

Contents

The Consumer Price Index

Mark A. Wynne and
Fiona D. Sigalla

The consumer price index (CPI) is probably the most closely watched indicator of inflation in the U.S. economy. In this article, Mark Wynne and Fiona Sigalla explain the construction of the CPI and evaluate some of its potential shortcomings as a measure of inflation. Specifically, they examine the discrepancies that arise between the CPI and the true cost-of-living index as a result of improvements in the quality of goods, the introduction of new goods, substitution on the part of consumers between different goods and retail outlets, and the difficulty of measuring the prices actually paid by consumers for the goods they purchase.

The authors review the literature that quantifies these discrepancies, with the objective of estimating the magnitude of the overall bias in the CPI. Wynne and Sigalla argue that, in fact, remarkably little is known about the extent or significance of the overall bias in the CPI. They conclude that biases in the CPI cause it to overstate inflation by no more than 1 percent a year, and probably less.

Mark A. Wynne

Senior Economist
Federal Reserve Bank of Dallas

- 302 -

Fiona D. Sigalla

Associate Economist
Federal Reserve Bank of Dallas

The Consumer Price Index

Without the notion of price there would be no economic science. The concept is of absolutely central significance. It is not as easy and trivial a concept as it appears to be at first sight. A satisfactory measurement of price is, as a consequence, a difficult undertaking, and it is not surprising that price statistics, abundant as they are, have to be approached with utmost caution.

—Oskar Morgenstern
*On the Accuracy
of Economic Observations*

The consumer price index (CPI) is one of the most closely watched price statistics published by the federal government. As an indicator of inflationary trends in the U.S. economy, the monthly CPI report undergoes intense scrutiny. For example, an upsurge in the rate of increase in the CPI in early 1993 prompted the Federal Reserve to lean toward raising interest rates in its policy deliberations. When the inflationary surge proved short-lived, the Fed reverted to a neutral stance.¹ Movements in the CPI also have a direct impact on the pocketbooks of many Americans. For instance, the annual increase in the CPI determines the rate at which the nominal payments to Social Security recipients will rise each year. The tax brackets that determine the income tax liability of most workers are indexed to the CPI as well. In this article, we review the construction of the CPI and discuss its potential shortcomings. We also examine how well the CPI measures what it is supposed to measure, namely, changes in the cost of living.

Our primary objective is to review and put into context what is known about the potential biases in the CPI. Inherent in the concept of bias is some notion of the true value of the CPI, which

the actual published value only approximates. The true value of the CPI is considered to be the true cost-of-living index, and so we begin with a discussion of the theory of the cost of living index. We progress to the construction of the actual CPI as it is reported every month, following the description of methodology in the Bureau of Labor Statistics' (BLS) *Handbook of Methods*. In the remainder of the article, we consider how well the CPI approximates the true cost-of-living index, paying particular attention to the problems of substitution, quality change, and the introduction of new goods, which are generally considered to cause the CPI to overstate the rate of increase in the cost of living or, alternatively, overstate the rate of inflation.² Some analysts have suggested that these measurement errors are so large that

We are grateful to Marshall Reinsdorf and Jack Triplett for detailed comments on this article. We also received helpful comments from Evan Koenig, Marilyn Manser, Steven Braun, Ken Emery, and Zsolt Becsi. This article is based on a longer paper, entitled "A Survey of Measurement Biases in Price Indexes" (Federal Reserve Bank of Dallas Research Paper no. 9340), that was prepared for the Joint Meeting of the System Committees on Business and Financial Analysis at the Board of Governors of the Federal Reserve System in Washington, D.C., September 22–24, 1993. We thank our discussant at that conference, Ana Aizcorbe, for comments on that paper.

¹ Note that some of this increase has since been revised away as a result of changes in the seasonal adjustment factors used to calculate the seasonally adjusted CPI. We do not review CPI seasonal adjustment procedures in this paper.

² These problems with the CPI are common to almost all official price indexes and have implications that extend beyond the measurement of inflation to such issues as the measurement of real output growth and productivity.

CPI inflation in the 3 to 4 percent range constitutes effective price stability. We argue that there is little evidence to support this belief and that, insofar as the CPI does overstate the rate of inflation, it probably does so by no more than 1 percent annually.

The conceptual basis of the CPI

The conceptual basis of the CPI is the theory of the cost-of-living index.³ The cost of living is a unique concept for each individual and is determined by the individual's preferences for different types of goods and services and the prices at which that person can purchase them. We can think of the preferences of each individual as being characterized by an intertemporal utility function that is defined over all the goods and services consumed today as well as all the goods and services the individual expects to consume in the future. The most general specification of this function includes items that have a price attached to them, such as automobiles, haircuts, and bananas, and items that are typically consumed without a user charge, such as roads, clean air, and leisure. Specifically, we can write

$$(1) \quad U = U'(q_1^1, q_2^1, \dots, q_M^1, q_1^2, q_2^2, \dots, q_M^2, \dots) \equiv U'(q),$$

where q_i^t denotes the quantity of the i th good or service consumed at date t .

Calculating a cost-of-living index requires that all items in the utility function be assigned a price. This task is relatively straightforward for commodities and services that have a user charge for consumption (meaning, they are purchased on markets) but is more difficult for items such as leisure and clean air that are typically consumed without a user charge. It is also difficult for durable goods such as houses and cars, which are typically purchased in one period but yield services to the consumer over several periods.

Because it is impossible to ascertain the price of every single good and service that is valued by consumers, the theory of the cost-of-living index as it is applied to the measurement of consumer prices typically focuses on a narrower set of goods and services, specifically, those that are purchased on markets in some time period. Focusing attention on this narrower set of goods and services, which we will denote as $x^t \equiv (q_1^t, q_2^t, \dots, q_N^t)$, where $N < M$, we can define a cost or expenditure function for the individual as follows:

$$(2) \quad e(p^t, u) = \min_{q_i^t} \sum_{i=1}^N p_i^t q_i^t : U(q^t) \geq u,$$

where $p^t = (p_1^t, p_2^t, \dots, p_N^t)$. The expenditure function gives the minimum cost to the individual of attaining some specified level of utility, u , when faced with a set of prices, p^t , for the goods and services that enter that person's utility function. The true cost-of-living index is then defined on the basis of the expenditure function. Specifically, it is the change in the cost of attaining some base level of utility, u^b , between a reference period, r , and a comparison period, c :

$$(3) \quad \frac{e(p^c, u^b)}{e(p^r, u^b)},$$

where p^c and p^r denote the prices faced by the individual in the comparison and reference periods respectively. If all prices in the comparison period p^c , are twice the reference prices, the index is 2; if all prices in the comparison period are one-half the reference period prices, the index is 0.5. If all prices in the comparison period are the same as prices in the reference period, the index is 1. The CPI is an approximation of this true cost-of-living index.

How the CPI is constructed

The purpose of the CPI is to measure the rate of change in the cost of living for urban consumers. It does this by calculating the average change in the prices paid by urban consumers for a fixed market basket of goods and services of constant quality. Since it is obviously impossible to obtain price data for all consumer transactions in the United States, the CPI is estimated from a

³ For a general review of the theory of the cost-of-living index, see Pollak (1989). For a discussion of the theory of the cost-of-living index as the conceptual basis for the CPI, see Gillingham (1974).

series of samples. These samples are designed with the objective of making the CPI representative of the prices paid by consumers in all U.S. urban areas for all goods and services. The use of a sample, of course, introduces a source of error into the index, but this is more than offset by the errors that sampling eliminates.⁴ Each year, the BLS selects new item and outlet samples for 20 percent of the primary sampling units on a rotating basis, with the intention of capturing new developments in the market for consumer goods through such rotation.

Individual commodities in the CPI are weighted by the share of expenditure on the item, as estimated in the Consumer Expenditure Survey. The outlets from which price quotes are obtained are determined on the basis of information collected in the Continuing Point-of-Purchase Survey, with an outlet's probability of selection being proportional to its share of consumers' expenditure for the good in question. Currently, the CPI is constructed using expenditure shares obtained from the 1982-84 Consumer Expenditure Surveys and is formally defined as

$$\begin{aligned}
 (4) \quad CPI^t &= 100 \times \frac{\sum_i p_i^t x_i^b}{\sum_i p_i^r x_i^b} \\
 &= 100 \times \sum_i \left(\frac{p_i^t}{p_i^r} \right) \frac{p_i^r x_i^b}{\sum_i p_i^r x_i^b} \\
 &= 100 \times \sum_i \left(\frac{p_i^t}{p_i^r} \right) \omega_i^b,
 \end{aligned}$$

where p_i^t is the price of the i th good in the comparison period, t , p_i^r is the price of the same good in the reference period, r , x_i^b is the quantity of the good consumed in the expenditure base period,

b , and $\omega_i^b = \frac{p_i^r x_i^b}{\sum_i p_i^r x_i^b}$ is the share of the i th good

in base period expenditures valued at reference period prices. When the expenditure base and the reference periods coincide, the result is the standard Laspeyres price index formula.⁵ Note that, in general, there is a difference between the base period for the expenditure weights and the numeric reference base period, although at present both are 1982-84 = 100. Indexes are estimated for wage

earners and clerical workers (CPI-W) and all urban consumers (CPI-U). The CPI-W is representative of the buying habits of about 32 percent of the U.S. population, while the CPI-U (which was introduced in 1978) covers about 80 percent of the U.S. population.

