BRENT R. MOULTON
Bureau of Labor Statistics

KARIN E. MOSES
Bureau of Labor Statistics

Addressing the Quality Change Issue in the Consumer Price Index

THE FINAL REPORT of the Advisory Commission to Study the Consumer Price Index (CPI) represents the most influential critique of the CPI in decades.[1] This report, in conjunction with a number of other reviews of CPI bias, has focused the attention of policymakers and economists on the limitations of price index numbers, in particular, and of other measures of economic activity, more generally.[2] The commission's estimate of overall bias in the CPI is 1.1 percent per year, of which 0.4 percent is attributed to the failure of the fixed weight index to account for consumer substitution as relative prices change, 0.1 percent is attributed to failure to account for discount stores and other innovations in retailing, and 0.6 percent is attributed to inadequate measurement of improvements in quality and of new goods. In contrast to the commission's estimates of substitution bias, which have been relatively uncon-

1. See U.S. Senate, Committee on Finance (1996). The members of the advisory commission were Michael J. Boskin (chair), Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches, and Dale Jorgenson.

2. Other recent reviews of CPI bias include Baker (1996), Congressional Budget Office (1994), Diewert (1996), Klumpner (1996), Lebow, Roberts, and Stockton (1994), Moulton (1996), Shapiro and Wilcox (1996), U.S. Senate, Committee on Finance (1995), and Wynne and Sigalla (1994).

troversial, the estimates of quality and new goods bias have been crit-
icized by several economists.[3]

There are two general categories of quality errors: failure to detect a
change in quality and failure to make the appropriate adjustment for a
change that has been detected. The new products bias can also be
considered in two forms: failure to include new products in the sample
without long lags and failure to include the consumer surplus generated
by a new product.[4] In contrast to the quality bias, which can, in prin-
ciple, go in either direction, the new product bias is theoretically known
to be an upward bias (although it may be offset to some extent by the
downward bias that occurs when a product disappears).

Many discussions of quality bias have begun with the premise that
much quality change goes undetected, that such quality change is pre-
dominantly improvement, and that the result is an upward bias of the
index. We concur with Jack Triplett that this is of doubtful validity as
a general proposition, although it may hold true in specific cases.[5] The
data collectors and commodity analysts at the Bureau of Labor Statistics

3. For example, Abraham (1997a, 1997b), Baker (1997), Bosworth (1997), Hulten
(1997), and Triplett (1997) suggest that the quality bias estimate may be too large,
whereas Diewert (1997, p. 95) describes the commission's estimates as ''perhaps a bit
conservative.''

4. These are not actually separate biases, since the consumer surplus generated by
the new good should, in principle, include the value generated by any price decline early
in the product's life cycle. The case of the new good may be treated in the theory of the
cost of living index by using the reservation price of the item; that is, the price at which
the consumer's demand for the item is just equal to zero. (See Hicks, 1940; Rothbarth,
1941; and Hausman, 1997.) Let P_{nt} represent the price of the new good after introduction
and $P^*_{t-1,n}$ represent its reservation price before introduction. Let the prices of the other
$n - 1$ goods be $p_t = (P_{t1}, P_{t2}, ..., P_{t, n-1})$, and similarly for p_{t-1}. The constant-utility
cost of living index, I, evaluated at the period t level of expenditures is

$$I(p_t, P_{nt}, p_{t-1}, P^*_{t-1,n}, u_t) = y_t / e(p_{t-1}, P^*_{t-1,n}, u_t),$$

where the expenditure function gives the minimum amount of expenditures, $y = e(p,$
$P_n, u_t)$, needed to achieve the level of utility, u_t, that arises from the corresponding
indirect utility function. The cost of living index is related to the equivalent variation,
EV, by

$$I(p_t, P_{nt}, p_{t-1}, P^*_{t-1,n}, u_t) = y_t / (y_t - EV).$$

The equivalent variation is negative in this case, because the introduction of the new
good is a welfare improvement. If the share of the new item is small relative to total
spending, the equivalent variation will closely approximate the ordinary Marshallian
consumer surplus. The bias from omitting the new good in the index calculation is
approximately equal to the ratio of consumer surplus to total expenditure.

