982 through 1984; hence, the base period denoted as  $B$  above is 1982–1984. By contrast, the link period denoted as  $L$  above currently is the end of 1986. Therefore, the second-stage base and link periods are separated by roughly three years. In the next comprehensive revision of the index, due for introduction in 1998, the BLS will update the base period to 1993–1995 and the link period to the end of 1997.

The not-seasonally-adjusted CPI is revised only under extraordinary circumstances.<sup>2</sup> In particular, it has been the policy of the BLS not to revise the index backward in time when it updates the composition of the marketbasket. Thus, for example, the monthly values of the index from January 1978 through December 1986 reflect the average marketbasket as of 1972–1973, whereas the monthly values from January 1987 forward reflect the average marketbasket as of 1982–1984. The use

2. The CPI is seasonally adjusted at a very detailed level of disaggregation. Seasonal factors are revised annually.

of different marketbaskets at different points in time is another way that the CPI departs from the standard Laspeyres index.

The quantity indexes that are used in the first stage of aggregation (within strata) are derived from two sources. First, the Census Bureau conducts a survey of households (known as the Point of Purchase Survey, or POPS) to determine the distribution of household expenditures across specific outlets. Based on the results of this survey, the BLS selects a sample of outlets (including, say, a particular grocery store). The probability of selection for any given outlet is proportional to that outlet's share in total expenditures in the survey area for the item in question. Once the sample of outlets is drawn, a BLS representative visits each selected outlet and chooses one or more specific items (e.g., a particular brand of breakfast cereal) from within the broader category of items (all breakfast cereals) to be priced. The probability of selection for any given specific item is proportional to its estimated share in the outlet's revenue.

This process of outlet and item selection is part of the continuous sample-refreshment procedure known as *sample rotation*. This process generates a sample of specific items, each of which had a probability of selection into the sample proportional to its share in nominal expenditure during a base period. About 20 percent of all PSUs undergo sample rotation in any given year; accordingly, the first-stage base and link periods differ across PSUs. At present, all items within a PSU are rotated simultaneously. The BLS plans to change this aspect of the sample rotation procedure when it introduces the next comprehensive revision in 1998. Under the revised procedure, the BLS will rotate about 20 percent of all items in all CPI areas simultaneously. This modification will allow the BLS the flexibility, for example, to rotate more frequently those items undergoing a more rapid pace of technological change.

All items brought into the index through the sample rotation process are treated as not directly comparable to those already included in the sample; that is, the BLS performs an implicit quality adjustment of the prices coming into the sample using the *overlap method*. We describe the overlap method and the other quality-adjustment methods used by the BLS in the next subsection.

### 2.3 ITEM SUBSTITUTIONS

BLS representatives aim to reprice exactly the same items from month to month. According to Armknecht and Weyback (1989), however, this was not possible in 3.95 percent of all pricing attempts during 1984, because the previously priced item was either sold out, discontinued, or otherwise unavailable. In a few categories, the frequency of substitution was very high indeed. For example, Armknecht and Weyback report that the

substitution rates in 1984 for women's suits, women's dresses, and girls' coats, jackets, dresses, and suits were all in excess of 40 percent.<sup>3</sup> In most cases in which an old item cannot be repriced, the BLS representative substitutes a new one.

When an item is substituted into the index, the BLS representative must evaluate whether the new item is sufficiently comparable to the old one to make a direct comparison of prices meaningful. For this purpose, the BLS has developed an item-specific set of guidelines spelling out the essential characteristics that items must share if a direct price comparison is to be allowed; items may differ in other less consequential ways and still be judged "comparable" for item substitution purposes. Armknecht and Weyback (1989) report that in 1984, about 43 percent of substitute items were judged to have been comparable; according to Armknecht, Lane, and Stewart (1994), this fraction has risen to 56 percent more recently.

If a substitute item is determined to be noncomparable, then the BLS makes one of several adjustments to the price of the new item, depending on what information is available.

1. If both the old and new varieties of the item can be priced in the same month (say, month  $t$ ), then the BLS uses the "overlap method" (see Fixler, 1993, p. 7). In this approach, the growth of the index from period  $t - 1$  to period  $t$  is calculated using the price of the old item, whereas the growth of the index from period  $t$  to period  $t + 1$  is calculated using the price of the new item. In effect, the difference in price between the old and new varieties in the overlap month is taken as reflecting the difference in quality between the two varieties, to the exclusion of all other possible influences. Aside from its application as part of the sample rotation process, the overlap method is seldom used because the BLS rarely observes the prices of both the old and new varieties in the same month (precisely because the need for item substitution usually is triggered by the disappearance of the old item).
  2. In some categories of items, manufacturers are asked to provide estimates of the cost of producing a given quality improvement. This cost (marked up to an estimated retail value) is then netted out of the observed increase in price to produce an estimated quality-adjusted increase in price. The most prominent application of this approach is in the area of motor vehicles (Triplett, 1988, p. 39).
  3. The BLS also makes some limited use of hedonic techniques in constructing the CPI. The first application of such techniques in the CPI was in the area of housing; since 1988, hedonic equations have been
3. More recently, the BLS has taken a variety of steps to reduce the noncomparable substitution rates in apparel. See Reinsdorf, Liegey, and Stewart (1995).



used to adjust rent quotations for the age of the rental unit (see Randolph, 1988). More recently, the BLS has begun to use hedonic equations in the pricing of apparel.<sup>4</sup> Although at one time hedonic techniques were viewed as holding great promise for widespread application in the CPI, the current consensus appears to be much more cautious, and views hedonics as probably ill suited for extremely complicated items such as motor vehicles [see, for example, Gordon (1993) and Triplett (1988)].

4. *Link pricing* is used when the new and old versions cannot both be priced in the same month (again, see Fixler, 1993, p. 7). Suppose the price of the old item is last observed in period  $t - 1$  and the price of the new variety is first observed in period  $t$ . In this case, the growth of the index from period  $t - 1$  to period  $t$  is estimated using the prices of closely related items (excluding both the price of the old variety in period  $t - 1$  and the price of the new variety in period  $t$ ).<sup>5</sup> The growth of the index from period  $t$  to period  $t + 1$  (and thereafter) is computed using the price of the new variety. As is the case with the overlap method, link pricing involves an implicit quality adjustment; in this case, the adjustment is given by the difference between an imputed price of the old variety in period  $t$  and the price of the new variety in period  $t$ .<sup>6</sup>

The importance of these techniques is illustrated by figures presented in Armknecht and Weyback (1989) and Armknecht (1984). As was noted above, item substitutions were made in the course of 3.95 percent of all pricing attempts during 1984. The official CPI-U for the items studied by Armknecht and Weyback increased 3.4 percent during 1984.<sup>7</sup> Of this amount, 3.26 percentage points of price increase was derived from pricing attempts which involved a substitution. To put it slightly differently, the CPI-U for all repriceable items within the purview of the Armknecht–Weyback study increased 0.14 percent during 1984.<sup>8</sup> Results presented in

4. Hedonic methods are used to price information-processing equipment in the PPI, but not in the CPI.
5. The BLS has recently refined this technique as it is applied to nonservice items other than food, so that the price change from  $t - 1$  to  $t$  is imputed using only the results from other pricing attempts in which an item substitution also took place, but in which the new item was judged comparable to the old, or a direct quality adjustment was possible. We discuss the reasons for this change in Section 4.6.
6. The imputed price of the old variety in period  $t$  is calculated as the price of the old variety in period  $t - 1$  extrapolated forward using the growth of the subindex in question.
7. Armknecht and Weyback excluded residential rent, homeowners' equivalent rent, used cars, health insurance, and magazines, periodicals, and books from their study.
8. This outcome of near-zero average price change for repriceable items probably reflects a mix of behaviors, with some items experiencing normal price increases and others being marked down sharply prior to discontinuation.

Armknacht (1984) for 1983 are slightly less dramatic but still very striking: In 1983, item substitutions were performed in 3.85 percent of all pricing attempts. The CPI-U for all items included in the study (same exclusions as in Armknacht and Weyback) increased 2.99 percent; of that amount, 1.83 percentage points were contributed from item substitutions.

The fact that measured inflation is concentrated in newly introduced products demonstrates that item turnover is a fundamental aspect of the inflation process. Something quite dramatic on the pricing front happens when an old variety of an item disappears and a new one is introduced. That a substantial majority of aggregate price change coincides with changes in some characteristics of items seems not to be widely known, and is certainly worthy of further study.

Table 1 (adapted from BLS (1992)) provides a selective chronology of major changes in the consumer price index. Among other things, the table shows that the methodology underlying the CPI is frequently modified to reflect developments in the marketplace and advances in technique.

Table 1 A SELECTIVE CHRONOLOGY OF MODIFICATIONS TO THE CONSUMER PRICE INDEX

<i>Date</i>	<i>Action</i>
1953	Weights adjusted to reflect 1950 spending patterns.
1964	Weights adjusted to reflect 1960–1961 spending patterns of single persons as well as families.
1967	Quality adjustment introduced for new-car prices.
1978	Weights adjusted to reflect 1972–1973 spending patterns. CPI-U introduced. Point-of-purchase survey introduced as mechanism for selecting outlets. Probability sampling within each outlet introduced as the method for selecting specific items.
1983	Rental equivalence introduced as concept for measuring homeowners' costs in CPI-U.
1985	Rental equivalence introduced as concept for measuring homeowners' costs in CPI-W, which is the index used as the escalator for social security benefits.
1987	Weights adjusted to reflect 1982–1984 spending patterns. Quality adjustment introduced for used-car prices.
1988	Depreciation adjustment for housing introduced.
1991	Use of hedonics for direct quality adjustment of apparel items introduced.
1992	Procedures for pricing of air fares modified to allow pricing of discount fares. Also, use of specialized subsample of items for imputing price change for substitutions.
1995	New method for pricing generic drugs introduced (see Section 4.4). Price "seasoning" introduced as method for improving the treatment of food purchased for consumption at home (see Section 4.3). Housing estimator changed from average-of-ratios to ratio-of-sums (see Section 4.3).

Table 2 MEASUREMENT PROBLEMS IN THE CONSUMER PRICE INDEX

- 
1. Problems associated with aggregation and choice of utility concept:
    - a. Across strata.
    - b. Within strata.
  2. Problems associated with maintaining the representativeness of the CPI sample:
    - a. New items.
    - b. New outlets.
  3. Problems associated with measurement of individual prices included in the CPI sample: quality change.
- 

### 3. *A Framework*

This section proposes a conceptual framework for the analysis of measurement problems in the CPI.<sup>9</sup> This framework is intended to represent an exhaustive and mutually exclusive organizational structure for problems with the CPI as currently defined.<sup>10</sup> We use the framework as a roadmap for our discussion below of the various imperfections in the CPI as a measure of the cost of living.

Table 2 presents our framework in schematic form. We divide the universe of possible problems with the CPI into several categories. The first category of issues pertains to the problem of aggregating individual prices and subindexes into the overall index. In economic terms, these issues correspond to the choice of a particular formula for the purpose of

9. Bryan and Cecchetti (1993) propose a similar framework.
10. There is a broader set of questions that we do not address in this paper pertaining to the overall design of the index. For example, should the index attempt to measure how much a representative consumer would have to *spend* in the current period in order to be as well off as she was in some base period, or should it attempt to measure how much she would have to *receive in income*? The difference is driven by direct taxes. Gillingham and Greenlees (1987) note that an expenditure-defined index (such as the current official CPI) will increase in response to a revenue-neutral swap of indirect taxes for direct taxes; this might be a matter of some concern, given that some plans currently being discussed in the political arena would entail such a swap. A second design-related issue concerns the coverage of medical care. If the CPI is intended to serve as a comprehensive index of the cost of living, then it should price all of medical care consumed, whether financed by employer-paid insurance or not. On the other hand, if the index is intended primarily to serve as an escalator for social security benefits, then it makes sense to follow current practice in excluding government-provided health care, although in this case the marketbasket and item selection presumably should be specifically tailored to the beneficiary population. Moreover, if the index were to be optimized for this purpose, and the objective of the Congress was to provide a benefit with constant purchasing power, then the index probably would ideally be reconstituted as an income-defined index with tax treatment targeted specifically at taxation of social security benefits.

aggregation, and the accuracy of the resulting index as an estimate of the true cost-of-living index.<sup>11</sup> The structure of the CPI makes it natural to discuss these issues at two different levels—a relatively aggregated level (e.g., food vs. apparel, medical care vs. automobiles) and a relatively disaggregated level (e.g., corn flakes vs. granola). We refer to these two effects as the “across-strata” and “within-strata” effects respectively.

The second category of issues pertains to the problem of maintaining the representativeness of the CPI sample. Because new items are constantly being offered to consumers for the first time, and new outlets are opening, the BLS must continually refresh the CPI sample in order to keep it representative of the transactions actually taking place in the economy. The issues we group under this heading stem indirectly from the BLS’s methodology for implementing this constant renewal of the sample. We refer to these issues as the new-items effect and the new-outlets effect.

The last group of issues pertains to the measurement of the prices of individual items that are included in the sample. Here, by far the most important issue has to do with quality change that is either undetected (and hence not controlled for) or detected but not handled accurately. We refer to this issue using the label “quality-change effect.” With this label, we mean to refer only to that element of quality change that is not already taken into account by the BLS’s procedures.

#### 4. *The Plumbing of Mismeasurement*

This section reviews available evidence on the sources and magnitudes of various imperfections in the CPI. We begin by describing our method for aggregating estimated magnitudes of imperfection across sources. Then we consider the imperfections themselves.

##### 4.1 AGGREGATION OF RESULTS

One way to describe the magnitude of a particular imperfection is to give a point estimate. A point estimate may be a good way of conveying a best estimate (or a conservative estimate) of the magnitude of a particular bias, but it is not a good way of describing the extent of one’s uncertainty about that estimate. Previous authors in this genre [e.g., Advisory Commission

11. Our focus on utility as the organizing concept for the CPI places us in the tradition of Fisher and Shell (1972), Pollak (1989), and many others who have analyzed price-index theory from the economic perspective. The competing paradigm is represented most prominently by Fisher (1922), and is based on the specification of axioms that a well-designed index should possess. [See Diewert (1987) for a modern treatment of the axiomatic approach.]

(1995) and Lebow, Roberts, and Stockton (1994)] have attempted to describe their uncertainty by specifying ranges. While ranges convey that there is uncertainty, they are not informative about whether the probability mass inside the range is concentrated (and if so, where), or about whether there is any probability mass outside the range. Moreover, ranges do not convey sufficient information to allow rigorous aggregation of magnitudes across different sources of imperfection.

We address these problems by presenting our results explicitly in terms of subjective probability distributions. Because we use numerical rather than analytical techniques to aggregate across the various sources of imperfection, we have considerable flexibility in the specification of our views. In particular, we are not restricted to the normal distribution; nor are we constrained to assume that the various effects are uncorrelated with one another.<sup>12</sup> A possible shortcoming of our approach is that it requires us to be very specific about the nature of our beliefs. We might prefer a technique that allowed us to be somewhat “fuzzy” in the specification of our beliefs, but we know of no such technique.

We now proceed to consider the various major sources of bias in the CPI.

#### 4.2 THE ACROSS-STRATA EFFECT

As relative prices change over time, consumers will generally find that the cost-minimizing strategy for achieving a given level of utility requires them to change the mix of their purchases. Other things equal (in particular, holding tastes and real incomes constant), consumers will tend to purchase less of items whose relative prices have increased and more of items whose relative prices have declined. A Laspeyres index ignores such shifts. In contrast, a Tornqvist index provides a second-order approximation to the true cost-of-living index, provided utility is homothetic (see Diewert, 1976), and so should be approximately free of any influence from the across-strata effect.