The prices used to calculate the CPI are collected from about 21,000 retail and service establishments in eighty-five urban areas across the United States. Data on rents are collected from about 40,000 landlords or tenants, and 20,000 owner-occupants are asked about their housing units. All price information is collected by Bureau of Labor Statistics field agents through visits or telephone calls. The CPI is published monthly, typically about two weeks after the end of the month to which it refers.

The problems that are typically thought to bias the CPI as a measure of inflation or the cost of living are substitution bias, quality adjustment bias, and new goods bias. The sample of goods and services that makes up the CPI is also criticized occasionally for being unrepresentative of the buying habits of all urban consumers. Substitution bias and new goods bias are aspects of this sampling problem, as is outlet substitution bias. The problems that arise from the use of list rather than transactions prices in constructing an index are less frequently discussed in relation to the CPI than they are in relation to the producer price index.⁶ We argue that the list-transactions problem is probably an issue in consumer price measurement, although for different reasons than in producer price measurement.

⁴ For estimates of the standard error of the CPI, see Leaver (1992).

⁵ A Laspeyres price index calculates the price change between two periods by comparing the cost of purchasing in each period the bundle of goods purchased in the initial period. A Paasche price index calculates the price change between two periods by comparing the cost of purchasing in each period the bundle of goods purchased in the final period. For a review of the theory of index numbers, see Diewert (1987).

⁶ For a review of some of the problems with the producer price index, see Wynne and Sigalla (1993).

We now illustrate how these various biases can arise in the CPI and review their quantitative importance.

Substitution bias

The CPI is estimated from a series of samples because of the impossibility of tracking the prices paid by every urban consumer for every purchase he or she makes. The first part of the sampling process is to figure out what it is that people buy and how their expenditures are distributed over the different goods and services that make up consumer spending. At present, the weights used to aggregate the prices of the different goods and services are based on expenditure patterns during the period 1982-84. The CPI is a fixed-weight index, meaning that the weights used to aggregate the prices of the different goods and services are held constant for relatively long periods. (For an alternative approach to evaluating the bias due to the use of fixed weights, see the box entitled "Statistical Approaches to Index Number Construction.") However, because the prices of all goods and services do not change at the same rate, and because consumers can substitute less expensive for more expensive goods, over time the weights used to combine the prices become increasingly unrepresentative of consumers' actual expenditure patterns. For example, if the price of beef rises more rapidly than the price of chicken, consumers will typically buy relatively more chicken and relatively less beef. Spending on beef becomes less important in consumers' budgets, while spending on chicken becomes more important. Because the CPI does not allow for these changing expenditure patterns, it tends to overstate the increase in the cost of living over time. The bias in the CPI due to this phenomenon is known as *substitution bias*.

Substitution bias is probably the best known and most frequently studied problem in the CPI. The most authoritative review to date of measurement problems in price indexes concluded that "estimates of substitution bias that have so far been made indicate that it is extremely small, so small that substitution bias cannot be viewed as an important empirical defect of fixed-weight consumption price indexes" (Triplett 1975, 66). The size of the substitution bias in a fixed-weight

cost-of-living index depends on two things: the extent to which households substitute between goods in response to relative price changes and the extent to which relative prices change over time. Absent either of these factors, a fixed-weight Laspeyres index will give an unbiased estimate of the true cost-of-living index. Thus, if household preferences are of the fixed-coefficient or Leontief type, no substitution occurs in response to relative price changes and the fixed-weight Laspeyres index is equal to the true cost-of-living index. Likewise, if all prices increase or decrease together, relative prices never change and again the fixed-weight Laspeyres index coincides with the true cost-of-living index. It is unlikely, however, that either of these conditions holds in practice.

At the time of Triplett's 1975 survey, the principal studies of substitution bias in the CPI were Noe and von Furstenburg (1972), Christensen and Manser (1976), and Braithwait (1980) (which circulated as a BLS working paper in 1975). The major conclusions of these studies, which are summarized in Table 1, were that

1. the size of the substitution bias was small, probably no more than 0.1 percent a year,
2. the estimated magnitude of the bias was relatively insensitive to the choice of functional form for household preferences,
3. the estimated magnitude of the bias increases with the level of commodity disaggregation, and
4. the size of the bias was greater during periods of high inflation when relative price fluctuations were greater.

Since Triplett's survey, Manser and McDonald (1988) have revisited the problem of substitution bias in fixed-weight Laspeyres-type indexes. Manser and McDonald used consumption data from the National Income and Product Accounts (NIPA) for 101 commodities over the period 1959-85 to obtain two sets of estimates of the substitution bias. First, using nonparametric methods and maintaining the assumption of homothetic preferences (which they test and are unable to reject) they calculate bounds on the size of the substitution bias in the Laspeyres index for the period 1959-85. They calculate a maximum possible bias of 0.22 percent per year and a minimum possible bias of 0.14 percent per year.

Table 1
Studies of Substitution Bias in the CPI

Study	Sample period	Categories covered	Size of bias
Noe and von Furstenberg (1972)	1964-70	All	.03 to .11
Christensen and Manser (1976)	1947-71	Meat and produce (four categories of each)	.1 to .2
Braithwait (1980)	1958-73	All	.1
Manser and McDonald (1988)	1959-85	All	.14 to .22
Reinsdorf (1993)	1980-89	Food and gasoline	.25 to 2
Aizcorbe and Jackman (1993)	1982-91	All	.2 to .27

NOTE: Bias is expressed in percentage points.

Second, using superlative index numbers to calculate changes in the cost of living, they estimate a substitution bias of about 0.18 percent per year over the same period.⁷

More recently, Aizcorbe and Jackman (1993) examined the issue of substitution bias using even more disaggregated data, specifically, the consumer expenditure data used to construct the CPI. They compare measures of price change over the period 1982-91 and arrive at an estimate of the substitution bias over this period of 2.6 index points (an average of 0.2 percent a year) when a fixed-weight Laspeyres index is compared with a fixed-weight Tornqvist index. The estimated substitution bias is somewhat higher (0.27 percentage points a year) when chain-linked index numbers are compared. Aizcorbe and Jackman argue that these estimates are small and conclude by noting the desirability of having some measure of whether the estimated bias is significant in a statistical sense.

The findings of these studies of substitution bias in the CPI are summarized in Table 1.⁸ The bottom line on substitution bias is that this particular form of bias in the CPI is probably relatively unimportant quantitatively, amounting to at most 0.2 percentage points a year. This conclusion is shared by Triplett (1975, 1988) and Gordon (1992), among others. However, recent work by Moulton (1993) shows that the issue is not yet settled. Moulton examined substitution effects within CPI product categories (or *strata*, as they are called by the BLS) and showed that substitution effects

within product categories are in some cases larger than those between categories. This is hardly surprising: it is plausible that consumers should be more willing to substitute between different types of fruit than between, say, fruit and meat. Moulton looks only at data for the period from June 1992 to June 1993, and it remains to be seen how his findings generalize to other sample periods.

Insofar as substitution bias is considered a problem in the CPI, it could be handled in one of two ways. One is to update the expenditure weights more frequently: currently the expenditure weights are updated about once a decade, the most recent revision occurring in 1987, when expenditure patterns from the 1982-84 Consumer Expenditure Survey replaced those from the 1972-73 survey. Updating the weights every year (or every quarter) would constitute a shift from a fixed-weight to a chain-linked or multiweighted index.⁹ The United Kingdom uses this approach in

⁷ A superlative index number is one that provides a second-order approximation to the true cost-of-living index for a large class of preference specifications. See Diewert (1976).

⁸ We do not review the results of two recent studies by Kokoski (1987) and Blisard and Blaylock (1991) that examine the relationship between demographic factors and substitution bias.

⁹ For a discussion of the merits of chain-linked price indexes, see Forsyth and Fowler (1981) and Szulc (1983).