5. Triplett (1971).

(BLS) work from checklists that are designed to provide detailed product descriptions—for example, they include model numbers. In general, quality changes are well detected for most commodities. The possibility of undetected quality change is presumably more important for services, especially complex, knowledge-based services, such as professional medical services and higher education. Quality adjustments are made for services, but gradual improvements or declines in quality may not be adequately detected by data collectors. For example, the CPI checklist for local telephone services provides for adjustments for special features such as call waiting, but not for gradual improvement in sound quality or gradual deterioration in the convenience of scheduling installation appointments. Changes in general retail services, such as the installation of credit card scanners at checkout counters or the disappearance of helpful salespersons, are generally not captured by the BLS checklists either.

In many cases, the CPI methods do detect quality change. Then the issue of quality error revolves, in part, around the adequacy or inadequacy of the quality adjustment methods used in constructing the index. It appears to be widely held that quality improvements are pervasive and that the BLS does very little to account for them. An alternative point of view, however, is that the BLS methods of adjusting for quality change already attribute a great deal of price change to quality improvement, so that any remaining quality bias could be either negative or positive. Consequently there is an important need, which we attempt to address in this paper, to develop useful measures of the amount of quality adjustment now applied by the BLS in constructing the CPI. However, such measures do not, by themselves, provide direct evidence of quality bias in either direction. The best test of quality bias is careful analysis of the data, item by item. Only a few item categories have received such detailed analysis. The advisory commission has reviewed available research and, in some cases, conducted its own analysis.

The problems associated with bringing new goods into the CPI samples have been highlighted in many recent studies.[6] Over the past two decades, the BLS has adopted procedures to bring new products into the sample more promptly, beginning in 1978 with a shift to probability-

6. For example, Berndt, Griliches, and Rosett (1993) demonstrate the importance of keeping the sample up to date for prescription drugs, and Hausman (1997) examines the potential consequences of the long lag in including cellular telephone service.

based methods for selecting samples. Subsequently, the regular rese-
lection of samples (currently, at five-year intervals) was introduced.
From 1999, the BLS will have the capability to schedule sample rota-
tions more frequently for selected categories of items, which should
further enhance its capacity to capture new goods.

In this paper, we first review the discussion of quality and new goods
in the commission's report. We then review BLS procedures for making
quality adjustments in the CPI and discuss how these might give rise to
quality change bias. Finally, we examine the quality adjustments that
occur when CPI sample items change or need to be replaced because
of sample attrition. We provide estimates of the quality and price
changes, both implicit and explicit, associated with those replacements.

Quality and New Products in the Advisory Commission's Report

In its analysis of the quality change and new products biases, the
advisory commission classifies the CPI into twenty-seven major cate-
gories of items and gives a separate bias estimate for each. This is the
first systematic analysis, category by category, of quality bias in the
CPI, and it is a noteworthy accomplishment. In general, the commis-
sion's approach to the problem of producing an overall assessment of
bias seems sensible, and this type of structure will likely prove to be
useful in the future. Hereafter, following the commission, we some-
times use the term ''bias'' without distinguishing between new product
and quality change bias.