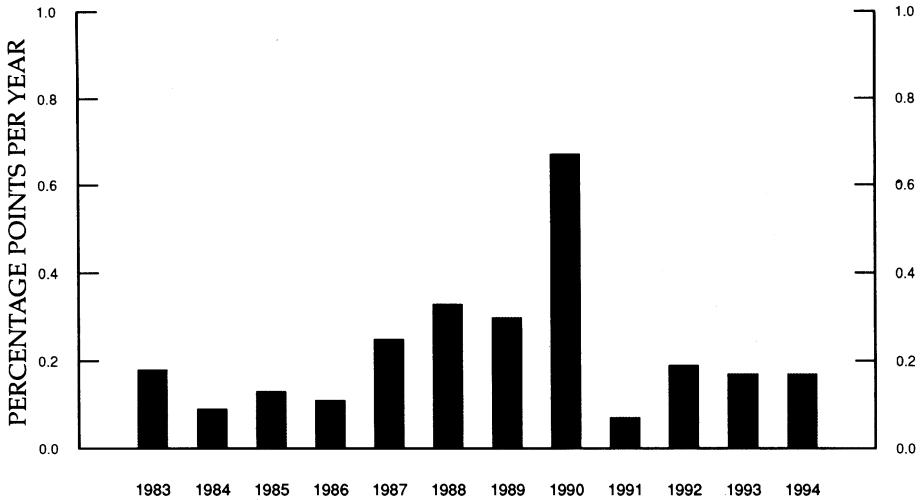
Aizcorbe and Jackman (1993 and updates) calculate annual differences in the rates of growth of a 1982-based Laspeyres index and a Tornqvist index.<sup>13</sup> Their results are shown in Figure 1.<sup>14</sup>

12. See Shapiro and Wilcox (1996) for a further description of our technique for computing the distribution of the aggregate bias. We are grateful to Frank Diebold for helpful suggestions in developing this method for summarizing our results. Stockton (1995) uses the vocabulary of probability to discuss in an informal manner how beliefs about the overall bias in the CPI could be expressed and interpreted.

13. A Tornqvist index calculates aggregate price change as a weighted geometric mean, where the weights are the arithmetic averages of the expenditure shares in the base and comparison periods.

14. At present, it is not possible to judge formally the statistical significance of the Aizcorbe–Jackman estimates, because associated sampling variances and autocovariances are not available. Such information would be useful and interesting.

Figure 1 THE AIZCORBE–JACKMAN ESTIMATES OF THE ACROSS-STRATA EFFECT



Source: Aizcorbe and Jackman (1993 and updates).

A striking feature of these estimates is that they fail to show an easily recognizable upward trend despite a widely held presumption that there should be such a trend.<sup>15</sup> This finding merits further research, because it casts doubt on the idea that the across-strata effect is likely to be especially large until the end of 1997—when the updated marketbasket is scheduled to be introduced—and on the idea that the introduction of the new marketbasket will do much to attenuate the size of the across-strata effect. One possible explanation of this finding is that it reflects a slow-down in the rate of drift of relative prices away from their base-period values (and hence a diminished scope for cost-reducing substitution) during the 1990s as compared with the 1980s. We investigated this hypothesis by constructing the following index of the cumulative drift of relative prices from the base period:

$$J_t = \sum_i w_i (\ln p_{it} - \ln \tilde{p}_i)^2,$$

15. The presumption that there should be such a trend derives from the observation that if the elasticity of substitution is greater than zero, a Laspeyres index calculated using period  $b$  as the base period will assign a lower weight to items whose relative prices have declined between period  $b$  and period  $t - 1$  than will a Laspeyres index calculated using period  $t - 1$  as the base period. If changes in relative prices are increasing, the Laspeyres index with fixed base year will therefore increasingly underweight the price changes of items whose prices are growing more slowly than the average.

where the  $w_i$ 's are nominal expenditure shares in 1982–1984 (taken from Mason and Butler, 1987), the  $p_{it}$ 's are the most detailed national-level indexes available on the BLS's public database, and  $\bar{P}_t$  is an aggregate price index calculated as the weighted geometric mean of the  $p_{it}$ 's. We then computed the 12-month changes in  $J_t$  and found, to our surprise, that they have essentially no explanatory power for the Aizcorbe–Jackman estimates.<sup>16</sup>

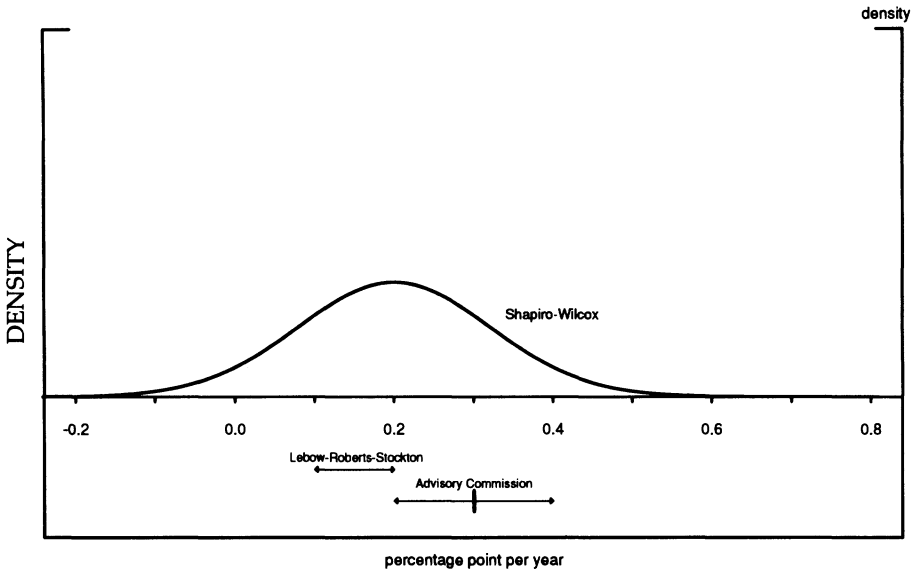
*Empirical Magnitudes* In their interim report, the Advisory Commission (1995) gave a point estimate for the average influence of the across-strata effect over the next decade of 0.3 percentage point per year, with a range extending from 0.2 to 0.4 percentage point per year. Lebow, Roberts, and Stockton (1994) give a range extending from 0.1 to 0.2 percentage point per year.

We assess the available evidence as suggesting that the average influence of the across-strata effect over the next decade or so is centered roughly around 0.2 percentage point per year (see Figure 1). Based on economic theory and available evidence, we are fairly confident that the substitution effect will be positive on average over the next decade or so. However, we do assign a low probability to the possibility that the across-strata effect will cause the CPI over the next decade to *understate* the rate of increase of the true cost-of-living index; this would occur if relative prices were to drift back toward their base-period (1982–1984) values between now and the introduction of updated weights in 1998.<sup>17</sup>

We summarize these beliefs using a random variable that is distributed according to the normal distribution with mean 0.2 percentage point per year and a 90 percent confidence interval extending from 0.0 to 0.4 percentage point. (We defer specification of correlation with other influences until those other influences are introduced.) Figure 2 displays our probability distribution along with the information provided by the Advisory Commission and by Lebow, Roberts, and Stockton.

16. In part, these results may reflect an idiosyncrasy in the weights used by Aizcorbe and Jackman in their published work. A separate unpublished table from the BLS shows the growth of a Laspeyres-type series calculated using weights defined over the three-year period 1982–1984 (conformable with the official CPI). The substitution bias in this series is somewhat greater overall, and somewhat smoother from year to year. If anything, however, it shows even less evidence of acceleration through the sample period.
17. For example, relative prices would be driven back toward their base-period values if oil prices were to increase substantially from their current levels.

Figure 2 ACROSS-STRATA EFFECT



Note: This graph shows the probability distribution or ranges for bias in component of the CPI. The Shapiro-Wilcox probability distribution is based on evidence discussed in the text. The horizontal lines represent the ranges of Lebow, Roberts, and Stockton and the Advisory Commission. The vertical line in the Advisory Commission's range represents its point estimate.

#### 4.3 THE WITHIN-STRATA EFFECT

As we noted in Section 2, the index that the BLS aims to construct as a measure of price trends at the stratum level is given by equation (1), which we repeat here for convenience:<sup>18</sup>

$$\frac{P_{it}}{P_{il}} = \frac{\sum_j q_{jb} p_{jt}}{\sum_j q_{jb} p_{jl}}. \quad (1)$$

The BLS observes nominal expenditures rather than real quantities; therefore, it is useful to rewrite equation (1) as follows:

$$\frac{P_{it}}{P_{il}} = \frac{\sum_j (\omega_{jb}/p_{jb}) p_{jt}}{\sum_j (\omega_{jb}/p_{jb}) p_{jl}}, \quad (3)$$

18. The discussion in this section draws heavily on Moulton (1993), Reinsdorf and Moulton (1996), and Reinsdorf (1996).



where

$$\omega_{jb} = \frac{e_{jb}}{\sum_j e_{jb}},$$

where  $e_{jb}$  is the nominal expenditure on item  $j$  during the base period, and  $\omega_{jb}$  therefore is the share of spending on item  $j$  in total expenditures during the base period  $b$ .

The BLS does not observe the base-period price  $p_{jb}$ . Since 1978, the BLS has been imputing a value for this variable using the following formula:

$$\hat{p}_{jb} = \frac{p_{jl}}{P_{il}/P_{ib}}. \quad (4)$$

As a result [as one can show by substituting (4) into (3)], the BLS since 1978 effectively has been calculating the price index for the  $i$ th stratum according to

$$\frac{P_{it}}{P_{il}} = \sum_j \omega_{jb} \frac{p_{jt}}{p_{jl}}. \quad (5)$$

Note that a true Laspeyres index would weight the price relatives in equation (5) using expenditure shares from the link period  $l$  rather than the base period  $b$ .<sup>19</sup>

If the objective is to construct the best possible approximation to the true cost-of-living index, the optimal form of the estimator for the stratum price index depends on the unknown utility function. In general, use of equation (5) as the estimator for the stratum price index will result in some bias. Given assumptions about the form of the utility function and the behavior of relative prices, one can estimate the size of this bias. For example, suppose that there exists a representative consumer and that the utility function of this consumer belongs to the constant-elasticity-of-substitution (CES) family with elasticity of substitution equal to  $\eta$ . Suppose also that the relative prices of individual items are given by

$$\ln p_{jt} = \epsilon_{jt}, \quad (6)$$

19. In practice, the BLS selects a random sample of items from the population of items indexed by  $j$ . The sampling probabilities are proportional to the expenditure shares  $\omega_{jb}$ . In the actual computation of the stratum indexes, the right-hand side of (5) is calculated as the unweighted sum over the items in the sample of the price relatives  $p_{jt}/p_{jl}$ . In this section we analyze the population analogue of (5), so the weights  $\omega_{jb}$  should be interpreted as sampling probabilities.

where  $\epsilon_{jt}$  follows a stationary process with normal innovations, unconditional variance  $\sigma_\epsilon^2$ ,

$$\text{corr}(\epsilon_{jt}, \epsilon_{js}) = \rho_{t-s},$$

and

$$\text{corr}(\epsilon_{js}, \epsilon_{kt}) = 0 \quad \text{for all } j \neq k \text{ and all } s \text{ and } t.$$

Then the probability limit of the true cost-of-living index is 1 (where the limit is taken with respect to the number of items priced in the stratum).<sup>20</sup> One can also show that

$$\text{Plim} \left( \frac{P_{it}}{P_{il}} \right) = \exp\{\sigma_\epsilon^2 [1 + (1 - \eta)(\rho_{t-b} - \rho_{t-b}) - \rho_{t-l}]\} \quad (7)$$

In most categories, the base-period relative price probably exhibits little or no correlation with either the link-period relative price or the current-period relative price, both because the average time span between the base and link periods is about two years, and because the base period for many items is a year or longer (so the base-period price is an average over an extended period of time). If  $\rho_{t-b}$  and  $\rho_{t-b}$  are small, equation (7) simplifies to

$$\text{Plim} \left( \frac{P_{it}}{P_{il}} \right) \approx \exp\{\sigma_\epsilon^2 [1 - \rho_{t-l}]\}. \quad (8)$$

Because the probability limit of the true cost-of-living index is 1, equation (8) can be interpreted as giving the approximate asymptotic bias inherent in the use of equation (5) as the estimator for the stratum subcomponent of the true cost-of-living index.

Previous authors in this literature (including us, in an earlier draft of this paper) have attempted to decompose the total bias in the circa-1995 stratum-level estimator used by the BLS into a substitution-related component and a “formula”-related component. Yet equation (8) implies that this bias was approximately invariant with respect to the elasticity of

20. More generally, if prices within the stratum are stationary around a common trend, one can show that the probability limit of the cost-of-living index is the common trend. See Moulton (1993) and Reinsdorf (1996).

substitution.<sup>21</sup> This fact suggests to us that efforts to decompose the bias in equation (8) probably have not been very productive. In this regard, we emphasize that equation (8) subsumes both within-stratum substitution and so-called formula bias.

In response to concern about the bias in the estimator given in equation (5), the BLS has taken a number of steps. First, in January 1995, the BLS implemented an improved procedure for the imputation of rent change for the owners'-equivalent-rent component of the CPI (see Armknecht, Moulton, and Stewart, 1995).

Second, also in January 1995, the BLS modified its treatment of the prices of food items purchased for consumption at home. Specifically, it began imputing the base prices in that category using a price reading from some month between  $b$  and  $l$  (call it  $n$  for "intermediate"). With this modification, the aggregation formula at the stratum level within this category began to read as follows:

$$\frac{S_{it}}{S_{il}} = \frac{\sum_j (\omega_{jb}/p_{jn}) p_{jt}}{\sum_j (\omega_{jb}/p_{jn}) p_{jl}}. \quad (9)$$

Under the same assumptions as were given above, the bias in this "seasoned" version of the CPI (so called because one can think of the BLS as taking one price reading in period  $n$  and then setting the item aside for a few months to let it "season" before bringing it into the index in period  $l$ ) is given by

$$Plim \left( \frac{S_{it}}{S_{il}} \right) \approx \exp\{\sigma_\epsilon^2 [\rho_{l-n} - \rho_{t-n}]\}. \quad (10)$$

If  $\rho_{l-n}$  (the autocorrelation of the relative price between periods  $l$  and  $n$ ) is small, this estimator should provide quite an accurate estimate of the rate of increase in the true cost-of-living subindex, regardless of the elasticity of substitution.

Third, in March 1996, the BLS announced that this approach to base-price imputation will be extended to all other items in the index starting in mid-1996.

Fourth, also in March 1996, the BLS announced a modification of its procedure for item substitution. "Except in rare and extreme cases," any item brought into the index as a substitute and determined not to be

21. The analogous expression for the bias in a true Laspeyres index (i.e., the index calculated using link-period rather than base-period expenditure shares) does show a role for the elasticity of substitution in determining the size of the bias.

comparable to its predecessor will simply inherit the same weight as was assigned to the predecessor item. (Previously, the BLS had computed a new weight for the substitute, using a procedure much like the one outlined above.) This modification is expected to eliminate any “formula” bias associated with item substitution.

Several recent papers [notably Moulton (1993), Reinsdorf and Moulton (1996), and Moulton and Smedley (1995)] have explored another alternative to the Laspeyres-based formula, namely the modified geometric-means estimator:

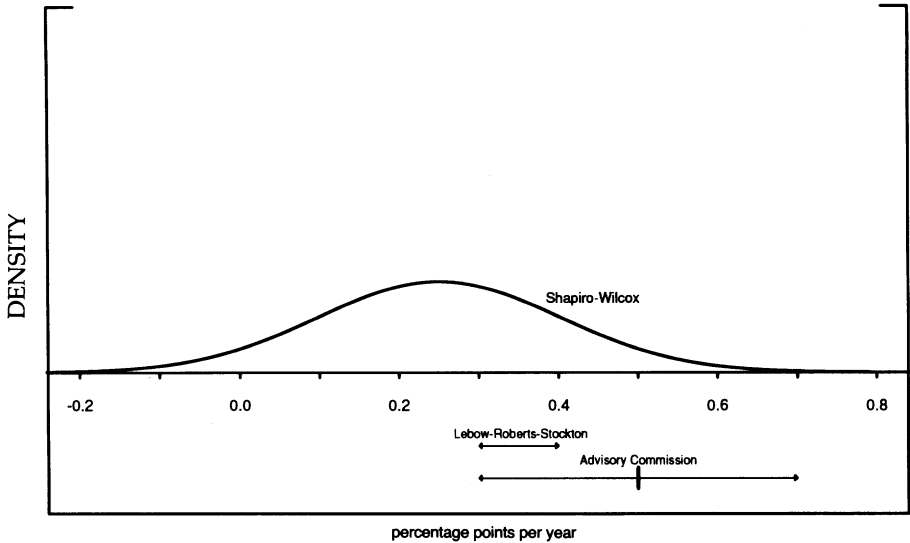
$$\frac{G_{it}}{G_{il}} = \exp \left\{ \sum_j \omega_{jb} \ln \left( \frac{p_{jt}}{p_{jl}} \right) \right\}, \quad (11)$$

where, again, the nominal expenditure shares pertain to the base period  $b$ . (A conventional geometric-means estimator would use nominal expenditure shares as of period  $l$ .) Under the same assumptions as we used above (CES utility, stationary distribution of relative prices, etc.), one can show that the modified geometric-means estimator is approximately asymptotically unbiased for the true cost-of-living index.<sup>22</sup> This result provides a crucial tool for assessing the magnitude of the within-stratum effect. The difference between the growth of the modified Laspeyres index (i.e., the CPI between 1978 and mid-1996) and the growth of the modified geometric-means index is an estimate of the bias in the modified Laspeyres index relative to the true cost-of-living index. Similarly, the difference between the seasoned version of the CPI and the geometric-means index is an estimate of the bias in the seasoned index relative to the true cost-of-living index.