Statistical Approaches to Index Number Construction

Bryan and Cecchetti (1993) adopt a novel approach to estimating the bias that arises from the use of fixed weights in the CPI. The essence of their approach is to estimate the common (inflation) factor in a number of the subcomponents of the CPI and interpret the difference between the estimated common factor and the CPI as an estimate of what they term the "weighting bias" in the CPI. Estimation of a common inflation factor is essentially the same as constructing a price index where the individual prices are weighted on the basis of the strength of their common inflation signal. The estimates of weighting bias they obtain average 0.60 percentage points a year over their full sample, which ranges from 1967:1 to 1992:12, with a standard error of 0.17 percentage points. They also estimate the bias for two subsamples, finding that the CPI overstated inflation by about 0.88 percentage points annually during 1967-81, whereas the bias was essentially zero (-0.07 percentage points with a standard error of 0.13 percentage points) during the 1982-92 period.¹ Insofar as substitution bias is a special form of weighting bias, their findings confirm the earlier results that substitution bias tends to be higher during periods of

high inflation, which typically are also periods of substantial relative price variability.

The Bryan and Cecchetti study is not really comparable to the other studies of substitution bias discussed in our article. As we have already noted, substitution bias is defined as the difference between the CPI as constructed and a true cost-of-living index. Bryan and Cecchetti are interested in the CPI as a measure of inflation rather than as a measure of the rate of increase in the cost of living. In many ways, their paper could be considered to be a contribution to the statistical approach to index number construction that had its origins in work by Jevons and Edgeworth but has since been largely abandoned in favor of the behavioral approach. For further details on the distinction between the statistical and behavioral approaches, see Diewert (1987). Other recent contributions to the statistical approach are Clements and Izan (1987) and Dow (1993).

¹ Bryan and Cecchetti rationalize the possibility of a negative, or downward, weighting bias in the CPI by pointing out that for some time periods and some goods, expenditure shares and price changes might be positively correlated if there is a taste shock that raises demand for the good.

constructing its measure of retail prices. Schmidt (1993) examines the consequences of recalculating the CPI using more recent expenditure patterns and finds that the differences for estimates of annual inflation rates are no more than 0.2 percent. The second alternative is to construct an index using a superlative index number formula, such as the Tornqvist or Fisher index number

formulas, instead of the Laspeyres formula currently used. Superlative index numbers have the desirable property of being exact for a class of utility functions that are second-order approximations for any utility function and are thus less susceptible to substitution bias.¹⁰

Both of these possibilities are now feasible since the Consumer Expenditure Survey has been conducted on an ongoing basis since 1980. Kokoski (1989) describes ongoing BLS research into using superlative and chain indexes to track price movements. However, use of chain indexes entails some sacrifice in terms of the timeliness with which the inflation estimates can be obtained, as it takes

¹⁰ For further details on Fisher and Tornqvist index numbers, see Diewert (1987).

more time to collect and process information on consumer expenditure patterns than it does to collect and process information on prices. Furthermore, depending on the behavior of relative prices, chain indexes may be just as prone to substitution bias as fixed-weight indexes and in some cases, more so.¹¹ Note that the Bureau of Economic Analysis (BEA) recently introduced chain-linked superlative price indexes for gross domestic product (GDP) and its major components.¹²

Outlet substitution bias

Once the BLS has determined how consumers allocate their expenditures across different goods and services, it then has to determine which varieties of the goods will be priced for the CPI. For example, suppose the BLS determines that urban consumers allocate some fraction of their expenditures to buying apples. The BLS then has to decide which varieties of apples it will price for calculating the CPI (in other words, Granny Smith, Golden Delicious, or whatever) and also where it will obtain these prices. The choice of outlet where prices are obtained is potentially very important. In an early study of the postwar growth performance of the U.S. economy, Denison (1962) argued that the price statistics used to deflate consumer spending were biased upward because the BLS typically relied on higher priced stores to obtain price quotes for various items in the CPI, neglecting the postwar shift in consumer spending patterns toward lower cost retailers and away from more traditional higher priced outlets. Persistent price dispersion in the retail market is essential for consumer gains from switching outlets not to be reflected in the CPI. Such persistence may arise due to market disequilibria or information costs. An index that tracks prices at incumbents is biased if quality-adjusted prices at incumbent retailers fail to decline to match those of the new retailer.

The BLS looked into this potential source of bias in the CPI nearly thirty years ago and concluded at that time that it was not a problem, claiming that the sample of outlets used for obtaining price quotes was quite representative of how consumers actually shopped.¹³ The concern that there might be an "outlet substitution bias" in the CPI has recently come to the fore once more because of the rapid growth in low-cost, high-

volume discount stores such as Wal-Mart, Sam's Club, Price Club, and so on during the 1980s. Further examples of this trend are the growth in home shopping and the increasing market share of factory outlet stores.¹⁴

Table 1 reports the estimates of the "outlet substitution bias" obtained by Reinsdorf (1993), since bias in the CPI as a result of consumer substitution toward lower price retail outlets is in some sense analogous to the standard substitution bias in a fixed-weight index as a result of commodity substitution. Reinsdorf offers two types of evidence on the size of the outlet substitution bias in the CPI. First, he compares the prices at incoming and outgoing CPI retailer outlet samples, and second, he compares the evolution over time of unlinked sample average prices and their (linked) CPI component counterparts.¹⁵

Comparing prices at new and old outlets, Reinsdorf obtains an estimate of upward bias in the food-at-home component of the CPI of 0.25 percent a year and a comparable figure for the motor fuel component. Reinsdorf notes that this figure may overstate the true size of the bias if average quality declines along with average prices. Reinsdorf's second test compares the growth of the Average Price series published by

¹¹ For details, see Forsyth and Fowler (1981) and Szulc (1983).

¹² For details, see Young (1992) and Triplett (1992a).

¹³ The issue of outlet substitution was also discussed by the Stigler Committee (NBER 1961, 58) and, in more detail, by Backman and Gainsbrugh (1966, 29-31). Backman and Gainsbrugh cite studies by the BLS and Willard Arant that put the magnitude of this bias at no more than a couple of tenths of a percentage point a year.

¹⁴ See, for example, Helliker (1991).

¹⁵ When new item-outlet combinations are rotated into the CPI, all the difference in price between the old item-outlet combination and the new item-outlet combination is assumed to be due to a difference in quality. This is the way in which the price quotes that go into the CPI are "linked" together. By contrast, for the Average Price program, all of the difference in price between old item-outlet combinations and new item-outlet combinations is assumed to be due to differences in price. Thus, the Average Price series are said to be "unlinked."

the BLS with that of the corresponding components of the CPI.¹⁶ For food during the 1980s, various components of the CPI rose more rapidly than the corresponding Average Price series, yielding an implicit estimate of outlet substitution bias of almost 2 percent a year. While quality-adjusting the Average Price food indexes might reduce some of the difference with the CPI, it would not eliminate it completely since the willingness of consumers to shift to low-cost retailers presumably indicates that the quality difference must be more than compensated for by the better prices. For gasoline the estimated bias is 0.3 percent per year. Reinsdorf notes that the estimates obtained from a comparison of Average Prices with the CPI should be taken as upper bounds on the amount of substitution bias, since no attempt is made to control for the possibility that the average quality of outlets may have declined.

In view of the importance that some analysts have attached to Reinsdorf's findings (see, for example, Gordon 1992 and *The Economist* 1993), it is important to be explicit about the caveats that accompany his results.¹⁷ One is that outlet substitution and variety substitution typically occur at the same time. To the extent that this allows the substitution in the CPI sample of, say, a (low-cost) store-brand item for a name brand item, some of the difference found by Reinsdorf may be simply due to switching brands. Also, it is important to try to quantify the quality difference between different retail outlets to get a proper handle on the size of the bias. Popkin suggests

the use of hedonic type regressions of the sort used to make adjustments in the apparel indexes to determine the appropriate quality adjustments for outlets. Fixler (1993) notes that comparison of Average Price series with corresponding CPI series does not provide direct evidence of outlet substitution bias because outlet switching is not the only source of difference between the two series.

Quality bias

If a poll were taken of professional economists and statisticians, in all probability they would designate (and by a wide majority) the failure of the price indexes to take full account of quality changes as the most important defect in these indexes. And by almost as large a majority, they would believe that this failure introduces a systematic *upward bias* in the price indexes—that quality changes have on average been quality improvements (NBER 1961, 35) [emphasis added].

There is a general perception among economists that quality bias is probably the most serious shortcoming of the CPI, causing it to overstate the true rate of increase in the cost of living. This perception probably reflects two other beliefs. First, that the average quality of all goods is increasing over time and, second, that the BLS does little or nothing to take quality improvement into account when calculating its price indexes. As we will see, neither of these statements is completely true.