The advisory commission's estimates of bias by category are shown
in table 1. Of the twenty-seven categories, the commission assigns eight
a quality bias of zero: fuels, housekeeping supplies, housekeeping ser-
vices, other private transportation, public transportation, health insur-
ance, entertainment services, and tobacco. It assigns each of the re-
maining nineteen categories an estimated bias that is positive; that is,
the commission concludes that price change is overstated because qual-
ity change is understated. For six of these nineteen categories (appli-
ances, including electronic; prescription drugs; professional medical
services; hospital and related services; entertainment commodities; and
personal care) the advisory commission reviewed existing studies of
bias in the price trends of specific items to draw inferences about likely

**Table 1. Advisory Commission's Estimates of Quality and
New Product Bias in the CPI**

Percent

Item category	*Share in CPI*	*Estimated bias at annual rate*[a]
Food and beverages		
Total	17.33	. . .
Food at home, other than produce	8.54	0.30
Fresh fruits and vegetables	1.34	0.60
Food away from home	5.89	0.30
Alcoholic beverages	1.57	0.15
Housing		
Total	41.35	. . .
Shelter	28.29	0.25
Fuels	3.79	0.00
Other utilities, including telephone	3.22	1.00
Appliances, including electronic	0.81	5.60
Other housefurnishings	2.64	0.30
Housekeeping supplies	1.12	0.00
Housekeeping services	1.48	0.00
Apparel and upkeep		
Total	5.52	1.00
Transportation		
Total	16.95	. . .
New vehicles	5.03	0.59
Used cars	1.34	0.59
Motor fuel	2.91	0.25
Other private transportation	6.15	0.00
Public transportation	1.52	0.00
Medical Care		
Total	7.36	. . .
Prescription drugs	0.89	2.00
Nonprescription drugs and medical supplies	0.39	1.00
Professional medical services	3.47	3.00
Hospital and related services	2.26	3.00
Health insurance	0.36	0.00
Entertainment		
Total	4.37	. . .
Commodities	1.98	1.20
Services	2.39	0.00
Other goods and services		
Total	7.12	. . .
Tobacco, smoking products	1.61	0.00
Personal care	1.17	0.90
Personal and educational expenses	4.34	0.20
Total	100.00	0.61

Source: U.S. Senate, Committee on Finance (1996, table 2).
a. Biases are estimated over different periods, all of which end in 1996.

bias for related but unstudied items within the category. These six
categories are important in the commission's accounting for quality
bias, since the appliance category and the three medical care compo-
nents together account for roughly half of the estimated quality and
new products bias.

For another four of the nineteen categories with positive estimated
bias (shelter, apparel and upkeep, new vehicles, and used cars), the
commission either conducted original research or presented detailed,
back-of-the-envelope calculations of the sources of bias.

For the other nine of the nineteen categories (food at home, other
than produce; fresh fruits and vegetables; food away from home; alco-
holic beverages; other utilities, including telephone; other house fur-
nishings; motor fuel; nonprescription drugs and medical supplies; and
personal and educational expenses), in the absence of empirical re-
search or data, the commission simply describes potential sources of
bias and gives an estimate of the magnitude of bias. The following
assessment of bias due to new products and variety for the category of
food at home is typical:

> There is little if any published evidence on the food category, other than
> [Jerry] Hausman's . . . attempt to establish the value for the introduction
> of a new variety of breakfast cereal. Perhaps more important than new
> varieties of packaged goods has been a wave of technological improve-
> ments that has greatly increased the variety of fresh fruits and vegetables
> available in the typical supermarket during the winter months, and a
> trend toward more services provided in supermarkets, eliminating the
> need to travel to small specialty shops, especially fresh fish markets and
> deli counters preparing fresh-cooked food. How much would a consumer
> pay to have the privilege of choosing from the variety of items available
> in today's supermarket instead of being constrained to the much more
> limited variety available 30 years ago? A conservative estimate of the
> value of extra variety and convenience might be 10 percent for food
> consumed at home other than produce, 20 percent for produce where the
> increased variety in winter (as well as summer farmers' markets) has
> been so notable, and 5 percent for alcoholic beverages where imported
> beer, microbreweries, and a greatly improved distribution of imported
> wines from all over the world have improved the standard of living.[7]

In several places, the report characterizes the commission's specific

7. U.S. Senate, Committee on Finance (1996, p. 28).

estimates of bias as "conservative," but it generally is not clear why this is believed to be so. The commission's standard, the cost of living index, is defined as a function of consumer preferences, so it is reasonable to ask whose preferences are being described and how they are assessed. Economists bring to the measurement of preferences expertise in the use of methods for drawing inferences about preferences from market data on observed consumer choices. It is therefore appropriate to ask what sort of analytical method or framework commission members used to determine their point estimates of bias.