*Empirical Magnitudes* The most important evidence on the magnitude of the within-stratum effect comes from comparisons between two indexes—one in which the elementary aggregates are computed using the official modified Laspeyres formula, and the other in which the elementary aggregates are computed as weighted geometric means. Moulton and Smedley (1995) perform such a comparison using data covering 96 percent of the weight of the overall CPI for the 30 months

22. If the distribution of relative prices is not stationary (as may be true in heterogeneous strata), then the elasticity of substitution is relevant for the size of the bias. For an elasticity of substitution of one, the geometric-means formula is the exact measure of the cost of living. At the substratum level, it is hard to see how the equilibrium value of the elasticity of substitution could be less than one.

Figure 3 WITHIN-STRATUM EFFECT



Note: Lebow, Roberts, and Stockton (1994) and the preliminary report of the Advisory Commission were written before—and therefore do not reflect—the changes in the substratum computation that BLS announced in March 1996. See also note to Figure 2.

from June 1992 to December 1994.<sup>23</sup> They find that the index based on geometric means increases 0.49 percentage point less per year than the index based on the modified Laspeyres formula that was in force during their sample period. This estimate overstates the magnitude of the within-stratum effect still remaining in the CPI because it does not reflect the modifications in technique described above. The BLS expects those changes taken together to reduce the growth of the overall index by about 0.25 percentage point.

We summarize our assessment of these various considerations in Figure 3, using a variable that is distributed normally, with a mean of 0.25 percentage point per year (i.e., the Moulton–Smedley estimate less the 25-basis-point reduction effected by the changes in technique already

23. In constructing their modified Laspeyres and geometric-means indexes, Moulton and Smedley aggregate the alternative sets of elementary price indexes using the same weights and Laspeyres aggregation formula at the superstratum level. They are building on the work of Reinsdorf and Moulton (1996), who performed similar calculations using a dataset covering 12 months and about 70 percent of the overall index. Moulton and Smedley state (1995, p. 13) that the 4 percent of the index not covered by their calculations consists of “items for which there are exceptional methods of calculating price change for the actual CPI, and where it would be inappropriate” to apply geometric means.

put in place) and a 90 percent confidence interval ranging from 0.0 to 0.5 percentage point per year. We also assume that there is no correlation between the within-strata effect and the across-strata effect, on the theory that the factors governing the magnitude of the within-strata effect (importantly, search) may have little bearing on the willingness of households to actually alter the composition of their consumption bundles in the face of changes in relative prices.

In their interim report, the Advisory Commission (1995) interpreted the difference between the geometric-means index and the modified Laspeyres index as an estimate of the magnitude of “formula bias;” they gave a point estimate for the influence of this bias over the past few years of 0.5 percentage point per year, with a range extending from 0.3 percentage point to 0.7 percentage point. Looking prospectively, the Commission assigned an estimate of zero to the influence of this effect, based on their expectation that the BLS would soon implement procedures to eliminate whatever influence from this effect remains in the index. Indeed, as we noted above, this expectation was at least partly fulfilled with the BLS announcement in late March 1996 concerning the extension of “seasoning” to all items in the CPI. Lacking any information on whether the Commission will regard the problem as having been solved by the recent BLS action, we compare our probability distribution with their retrospective assessment, noting that their interim report was written before the recent BLS press conference and so did not reflect the change in technique announced there.

Lebow, Roberts, and Stockton interpreted the same evidence on the difference between geometric means and modified Laspeyres indexes as bearing on the strength of a pure substitution effect within strata; based on this evidence, they specified a range extending from 0.3 percentage point to 0.4 percentage point per year. This specification also pre-dates the recent BLS announcement regarding within-strata aggregation.

#### 4.4 THE NEW-ITEMS EFFECT

New items generally are brought into the CPI sample in a way that guarantees that their prices will have no effect on the level of the index in the first month of their inclusion.<sup>24</sup> In effect, the *levels* of these prices are stripped of any implication for the index, and only the *changes* from the date of inclusion forward matter. In economic terms, this approach can be thought of as building into the index the assumption that access to new items—at the prices at which they are brought into the index—does

24. New cars represent an important exception to this general rule. As we discussed in Section 2, the BLS makes direct adjustments to the prices of new cars based on manufacturers’ cost estimates (marked up to the retail level).

not reduce the minimum cost of attaining the benchmark level of utility. Put slightly differently, these new items are assumed not to create any consumer surplus at the prices at which they are brought into the index. For the bulk of new items that are close substitutes for others already represented in the index, the current approach probably works reasonably well, but this may not be so for the rare new item that delivers services radically different from anything previously available. For example, even the earliest generation of personal computers allowed consumers to undertake tasks that previously would have been prohibitively expensive.

This problem can be solved only by estimating the consumer surplus created by the introduction of each new item. Hausman (1994) argues that this must involve explicit modeling of the demand for each new item. In principle, such modeling allows the researcher explicitly to estimate a reservation price for each new item, and thus to calculate the consumer surplus it produces even at its introductory price. Hausman applies this approach to the market for breakfast cereals, and concludes that the CPI overstates the true rate of increase of a cost-of-living subindex for breakfast cereals by about 20 percent, or 0.8 percentage point per year if the measured average annual rate of inflation in this category is 4 percent.<sup>25</sup> Although explicit modeling of demand may be of dubious practicality for widespread implementation in the CPI, strategic application in a few selected cases might be worthwhile.

The difficulty of analyzing the impact of new items on the CPI is compounded by the fact that such items are not brought into the index immediately upon their introduction into the market, but only with a lag. According to conventional wisdom, many items experience large price declines early in their life cycles. If this conventional wisdom is right, then the delay in incorporating new items into the index causes them to be linked in at a lower price, and hence with a larger amount of omitted consumer surplus.<sup>26</sup> [See, for example, Gordon (1993, p. 238) for

25. Hausman's results may be overstated to the extent that the constant introduction of new varieties of cereal reflects changing tastes rather than utility gain for given tastes. Separately, there is a difference of opinion between Hausman (1994) and Fisher and Griliches (1995): Whereas Hausman models the *market* demand curve and advocates use of the intercept in the price index, Fisher and Griliches argue that the tightest lower bound on the rate of growth of the true cost-of-living index is obtained by using the quantity-weighted average of the intercepts from each *individual's* demand curve.

26. Sample rotation alleviates this problem because it brings new products into the index more rapidly than would be the case if the BLS refreshed the sample only in the course of a comprehensive (roughly decennial) revision. Even under the best of circumstances, however, new items still attain only 40 percent of their steady-state representation in the index after about 4 years given the current sample-rotation scheme, and 100 percent after about 7 years. And if a new item is so dissimilar from anything previously

a statement of the general problem, and Scherer (1993, pp. 102–103) for specific application to the case of pharmaceuticals.]

We stress, however, that earlier incorporation of new items into the index—even in combination with explicit modeling of the demand for new items—would not solve this problem. In fact, a U-shaped pattern of prices over the life of a typical item creates a dilemma that cannot be resolved within the context of a Laspeyres-type index. Early incorporation of new items into the CPI will cause them to be underrepresented in the index because they will not have won a significant share of the market compared with the share that they may attain later in their life cycle. On the other hand, late incorporation will cause the period of supernormal decline in relative price to be missed entirely. The only way out of this dilemma is to combine explicit modeling of the demand for new items with abandonment of the Laspeyres framework.<sup>27</sup>

A second factor complicating the analysis of new items is the disappearance of old ones. The common presumption (shared by us) is that the loss of the consumer surplus associated with the disappearance of old items does not fully offset the gain in consumer surplus associated with the appearance of new ones. Although such presumptions may well be valid, models presented in Dixit and Stiglitz (1977) and Spence (1976) indicate that there is no theoretical proof that this must be the case.

Griliches and Cockburn (1994) illustrate the importance—until recently—of the new-items issue in the market for prescription drugs.<sup>28</sup> Until January 1995, newly available generics were not represented in the CPI unless and until they were brought in through regular sample rotation or other item replacement. Consistent with the BLS's usual methodology, any generic that was brought into the index through either of those mechanisms influenced the index only to the extent that the price of the generic changed subsequent to its inclusion, and no account was taken of any gap between the price of the generic and the

---

available in the marketplace as not to fit naturally within any existing item stratum, the delay can be much longer.

27. Our analysis here is similar in spirit to that of Griliches and Cockburn (1994, p. 1229). They construct several different price indexes for the drug cephalexin, including a Laspeyres index which suffers from "late inclusion of generics with too low and too fixed a weight." A further important complication in this regard involves the slow diffusion of knowledge about a new product. Griliches and Cockburn present evidence suggesting that knowledge about newly available generic drugs may take about 6 months to diffuse through the economy.
28. Strictly speaking, Griliches and Cockburn tailored their discussion to the treatment of prescription drugs in the producer price index, but qualitatively the same critiques could have been made of the treatment of prescription drugs in the CPI.



price of its branded counterpart as of the date of inclusion.<sup>29</sup> In January 1995, the BLS implemented a new approach which does allow for direct comparison between the prices of generic and branded versions of a given drug. We have no presumption that the new approach results in any bias in either direction.<sup>30</sup>

Pakes, Berry, and Levinsohn (1993) present estimates of price indexes for new cars based on an estimated-demand system. The demand system takes into account heterogeneity in preferences and the discrete nature of the car-purchase decision. They compare two price indexes—a Laspeyres-type index that reprices the same marketbasket of autos over time, and an alternative that allows the mix of auto purchases to evolve in line with historical experience (and in particular allows for introduction of new models).<sup>31</sup> They find that the Laspeyres-type index increased twice as rapidly during the 1980s as did the more flexible alternative. While Pakes, Berry, and Levinsohn caution against reading too much into their calculation, their work provides both an illustration of a case in which the new-items effect appears to have been important and a demonstration of one procedure for overcoming it. Since new cars are, however, one stratum where the BLS does attempt to make adjustments for new goods, the difference between the CPI and the Pakes–Berry–Levinsohn index is not a measure of the new-goods effect *per se*. Yet, their index is suggestive for the magnitude of the effect and provides a possible model for taking into account new goods in other strata.

*Empirical Magnitudes* In its interim report, the Advisory Commission penciled in a point estimate of 0.3 percent for the new-items effect, with a

29. While Griliches and Cockburn demonstrate a substantial upward bias in price indexes for two generic drugs based on BLS methodology, they do not provide an estimate of the bias in the overall prescription-drug component of the CPI.

30. Under the new methodology (see Armknecht, Moulton, and Stewart, 1995), the BLS monitors the expiration of all prescription-drug patents. Six months after the expiration of the patent for any prescription drug in the CPI sample, a BLS representative will survey each outlet where that drug was priced, and ascertain the distribution of quantities dispersed at that outlet as between the branded drug and any generic substitutes. Based on sampling probabilities proportional to those quantities, the representative will then designate either the branded drug or one of the generics as the item to be priced henceforth at that outlet (until the next sample rotation). (Contrary to the description in Armknecht, Moulton, and Stewart, the sampling is with respect to quantities, not revenue shares.) If a generic version is selected, any gap between its price and the price of the branded version will be fully reflected in the CPI, contrary to prior practice. According to Armknecht, Moulton, and Stewart, this adjustment in procedure “will have the effect of slightly slowing the rate of growth in the CPI prescription drugs component” (p. 18).

31. The latter index is based on the equivalent variation from their estimated-demand system.

range extending from 0.2 percent to 0.7 percent.<sup>32</sup> The Commission offered no empirical evidence in support of these estimates. Instead, the Commission suggested the following question as a “thought experiment” that might be useful for determining the magnitude of the new-items effect: “How much more income would you require to be as satisfied with the 1995 basket and prices as with the 1970 basket and prices?” (p. 24). The Commission conjectured that the percentage increase in income required would be substantially less than the percentage increase in the CPI. The Commission then attributed the difference to new items (“the many benefits of modern life”). The difficulty with this thought experiment is that all of the other problems we and the Commission analyze also caused the CPI to misstate the cost of living, so it is inappropriate to attribute all the difference to new items. In any event, the Commission promised further analysis of this issue in its final report, so its interim estimates should be regarded as tentative.

Lebow, Roberts, and Stockton (1994) make “some rather extreme assumptions” (p. 11) to calculate an upper bound on this effect: They begin by judgmentally identifying those categories of consumer expenditures in which introduction of new goods is likely to be most important. These categories had a combined relative importance weight in December 1993 of 2.4 percent. Lebow, Roberts, and Stockton then assume that a like share of households’ *true* marketbasket at any given moment is spent on items that are not yet represented in the index. Finally, they assume that the relative price of the unrepresented portion of the marketbasket is declining at a 20 percent annual rate (roughly in line, as Lebow, Roberts, and Stockton point out, with the rate of decline of the relative price of information-processing equipment). If all of these assumptions were true, the new-items effect would be adding 0.5 percentage point per year to the growth of the overall index. Lebow, Roberts, and Stockton believe their assumptions “surely make the estimate an upper limit on this effect” (p. 11).<sup>33</sup>

As the preceding discussion should make clear, the scientific basis for making a judgment about the magnitude of the new-items effect is particu-

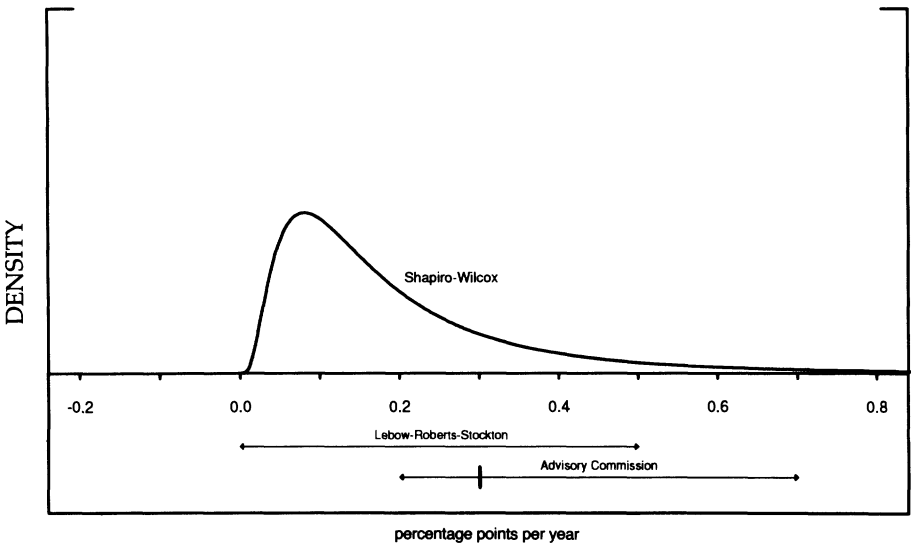
32. Under the rubric of “new product bias,” the Commission also included that the BLS does not build into the index direct price comparisons between old and new items that provide similar services. (Among other examples, the Commission cites the fact that the CPI does not recognize video rentals as a substitute for trips to movie theaters.)