While it is the case that the quality of most products does seem to improve over time, we should not overlook the obvious examples where quality seems to deteriorate—for example, the increased use of graduate students for undergraduate instruction at major universities, the disappearance of full-service gas stations, and the decline in the quality of in-flight service on some airlines.¹⁸ Second, the BLS has a number of methods for dealing with quality change. The question then becomes how adequately do these methods capture changes in the quality of the products sampled for the price indexes. We will see that in some cases the BLS may in fact *overadjust* for quality change in calculating the price indexes, causing them to understate the true rate of inflation.

¹⁶ Average Prices are estimated from CPI data for a limited number of goods and are calculated as weighted averages of price quotes obtained for a representative variety of the good in question. For further details, see U.S. Department of Labor (1992, 199).

¹⁷ See, for example, Popkin (1993).

¹⁸ On the latter, see, for example, Hirsch (1993), who notes, "Seeking to stanch their financial hemorrhaging, most airlines are putting fewer attendants on board their aircraft. The result: Passengers wait longer for meals and beverages; meal carts clog the aisles longer; dirty trays stack up; and obtaining the little extras of life aloft—a pillow, a magazine, a drink of water—is often a do-it-yourself experience."

How the BLS deals with quality change

The quality adjustment problem in constructing a price index may be stated as follows.¹⁹ Suppose some particular variety of a good is selected for inclusion in the CPI. Suppose further that at some later date, the chosen variety disappears and is replaced by a newer model. How do we compare the price of the old variety of the good in the earlier period with the price of the new variety in the later period? As a concrete example of this problem, consider how we would compare the price of an older model VCR without remote control capability with the price of a newer model VCR that has remote control capability. Since higher quality goods are typically more expensive than lower quality goods, a direct comparison of the prices of the two varieties would result in an overestimate of how much prices have increased.²⁰ The appropriate comparison of prices requires that we compare prices for goods of the same quality. In the VCR example, we need to somehow adjust the price of the model with remote control capability for the improvement in quality that remote control represents.

In practice, the BLS has a number of methods for dealing with quality change. The BLS categorizes the methods it uses to deal with new product varieties or quality change as follows: (1) direct comparison, (2) direct quality adjustment, and (3) imputation.

In direct comparison, if the two varieties are judged to be sufficiently similar in terms of quality in some well-defined sense, all the observed difference in price between the new and old varieties of the product is counted as a price change, and nothing more is done. The risk here, of course, is that some unnoticed quality change is inadvertently being treated as a price change, imparting an upward bias to the CPI.

If, however, the varieties are judged to be different in some meaningful sense, the BLS makes some form of direct quality adjustment using one of a number of different methods.²¹ The simplest case is when the two varieties are observed in some common period, in which case it is possible to obtain overlapping price quotes for the two varieties. In this situation, the ratio of the two prices in the period of overlap is taken as the quality adjustment. For example, if a VCR without

remote control costs \$200 while one with remote control costs \$220, the \$20 price difference between the two varieties can be considered an estimate of the value of remote control capability to consumers. The problem with this adjustment procedure is that we rarely observe overlapping prices for new and old varieties.

An alternative is to make an adjustment based on the manufacturer's production cost differences for the two varieties. In this case, the manufacturer is asked to estimate the cost difference for the two varieties, which is then scaled up to the retail level and added to the price of the old variety in the initial period to obtain an estimated price for the new variety in the previous period. This form of quality adjustment, in use since around 1960, is used most frequently in valuing quality changes in automobiles.²²

Another form of quality adjustment, and the one that serves as the basis of most studies of

¹⁹ This discussion follows Triplett (1988). See also Armknecht and Weyback (1989), Armknecht and Ginsburg (1992), U.S. Department of Labor (1992), Gordon (1992), and Kokoski (1993).

²⁰ Computers seem to be an exception to this rule, and as a result, the problem of quality-adjusting computer prices has attracted a lot of attention. Some of this research is reviewed in Wynne and Sigalla (1993). For more detailed reviews, see Triplett (1989) and also the chapter on computer prices in Gordon (1990).

²¹ Triplett and McDonald (1977) use a hedonic price function to evaluate the judgment of the BLS commodity specialists as to whether specification changes for refrigerators are "major" or "minor." During their sample period (1960-72), there were seventy-six changes in the specifications of refrigerator-freezers priced for the Wholesale Price Index (WPI). In forty-six of these cases the BLS judged that the changes were minor, and no quality adjustments were made. Triplett and McDonald apply a quality adjustment to twenty of these forty-six cases. Of the remaining thirty cases that were judged by the BLS to be major, Triplett and McDonald apply a quality adjustment to twenty-five.

²² Gordon (1981) notes that a problem with this procedure is that a manufacturer may overstate the cost of a quality improvement to disguise some part of an actual price increase, especially when price controls or guidelines are in force.

quality bias in the CPI. entails estimating a hedonic regression that relates the prices of different varieties of a good to the characteristics of the different varieties.²³ The estimated parameters from such a regression provide implicit prices for each of the price-determining characteristics of the good. Thus, when a new variety of a product that differs from the existing variety in terms of some or all of the relevant characteristics becomes available, it is straightforward to make a quality adjustment on the basis of the hedonic regression. Despite the initial promise of hedonic techniques, however, these techniques have not proven to be a panacea in dealing with the problem of quality change.²⁴ Specifically, hedonic techniques are not able to deal with quality changes that are not easily quantified (such as the handling characteristics of a car, the multitasking ability or portability of a personal computer, the quality of care during a hospital stay, or whether an item of clothing is in or out of fashion).

In some situations, the BLS has not yet determined how best to make quality adjustments. New product varieties that can be neither directly compared nor quality adjusted are called *non-comparable*, and in these situations, the BLS estimates the constant-quality price change by imputation.

The most common type of imputation for noncomparable substitutions in the food and services categories consists of setting the rate of price change for the new and old varieties equal to the average price change for similar goods. The implicit assumption that treating noncomparable substitutions in this manner is benign is questionable. For example, it is quite likely that the new product is in the early stage of its product cycle and experiencing substantial price declines, while the products used to impute the price change are

probably mature products that are experiencing price increases.

The inherent difficulty of deciding whether new products, or new varieties of existing products, are comparable to old is the essential source of quality error in the CPI. When a new (higher priced) product is deemed comparable with an old product, some quality change may be incorrectly treated as price change, leading to an upward bias in the CPI. When the new product is deemed noncomparable, some price change may be incorrectly treated as quality change, leading to a downward bias in the CPI.²⁵

How big is the quality bias in the CPI?

Since the Stigler Committee Report in 1961, numerous studies have attempted to estimate the extent of quality bias in the CPI. It is not a simple matter to use the results of these studies to infer the extent of quality bias in the CPI as it is currently constructed because the methods used by the BLS have evolved over time, partly in response to these studies.

Triplett (1975) concluded his survey of quality bias in the CPI by noting that "...the research results imply that no simple estimate of the overall quality error can be made, and, moreover, even the sign of the error is in doubt" (Triplett 1975, 48). In concluding a more recent review of research on price indexes, Triplett was even more emphatic about the uncertainty surrounding the sign and magnitude of the quality bias in the CPI, noting that "...because a number of large CPI components appear quite clearly downward biased, I suspect that the CPI has, if anything, understated inflation in the last several years" (Triplett 1988, 67).

What components of the CPI are downward biased, and how large are the biases? The first component of the CPI that Triplett suggests contained a downward bias was housing. Before 1988, when estimating the shelter cost component of the CPI, the BLS did not take into account the deterioration in housing stock quality as a result of aging and depreciation. Randolph (1988) shows that failure to allow for depreciation in the housing components of the CPI cause the shelter cost indexes to be downward biased by as much as 0.3 to 0.4 percent annually. However, since 1988

²³ Hedonic methods are reviewed in Triplett (1987, 1986) and Griliches (1971). Griliches (1961) is the seminal application of hedonic methods to evaluating the quality bias in the CPI.

²⁴ See, for example, Triplett (1988) and Gordon (1990).