We consider the commission's bias estimates for selected item categories, sometimes presenting simple back-of-the-envelope calculations based on available, though often incomplete, data. We find that most of our calculations imply significantly smaller estimates of bias than those of the advisory commission. We do not intend that these estimates be interpreted as a complete response to the commission's estimates of bias and are aware that our conclusions could be altered by further analysis. We view them, rather, as starting points for analysis; their advantage over some of the commission's estimates is that they are based on data or assumptions about consumer behavior—such as demand elasticities—that, at least in principle, can be tested.

Fresh Fruits and Vegetables

The quote cited above indicates that the advisory commission attributes a bias of 20 percent over the period 1967–96 due to increased seasonal availability and variety. It is reasonable to think that, to the extent that consumers value the increased seasonal availability of produce, they will consume more of it. Our analytical framework is to consider the "November strawberry" to be a new good, distinct from the "June strawberry," and measure the consumer surplus associated with the new good.[8]

Among the various methods that have been proposed for incorporating new goods in a cost of living index, Jerry Hausman's suggestion of calculating the consumer surplus from a linearized demand is particularly easy to apply to back-of-the-envelope calculations.[9] Hausman's

8. See Diewert (1983).
9. Hausman (1997) refers to his linearized method as a "lower bound" on the consumer surplus, but it is unclear to us whether the conditions for the method to be a lower bound—that is, a convex demand curve—necessarily hold in all cases.

linearized method implies that the percentage bias of the price index from failure to incorporate the consumer surplus from a new good, n, is approximately

(1) $$\text{bias} = -0.5 \times S_n/\delta_n,$$

where S_n is the percentage expenditure share of the new good after introduction and δ_n is its price elasticity of demand. Thus the calculation of consumer surplus and bias can be inferred from information on the expenditure share, which is often readily available, and the elasticity of demand, which can be estimated or inferred from elasticity estimates for similar goods.[10]

New varieties or seasonal availability of fresh fruits and vegetables face many substitutes, not only from other fresh produce, but also from frozen fruits and vegetables. We assume a value of -1.0 for δ_n. Under these assumptions, equation 1 implies that the increased consumption of new seasonal items and varieties as a share of current fruit and vegetable consumption would need to be quite large—about 40 percent of 1996 expenditures—to be consistent with the advisory commission's estimated index bias of 20 percent.[11]

Table 2 presents U.S. Department of Agriculture data on changes in per capita consumption of fresh fruit from 1975 to 1995. The change in consumption is shown, somewhat unconventionally, as a percentage of 1995 consumption, because the shares in equation 1 refer to current period consumption. As the advisory commission observes, per capita consumption of many fruits has indeed increased substantially over this period: in particular, limes, cranberries, grapes, kiwifruit, mangos, papayas, and strawberries. Despite these large increases, however, most of these items continue to represent a small percentage of overall fruit consumption, so that the total increase in per capita fruit consumption as a share of 1995 consumption (measured in pounds) is only 14 percent. The largest absolute increase in consumption of fruit is that for bananas. We are confident that there was no important improvement in seasonal availability of bananas and that there were only minor

10. If a new variety fully replaces an old one, the consumer surplus calculation should deduct the lost surplus of the disappearing variety from the surplus gained from the new variety.

11. Ideally, one would examine monthly consumption data to isolate seasonal changes in consumption, but such data do not appear to be available.