33. By focusing exclusively on the declining relative price of new items and making no assumption about their rate of introduction into the marketplace, the amount of consumer surplus created upon their introduction, and their rate of incorporation into the CPI, Lebow, Roberts, and Stockton implicitly are assuming that new items are introduced into the marketplace at zero increment to consumer surplus. This assumption is considerably less restrictive than the similar assumption of the BLS: the former stipulates zero consumer surplus at the date of introduction into the marketplace, whereas the latter stipulates zero consumer surplus at the date of introduction into the index.

larly thin. Nonetheless, we find the conventional arguments plausible, and we find the arithmetic of Lebow, Roberts, and Stockton suggestive. Although we are highly uncertain about the magnitude of this effect, we are quite confident it should be positive. We attempt to convey the gist of these beliefs using a variable that is distributed lognormally, with mean 0.20 percentage point per year and 90 percent of its mass to the left of 0.4 percentage point. This calibration puts nearly all of the probability mass below the top end of Lebow, Roberts, and Stockton's range, consistent with their view that the top end of their range is a very loose upper bound on the true magnitude of the effect. Figure 4 compares this assumption with those of the Advisory Commission and of Lebow, Roberts, and Stockton.

The more responsive consumers are to changes in relative prices, the more they might substitute to new goods that have the effect of allowing them to achieve the benchmark level of utility more cheaply. Hence, if we are underestimating the within-strata effect, we will also be tending to underestimate the new-goods effect. To capture this correlation between the magnitudes of the two effects, we assume that there is a correlation equal to 0.25 between the within-strata effect and the new-items effect. This correlation merely reflects our subjective prior, not

Figure 4 NEW-ITEMS EFFECT



Note: See note to Figure 2.

specific empirical evidence.<sup>34</sup> All we have to go on is our presumption that the correlation is positive (because both effects depend positively on consumers' elasticity of substitution) and that it is not one (i.e., there is some independent uncertainty about the two effects).

#### 4.5 THE NEW-OUTLETS EFFECT

In many respects the "birth" of a new outlet is analogous to the introduction of a new item into the marketplace. Under certain circumstances, such a birth may create consumer surplus. And under certain circumstances, such consumer surplus will not be captured in the CPI. We analyze the various possibilities by considering five cases.<sup>35</sup>

In the first case, an entrepreneur discovers a technological innovation that allows her to deliver some item to consumers at a lower price than is offered by incumbent outlets. This entrepreneur goes into business. Knowledge in consumer markets is less than perfect, however, so although some consumers chance upon the new outlet and purchase the item there, not all consumers make this discovery. A few of the incumbent outlets may go out of business, but not all do, and the ones that remain keep their prices for the item at the same level as before. Eventually, the new outlet is brought into the CPI sample.

In this case, the birth of the new outlet creates consumer surplus. Moreover, that surplus will not be captured in the CPI, because when the new outlet is brought into the sample, its prices will be linked into the index in a way that will guarantee no impact on the level of the index in the first month (exactly as is the case with new items). Therefore, in this case, the index will be biased upward.

The suppositions for the second case are the same as for the first, except that in this case knowledge is perfect, and all consumers discover the new outlet. In response, incumbent outlets cut their prices to match the entrant's price, possibly by copying the entrant's technological innovation. In this case, the current methodology works perfectly. Consumer surplus is created, and the index captures it. Competition forces the impact of the technological innovation to be fully reflected in the prices

34. In Shapiro and Wilcox (1996) we explore the robustness of the distribution of the total bias for different assumptions about this correlation. The higher the correlation, the wider the dispersion of the estimate of the total. (Correlated effects "average out" more slowly.) The results for the total are not, however, highly sensitive to the magnitude of the presumed correlation.

35. Most discussions of outlet substitution fail to emphasize that the relevant outlet is a new one. If consumers merely are switching between existing outlets in response to a change in relative prices, they are only engaging in within-strata substitution. Any bias in the index resulting from this behavior should be corrected by adopting a modified geometric-means formula for the first stage of aggregation.

offered at incumbent outlets (that is, reflected in the prices in the CPI sample). Therefore, there is no distortion stemming from initial exclusion of the entrant, or from the linking in of its price.

In the third case, a new outlet enters the market offering the item at a lower price, but only because it offers an inferior mix of other attributes (e.g., service, store location, etc.). Consumers have homogeneous tastes, and knowledge is perfect. In this case, a price differential between incumbents and the entrant is established, and that differential exactly reflects the market valuation of the difference in satellite characteristics. No consumer surplus is created, and none is recorded under current procedure. Once again, the current procedure works exactly correctly. The common thread of the second and third cases is that the law of one price holds at every instant, so price differentials reflect quality differentials.

In the fourth case, a new outlet enters the market at the *same* price as the incumbents, but offers a different mix of other services. Consumers have heterogeneous tastes, which cause different relative preferences for the two outlets. For example, some consumers may appreciate attentive service, while others prefer to browse undisturbed. In this case, consumer surplus is created because variety in shopping experience is valued in the marketplace. However, no increase in surplus will be recorded in the CPI, because no price change has occurred.

In the fifth case, a new outlet enters with lower price and lower-quality service than are offered by the incumbents. The incumbents meet the competition by imitating both the lower price and the lower-quality service. The CPI will register the price decline, but will not reflect the decline in quality. As a result of that omission, the CPI will overstate the true rate of decline of the cost of living.

Under some circumstances, the evidence reported in Reinsdorf's (1993) paper is useful for gauging the magnitude of the new-items effect. For certain food and fuel items, Reinsdorf compared the average price among outlets rotating into the sample with the average price among outlets rotating out. He found that, for the set of items he studied, the average difference was about  $1\frac{1}{4}$  percentage points. Because sample rotation takes place at the frequency of once every five years, he converted this to a bias in the rate of change equal to one-quarter percentage point per year.

If either of the first two cases is the relevant one, Reinsdorf's experiment provides an exact reading on the consumer surplus created by the birth of a new outlet: In the first case, the CPI is biased upward, and Reinsdorf's evidence shows by exactly how much; in the second case, the CPI is not biased, and Reinsdorf's approach correctly would suggest none.

In the other three cases, unfortunately, Reinsdorf's evidence scores

less well. The third case involves no creation of consumer surplus, but Reinsdorf's evidence would suggest some increment to surplus. In the fourth case, Reinsdorf's approach would show no difference between incoming and outgoing samples even though the CPI was biased upward. And in the last case, Reinsdorf's approach would show nothing even though the CPI was biased downward.<sup>36</sup>

As this discussion should make clear, the new-outlets effect would be difficult to correct—at least as difficult, in our estimation, as the new-items effect. The only avenue to a solution appears to involve explicit modeling of preferences across outlets.

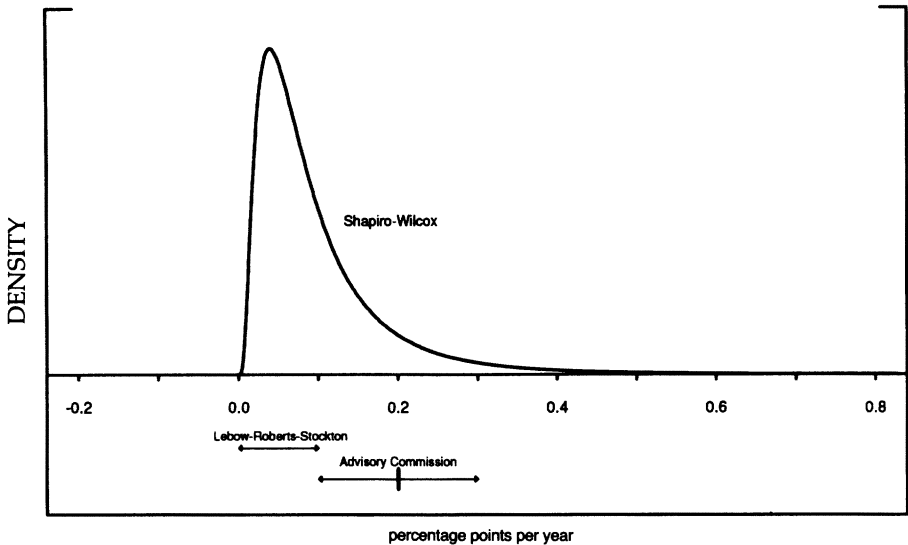
*Empirical Magnitudes* In their interim report, the Advisory Commission assigned a point estimate of 0.2 percentage point per year to the impact of "outlet bias," with a range extending from 0.1 to 0.3 percentage point. Lebow, Roberts, and Stockton (1994) developed their estimate of an upper bound on the magnitude of this problem by building on Reinsdorf's (1993) estimate. Specifically, they judgmentally identified all the categories of the CPI (including those studied by Reinsdorf) for which, in their opinion, outlet substitution might be operational. These categories amount to 40 percent of the overall weight of the index. They used the resulting figure of 0.1 percentage point ( $0.4 \times 0.25$ ) as the top end of their range, and zero as the bottom end.

Possible shortcomings notwithstanding, Reinsdorf's evidence still is the best available for the purpose of gauging the magnitude of the new-outlet effect. It persuades us that the new-outlet effect is small. And however big it may be, we are willing to assume that it is positive.

In light of these considerations, we summarize our understanding of the magnitude of this effect using a variable that is distributed log-normally, with mean equal to 0.1 percentage point per year, and 90 percent of its mass to the left of 0.2 percentage point per year. We further specify that this effect is positively correlated with both the within-strata effect and the new-items effect, with coefficient equal to 0.25. To be clear, we have no empirical basis for this last assumption, but a fairly strong presumption on theoretical grounds that zero is not the right answer because all three effects involve the sensitivity of consumers to incentives provided by changes in relative prices. Figure 5 compares our assumption about the marginal distribution for the new-outlet effect with those of the Advisory Commission and of Lebow, Roberts, and Stockton.

36. To be clear, we are asking Reinsdorf's evidence to do more than Reinsdorf himself had in mind.

Figure 5 NEW-OUTLETS EFFECT



Note: See note to Figure 2.

#### 4.6 THE QUALITY-CHANGE EFFECT

The operating characteristics of existing goods and services are continually being changed—and generally for the better. Quality change must be controlled for in the course of calculating the CPI. Failure to do so would induce an upward bias in the index assuming that new varieties are better than old ones on average. (For example, it is clear that one should not compare the price of a 1970 Chevrolet with the price of a 1995 Chevrolet without taking account of all the added features in the later model.)

Contrary to widespread misimpression, the BLS does not ignore quality change, even outside automobiles. In fact, as we discussed in some detail in Section 2, the BLS uses several methods for dealing with quality change. Despite these extensive efforts on the part of the BLS, many analysts believe that there have been and continue to be serious quality-change-related problems in the index that cause it to overstate the true rate of change of the cost of living. Gordon's monumental 1990 volume is the foremost piece of empirical work on the magnitude of the quality-change effect in BLS price series. While the bulk of Gordon's attention was directed toward constructing alternative deflators for producers' durable equipment, 11 of the 105 product indexes described in his vol-

ume also are relevant for the issue of quality bias in the CPI. Gordon constructs two indexes using the Tornqvist aggregation formula—one using the official CPI detailed components and the other substituting his 11 price series for the corresponding official series.<sup>37</sup> On the basis of a comparison of these two aggregates, Gordon concludes that the CPI for durable goods overstates the true rate of inflation for those goods by at least  $1\frac{1}{2}$  percentage points per year on average over the period 1947–1983. Measurement problems in this area were especially severe prior to 1960; for the last decade of his sample (1973–1983), Gordon estimates an average bias of at least 1 percentage point.<sup>38</sup>

Gordon finishes the opening chapter of his book with a list of factors which even he did not take into account in constructing his indexes. This list aptly conveys the difficulty of quality adjustment. In particular, among many other factors, Gordon cited his own inability to adjust for:

“Improved design of power lawn mowers, which has resulted in an order-of-magnitude reduction in injuries since the mid-1970s; . . .  
Improved cleaning ability of automatic washing machines and dishwashers; . . .  
And finally, immeasurably better picture quality of color television sets.”  
(p. 39)

Gordon’s inability to take account of these and the many other factors he lists, as well as his assumption that all of the other elements of consumer durables he did study were not mismeasured, implies that his estimate of 1.0 percentage point per year for the average bias in the CPI for durable goods probably is too low for the period he studied.

Many specific cases of quality change can be thought of as reflecting improvements in the efficiency with which a particular item priced by

37. Gordon estimates that his series cover about half of the weight of the CPI durables index. For the other half of the index, about which he has no evidence, Gordon assumes that the CPI measures quality change without error. This consideration suggests that his results may understate the quality-change effect in the CPI for durable goods over the period he studied.

38. In basing his results on a comparison of Tornqvist aggregates, Gordon can be interpreted as filtering out the separate contribution of substitution bias. Even so, the difference between Gordon’s alternative index and the CPI reflects a number of different effects, including quality change, the effect of introducing new products more promptly, and the difference between the pricing at the outlets he samples and the average outlet. We are assuming that the preponderance of the difference is quality change. The uncertainty about the relative importance of quality change and the other effects does, however, contribute to the uncertainty about the magnitude of the quality-change effect.



the CPI produces the service which is actually valued by the consumer. Below, we use the example of cataract surgery to illustrate this point.

Over the years, a few sources of downward bias have also been identified. For example, the BLS's technique for linking replacement items into the index—as it was implemented prior to 1993—caused the CPI to understate the true rate of inflation whenever a bona fide price increase coincided with the introduction of a new variety. Bias occurred in this situation because the prices of repriceable items behaved differently from the prices of nonrepriceable items. For example, Armknecht and Weyback (1989, pp. 114–115) report that the average month-to-month price change during 1983 for repriceable men's suits was only 0.3 percent. By contrast, the average price change for substitute suits judged close enough in quality to their predecessors to be “comparable” (and therefore requiring no quality adjustment) was 15 percent.<sup>39</sup>

The BLS addressed this problem by refining its method for imputing the missing price change: Rather than using the prices of all repriceable items to impute the missing price, the BLS began to limit the information set to include only those pricing attempts in which an item substitution took place, but in which the new item was judged comparable to the old or a direct quality adjustment (e.g., an adjustment based on manufacturer's cost) was feasible. The BLS applied this improved method of imputation to the pricing of new cars beginning in October 1989, and extended the use of this technique to other nonfood, nonservice items in the CPI beginning in December 1992.

The BLS evaluates the entire body of evidence on the quality-change effect as ambiguous, and maintains that “the total magnitude—and even the direction—of quality change effects on prices not accounted for by [the BLS's] current procedures is unknown” (Bureau of Labor Statistics, 1995).

*Empirical Magnitudes* Quality change is the house-to-house combat of price measurement. There is no simple formula that one can apply to deduce a magnitude of the problem, nor any simple solution. Unfortunately, there is no substitute for the equivalent of a ground war: an eclectic case-by-case assessment of individual products.

In its interim report, the Advisory Commission placed a range of 0.2 to 0.6 percentage point around the quality-change effect, and put the point estimate at the bottom of this range. Lebow, Roberts, and Stockton (1994) developed their estimate of the quality-change effect in the follow-

39. Diewert (1995, p. 30) characterizes the occurrence of real price *declines* upon introduction of a new variety as the “typical” case. A comprehensive summary of available evidence on this issue would be useful.

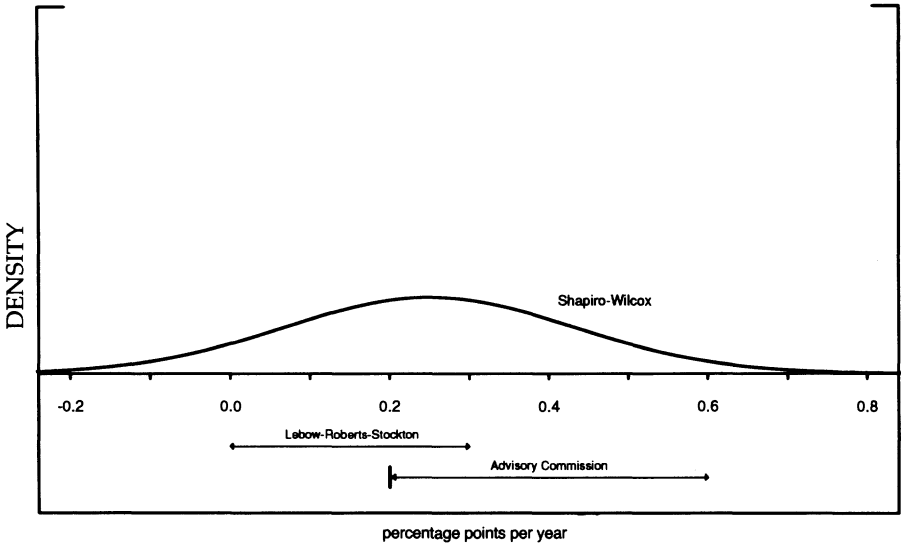
ing manner: First, they judgmentally identified “those categories [of the CPI] where year-to-year quality adjustment difficulties appear to be most acute” (p. 10). These categories collectively had a relative importance weight in the index of about 23 percent at the end of 1993. They then assumed that Gordon’s (1990) estimate of quality-change bias for durable goods over the period 1947–1983 (1.5 percentage points per year) could be applied to this broader aggregate. These assumptions yield their “high end” estimate of 0.3 percentage point per year. Lebow, Roberts, and Stockton used zero for their “low end” estimate.