²⁵ See Armknecht (1984, 58).

the BLS has made adjustments to the shelter components of the CPI to allow for aging.²⁶

Another source of downward bias in the CPI is in the indexes for apparel and clothing. Here Triplett cites the almost impossible task of separating taste or fashion changes from quality changes and the strong seasonal pattern in clothing styles as sources of error in the CPI treatment of clothing. One piece of evidence of downward bias in the apparel indexes comes from a comparison of the rates of inflation for infants' and toddlers' apparel, which presumably is less subject to fashion cycles, with those for men's and boys' apparel and women's and girls' apparel. In the 1967-87 period, the index for infants' and toddlers' apparel grew at a 6-percent average annual rate, while those for men's and boys' apparel and women's and girls' apparel grew at 3.4 percent and 2.9 percent, respectively, suggesting that the BLS may have overadjusted for quality change in these categories. Armknecht and Weyback (1989) and Liegey (1993) provide further evidence of the likely downward bias in the apparel components of the CPI.

Triplett also argues that auto prices are downward biased because of the manner in which the BLS treats mandatory safety and pollution devices. Specifically, the BLS treats these changes as quality improvements rather than price increases, as the theory of the cost-of-living index indicates they should be treated.²⁷ Triplett (1992b) presents evidence to suggest that frequently these mandatory changes account for a significant proportion of the estimated quality change in autos.²⁸ Further insight on the quantitative significance of these mandatory changes can be obtained from Gordon (1990), who disentangles the various components of the quality adjustments applied to the auto component of the CPI (see Gordon 1990, Table 8.10). Gordon shows that over the period 1967-83, quality adjustments reduced the average annual rate of increase in the auto component of the CPI from 7 percent to 4.3 percent. Furthermore, safety and environment related quality adjustments account for 2 percentage points of the 2.7 percent difference between the unadjusted and adjusted CPIs for autos.

The candidates for upward bias Triplett identifies are services, used cars, and "miscellaneous other." The problems with measuring the

prices of services accurately, and especially the price of medical care, are well known. We will return to the measurement of service prices in more detail below.

For used cars, the problems stem from the inadequate quality adjustments made to price observations for used cars, although since 1987 the BLS has been making adjustments based on the quality adjustments for new cars. As for the "miscellaneous other" category, Triplett cites the difficulty of dealing with the subtle substitutions between restricted and unrestricted airline fares as a source of possible upward bias in the CPI. Until recently, the BLS priced unrestricted full-fare airline tickets in calculating the CPI, although relatively few people purchased such tickets.

Table 2 summarizes the principal studies of quality bias in the CPI that have appeared in the past five years.

Probably the single most important recent piece of research on the problem of quality change and price measurement is Gordon's (1990) study of durable goods prices. While Gordon's primary objective is to obtain improved estimates of the prices of producer durables for better deflation of the producers durable equipment (PDE) component of investment in the national accounts, he also looks at selected categories of consumer

²⁶ These adjustments are described in more detail in Lane, Randolph, and Berenson (1988).

²⁷ This is probably too extreme. Presumably some consumers would be willing to pay for extra safety features even if they were not mandated by the government, and for these consumers the safety improvements do constitute quality improvements. We thank Evan Koenig for this point.

²⁸ See Triplett (1992b) Table 7.1. Most recently (November 1993), the BLS estimated that the retail equivalent value of quality changes for 1994 model passenger cars averaged \$363.63, just under two-thirds the average increase in manufacturers' suggested list prices of \$612.74. The \$363.63 estimated retail value of quality changes could be further broken down into \$223.53 associated with changes in accordance with mandated pollution and safety features and \$140.10 for other quality improvements, such as powertrain improvements, corrosion protection upgrades, and changes in levels of standard or optional equipment.

Table 2
Recent Studies of Quality Bias in the CPI

Study	Categories studied	Sample period	Estimated bias
Noe and von Furstenberg (1972)	1964-70	All	.03 to .11
Armknrecht and Weyback (1989)	Women's coats and jackets, women's suits	1987-88	-3.4 to 3.5
Randolph (1988)	Housing	1983	-.3 to -.4
Gordon (1990)	Durables	1947-83	1.54
Liegey (1993)	Women's coats and jackets, women's suits	1989	-1.3 to 6

NOTE: A negative bias means that the CPI understates the rate of inflation of the item.

durables.²⁹ Gordon concludes that the CPI overstated the rate of increase in durables prices by an average of 1.54 percentage points a year over the full 1947-83 sample, with the largest errors occurring before 1960. Table 3 summarizes the details of Gordon's findings on the biases in the official estimates of consumer durables prices.

The figures in Table 3 are the average annual rate of drift of the ratio of Gordon's alternative price indexes for the indicated categories to the official price index, where both the alternative and official indexes are constructed using the Tornqvist index number formula. (For a description, see the box entitled "Tornqvist Index Numbers.") The negative drift in all the price ratios over all the sample periods Gordon examines suggests that the CPI significantly overstated the rate of price increase for these categories of goods for the indicated periods. Gordon notes that while the rate of drift, or error, for the appliance and radio-TV categories of durables may appear surprisingly high, the data used to construct these indexes are among the most accurate and comprehensive parts of his study. The difference between Gordon's price

series and the official price series is attributable to Gordon's use of alternative sources for his price data in conjunction with a more rigorous application of quality adjustments when such adjustments are called for. Gordon uses hedonic methods to quality-adjust prices when data allow their use, but for most of the data, quality adjustment is carried out using existing BLS techniques.

Again, we need to interpret these results with caution. Triplett (1993) argues that by following the BLS convention on the treatment of government-mandated pollution and safety features on durable goods, Gordon omits a significant source of downward bias in automobile prices. Thus, Gordon overestimates the size of the overall quality bias in durable goods prices, possibly by as much as one-third to one-half.

New goods bias

In some respects, the new goods problem is simply another version of the quality adjustment problem: the distinction between a new variety of an existing product and an entirely new product is not always obvious. For example, a personal computer can be considered a new product, or it can be viewed as an extraordinarily efficient combination calculator and typewriter. Fixler (1993) suggests that a quality change be defined as a change in a product's characteristics and a new good be defined as a rebundling of a product's characteristics or the addition of new characteristics.

²⁹ Indeed, it is arguable that, given the data sources Gordon relied upon (the Sears catalog, Consumer Reports), his results are more relevant for assessing potential problems in the CPI than the PPI.

Table 3
Drift of Ratio of Tornqvist Indexes
Gordon's study and Corresponding NIPA Implicit Deflators for Selected Consumer Durables, 1982 base

	Annual growth rates			
	Full sample	1947-60	1960-73	1973-83
Motor vehicles and parts	-1.71	-2.39	-1.69	-.85
Furniture and household equipment	-1.79	-2.52	-1.26	-1.55
Kitchen and other household appliances	-3.22	-4.39	-2.37	-2.83
Radios and TV's	-5.94	-9.07	-3.77	-4.69
Total consumer durables	-1.54	-2.21	-1.24	-1.05

SOURCE: Gordon (1990), Table 1.2.

NOTE: The table shows the drift in the ratio of Gordon's alternative price indexes to the official price indexes for different periods. Thus, negative numbers are interpreted as showing that the official series are upward biased; that is, they overstate inflation.

The essence of the problem posed by the arrival of new goods is as follows.³⁰ Suppose we are trying to track a measure of the price level over time, and we have price and quantity data for $N-1$ commodities in periods 1 and 2, p_n^1 and x_n^1 for $t = 1, 2$ and $n = 1, \dots, N-1$. Suppose, in addition, that x_n^2 units of a new good are sold at price p_n^2 in period 2. How are we to calculate the bilateral price index $P(p^1, p^2, x^1, x^2)$ when we do not know p_n^1 , the price of the new good in period 1? Stated this way, the similarity between the problem posed by new goods and quality change in existing goods becomes clear.³¹

While there do not appear to be empirical studies of the new goods problem as such, an example from Diewert (1987) is illuminating and gives us some sense of the potential magnitude of the problem. Diewert estimates that, depending on the fraction of expenditures allocated to new products not covered by the price index and the price profile of new products (typically, new products experience rapid price declines following their introduction), the price level would be overstated by between 1 and 5 percent annually when calculated using the Laspeyres formula.

Lebow, Roberts, and Stockton (1992) attempt to put more concrete numbers on the size of the new goods bias by making some simple assump-

tions. They isolate the CPI categories in which they think rapid product innovation is most likely to be important and find that the relevant categories account for about 2.4 percent of the CPI. Assuming that new products experience price declines at a rate comparable to that of computers—that is, about 20 percent a year on average—they arrive at an estimate of new goods bias in the overall CPI of 0.5 percent a year. Insofar as new goods are important in categories other than appliances, lawn equipment and power tools, and medical care commodities, the 0.5 percent figure is an underestimate of the size of the new goods bias.

³⁰ This discussion follows Diewert (1987).