Table 2. Per Capita Consumption of Fresh Fruits, by Type, 1975–95
Units as indicated

| | Pounds per capita | | | Change, 1975–95 | |
| | | | | As percentage of | In |
Type of fruit	1975	1985	1995	1995 consumption	pounds
Citrus					
Oranges and temples	15.9	11.6	12.3	−29.6	−3.6
Tangerines and tangelos	2.6	1.5	2.0	−27.9	−0.6
Lemons	2.0	2.3	2.9	32.1	0.9
Limes	0.2	0.6	1.2	81.7	1.0
Grapefruit	8.4	5.5	6.0	−38.4	−2.3
Total	29.0	21.5	24.4	−18.9	−4.6
Noncitrus					
Apples	19.5	17.3	18.9	−3.0	−0.6
Apricots	0.1	0.2	0.1	20.0	0.0
Avocados	1.2	1.8	1.4	10.9	0.2
Bananas	17.6	23.5	27.4	35.6	9.8
Cherries	0.7	0.4	0.2	−187.5	−0.5
Cranberries	0.1	0.1	0.3	53.3	0.2
Grapes	3.6	6.8	7.6	52.7	4.0
Kiwifruit	. . .	0.1	0.5	. . .	. . .
Mangos	0.2	0.4	1.1	85.8	1.0
Peaches and nectarines	5.0	5.5	5.4	8.5	0.5
Pears	2.7	2.8	3.4	19.4	0.7
Pineapples	1.0	1.5	1.9	46.6	0.9
Papayas	0.2	0.2	0.4	56.8	0.2
Plums and prunes	1.3	1.4	0.9	−41.5	−0.4
Strawberries	1.8	3.0	3.8	52.1	2.0
Total	55.1	65.1	73.5	25.0	18.4
Total	84.1	86.5	97.9	14.1	13.8

Source: U.S. Department of Agriculture (1996a, table F-29).

increases in consumption of new varieties of bananas over this period. In addition, consumption of apples did not change significantly and consumption of oranges decreased. We wonder whether the use of apples for baking may have decreased during this period, masking a possible increase in the consumption of raw apples.

We do not attempt to calculate the overall bias using equation 1 because doing so would require average price or expenditure data for each of the detailed categories, which we have not been able to assemble. As mentioned above, under Hausman's model and our earlier as-

sumptions, to be consistent with the commission's bias estimate consumption of new varieties and seasonal items would need to increase by about 40 percent over thirty years, which annualizes to 25 percent over the twenty years for which we have data. If increased consumption of seasonal varieties was relatively unimportant for apples, bananas, and citrus fruits, which, according to the Consumer Expenditure Survey (CEX) of the Bureau of Labor Statistics, together represent 61 percent of dollar expenditures on fresh fruit in 1995, it would be difficult for increased seasonal consumption of the other fruits to produce an estimated bias as large as the commission proposes.

Table 3 shows changes in consumption of vegetables from 1972 to 1995. Unlike the data for fruit, the data for vegetables show important increases in consumption for many items and thus appear, at first glance, to be consistent with the advisory commission's estimates of bias. Under the assumptions stated above, our consumer surplus calculations indicate that for the commission's estimate to hold, the growth in consumption over thirty years would need to be about 40 percent of current consumption, which annualizes to 29 percent over the twenty-three years for which we have consumption data. This is, in fact, very close to the overall increase over this period: 27 percent. We are skeptical, however, about concluding that the increase in consumption derives entirely from improved seasonal availability. A BLS food specialist, Bill Cook, has suggested that the increase in seasonal availability of fresh vegetables mostly occurred before 1985, as evidenced by a 1984 internal BLS study showing that 91 percent of the CPI price quotes for the "other fresh vegetables" category were by then available year round.[12] Table 3 shows, however, that almost half of the increase in consumption of fresh vegetables occurred after 1985. Part of the increase appears to have been driven by shifts in preferences, perhaps as a response to improved knowledge about the health benefits of fresh vegetables.