In specifying our probability distribution for the quality-change effect, we modify Lebow, Roberts, and Stockton’s calculations in two respects. First, we use Gordon’s estimate of the bias in the CPI for durable goods over the last decade of his sample (1973–1983) rather than his estimate for the entire 1947–1983 period, in the belief that the more recent evidence provides a better indicator of the quality-change effect still remaining in the durable-goods component of the CPI. Over the later period, the average bias computed by Gordon was 1 percentage point. (We would have preferred to have used still more recent evidence, but neither Gordon nor anyone else to our knowledge has updated his series beyond 1983.) Second, we use data from the Consumer Expenditure Surveys for 1993 and 1994 to recompute the relative importance weights for the categories designated by Lebow, Roberts, and Stockton as susceptible to the quality-change effect.<sup>40</sup> This results in a tiny upward revision to the relative importance weight of the designated categories, to 24.5 percent. The combination of these modifications yields an estimate of 0.25 ( $1.0 \times 0.25$ ) percentage point. Partly on the basis of our preliminary exploration of the medical care area (see Section 5), we are inclined to treat this estimate as a mean rather than an upper bound.

These considerations lead us to summarize our beliefs concerning the size of the quality-change effect using a variable that is distributed normally, with mean 0.25 percentage point per year and 90 percent confidence interval extending from  $-0.05$  to  $0.55$  percentage point. We place nonzero probability mass in negative territory in light of the fact that examples have occurred in the past in which quality-adjustment problems contributed a downward bias to the index. Figure 6 compares our assumption with those of the Advisory Commission and of Lebow, Roberts, and Stockton.

40. We would have calculated an average for 1993–1995 in conformity with the planned base period to be introduced in 1998, but the data for 1995 are not yet available. We are grateful to Stephanie Shipp of the BLS for supplying detailed tabulations of the 1993 and 1994 CEX.

Figure 6 QUALITY-CHANGE EFFECT



Note: See note to Figure 2.

#### 4.7 THE TOTAL BIAS FROM ALL SOURCES

In its interim report, the Advisory Commission calculated its point estimate for the overall bias in the CPI by summing the point estimates it specified for each of the individual imperfections described above. Similarly, the Commission calculated an upper bound on the total bias by taking the sum of the upper bounds it specified for the individual imperfections, and likewise for a lower bound on the total bias.

By this means, the Commission arrived at a point estimate of 1.5 percentage points per year for the total bias in the CPI during the last few years, with a range extending from 1.0 to 2.7 percentage points. Looking ahead, the Commission assumed that the BLS would soon take action to eliminate the within-strata effect from the CPI. As a result, it estimated the likely total bias in the CPI over the next decade or so at 1.0 percentage point per year, with a range extending from 0.7 to 2.0 percentage points. As we noted earlier, the BLS has announced that it will implement new procedures beginning in mid-1996 that should reduce the severity of the within-strata problem. As of this writing, the Commission has not issued a revised assessment of the size of the overall bias taking into account the recent BLS announcement. We address this situation by comparing our distribution for the overall bias

with the Commission's backward-looking range (i.e., the one that makes nonzero allowance for the within-strata effect), while noting that the Commission's backward-looking range does not yet incorporate the Commission's thinking with regard to the BLS's recent actions.

In both the forward-looking and the backward-looking versions of the Advisory Commission's specifications, the point estimate is less than the midpoint of the range. One possible interpretation of this circumstance is that the Commission filtered its point estimates for the individual effects through an asymmetric loss function which penalized estimates that turn out to be too high more heavily than it penalized estimates that turn out to be too low.<sup>41</sup>

Like the Advisory Commission, Lebow, Roberts, and Stockton calculated their range for the overall bias in the CPI by summing the ranges for the individual estimates. On the basis of information available to them as of their writing, they therefore declared an overall range extending from 0.4 to 1.5 percentage points per year. Their paper also predated the recent BLS announcement described in Section 4.3. In parallel with our treatment of the Commission's range, we show in Figure 7 Lebow, Roberts, and Stockton's range as given in their paper, and simply note that it does not reflect any adjustment for the recent BLS action.

To calculate the distribution for the total bias, we construct a random variable equal to the sum of the effects whose distributions are shown in Figures 2 through 6. Figure 7 shows the distribution of this total bias.<sup>42</sup> It also compares our distribution with the estimates of the Advisory Commission and Lebow, Roberts, and Stockton. We estimate that there is a 90 percent probability that the total bias in the CPI is greater than 0.6 percentage point per year, and a 90 percent probability that it is less than 1.5 percentage points per year. The median of our distribution occurs at just under 1.0 percentage point per year, and the mean at 1.0 percentage point per year. The slight skewness in the distribution reflects our specification of lognormal distributions for the new-items and new-outlets effects.

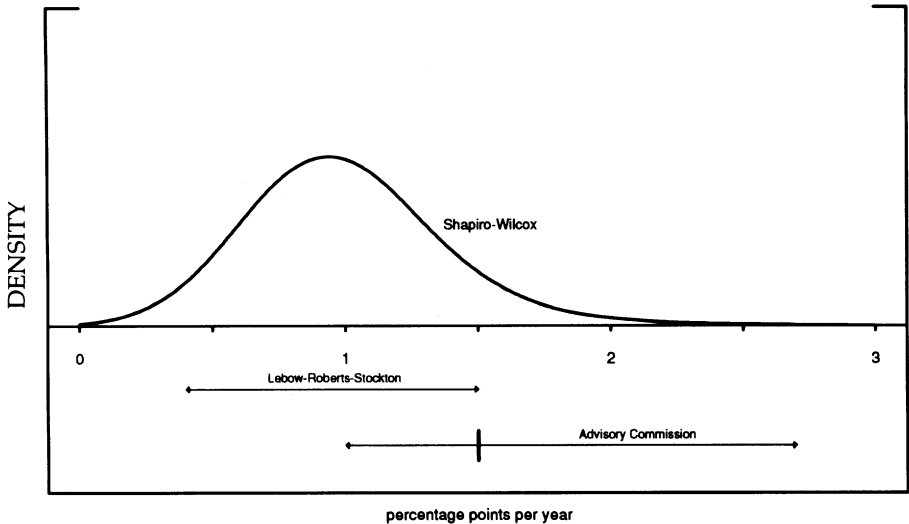
#### 4.8 USING THE DISTRIBUTIONS FOR POLICY FORMULATION

Some policymakers have suggested that indexation of items in the federal budget be modified in light of the overstatement of the increase in

41. Neither the Advisory Commission nor Lebow, Roberts, and Stockton specified whether their ranges (or, in the case of the Commission, its point estimate) could be given a formal interpretation in terms of probability theory. Nor did either group specify whether the interpretation of the range for the overall bias was necessarily the same as the interpretation of the ranges for the individual effects.

42. Because the total bias is a sum of normals and lognormals and because we allowed the elements in the sum to be correlated, we carried out this calculation numerically. See Shapiro and Wilcox (1996) for a description of this method.

Figure 7 OVERALL BIAS IN THE CPI



Note: The Shapiro-Wilcox distribution is based on aggregating the component effects reported in Figures 2–6. See Shapiro and Wilcox (1996) for the method of calculating the distribution of the aggregates consistent with the distribution of the components. The ranges of Lebow, Roberts, and Stockton and the Advisory Commission are the sums of the lower and upper bounds of the ranges in Figures 2–6. Neither of these ranges reflects the changes in the substratum calculations announced by the BLS in March 1996. See also note to Figure 2.

the cost of living by the CPI.<sup>43</sup> We do not take a stand on these proposals. We do note, however, that while an accurate measure of the cost of living may be necessary for the optimal design of these policies, it is certainly not sufficient. For example, even a perfect cost-of-living index would not guarantee that the redistributive properties of the social security system are as intended or that the system is sustainable; an assessment of issues such as those lies far beyond the boundaries of this paper.

Nonetheless, the probability distribution we provide for the overall bias in the CPI will be relevant for policymakers wrestling with issues related to indexation. An adjustment to indexation runs the risk of overadjusting—that is, having benefits and tax brackets increase less rapidly than the cost of living—as well as the risk of underadjusting. A

43. See Daniel Patrick Moynihan, "The CPI: An Easy Fix. . . .," *Washington Post*, September 26, 1995, opinion page. It is not clear from this op-ed piece whether Senator Moynihan is suggesting technical adjustments in the CPI that would reduce the average rate of increase in the CPI or legislative adjustments in indexation formulas relative to the CPI. Several state governors have also endorsed a change in CPI indexation as part of a budget deal (see Judith Havemann, "Governors Recommend CPI Changes," *Washington Post*, December 5, 1995, p. A9).

policy maker who wanted to balance these risks might vote for adjusting benefits and tax brackets by the rate of increase in the CPI less an adjustment factor equal to the median of our distribution. On the other hand, a policy maker might prefer to run a lesser risk of undercompensating beneficiaries (and, accordingly, run a greater risk of overburdening contributors to the relevant benefit program). The distribution of the overall CPI bias can be used to design a policy of this type. For example, our distribution implies that an adjustment in indexation of 0.6 percentage point would stand only a 10 percent chance of being too large.

### *5. A Case Study of Quality Change: The Price of Treatment for Cataracts*

One clear message from the theory of the cost of living is that the most straightforward way to build a cost-of-living index is to price the proximate causes of consumer utility. This is easier said than done, however, and—as Nordhaus (1994) notes—for a variety of practical reasons, the BLS in a large number of areas prices goods and services that are one step removed from the items that directly produce consumer satisfaction.

In principle, the pricing of inputs rather than outputs is not fundamentally inconsistent with adequate adjustment for quality change; one can still obtain an accurate index of the cost of living by adjusting the prices of inputs for changes in their efficiency in delivering consumer satisfaction. Relatively few such adjustments are made. If the efficiency of inputs increases over time and no compensating adjustment is made, resulting price indexes will overstate the true rate of increase of the cost of living.

Nordhaus studies one example of this phenomenon—the pricing of household lighting. Whereas consumers presumably derive satisfaction from the intensity and reliability of the lighting services they purchase, the CPI prices the inputs that produce those services (e.g. light bulbs, fixtures, and electricity). Nordhaus constructs a proxy for the true price of lighting, and finds that it increases much more slowly than the most comparable elements of the CPI.

By far the most important example of this problem occurs in the area of medical care. Here, the CPI prices inputs (an hour of a physician's time, a day in the hospital, a basket of prescription drugs) rather than treatments (the restoration of eyesight impaired by cataracts, the repair of a broken bone, the treatment of a psychosis, and so forth). The notion that relatively little quality adjustment is performed in the medical area is supported by figures reported in Armknecht and Weyback (1989, p. 110) showing that in 1983 and 1984, only about  $1\frac{1}{2}$  percent of

attempts to price medical care goods and services resulted in non-comparable item substitutions—less than in any other major category other than food and beverages.

This section presents a prototypical index of the price of treatment for cataracts. We hope this index will not only be of interest in itself, but also serve as a model of how to improve the pricing of medical care.

Several cautions need to be raised at the outset. First, we chose to examine cataract surgery in part because we knew there had been dramatic changes in technique in that form of surgery. The bias that we uncover thus is not representative of the bias in the medical care component of the CPI—much less in the overall CPI. Second, we calculate our prototypical index by combining two existing components of the CPI (one for physicians' services, and other for hospital services). Therefore, any inadequacies in these series will affect our calculations. Third, our information about changes in technique is based on interviews with medical personnel about typical practice at different points in time. Thus, our index should capture broad trends in the cost of cataract surgery, but not year-to-year or area-by-area variation.

Given the prevalence of third-party payment for surgical procedures, we need to address the issue of whether a study of the price of cataract treatment is relevant for the consumer price index. We believe that it is. On a practical level, the CPI covers all medical-care purchases financed by households (whether directly or through insurance paid for by them), and even Medicare-eligible patients (who constitute the bulk of the population having cataract surgery) finance some of their own treatment. On a theoretical level, one might further argue that the *whole* of medical care expense would be relevant if the objective were to construct a comprehensive index of the cost of living.

## 5.1 BACKGROUND ON CATARACT SURGERY

The lens of the eye focuses light onto the retina. A cataract is a cloudy lens, and this cloudiness impairs vision. Cataracts are removed surgically. Until recently, no other lens was inserted into the eye, so anyone whose cataracts had been removed required thick glasses or contact lenses to provide focus. Since the late 1970s, however, surgeons in the United States routinely have been inserting an intraocular lens (IOL) as a replacement for the defective natural lens. IOLs eliminate the need for thick glasses or contact lenses, and leave the patient with much better postoperative vision than they could have obtained under the old regime.

At the same time as outcomes have been improving, there have also been dramatic changes in the way those outcomes have been achieved—

mainly centering on the techniques used to make the incision in the eye, extract the defective lens, and close the incision. Together, these improvements have allowed the typical patient to be ambulatory much more quickly, and so have facilitated a dramatic reduction in the typical length of stay in a hospital from seven nights in the 1950s, to one night in the 1970s, and none currently—surgery for cataracts now being performed almost universally on an outpatient basis. The new techniques also have reduced the rate of complication and the number of required follow-up office visits. See Table 3 for a summary of the evolution of cataract treatment and estimates of the number of hospital days the treatment typically required at each stage of its evolution.

## 5.2 HYPOTHETICAL CPI VERSUS PROTOTYPICAL PRICE INDEX

The CPI does not price treatment for cataracts per se, but instead prices hospital services and physician services, among other items. In turn,

Table 3 A BRIEF CHRONOLOGY OF TYPICAL TREATMENT FOR CATARACTS

<i>Year</i>	<i>Procedure</i>	<i>Average Length of Hospital Stay (Nights)</i>	<i>Comments</i>
1947	Extracapsular extraction	7	Cataract removed mechanically or by irrigation Improved methods of extraction and of suturing; also, routine use of operating microscope
1952	Intracapsular extraction	7	
1969	Intracapsular extraction	3	
1972	Extracapsular extraction	1	Modern extracapsular extraction pioneered with phacoemulsification; typical extraction mechanical and suction
1979	Extracapsular extraction with intraocular lens (IOL)	1 or outpatient	Techniques to lessen complications; improvement in incisions and placement of IOL
1985	Extracapsular extraction with IOL	Outpatient	
1990	Extracapsular extraction with IOL	Outpatient	Phacoemulsification now common for extraction
1995	Extracapsular extraction with IOL	Outpatient	Reduced size of incisions



these component indexes are combined into an index of medical prices using base-period relative importance weights—currently, weights for 1982–1984. If current practice is maintained, the BLS will reweight the basket of medical inputs in 1998 according to expenditure shares in 1993–1995, and will then compute changes in the index from 1998 forward as weighted averages of the changes in the prices of the inputs.

This approach has a startling implication: Technological change that increases the efficiency of inputs in delivering consumer satisfaction affects only the *rate of change* of the index of medical prices, but not in the first instance the *level* of that index. Thus, for example, if the BLS (counterfactually) used this methodology to construct an index of the price of cataract treatment, the sharp decline in the average length of hospital stay would eventually cause hospital services to receive a lower weight in the marketbasket used to determine the *growth* of the cataract index from the base period forward, but it would never be reflected in a downward adjustment to the *level* of the series.