³¹ A formal solution to this problem was proposed by Hicks (1940). Simply calculate the "shadow" price that would just make the consumer's demand for the good in period 1 equal zero. The calculation of this shadow price requires knowledge of the consumer's preferences, which might be obtained by econometric techniques. In practice, however, it is too costly to resort to such techniques, and official indexes frequently ignore new goods. Cars were not introduced into the CPI until 1940, while the PPI did not include computer prices until 1990.

Tornqvist Index Numbers

A Tornqvist price index is defined as

$$P_T = \prod_i \left(\frac{p_i^2}{p_i^1} \right)^{s_i}$$

where

$$s_i = 0.5 \left[\frac{p_i^1 x_i^1}{\sum_i p_i^1 x_i^1} + \frac{p_i^2 x_i^2}{\sum_i p_i^2 x_i^2} \right]$$

The interpretation of an index number of this type is a little clearer if we rewrite it in logarithmic form:

$$\log(P_T) = \sum_i s_i (\log(p_i^2) - \log(p_i^1)).$$

That is, the (log of the) Tornqvist index is a weighted average of the growth rates of the individual prices, with weights equal to the average of the expenditure shares in the two periods used to compute the growth rates. The choice of the Tornqvist index number formula is motivated by the fact that it belongs to the class of "superlative" index numbers identified by Diewert (1976). Superlative index numbers have the property that they are exact for utility functions that are second-order approximations for any utility function and are thus less susceptible to substitution bias.

If most new goods do not experience price declines comparable to those experienced by computers, the 0.5 percent figure is an overestimate of the size of the new goods bias. The main reason computer prices seem to have been examined so frequently is because they have declined at such extraordinarily rapid rates. New goods bias could well be a lot higher or a lot lower than their calculation suggests; we simply do not know.

List versus transactions prices

As we have already noted, the price information that goes into the CPI is collected by BLS field representatives through visits or telephone calls. The BLS puts great emphasis on obtaining price quotes that reflect the actual prices paid by consumers, and to this end, it makes a number of adjustments to some of the raw price data to

obtain better estimates of transactions prices. For example, in pricing new cars, the BLS agents obtain estimates of the base price for the vehicle, along with estimates of the prices of various options, dealer preparation, and delivery. The BLS agents also obtain estimates of the average concession or markup during the previous thirty days to arrive at an estimate of the transaction price of the vehicle. The BLS also tries to take account of manufacturers' rebates, bonus merchandise, quantity discounts, and utility refunds when pricing many other goods and services.³² However, no adjustment is made for the use of cents-off coupons by consumers, except when the coupons are attached to the product for immediate redemption at the time of purchase.

There seems to be no research on how accurately the prices that are used to construct the CPI reflect the actual prices paid by consumers.³³ It would appear that the BLS does make a reasonable attempt to ensure that the prices are accurate, but the failure to account for the use of cents-off coupons does raise some questions. Data on coupon use is difficult to come by, although we can get some sense of their potential importance from Nielsen Clearing House (1993). According to the Nielsen Report, in 1992 consumers redeemed about

³² U.S. Department of Labor (1992, 193-96).

³³ In contrast, a number of studies address the list-transaction price problem for the PPI.

7.7 billion manufacturer-issued coupons, whose face value averaged 58 cents, for total "savings" of \$4.5 billion, which is slightly less than 1 percent of consumer spending in the relevant categories. Somewhat more than 14 percent of total grocery volume was purchased with a coupon in 1992. Growth in the average face value of coupons redeemed has consistently exceeded growth in the CPI since 1980. It is an open question whether failure to allow for the use of cents-off coupons by consumers leads to an upward bias in the CPI and, if so, by how much.

Treatment of durable goods

The theoretical basis of the cost-of-living index and the CPI is essentially a static theory. The appropriate treatment of durable goods in such an index requires the measurement and pricing of the flow of services obtained by the consumer from the good over time. That is, since a durable good yields a flow of consumption services valued by the consumer over several time periods, we do not want to price the purchase of the good but rather, the flow of services that it yields each time period.

In 1983, the BLS switched to the rental equivalence concept to measure housing costs in the CPI (1985 for the CPI-W). Part of the impetus for this change was the large discrepancy that emerged in the late 1970s between the CPI and the deflator for personal consumption expenditures (PCE) in the national accounts due to their different treatment of housing costs. It became generally recognized that the rental equivalence approach employed in the construction of the PCE deflator was superior on theoretical grounds.³⁴ Before the change, the BLS was accused of mixing the consumption and investment components of housing costs. The appropriate concept for a cost-of-living-based index is the cost of the flow of housing services consumed over the measurement interval. However, the rental equivalence approach is not without its problems. There are important differences between the markets for owner-occupied homes and the markets for rental units, not least of which is the quality. Prior to the 1987 revision of the CPI, the BLS simply took figures from the CPI's rental component to estimate the implicit rent of owner-occupiers. Since 1987 the BLS has used

rents from houses in the same geographic area and with similar characteristics to those of owner-occupied houses to calculate the owners' equivalent rent index.

Armknrecht and Ginsburg (1992) describe research currently being undertaken at the BLS to shift the treatment of autos in the CPI to the theoretically more appropriate flow-of-services approach. Under ideal circumstances, all durable goods would be priced on a flow-of-services basis. In reality, this is not possible, primarily because there are no rental markets for many types of durable goods.³⁵ For autos, however, two very active and distinct rental markets may facilitate the adoption of a rental equivalence approach. Armknrecht and Ginsburg point out that car use in the traditional (short term) rental market is very different from normal use, making it unrepresentative of the general population. However, the long-term rental market for auto leases may provide more appropriate measures of the rental equivalence of auto services. The BLS currently is examining the feasibility of pricing auto transportation services on this basis.

Measuring the prices of services

Triplett's 1975 survey reviews a small number of studies that attempt to assess the quality of the CPI service price indexes. Four of the six studies he reviews examine the medical services component of the CPI, and three of these four find upward bias in the CPI components. However, the main conclusion that Triplett draws from these studies is that the appropriate pricing concept in the medical services area is not very well defined. Should we be pricing treatments or cures? Pricing a cure may well be the appropriate approach, but, as Triplett notes, cures have multiple characteristics that are difficult to value. For example, how should we compare the cost of treating appendicitis by

³⁴ See, for example, *Blinder (1980), Gordon (1981), and Dougherty and Van Order (1982)*.

³⁵ As *Armknrecht and Ginsburg* note, there is no rental market for shoes.

surgery versus treating it with drugs when the latter treatment is accompanied by an increased risk of a ruptured appendix? Triplett concludes his 1988 review by noting that "...existing research on the subject is insufficient to indicate whether the medical care components are upward biased" (Triplett 1988, 70). However, the BLS handbook quite explicitly states that in many instances quality changes are treated as price changes, either because the BLS is unaware of the quality change or has no method for dealing with it. (U.S. Department of Labor 1992, 193). In instances where quality adjustment is feasible, it is carried out.³⁶

Kroch (1991) reviews the problems of price measurement in the service sector and argues that the true rate of service-sector inflation is probably lower than the measured rate and closer to the rate in the goods sector. Kroch argues that the slower rate of inflation for medical equipment (which is adjusted for quality changes) than for medical services (for which relatively few quality adjustments are made) is suggestive of an upward bias in the CPI medical services category. Kroch's

comparison of the inflation rates of the two series suggests that medical services inflation was overstated by as much as 1 percent a year during the 1980s. However, it is not clear what one can infer from a comparison of the rates of increase of input and output prices. For example, the rate of inflation for capital equipment used in the auto insurance industry, which is a heavy user of computers, is falling rapidly, which is not generally true of auto insurance premiums.³⁷

Kroch also argues that the educational price category overstates inflation when compared with an index of tuition for higher education. The last category of services that Kroch considers is the rental equivalence measure of owner-occupied housing, and while he suggests that the failure of the rental equivalence index to track house prices in recent years may mean that the CPI is overstating inflation in the housing services category, he refrains from drawing a conclusion.³⁸

It is, of course, important to remember that despite the difficulties that may accompany price measurement for many services, the influence of these problems on the overall CPI is determined by the importance of the problem categories in the consumers' budget. Thus, even if it were true that inflation in the medical care component of the CPI is overstated, the fact the medical care only accounts for 4 percent of consumers' expenditures would greatly limit the influence of mismeasurement in this component on the overall CPI.³⁹

The categories of the CPI that Kroch argues may be overmeasuring inflation together account for only 5.6 percent of budget outlays in the base period.

Conclusions about measurement bias in the CPI

The point of departure for this review of the CPI is the earlier survey by Triplett (1975), with its conclusion that, as of the mid-1970s, not enough was known to determine whether there was a clear overall bias in the CPI or determine its sign. Triplett repeats this opinion in his unpublished 1988 survey. Have we learned anything in the intervening period that would lead us to draw different conclusions?

In the almost twenty years since Triplett's first survey, there has been remarkably little new

³⁶ Recently, Armknecht and Ginsburg (1992) have reviewed the manner in which the BLS has tried to deal with the increasing importance of services in consumers' budgets. They also have reviewed research currently under way to improve measurement and quality adjustment of services prices. They discuss research that attempts to apply hedonic techniques to the measurement of the price of hospital room stays and note the problem of properly accounting for and measuring differences in the level of nursing care among hospitals as a major obstacle to the implementation of hedonic methods in this area.

³⁷ This example was suggested by Jack Triplett.

³⁸ However, the study by Rogers, Henderson, and Ginsburg (1993) suggests that the housing component of the CPI has performed quite well in recent years.

³⁹ Kroch notes that nearly two-thirds of health care services consumed in the United States are not part of the consumer budget because they are provided through employer-paid health plans rather than paid for directly by consumers. Hence, medical care prices are given only a 4-percent weight in the CPI. This figure includes household payments for health insurance. By contrast, household consumption of health services accounts for more than 12 percent of personal consumption in the National Income and Product Accounts. See Kroch (1991, 32).

research on the problems of price measurement. Probably the most important single contribution to the field of price measurement in recent years is Gordon's (1990) study of producer durables prices, which also includes some analysis of consumer durables prices. For the CPI, the most significant recent studies have been those of Manser and McDonald (1988) on substitution bias and Reinsdorf (1993) on outlet substitution bias. There are no recent studies of quality bias for the nondurables and services components of the CPI. For example, we still have no sense of how large the potential bias is in the measurement of health care costs.⁴⁰

It seems clear that the issue of substitution bias is the closest to being settled. We probably can conclude with some confidence that the substitution bias arising from the use of the fixed-weight index is currently 0.2 percent a year at most. Recently, the issue of outlet substitution bias has received a lot of attention, primarily as a result of the work of Reinsdorf (1993). In view of this, we give it more detailed coverage in our review of the recent literature and argue that it needs to be backed up by further work before the figure of an upward bias of as much as 2 percent a year can be accepted as valid.⁴¹ Note that the 2 percent figure is for the categories studied by Reinsdorf (food at home and gasoline) and does not apply to the CPI as a whole.

We note that some categories of the CPI that Triplett cites as having potential downward bias, such as housing, are now treated differently, and aging bias is less likely to cause inflation of housing costs to be understated. Apparel remains a problem, although it is interesting that the studies by Armknecht and Weyback (1989) and Liegey (1993) find both upward and downward bias in this category. Auto prices are still biased downward, for the reasons Triplett states, and will remain so until the BLS changes its methodology. We have no firm evidence on the size and nature of the biases in pricing medical care and so cannot draw any firm conclusions about this category.

Lebow, Roberts, and Stockton (1992) conclude their survey of measurement bias in the CPI by noting that under *extreme assumptions*, the upper bound on measurement bias is about 1.8 percent a year. They arrive at this figure by adding the various biases that have been identified and quantified by other authors. While calculations of

this sort are suggestive, it is important to consider the limitations and caveats regarding them.

Specifically, we need to ask whether calculations of this sort may give us figures on the overall bias in the CPI that are too high because of double counting of some of the biases. Can we simply add together estimates of the quality adjustment bias and the new goods bias, given that the distinction between the two is elusive? Is it possible that traditional substitution bias and quality adjustment bias are also aspects of the same phenomenon? The same question can be raised for the outlet substitution bias discovered and quantified by Reinsdorf (1993): how do we disentangle this effect from other, more traditional forms of bias? How much of the difference in prices between conventional and low-cost retail outlets is due to a list-transactions price problem that stems from the BLS's failure to account for the use of cents-off coupons at conventional outlets?

As Gordon (1990) notes in his study of producer durable prices, and Triplett emphasizes in his 1975 and 1988 surveys, the problem is that many potential pitfalls of the different price indexes are considered in isolation from one another, without any regard to the possible interaction among them. This comment is not to criticize the calculations carried out by Lebow, Roberts, and Stockton. Such calculations are essential if any sort of conclusions are to be drawn about the potential biases in the most closely watched price index. However, it is important to be aware of, and at some point do something about, the limitations that surround calculations of this type.

In view of the paucity of evidence on the various potential biases in the CPI, we are inclined to think that it is better to err on the side of con-

⁴⁰ Thus, Tregarthen (1993) points out that recent concerns over the rising cost of health care may be mistaken, in part because of the failure of the BLS to properly account for quality change in health care and in part because of the reliance of the BLS on list rather than transaction prices in pricing the health components of the CPI.

⁴¹ It is important to note that Reinsdorf himself considers the 2 percent figure to be at best a ballpark estimate rather than a point estimate and that it may be capturing phenomena other than outlet substitution.

servatism in guesstimating the size of the overall bias. A figure of less than 1 percent thus strikes us as a plausible estimate of the overall bias. The true figure may be a lot larger or a lot smaller; at present we simply do not know. The more agnostic position that Triplett has adopted in his surveys is harder to defend, as the evidence seems to indicate more instances of upward than of downward bias in the CPI. However, Triplett's arguments are an important antidote to those who are inclined to accept uncritically the position that the CPI consistently overstates inflation.

References

- Aizcorbe, Ana M., and Patrick C. Jackman (1993), "The Commodity Substitution Effect in CPI Data, 1982-91." *Monthly Labor Review* 116 (December): 25-33.
- Armknrecht, Paul A. (1984), "Quality Adjustment in the CPI and Methods to Improve It," *Journal of the American Statistical Association: Proceedings of the Business and Economic Statistics Section*, 57-63.
- and Donald Weyback (1989), "Adjustments for Quality Change in the U.S. Consumer Price Index," *Journal of Official Statistics* 5 (February): 107-23.
- , Donald Weyback, and Daniel H. Ginsburg (1992), "Improvements in Measuring Price Changes in Consumer Services: Past, Present, and Future," in Zvi Griliches (ed.), *Output Measurement in the Service Sectors*, NBER Studies in Income and Wealth, vol. 56 (Chicago: University of Chicago Press).
- Backman, Jules, and Martin R. Gainsbrugh (1966), *Inflation and the Price Indexes*, U.S. Congress, Joint Economic Committee, Subcommittee on Economic Statistics, 89th Congress, 2nd Session, July (Washington, D.C.: Government Printing Office).
- Blinder, Alan S. (1980), "The Consumer Price Index and the Measurement of Recent Inflation," *Brookings Papers on Economic Activity*, 539-65.
- Blisard, William Noel, and James R. Blaylock (1991), "Construction of True Cost of Food Indexes from Estimated Engel Curves," *American Journal of Agricultural Economics* 73 (August): 775-83.
- Braithwait, Steven D. (1980), "The Substitution Bias of the Laspeyres Price Index," *American Economic Review* 70 (March): 64-77.
- Bryan, Michael F., and Stephen G. Cecchetti (1993), "The Consumer Price Index as a Measure of Inflation," *Federal Reserve Bank of Cleveland Economic Review*: April, 15-24.
- Christensen, Laurits R., and Marilyn E. Manser (1976), "Cost-of-Living Indexes and Price Indexes for U.S. Meat and Produce, 1947-1971," in Nestor Terleckyj (ed.), *Household Production and Consumption*, NBER Studies in Income and Wealth, vol. 40 (New York: Columbia University Press).
- Clements, Kenneth W., and H.Y. Izan (1987), "The Measurement of Inflation: A Stochastic Approach," *Journal of Business and Economic Statistics* 5 (July): 339-50.
- Denison, Edward F. (1962), "The Sources of Growth in the United States and the Alternatives Before Us," Supplementary Paper No. 13 (New York: Committee for Economic Development).
- Diewert, W. Erwin (1987), "Index Numbers," in John Eatwell, Murray Milgate, and Peter Newman (eds.), *The New Palgrave: A Dictionary of Economics*, vol. 2 (London: Macmillan).
- (1976), "Exact and Superlative Index Numbers," *Journal of Econometrics* 4 (May): 114-45.
- Dougherty, Ann, and Robert Van Order (1982), "Inflation, Housing Costs, and the Consumer Price Index," *American Economic Review* 72 (March): 154-64.
- Dow, James P., Jr. (1993), "Measuring Inflation Using Multiple Price Indexes" (Paper presented to the 68th Annual Conference of the Western Economic Association, Lake Tahoe, Nev., June).
- The Economist* (1993), "What Happened to Inflation?" October 30, 19-20.
- Fixler, Dennis (1993), "The Consumer Price Index: Underlying Concepts and Caveats," *Monthly Labor Review* 12 (December): 3-12.