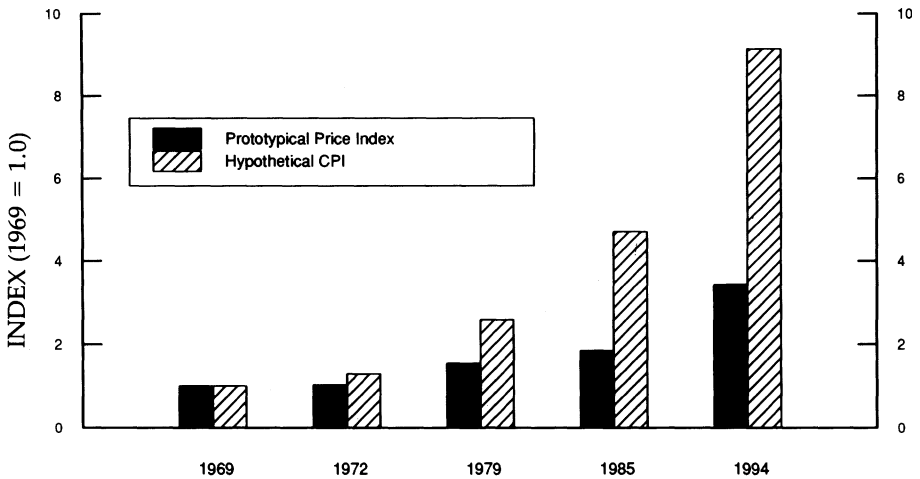
To illustrate this problem, we have constructed a hypothetical CPI for cataracts. Our hypothetical CPI for cataract treatment is based on the information in Table 3 as well as the CPI components for physician and hospital services.<sup>44</sup> We construct the hypothetical index by first estimating relative importance weights in hypothetical benchmark years for the physician services and hospital services required to treat a standard cataract patient. We then use these relative importance weights to aggregate the CPI components for physician and hospital services. The resulting time-series for selected years is shown as the cross-hatched bars in Figure 8. According to this input-based measure, the price of cataract treatment increased by a factor of nearly 10 between 1969 and 1993. This hypothetical CPI is meant to capture the price change the BLS would report were it to construct a CPI for cataracts using its current procedures.

We now describe the method we used to construct a prototypical index. This index measures the price of cataract surgery. It uses the same data as the hypothetical CPI—the quantity of hospital and physician services and the BLS indexes of their prices. Importantly, however, the prototypical price index reflects the decline in the level of hospital services required for cataract surgery.<sup>45</sup> The result is shown as the solid bars in Figure 8. According to the prototypical index, the price of cataract treatment increased over our sample period by a factor of only about 3.

44. Thus, we are ignoring other components, including office visits, anesthesiology, and glasses, contact lenses, or intraocular lenses.

45. We constructed this alternative as  $(\sum q_{jt}p_{jt})/(\sum q_{jb}p_{jb})$ , where  $q_{jt}$  is the quantity of input  $j$  required to treat a standard patient using standard techniques in period  $t$ , and  $p_{jt}$  is a CPI detailed component for item  $j$ .

Figure 8 PROTOTYPICAL PRICE INDEX VS. HYPOTHETICAL CPI FOR CATARACT SURGERY



Note: Authors' calculations. The hypothetical CPI is an estimate of what the BLS would report if it were to compute a price index for cataract surgery using the methodology of the CPI. The prototypical price index is an estimate of the price of cataract surgery. See Section 6.

Thus, from 1969 to 1994, our price index for cataracts rose only 5.1 percent per year, while the hypothetical CPI for cataract prices rose 9.2 percent per year.<sup>46</sup> Hence, BLS procedures for pricing medical care dramatically overstate the increase in price of procedures that can be accomplished with reduced levels of physician or hospital services.

### 5.3 CONCLUSIONS

The failure to take into account the decline in the level of inputs needed to deliver a service can lead to a dramatic overstatement of its price. Cataracts provide one striking example. Cutler *et al.* (1996) undertake a similar examination of the treatment of heart attacks, another very common medical condition whose treatment has been subject to substantial technical change. They find that from 1983 to 1994, the real cost of treating a heart attack increased less than 1 percent per year, compared to 2.4 percent or 3.3 percent per year for a hypothetical CPI, depending

46. This index is a major step toward pricing the product of cataract surgery, but it relies on the BLS indexes for the broad components determining the price. In work in progress, we are attempting to price directly a cataract operation—e.g. the ophthalmologist fee and the hospital charges for the specific operation. The results of this exercise will differ from our expenditure index to the extent the prices for these specific doctor and hospital services diverged from the averages captured by the BLS index.

on how it is calculated. When they adjust for quality change using declines in mortality rates, they find that the real price of treating a heart attack fell between 1983 and 1994.

Our prototypical price index for cataracts addresses one shortcoming of the CPI for medical care—its failure to take into account the reduction in the amount of inputs required to treat the condition. There are additional factors which make the wedge between the CPI and the actual price of care yet wider. First, the BLS series we use for physicians' services and hospital services probably overstate the rate of increase in the prices of those services (for example, by inadequately taking into account the growth in discounts for medical services). Therefore, the difference between the hypothetical CPI for cataract treatment and the actual price of treatment is probably even greater than is indicated by our results. Second, our index ignores improvements in the quality of the medical outcome. These include lower complication rates, shorter hospital stays, faster recoveries, better postoperative optical results, and no need for thick glasses. For this reason as well, the results shown in Figure 8 understate the difference between the quality-adjusted price and the hypothetical CPI for cataract surgery.

Current BLS procedure could be improved upon by pricing the treatment of conditions rather than a fixed-weight bundle of inputs. Specifically, a preferable approach would involve obtaining prices for the treatment of patients with standardized diagnoses. Posted prices are frequently discounted, so the BLS should attempt to measure the amount that healthcare providers actually receive, not what they bill.

## *6. Consequences of Mismeasurement*

The consequences of CPI mismeasurement for policymakers are fairly straightforward to enumerate. On the fiscal side, CPI mismeasurement matters because social security benefits, federal civilian and military pension benefits, veteran's benefits, tax brackets, personal exemptions, the standard deduction, the amount of investment income a child can receive tax-free, and school lunch prices are all indexed to the CPI. As we noted in the introduction, the consequence of this indexation, according to the CBO, is that a permanent one-half percentage point reduction in the annual rate of growth of the CPI, relative to baseline and starting in 1996, with all other factors in the economic environment held constant, would reduce the Federal deficit by \$26 billion in 2000, and nearly \$67 billion cumulatively over the five years ending in 2000, including the consequent reduction in debt service payments (Congressional Budget Office, 1995, pp. 2–3).

Duggan, Gillingham, and Greenlees (1995) point out an important feature of the indexation of social security benefits: The initial benefit entitlement does not depend on the CPI. (Each individual's nominal wage history is adjusted using a national-average wage series developed for this purpose.) Only the *growth* of the benefit subsequent to initial receipt depends on the CPI. The important implication is that measurement errors in the CPI have only temporary (albeit highly persistent) effects on outlays for social security benefits.<sup>47</sup> CBO estimates of the budgetary impact of changes in the CPI properly allow for this aspect of the social security system.

CPI mismeasurement also matters for the conduct of monetary policy. The Federal Reserve has made clear that its long-run policy objective is the attainment of price stability. Possibly for reasons related to the issues motivating this paper, the Fed has made clear that price stability would not necessarily correspond to a zero rate of increase in any particular existing price index. The existence of upward bias in the rate of growth of the CPI suggests that true price stability will correspond to positive *measured* CPI inflation.

For short- to medium-term monetary policy, it may be that the most important aspect of the bias in the CPI may be its variation from year to year. A bias that was both highly variable and difficult to observe or estimate would complicate the job of judging the appropriateness of the stance of monetary policy at any given moment. Unfortunately, we have been able to develop very little evidence on the year-to-year variation in the bias [the one exception concerning the across-strata effect, where estimates reported in Aizcorbe and Jackman (1993) were suggestive of significant year-to-year variation].

In addition to these implications for fiscal and monetary policy, mismeasurement in the CPI affects official statistics. CPI mismeasurement feeds through into the national income accounts because the Bureau of Economic Analysis uses detailed components of the CPI in constructing various elements of its price index for personal consumption expenditures. The effect of CPI bias on the measured rate of growth of real GDP is, however, less than one-for-one, for two reasons: First, consumption is only about two-thirds of GDP. Second, real GDP now is calculated using Fisher's ideal aggregation formula; as a result, real GDP should not suffer from bias induced by substitution across relatively aggregated categories. Together, these factors imply that the

47. Duggan, Gillingham, and Greenlees apply this insight to the estimation of the budgetary implications of the mistreatment of homeowners' costs in the CPI during the 1970s and early 1980s, and show that simple back-of-the-envelope calculations based on the assumption that measurement errors have permanent effects are seriously misleading.

mean of our subjective probability distribution over the bias in the growth of real GDP is on the order of one-half percentage point per year (that is, about two-thirds of 1.0–0.2).

CPI bias also affects the measured growth of productivity (by about as much as it affects real GDP) and the measured growth of real wages (by the full amount of the bias in the CPI).

In official U.S. statistics, the poverty line currently is calculated as three times the minimum cost in 1965 of an adequate diet, adjusted upward by the cumulative increase in the CPI. Hence, the CPI (and any biases in it) have a mechanical effect on official poverty statistics. There is, however, a growing consensus that this measure—*notwithstanding its linkage to the CPI*—understates the current poverty level (see National Research Council, 1995).

We also highlight two issues for which CPI mismeasurement is *not* particularly important, frequent claims to the contrary notwithstanding. First, there is very little evidence that CPI mismeasurement helps explain the apparent slowdown in growth during the 1970s. As Reinsdorf (1996) points out, the within-strata effect may have increased in size in 1978 when the current method of price sampling was introduced, but available evidence suggests that this effect is small compared to the slowdown in trend output growth. Furthermore, Gordon's (1990) evidence goes in the other direction: The quality-change effect appears to have been somewhat bigger before 1973 than after.

It is tempting to imagine that the pace of unmeasured technological change or productivity improvement must have increased in recent years given the ongoing shift toward intangible (especially information-intensive) forms of output. But it is important to bear in mind that there were dramatic changes in the 1950s and 1960s, including the harnessing of the atom and the space race. While we do not want to minimize how electronics have changed consumer goods recently, one should not forget Teflon, nylon, penicillin, and the automatic dishwasher.

Second, CPI mismeasurement has no bearing on the current debate over whether the economy can be allowed to grow more rapidly without overheating. Upward bias in the growth of the CPI would imply that "potential" output has been growing more rapidly than current official statistics would lead one to believe. But it would also imply that *actual* output has been growing more rapidly as well. Therefore, CPI mismeasurement has essentially no implication for the gap between actual and potential output, or between the "natural" and actual rates of unemployment, and hence no implication for the stance of monetary policy or other aggregate demand policy.

## *7. Looking to the Future*

The consumer price index is not a static construct. Over the years, the BLS has taken many important steps to improve the index (recall the selective listing of changes given in Table 1), and we fully believe that this improvement will be continued in the future. Many further improvements will be made as part of the BLS's comprehensive CPI revision, which spans the period from now through 2000.<sup>48</sup> As noted above, in 1998 the CPI will incorporate a new set of expenditure weights based on CEX data for 1993–1995. Other revision activities will include introducing new geographic and housing samples based on the 1990 census, updating the housing estimation and processing system to improve the accuracy of the CPI shelter indexes, and using computer-assisted technology to improve the speed and accuracy of data collection.

Also as part of the revision, the POPS survey of households (used to determine shopping patterns across outlets) will be restructured using telephone interviewing to permit more efficient sample rotation. Instead of revising all samples in 20 percent of areas each year, approximately 20 percent of item strata will be resampled in each area every year. This will add the potential for more frequent resampling of item strata that exhibit higher rates of product or outlet turnover.

The BLS is also developing a broader array of experimental indexes to evaluate the importance of substitution and other issues. For example, indexes based on the Aizcorbe–Jackman approach are being constructed using different three-year base periods, and using both fixed-weight and superlative formulas for aggregating stratum indexes. Another experimental index under development will employ a weighted geometric mean formula at the substratum level.

Within the CPI medical care component, the BLS is engaged in a variety of research activities and other enhancements, including changing the item structure and data collection forms for hospitals to better reflect the shifting mix of inpatient and outpatient care and the increasing divergence of transaction prices from list prices.

The main purpose of the rest of this section is to advance a few suggestions of our own for improving the CPI. The structure of this section is patterned after the framework we outlined in Section 3 and in Table 2. Our objective is to propose changes that would bring the CPI more closely in line with the theoretical benchmark of a true cost-of-living index. We recognize that most or all of these suggestions would have to

48. We thank John Greenlees for supplying the following description of the BLS's plans for the comprehensive revision.

be explored and developed further before they could be implemented; that process would no doubt require time and resources.

To address concerns related to the issue of substitution, we suggest that the BLS abandon the modified Laspeyres formula as the aggregation concept it is aiming to implement. One alternative that strikes us as well motivated theoretically would involve a hybrid of structures designed to exploit *a priori* theoretical restrictions and availability of information at each level of disaggregation. Specifically, the BLS might consider constructing the CPI as a modified geometric mean at the substratum level, a Tornqvist index within geographic areas at the superstratum level, and a Laspeyres index across geographic areas. This approach would have the virtue of adjusting the underlying utility construct at both the substratum and superstratum levels toward a benchmark that is more plausible within geographical areas than the current Leontief benchmark, and yet still preserve the assumption of no substitutability across geographic areas.

At least two points would have to be explored further before such a structure could be put in place. First, the Tornqvist formula is not implementable in real time because the data on expenditure shares become available only with about a two-year lag. Therefore, further research would be required to determine whether there might be a feasible real-time approximation to the true Tornqvist formula, possibly based on a forecast of expenditure shares. In this connection, the BLS might reconsider (and not only for this reason) its current policy of never revising the CPI, although we recognize that a host of issues would be raised by any move away from that policy; or, following a recommendation of the Advisory Commission in its interim report, the BLS might consider publishing one index that is never subject to revision and another that is.<sup>49</sup> Second, some thought would have to be given to the fact that 12 of the 44 strata currently do not pertain to a single geographical location, so a pure geometric-means formula might not be the most appropriate for those strata. Despite these significant conceptual hurdles, we believe that an alternative index formulated along these lines would probably represent a significant step forward.

As for concerns about new goods and new outlets, one useful (albeit expensive) step might be to put the sample rotation process on a once-every-three-years basis (as originally planned) rather than the once-

49. The Advisory Commission (1995, p. 21) suggested that the BLS consider publishing two versions of the CPI, one resembling the current index, "dedicated to timely measures of month-to-month price changes, and a second supplementary index produced with a greater time lag and subject to periodic revision, dedicated to accurate measurement of price changes over years and decades."

every-five-years basis it currently is on. An adjustment along these lines would cause new items to be brought into the sample more rapidly, and so would probably cause a larger fraction of the total consumer surplus created by such items to be captured. In this regard, however, we should stress an important linkage: The pace of the point-of-purchase survey probably should not be stepped up unless and until the aggregation formula at the substratum level has been adjusted. The current Laspeyres formula performs relatively poorly when the index is chained (as it is at sample rotation time), but an alternative aggregation formula such as the modified geometric mean probably would be much more robust to chaining. In addition, some explicit modeling of consumer demand might be undertaken on an exploratory basis; this is the only avenue we are aware of for addressing the problem of the surplus created both by new items and by the birth of new outlets.

Also related to sampling, an explicit linkage between the CPI, employment, and retail sales and inventories samples (the last of which is currently maintained by the Census Bureau) might yield some operating efficiencies, reduction in aggregate respondent burden, and cross-fertilization of ideas between agencies. Such a linkage would be very interesting substantively if it resulted in prices, wages, sales, inventories, and employment being measured at exactly the same outlets. A coordinated dataset of this type might yield dramatically new insights into the dynamics of adjustment at the microeconomic level, much in the same way that the Census Bureau's Longitudinal Research Database has done for the manufacturing sector.

On the quality front, there seems to be no alternative but to undertake detailed case studies of the type performed by Gordon (1990) for a subset of consumer durables, Griliches and Cockburn (1994) for two generic drugs, and us for cataract surgery. Probably hundreds of useful and interesting case studies remain to be executed. This is an area where academic researchers can—and ought to—make a constructive contribution to the efforts of the BLS. Our sense is that many of the most interesting case studies will bear on the pricing of medical care commodities and services. Such case studies will have the greatest influence if they attempt to construct prototypes of indexes that could actually be implemented by the BLS using reliable data sources available in real time. Ideally, the structure of the CPI should be flexible enough to allow yesterday's best thinking on any given item to be supplanted according to today's latest research. Finally, there should be at least a "research" version of the CPI that incorporates these quality-adjusted prices on a consistent basis as far back as possible.

On a related note, the BLS should consider changing its methodology



for pricing medical care along the lines of our prototypical index for the price of cataract treatment. Specifically, it should move toward pricing the treatment of diagnoses rather than a fixed bundle of medical goods and services.

Much interesting work also remains to be done in the realm of basic research. On the empirical front, the gaps in evidence are obvious and widespread. More extensive investigation using longer sample periods should be undertaken of the within-strata effect. An attempt should be made to develop standard errors for existing estimates of the across-strata effect. Strategies for implementing non-Laspeyres aggregation schemes in real time should be explored. On the theoretical front, our sense is that there is further work to be done in spelling out the consequences of heterogeneity in preferences among households for the construction of aggregate price indexes.

To facilitate all this research, the BLS should assign a high priority to the further development of a longitudinal database—recently established but still relatively inaccessible—housing all of the information used to construct the CPI each month, including the individual price quotes and comprehensive data on item substitutions and quality adjustments. An easy-to-use dataset could serve as a laboratory for testing new theories and methods, and hence redound rather quickly to the benefit of the CPI. If there are concerns about confidentiality associated with such a database, then perhaps non-BLS researchers could be limited to on-site use of the data. Much of the excellent research performed by BLS staff has been undertaken *despite* the lack of ready access to detailed data, with the consequence that a considerable portion of our evidence on key questions is based on sample periods of three years or less. In the future, it should be the case that additional research is performed *because of* ready access to such data.

## REFERENCES

- Abraham, K. G. (1995). Prepared statement. In *Hearings Before the Committee on Finance, United States Senate*, Senate Hearing 104–69. Washington: U.S. Government Printing Office.
- Advisory Commission to Study the Consumer Price Index (1995). Toward a more accurate measure of the cost of living: Interim report to the senate finance committee. Washington: Senate Finance Committee. Mimeo.
- Aizcorbe, A. M., and P. C. Jackman. (1993). The commodity substitution effect in CPI data, 1982–1991. *Monthly Labor Review* 116(December):25–33.
- Armknrecht, P. A. (1984). Quality adjustment in the CPI and methods to improve it. *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, 57–63.

- \_\_\_\_\_. W. F. Lane, and K. J. Stewart. (1994). New products and the U.S. CPI. Mimeo, Bureau of Labor Statistics.
- \_\_\_\_\_. B. R. Moulton, and K. J. Stewart. (1995). Improvements to the food at home, shelter, and prescription drug indexes in the U.S. Consumer Price Index. Bureau of Labor Statistics. Working Paper 263.
- \_\_\_\_\_. and D. Weyback. (1989). Adjustments for quality change in the U.S. Consumer Price Index. *Journal of Official Statistics* 5(2):107–123.
- Bryan, M. F., and S. G. Cecchetti. (1993). The Consumer Price Index as a measure of inflation. *Economic Review*, Federal Reserve Bank of Cleveland, 29(4):15–24.
- Bureau of Labor Statistics (1992). *BLS Handbook of Methods*, Washington: Government Printing Office. Bureau of Labor Statistics Bulletin 2414. September.
- \_\_\_\_\_. (1995). Seminar on current CPI issues. Mimeo.
- Congressional Budget Office (1995). Statement of June E. O'Neill, Director, on the Consumer Price Index, before the Committee on Finance, United States Senate, March 13, 1995. Washington.
- Cutler, D., M. B. McClellan, J. Newhouse, and D. Remler. (1996). Are medical care prices declining? Cambridge, MA: Harvard University. Mimeo.
- Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics* 4:115–145.
- \_\_\_\_\_. (1987). Index numbers. In *The New Palgrave: A Dictionary of Economics*, Vol. 2, J. Eatwell, M. Milgate, and P. Newman (eds.). London: The Macmillan Press, pp. 767–780.
- \_\_\_\_\_. (1995). Price and volume measures in the system of national accounts. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 5103.
- Dixit, A. K., and J. E. Stiglitz. (1977). Monopolistic competition and optimum product diversity. *American Economic Review* 67(3, June):297–308.
- Duggan, J. E., R. Gillingham, and J. S. Greenlees. (1995). Housing bias in the CPI and its effects on the budget deficit and the social security trust fund. Washington: U.S. Department of the Treasury, Office of Economic Policy. Mimeo.
- Fisher, F. M., and Z. Griliches. (1995). Aggregate price indices, new goods, and generics. *Quarterly Journal of Economics* 110(1, February):229–244.
- \_\_\_\_\_. and K. Shell. (1972). *The Economic Theory of Price Indices*. New York: Academic Press.
- Fisher, I. (1922). *The Making of Index Numbers*. Boston: Houghton Mifflin Co.
- Fixler, D. (1993). The Consumer Price Index: Underlying concepts and caveats. *Monthly Labor Review*, December, 3–12.
- Gillingham, R. F., and J. S. Greenlees. (1987). The impact of direct taxes on the cost of living. *Journal of Political Economy* 95(4, August):775–796.
- Gordon, R. J. (1990). *The Measurement of Durable Goods Prices*. Chicago: University of Chicago Press.
- \_\_\_\_\_. (1993). Measuring the aggregate price level: Implications for economic performance and policy. In *Price Stabilization in the 1990s: Domestic and International Policy Requirements*, K. Shigehara (ed.). Macmillan, London, pp. 233–268.
- Griliches, Z., and I. Cockburn. (1994). Generics and new goods in pharmaceutical price indexes. *American Economic Review* 84(5, December):1213–1232.

- Hausman, J. A. (1994). Valuation of new goods under perfect and imperfect competition. Cambridge, MA: National Bureau of Economic Research. NBER Working Paper 4970. Forthcoming in *The Economics of New Goods*, T. Bresnahan and R. J. Gordon (eds.), Cambridge, MA: National Bureau of Economic Research.
- Lebow, D. E., J. M. Roberts, and D. J. Stockton. (1994). Monetary policy and "the price level." Washington: Federal Reserve Board. Mimeo.
- Mason, C., and C. Butler. (1987). New basket of goods and services being priced in revised CPI. *Monthly Labor Review*, January, 3–22.
- Moulton, B. R. (1993). Basic components of the CPI: Estimation of price changes. *Monthly Labor Review*, December, 13–24.
- , and K. E. Smedley. (1995). A comparison of estimators for elementary aggregates of the CPI. Washington: Bureau of Labor Statistics. Mimeo.
- National Research Council. (1995). *Measuring Poverty: A New Approach*, C. F. Citro and R. T. Michael (eds.). Washington: National Academy Press.
- Nordhaus, W. D. (1994). Do real output and real wage measures capture reality? The history of lighting suggests not. Cowles Foundation Discussion Paper No. 1078. Forthcoming in *The Economics of New Goods*, Timothy Bresnahan and Robert J. Gordon (eds.). Cambridge, MA: National Bureau of Economic Research.
- Pakes, A., S. Berry, and J. A. Levinsohn. (1993). Applications and limitations of some recent advances in empirical industrial organization: Price indexes and the analysis of environmental change. *American Economic Review Proceedings* 83(2, May):241–246.
- Pollak, R. A. (1989). *The Theory of the Cost of Living Index*. New York: Oxford University Press.
- Randolph, W. C. (1988). Housing depreciation and aging bias in the consumer price index. *Journal of Business and Economic Statistics* 6(3, July):359–371.
- Reinsdorf, M. (1993). The effect of outlet price differentials on the U.S. Consumer Price Index. In *Price Measurements and Their Uses*. M. Foss, M. Manser, and A. Young (eds.). NBER Studies in Income and Wealth, Vol. 57. Chicago: University of Chicago Press.
- . (1996). Formula bias and seller substitution bias in the U.S. CPI. Washington: Bureau of Labor Statistics. Mimeo.
- , P. Liegey, and K. Stewart. (1995). New ways of handling quality change in the U.S. Consumer Price Index. Washington: Bureau of Labor Statistics. Mimeo.
- , and B. R. Moulton. (1996). The construction of basic components of cost of living indexes. Forthcoming in *The Economics of New Goods*, T. Bresnahan and R. J. Gordon (eds.). Cambridge, MA: National Bureau of Economic Research.
- Scherer, F. M. (1993). Pricing, profits, and technological progress in the pharmaceutical industry. *Journal of Economic Perspectives* 7(3, Summer):97–115.
- Shapiro, M. D., and D. W. Wilcox. (1996). Generating Non-Standard Multivariate Distributions with an Application to Mismeasurement in the CPI. Cambridge, MA: NBER Technical Working Paper 196.
- Spence, M. (1976). Product selection, fixed costs, and monopolistic competition. *Review of Economic Studies* 43(2, June):217–235.
- Stockton, D. J. (1995). The Consumer Price Index and public policy. Washington: Federal Reserve Board. Mimeo.
- Triplett, J. E. (1988). Price index research and its influence on data: A historical review. Washington: Bureau of Economic Analysis. Mimeo.

---

## Comment

JOHN S. GREENLEES  
Bureau of Labor Statistics

---

The paper by Shapiro and Wilcox is a particularly careful and thoughtful summary of the issues and state of knowledge concerning potential biases in the Consumer Price Index (CPI). There is a valuable emphasis on the difficulty of some of the problems involved in eliminating those biases. I have only a few reservations about the reasoning in the paper, and after reviewing these I will devote much of my time to providing an "insider's view" of the issues raised by the authors.

In my opinion, the major limitation of the paper is that it gives too little attention to the basic question of *why* the authors are trying to quantify the CPI bias. The meticulous effort to construct probability distributions begs the question of how or why these distributions might be used. Although the paper's title refers to both the causes and consequences of CPI imperfections, the authors give much more attention to the former than to the latter.

My purpose in mentioning this point is that I believe if more attention had been given to the purpose of measuring CPI bias, a different estimated distribution of total bias would have been obtained. The most important example of this point concerns social security. As the authors recognize, the indexation of federal taxes and spending is the most obvious reason that CPI mismeasurement matters, and social security is the most important component of that indexation. But if the question had been asked, "What is the bias of the CPI relative to a cost-of-living index for social security recipients?" it would have been natural to consider such issues as whether the expenditure weights used in the CPI (which are based on the 80% of the U.S. population that is civilian, urban, and noninstitutional) accurately reflect the purchasing patterns of the elderly.

This is an issue that has been raised by such groups as the American Association of Retired Persons and the National Council of Senior Citizens. It is a legitimate index-number issue, and one on which there is some limited evidence. The BLS produces an experimental index we call the CPI-E, with expenditure weights based on the spending of consumer units with heads aged 62 or over. The CPI-E has many limitations, but it is certainly suggestive that it has risen more rapidly than the official CPI in recent years. One way of looking at this issue would be to define another component of CPI bias, in addition to those listed by Shapiro

and Wilcox's, that would be the bias from using the wrong population target. Whether or not our estimate of this bias would be negative, it would at least contribute a component of variance to the authors' aggregate probability distribution.

Some also argue that social security recipients do not benefit from new goods such as personal computers and cellular phones, and new outlets such as warehouse clubs, to the same extent that the general population does. Again this should contribute some variance to the authors' component bias distributions. Finally, note that parallel considerations would arise if we took as the objective to measure the bias of the CPI relative to the best index for some other purpose: indexing supplemental security benefits, for example, or indexing individual income-tax brackets.

There is another issue that the authors address only in passing. The CPI is measured exclusive of (or conditional on) numerous prices or quantities that would be included in a comprehensive cost-of-living index. These include such factors as crime levels and public-school quality that would point in the direction of a downward bias in the CPI as a measure of the true cost of living. The paper's only discussion of a more comprehensive cost-of-living index, however, is in reference to the question of whether employer-financed health insurance should be included. Once again, the (necessarily) limited scope of the CPI should constitute another component of potential bias, possibly with a wide confidence interval and perhaps with a negative or zero point estimate.

Of course, it is convenient from the BLS point of view that the authors implicitly take as given the CPI's scope and population target. Their analysis provides evidence that we can use to evaluate our success in approximating our own measurement objective. On the other hand, I still am dubious about the need for new best "guesstimates" on sources of bias like new goods, where the evidence is extremely weak and where there is no clear guide as to what should be done. I have to be concerned that the bias estimates will be used by the general public as a measure of the BLS's competence, and that the authors' caveats may not be given sufficient attention.

Regarding the authors' specific estimates of bias, I have relatively little to say. As I have noted, I think there should be additional sources of uncertainty for most specific purposes. My colleague Tim Erickson has suggested, and I agree, that it would be helpful to present some results showing the sensitivity of the total bias estimate to the individual component estimates and to the assumptions about covariance. Also, the authors present only the 80% confidence interval on the total bias; it would be useful to have more information, such as the 90% interval.

Quantitatively, I was surprised that despite the huge array of evidence

on across-strata substitution bias and the virtual absence of evidence on new-items bias, the authors assign them the same expectation and the same upper bound of the 90% confidence interval. I find it especially hard to accept that we understand quality bias as well as would be indicated by its confidence interval. Quality adjustment problems are different in every sector, and generalizations from individual product studies to the rest of the CPI are very problematic.

Now for the “insider’s view” of the issues and problems raised by Shapiro and Wilcox. (I emphasize that the comments by no means constitute an official position of the BLS.) Given that the BLS has announced plans for eliminating the remaining “formula bias” in the CPI, there are two categories of problems that we are most concerned with at present.

The first set of issues concern the biases due to across-strata and within-strata substitution. In principle, it should be possible to minimize or eliminate these biases given more data and more time. The BLS is working to expand our experimental family of annual superlative indexes. Bringing a superlative index up to what we would call “production quality” presents a variety of interesting but presumably tractable problems such as determining the length of the expenditure base period and choosing between an annual lagged index and a monthly “real time” index subject to annual revision. Ultimately, the wider use of scanner data should make available current quantity information and permit at least some use of superlative formulae at the within-stratum level.

In contrast to substitution bias, there do not seem to be clear remedies for the biases that the authors classify under the headings of new outlets, new items, and quality change. Without question, we have to increase our emphasis on hedonic regression and other means of quality adjustment. We also need to think about how we can keep our sample of items and goods more current than we do now. I would like to emphasize, however, the size of the problem. In December 1995, we collected about 70,000 commodity and service prices. About 2000 of these prices were for substitute items, and in about 1000 of those cases the substitution was judged comparable. As we move away from our implicit assumption that linking is an unbiased technique for handling substitutions, we have to identify an alternative technique that can deal with such a large volume of prices.

In the CPI program we only have about 35 commodity analysts, who have only a few days each month in which to make comparability and quality-adjustment decisions. Hedonic regressions are not a panacea, at least not in real time and using actual CPI data as has been our practice. The case sometimes mentioned is that of wired and wireless remote

controls. By the time we could have gathered CPI data on the relative prices of TVs and VCRs with the different kinds of remotes, wired remotes had disappeared from the market. We may have to find a different approach—perhaps a greater use of informed judgment by the commodity analysts, or greater reliance on secondary data. What is clear is that this is of critical quantitative importance; as noted in the 1989 paper by Armknecht and Weyback, movements in the all-items CPI are largely driven by the treatment of substitutions.

In my view, the implications for the CPI of the arguments and evidence reviewed by Shapiro and Wilcox are very fundamental and severe. This can be seen by recalling the authors' discussion of the statistical model of price change used in papers by Brent Moulton and Marshall Reinsdorf to demonstrate "formula bias." In that model, individual prices within a stratum move according to a common trend represented by a term  $\pi_t$ , although there are permanent item-specific and also transitory stochastic deviations from that trend value. The measurement problem is to use the individual price data to estimate  $\pi_{t+1} - \pi_t$ , the movement in the common trend. Despite the sophistication of this model and its value in evaluating different stratum-level index formulas, it is interesting to note how it abstracts from many of the most critical bias issues raised in Shapiro and Wilcox's paper.

In the Moulton–Reinsdorf model, linking could still be a reasonable procedure for dealing with substitution, because nothing in the model indicates that the expected price change of disappearing or appearing items will differ from the rest of the population. By contrast, the arguments about U-shaped product life cycles imply that newly introduced goods will probably decrease in relative price for some period of time. Also, the quality-change arguments reviewed by Shapiro and Wilcox suggest that new product models may enter the sample at systematically lower (or sometimes, as in automobiles and apparel, systematically higher) quality-adjusted prices. Finally, the new-goods bias largely derives from the presumption that the total number of items is increasing, and that this by itself contributes to consumer surplus.