- Forsyth, F.G. and R.F. Fowler (1981), "The Theory and Practice of Chain Price Index Numbers," *Journal of the Royal Statistical Society* 144 (pt. 2), 224-46.
- Gillingham, Robert F. (1974). "A Conceptual Framework for the Consumer Price Index," *Proceedings, Business and Economic Statistics Section, American Statistical Association*, 246-52.
- Gordon, Robert J. (1992), "Measuring the Aggregate Price Level: Implications for Economic Performance and Policy," NBER Working Paper Series, no. 3969.
- (1990), *The Measurement of Durable Goods Prices* (Chicago: University of Chicago Press).
- (1981), "The Consumer Price Index: Measuring Inflation and Causing It," *The Public Interest*, Spring, 112-34.
- Griliches, Zvi (ed.) (1971), *Price Indexes and Quality Change: Studies in New Methods of Measurement* (Cambridge, Mass.: Harvard University Press).
- (1961), "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," in *The Price Statistics of the Federal Government* (New York: National Bureau of Economic Research), reprinted in Griliches 1971.
- Helliker, Kevin (1991). "Thriving Factory Outlets Anger Retailers as Store Suppliers Turn into Competitors," *Wall Street Journal*, October 8, B1.
- Hicks, John R. (1940). "The Valuation of Social Income" *Economica* 7 (May): 105-24.
- Hirsch, James S. (1993). "With Fewer Attendants Aboard Jets, Mood of Passengers Turns Turbulent," *Wall Street Journal*, July 23, B1.
- Kokoski, Mary F. (1993). "Quality Adjustment of Price Indexes," *Monthly Labor Review* 116 (December): 34-46.
- (1989). "Experimental Cost-of-Living Indexes: A Summary of Current Research," *Monthly Labor Review* 112 (July): 34-39.
- (1987). "Problems in the Measurement of Consumer Cost-of-Living Indexes," *Journal of Business and Economic Statistics* 5 (January): 39-46.
- Kroch, Eugene (1991), "Tracking Inflation in the Service Sector," Federal Reserve Bank of New York *Quarterly Review*, Summer, 30-35.
- Lane, Walter F., William C. Randolph, and Stephen A. Berenson (1988), "Adjusting the CPI Shelter Index to Compensate for Effect of Depreciation," *Monthly Labor Review* 111 (October): 34-37.
- Leaver, Sylvia G. (1992), "Estimating Variances for the U.S. Consumer Price Index for 1987-1991," *Proceedings of the Survey Research Methods Section American Statistical Association*, 740-45.
- Lebow, David E., John M. Roberts, and David J. Stockton (1992), "Understanding the Goal of Price Stability," mimeo, Board of Governors of the Federal Reserve System, October.
- Liegey, Paul F., Jr. (1993), "Adjusting Apparel Indexes in the Consumer Price Index for Quality Differences," in Murray F. Foss, Marilyn E. Manser, and Allan H. Young (eds.), *Price Measurements and Their Uses NBER Studies in Income and Wealth*, vol. 57 (Chicago: University of Chicago Press).
- Manser, Marilyn E., and Richard McDonald (1988), "An Analysis of Substitution Bias in Measuring Inflation, 1959-1985," *Econometrica* 56, (July): 909-30.
- Morgenson, Gretchen (1993), "The Fall of the Mall," *Forbes*, May 24, 106-12.
- Moulton, Brent R. (1993), "Basic Components of the CPI: Estimation of Price Changes" *Monthly Labor Review* 116 (December): 13-24.

- Morgenstern, Oskar (1963). *On the Accuracy of Economic Observations*. 2d ed. (Princeton, N.J.: Princeton University Press).
- NBER (1961). *The Price Statistics of the Federal Government: Review, Appraisal, and Recommendations*, General Research Series, 73 (New York: National Bureau of Economic Research).
- Nielsen Clearing House (1993). *Worldwide Coupon Distribution and Redemption Trends* (Chicago: A.C. Nielsen Co.).
- Noe, Nicholas N., and George M. von Furstenberg (1972). "The Upward Bias in the Consumer Price Index Due to Substitution," *Journal of Political Economy* 80 (November/December): 1280-86.
- Pollak, Robert A. (1989). *The Theory of the Cost of Living Index* (New York: Oxford University Press).
- Popkin, Joel (1993). "Discussion of the Effect of Outlet Price Differentials on the U.S. Consumer Price Index by Marshall Reinsdorf," in Murray F. Foss, Marilyn E. Manser, and Allan H. Young (eds.), *Price Measurements and Their Uses NBER Studies in Income and Wealth*, vol. 57, (Chicago: University of Chicago Press).
- Randolph, William C. (1988). "Housing Depreciation and Aging Bias in the Consumer Price Index," *Journal of Business and Economic Statistics* 6 (July): 359-71.
- Reinsdorf, M. (1993). "The Effect of Outlet Price Differentials on the U.S. Consumer Price Index," in Murray F. Foss, Marilyn E. Manser, and Allan H. Young (eds.), *Price Measurements and Their Uses, NBER Studies in Income and Wealth*, vol. 57 (Chicago: University of Chicago Press).
- Rogers, R. Mark, Steven W. Henderson, and Daniel H. Ginsburg (1993). "Consumer Prices: Examining Housing Rental Components," Federal Reserve Bank of Atlanta *Economic Review*, May/June, 32-46.
- Schmidt, Mary Lynn (1993). "Effects of Updating the CPI Market Basket," *Monthly Labor Review* 116 (December): 59-62.
- Szulc, Bohdan J. (1983). "Linking Price Index Numbers," in W.E. Diewert and C. Montmarquette (eds.), *Price Level Measurement: Proceedings from a Conference Sponsored by Statistics-Canada* (Ottawa: Statistics Canada).
- Tregarthen, Suzanne (1993). "Statistics Overstate Health Care Costs," *Wall Street Journal*, August 18, A10.
- Triplett, Jack E. (1993). "Review of Robert J. Gordon 'The Measurement of Durable Goods Prices,'" *Journal of Economic Literature* 31 (March): 244-46.
- (1992a). "Economic Theory and the BEA's Alternative Quantity and Price Measures," *Survey of Current Business* 72 (April): 49-52.
- (1992b). "Hedonic Methods in Statistical Agency Environments: An Intellectual Biopsy," in Ernst R. Berndt and Jack E. Triplett (eds.), *Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth, NBER Studies in Income and Wealth*, vol. 54 (Chicago: University of Chicago Press).
- (1989). "Price and Technological Change in a Capital Good: A Survey of Research on Computers," in Dale W. Jorgenson and Ralph Landau (eds.), *Technology and Capital Formation* (Cambridge, Mass.: MIT Press).
- (1988). "Price Index Research and Its Influence on Data: A Historical Review" (Paper presented to the Fiftieth Anniversary Conference of the Conference on Research on Income and Wealth, Washington D.C., May 12-14).
- (1987). "Hedonic Functions and Hedonic Indexes," in John Eatwell, Peter Newman, and Murray Milgate (eds.) *The New Palgrave: A Dictionary of Economics*, vol. 2 (London: Macmillan).

- (1986). "The Economic Interpretation of Hedonic Methods," *Survey of Current Business* 64 (January): 36-40.
- (1975). "The Measurement of Inflation: A Survey of Research on the Accuracy of Price Indexes," in Paul H. Earl (ed.), *Analysis of Inflation* (Lexington, Mass.: Lexington Books).
- (1971). "Quality Bias in Price Indexes and New Methods of Quality Measurement," in Zvi Griliches (ed.), *Price Indexes and Quality Change: Studies in New Methods of Measurement* (Cambridge, Mass.: Harvard University Press).
- , and Richard J. McDonald (1977). "Assessing the Quality Error in Output Measures: The Case of Refrigerators," *Review of Income and Wealth* 23 (June): 137-56.
- U.S. Department of Labor (1992). *BLS Handbook of Methods*. Bureau of Labor Statistics Bulletin 2414 (Washington, D.C.: Government Printing Office, September).
- (1987). *The Consumer Price Index: 1987 Revision*, Bureau of Labor Statistics Report 736 (Washington, D.C.: Government Printing Office, January).
- Wynne, Mark A., and Fiona Sigalla (1993). "A Survey of Measurement Biases in Price Indexes," Federal Reserve Bank of Dallas Research Paper no. 9340, (Dallas, October).
- Young, Allan H. (1992). "Alternative Measures of Change in Real Output and Prices," *Survey of Current Business* 72 (April): 32-48.