All these considerations lead to the conclusion that even if there is a common trend  $\pi_t$ , the CPI measurement objective has to be something different. That is, the period-to-period change in average quality-adjusted price is probably lower than the trend in prices of items continuing in the sample, and because of the proliferation of goods and varieties the change in the cost of living is lower still. Since tracking the prices of continuing items is still the basic logic of CPI sampling and estimation, we may need to consider fundamental changes in the way we view and process the data collected in our CPI item sample.

In closing, I would like to invite the academics in the audience to consider suggesting to their graduate students that they spend time at the BLS studying CPI measurement issues. The issues are numerous and interesting, and we need as much help as we can get. In particular, I hope that in addition to intensive studies of particular product categories, research can lead to progress on new general solutions to the problems outlined by Shapiro and Wilcox.

## Comment

ZVI GRILICHES  
Harvard University

This is a very good survey. I have no fundamental disagreements with it, and I expect that the CPI Commission will borrow heavily from it for its final report.

I have only a few comments on some places where my emphasis would be different and on topics which were not covered by this paper: What *should* we index and how?

Shapiro and Wilcox ask whether there is a bias in the CPI and conclude that indeed there is. Their overall estimate is slightly lower than the Commission's "interim" one, but the latter falls comfortably within their "uncertainty" range. The question whether the bias today is larger than it was in the past is impossible to answer with the currently available data. Only one component of the overall potential bias in the CPI is relatively new, the "formula bias" introduced in 1978 by the move (itself a major advance) to probability sampling. This is a by-product of the sampling procedure, which is biased in favor of picking lower-priced items (volume sellers) at the time of introduction, replacement, or sample revision.

Another technical comment is their (and our) estimate of the between-strata substitution effects. The major source of data here is the BLS computations that are based on 207 (strata) and 56 (primary sampling areas)  $\approx$  11,000+ cells. But there is no sense in which there should be substitution *across* areas in the same sense as there is across commodities. It is not clear what amount of noise is introduced by this cross-classification and how it may affect their and our estimates.

On the within-strata substitution effects, I would make two comments:

1. The underlying price elasticities at this level are surely higher than unity (e.g., as far as different brands of cereal, or different cuts of



meat, or different types of men's suits are concerned). This would make such effects larger, and a geometric mean would not capture them all.

2. Implicitly, by choosing outlets based on the POPS survey and items based on estimated sales at the store level, the weights used at this level are more up-to-date—lag less—than those used in the across-strata aggregation. This should reduce the resulting biases somewhat.

The net effect of these two conflicting forces is unclear, unfortunately.

The big items are new goods and quality change, which is really mostly new goods within the current CPI procedures. There are two problems here, which I will call: (1) "too late" and (2) "too particular." A new good which does not fit an existing stratum definition will not appear in the index at least until the next full 10-year revision. Thus, neither the PC nor the VCR was in the CPI before 1987, even though they came to the market in serious numbers about a decade earlier and experienced enormous price declines in the interim. Moreover, once inside, a chosen model is not changed until it is rotated out (on average after 3 years in the sample—or is it 5 years in steady state?) or disappears and has to be replaced. If "old" items had the same price history as new ones, this wouldn't matter, but for many (most?) durable goods, items whose market share is declining do not reduce their prices accordingly, but rather exit. As the result, the observed price history is not representative of a more inclusive "average" price history. Also, the current rotation policy will miss a whole generation of items where turnover is rapid, as in computer models, and underweight those models that it will catch, since they will not get full weight until they are at least 5 years old.

But the big problem is that the new models are rarely compared with the old. Since the CPI does *not* use hedonics for PCs, it has no way to evaluate and incorporate the implicit price decline that happened from the appearance, successively, of the 386, 486, and Pentium PC models. All of that is "linked out," since old models, by the time they have entered the index, do not decline much more, but rather disappear.

This is not just a problem for fancy technological items such as cellular phones and satellite dishes, but also for the treatment of grapes and raspberries from Chile in January, and the impact of Walmart, of new bakeries, and new sources (and types) of fish made possible by the decline in real transport costs and trade barriers.

Thus, I believe that the combined quality-change–new-goods bias may be significantly higher than is indicated by them. I have been trying to produce examples of new "bads" to counterbalance this conclusion, with little success to date.

Their example of cataract surgery, which I hope that they'll expand on, illustrates this point. They show a sharp reduction in inputs used (primarily hospital services) per cataract surgery. Nothing in the current procedures or in the proposed revisions to them for the 1998 revision will eliminate such biases. In fact, it seems reasonably clear (see Fuchs 1996, Cutler and McClellan, forthcoming) that the bulk of the increase in health costs in the last decade or so has come from an increase in the quantity of inputs used in the health sector, from additional new procedures, and not primarily from a rise in factor prices. But more importantly, neither Shapiro and Wilcox nor Cutler and McClellan are able to estimate the implicit gains in healthiness (consumer surplus) and the associated implicit declines in the "real" price of health per constant quality unit. My guess is that these have been quite large and would dwarf the other estimates in this paper.

I will skip the discussion of how the CPI could be improved in the intermediate run. That is a most important issue that the Commission is currently grappling with. But I do want to note and comment briefly on three major questions that are only alluded to in Shapiro and Wilcox's paper: (1) Who is the "representative" consumer? How representative is she really? (2) What is "living" and what should be included in its "cost"? Also, what is the "income" that it should be compared with? And (3) what should be "indexed," and how?

A major problem with the conceptual model for the CPI is its reliance on the paradigm of a representative or average consumer. Consumers are very heterogeneous, with widely differing tastes. A study of quality change, of the gains from the introduction of new products (and the losses from the disappearance of "old" products) forces us to pay direct attention to such taste, income, and opportunity differences, something that we are ill equipped to do. [See Fisher and Griliches (1995) and Griliches and Cockburn (1994) for more discussion of such issues.] There are two basic points to come out from such considerations:

1. We may need to distinguish between different groups of consumers in making our "bias" computations. Not all of them may be "us," have our tastes.
2. "Quality" is rarely a sharply defined concept, to which we can attribute a fixed valuation. A particular new product will be valued differently by different people, and this value will change over time, as knowledge spreads, complementary inputs are developed, and its use spreads to lower-value activities. Thus, how it is evaluated depends crucially on the time at which it is introduced into the index, and a full, correct evaluation of its role is impossible without a more

or less complete history which would allow us to integrate the area under the revealed demand curve for it.

The main group of "others" that has been singled out in recent discussions are the retired elderly. Of course, we all will be these "others," sooner or later. As of the moment, I believe that the biases we discussed apply to them also, or to say it differently, I doubt that their "true" CPI has risen faster. If anything, it is likely to have gone in the opposite direction. The elderly are "beneficiaries" of two flaws in the current CPI, one of concept and the other of measurement. The first refers to the treatment of home ownership as "rent equivalence" in the CPI. That is a useful approximation, but it does not go further and include the resulting capital gains in the definition of income to be indexed. In other words, homeowners are hedged against housing price inflation, and there is no reason to "compensate" them for such price rises. (The total effect of this objection is muddled somewhat by the fact that the current rent-equivalence concept includes maintenance and repair costs within its definition, which would still be there if pure ownership costs were eliminated from it.) Also, the other main component where their consumption weights are higher, medical care, is the area with much quality change and new goods (from bypasses and hip replacements, to Prozac, Zantac, and many other drugs) and almost no adjustment for it in the official indexes. Nevertheless, the elderly may have a case, but it is outside the current concept of the CPI.

Some of the problem can be seen in the treatment of taxes and fringe benefits. Currently, if my employer reduces his contribution to my health plan, this will not show up as a rise in medical insurance prices to me. It will be an increase in my expenditures which will only show up in an increased weight given to medical insurance payments in the next CPI revision. But in fact, I had a decline in my real wage, though the data may show an increase in my "real consumption," as my expenditures rise and they are not "deflated" away. Similarly, a rise in Medicare expenditures may lead to a rise in Medicare premiums, which are treated as a tax rather than as a price. This too would not show up in the CPI even though it would lead to a decline in the real income of Medicare recipients. It is tempting to advocate that all such fringe benefits and services in kind be imputed to total consumption, but finding appropriate prices for them will not be an easy task.

The health area raises a variety of conundrums. Consider an unanticipated "breakthrough" that allows one to live another 6 months, aware but largely immobile, at a cost of, say, \$2000 per day. We are better off, since we now have an option we didn't have before. We could decline it,

but we (and our children) are unlikely to be willing to do so. We will follow this technological imperative, remain alive, and feel and be impoverished at the same time! Even if much of it is paid by insurance, all of the various health options available today to the elderly have put significant strains on their budgets. Of course, the alternative is worse, but it is not inconsistent of them to complain about the "rising cost of living," which does rise with age, even if the CPI is not rising, since it is defined for an average, unchanging, unaging consumer unit.

The CPI is only one tool in the measurement of our standard of living. For a complete accounting we need estimates of total consumer expenditures, including imputations of various in-kind services; estimates of total incomes, including a valuation of fringe benefits; estimates of asset accumulation and the associated capital gains and losses; and also (and that is the most difficult) estimates of changes in our environment, physical, economic, and social. Many of our expenditures, including some of the health expenditures discussed above, may not be producing net *increases* in utility but rather responding to certain deteriorations in the environment, such as rising crime, or new diseases. One measuring instrument such as the CPI cannot solve all such problems, but it is worth bearing in mind both what it does do well and what it cannot do given our current state of knowledge as we hand out "grades" to it.

Finally, a few words about indexing. In principle, private contracts could be indexed to anything, even to CPI-1%, and the parties would still be exposed to some "basis risk," i.e., that the relevant concept for them does not move exactly as the index formula that they have settled on. But when we, as a society, decide to index a certain stream of payments, we need to be clearer as to why and how we are doing it. The simplest rationale is to compensate for *monetary* inflation, where those on nominal contracts may be losing while others in society are gaining. This is a redistribution argument, where we, who are receiving "flexible" wages, tax ourselves to compensate those whose pensions have been fixed in nominal terms. The point that I am making is that in such a context there are "gainers" who have something to give up to the "losers." But many changes in the CPI are not of this form. Consider an OPEC-induced rise in energy prices. This is an external tax imposed on our economy. We are *all* poorer for it. There is no way in which all of us can be compensated for it. Moreover, it is not clear that one group, say the elderly, are more deserving and should be fully compensated for it. That implies that real income of the rest of the population should fall even further! Why shouldn't the burden of such changes be shared somehow?

In short, what should be used in such compensation arrangements is a

price index based on the domestic value-added components of the various consumption goods. In fact, I would suggest the median wage (including fringe benefits) or per capita nominal income as a more appropriate concept for these purposes than the CPI.

The other point I want to make is that the discussion has concentrated on the effect of the CPI on the rate of growth in entitlements, not on their levels. But the fundamental question to be asked is whether the *levels* are right. Adjusting rates of growth will not get us, necessarily, to the right levels. What is the minimum safety-net level at which we want to protect the elderly or the disabled? Is the current level too high or too low? Only then does it make sense to worry about whether the escalation formulae are distorting these levels. They may have first-order budget implications, but in terms of what is the right social policy to pursue, they are distinctly second-order questions.

## REFERENCES

- Cutler, D., and M. McClellan. What is technological change? In *The Economics of Aging*, David Wise (ed.). University of Chicago Press, forthcoming.
- Fisher, F. M., and Z. Griliches. (1995). Aggregate price indices, new goods, and generics. *Quarterly Journal of Economics* 110(1):229–244.
- Fuchs, V. (1996). Economics, values, and health care reforms. *American Economic Review* 86(1):1–24.
- Griliches, Z., and I. Cockburn. (1994). Generics and new goods in pharmaceutical price indexes. *American Economic Review* 84(5):1213–1232.

## Discussion

---

The authors first responded to some of the points raised by the commentators. Matthew Shapiro stressed that the paper's goal was the relatively narrow one of clarifying the methodology used by the Bureau of Labor Statistics and quantifying the biases in the CPI. He agreed, though, that the importance of the various issues raised by the paper, and the appropriate responses, depend on what the price index is to be used for: If the primary objective is to find the ideal index for social security benefits, for example, the issues are quite different than if the main purpose is improved historical analysis. David Wilcox defended the focus of the paper by arguing that their findings would be useful in many contexts; for example, although social security indexation per se was outside the scope of their research, the probability distributions they constructed for the bias in the CPI could help analysts calculate the risks of over- or underindexation inherent in any specific indexation scheme. Shapiro

added that it is important to know the extent of bias in the CPI so that better decisions can be made about how available resources for improving the index should be allocated.

Greg Mankiw asked whether the bias in the CPI detected by the authors had any specific time-varying patterns. For example, are there cyclical fluctuations or long-term trends in the bias, or is subtracting 1% from each historical observation of CPI inflation an adequate correction? Zvi Griliches replied that this question is difficult to answer, as much of the uncertainty about the bias comes from quality change and there is little direct evidence on how the rate of quality change has evolved over time. He thought that improving the sampling program beyond the probability sampling introduced in 1978 would be an important step toward reducing bias due to new items and quality change; he argued also that the rotation policy should be adjusted to allow the CPI to capture price changes associated with normal product life cycles.

David Wilcox noted that the paper by Aizcorbe and Jackman provides some evidence for year-to-year variation in the CPI bias, at least in the component arising from across-strata substitution. These changes are not necessarily trivial: For example, between 1990 and 1991 the Aizcorbe–Jackman estimates of the substitution bias dropped by half a percentage point, a large enough change conceivably to affect the decision process of the Federal Reserve Board and other policymakers. Wilcox emphasized the importance of solving the data availability problems that prevent the calculation of year-to-year variation in other components of the bias.

Shapiro added that Robert Gordon had found evidence for a larger bias in the earlier part of the sample. Shapiro also raised the question of whether there is some covariation between the magnitude of the bias and the level of inflation. Since it appears that relative price variability increases with inflation, it could be the case that higher inflation increases the substitution bias. There is no direct evidence on this point, however.

Olivier Blanchard wondered whether the improvements to the index suggested by the paper would be feasible on a real-time basis. He suggested that before departing from the Laspeyres index we should assess whether alternative indexes could be constructed with readily available data. Wilcox answered that most of the information needed to construct alternative indexes currently exists, although in some cases timeliness is an issue. He pointed out that, for example, the across-strata substitution bias could presently be dealt with to some degree by utilizing a geometric-means index, which would be an exact cost-of-living index if utility is Cobb–Douglas. He argued that this assumption is theoretically more attractive than those underlying the Laspeyres index.

Several participants pursued the question, raised by both discussants, of what the CPI is for. Mankiw elaborated on the idea, discussed by Griliches, of indexing social security benefits to the median wage. He argued that, from the perspective of social risk sharing, it is highly unlikely that the CPI is the optimal index for retirement benefits. Indexes like the median wage, or nominal income per person, would lead to greater intergenerational risk sharing, while avoiding measurement issues like quality adjustments. Mankiw expressed the view that the current debate about the CPI was really a political debate about how, and by how much, to cut real entitlements.

Robert Shiller said that, according to his own research, the construction of wage-based indexes involved problems of its own, including the usual difficulties of aggregating across people and across occupations. Peter Diamond called attention to the fact that the CPI plays no role in the level of benefits received by a newly retired person; instead, the new retiree's benefits are determined by what amounts to a mixture of wage indexation and no indexation. He also raised the idea, which he attributed to Robert Merton, of indexing retirement benefits to aggregate consumption; however, Diamond noted, the exact relation of benefits to aggregate consumption should in principle depend on how society chooses to allocate the "risk" of increasing consumption needs implied by longer life spans. Griliches added that neither wage- nor consumption-based indexation schemes deal adequately with the fact that the "cost of living"—taken literally, to include medical expenses—depends very much on age and on largely unforeseeable developments in medical technology.

Ben Bernanke asked about the relationship between the bias in the PCE deflator and the CPI bias. Wilcox replied that the PCE is constructed by reaggregating components of the CPI, and so inherits many of its sources of bias. However, since now Fisher ideal formulas are being used in the construction of national income accounts, at least the PCE deflator will no longer suffer from the across-strata substitution bias at the very aggregate level.