Shelter

Because rent and owners' equivalent rent have a very large weight in the index, any quality bias in shelter is particularly important. Con-

---

12. Internal memorandum from William L. Weber to Dan Ginsburg, U.S. Bureau of Labor Statistics, May 25, 1984.

Table 3. Per Capita Consumption of Fresh Vegetables, by Type, 1972–95

Units as indicated

	Pounds per capita			Change, 1972–95	
				As percentage of	*In*
Type of vegetable	*1972*	*1985*	*1995*	*1995 consumption*	*pounds*
Asparagus	0.4	0.5	0.6	33.3	0.2
Broccoli	0.7	2.6	3.2	78.1	2.5
Carrots	6.5	6.5	10.1	35.6	3.6
Cauliflower	0.8	1.8	1.3	38.5	0.5
Celery	7.1	6.9	6.4	− 10.9	− 0.7
Sweet corn	7.8	6.4	7.8	0.0	0.0
Bell peppers	2.4	3.8	5.8	58.6	3.4
Onions	10.7	13.6	17.7	39.5	7.0
Tomatoes	12.1	14.9	16.6	27.1	4.5
Cabbage	8.5	8.8	9.1	6.6	0.6
Spinach	0.3	0.7	0.6	50.0	0.3
Cucumbers	3.0	4.4	5.6	46.4	2.6
Artichokes	0.4	0.7	0.4	0.0	0.0
Snap beans	1.5	1.3	1.6	6.3	0.1
Eggplant	0.4	0.5	0.4	0.0	0.0
Escarole or endive	0.6	0.4	0.2	− 200.0	− 0.4
Garlic	0.4	1.1	2.1	81.0	1.7
Lettuce					
Head	22.4	23.7	21.6	− 3.7	− 0.8
Leaf or romaine	. . .	3.3	6.0	. . .	. . .
Watermelon	12.3	13.5	15.9	22.6	3.6
Cantaloupe	7.0	8.5	9.9	29.3	2.9
Honeydews	1.0	2.1	2.4	58.3	1.4
All others	0.8	0.8	0.7	− 14.3	− 0.1
Total	107.1	126.8	146.0	26.6	38.9

Source: U.S. Department of Agriculture (1996b, table 14).

sequently the section on shelter is one of the most detailed in the advisory commission's report and includes many citations to data about changes that have occurred in the characteristics of housing. In our opinion, however, the commission's analysis of the data on shelter is misleading and its conclusions on bias are invalid.

The commission's reasoning is, essentially, as follows. Over the period 1976–93, the median rent increased about 1 percent per year faster than the CPI rent index. This fact might suggest that the quality adjustments embedded in the index are substantial. According to the advisory commission, however, these quality adjustments remain inadequate because of a supposed dramatic increase in average apartment

size: ''From the evidence we have examined, we believe that 20 percent is a low-end estimate of the increase in the average size of apartments [between 1976 and 1993], which would support the conclusion that the average rent per square foot has increased no faster than the CPI.''[13] In addition, the commission estimates that other improvements, including ''appliances, central air conditioning, and improved bathroom plumbing, and other amenities,'' amount to 10 percent over the past forty years, giving a net upward bias of 0.25 percent per year. Several observers have pointed out that rents generally increase less than proportionally to apartment size, which implies that the advisory commission's proportional adjustment for apartment size overstates the increase in rents. A more fundamental problem, however, is that the commission's factual premise—that average apartment size increased by 20 percent from 1976 to 1993—is clearly wrong. Although data giving an exact measure of the growth in size of rental units since 1976 are not available, one careful examination of data from several sources concludes that the increase was probably about 6 percent; that is, the commission's estimate is too high by, roughly, a factor of three.[14] This estimate is based on the following evidence:

—The American Housing Survey, conducted by the Census Bureau, has published the median size of single detached and mobile home rental units since 1985. From 1985 to 1993, the median unit size increased by 2.2 percent, an average of less than 0.3 percent per year.[15]

—The Energy Department's Residential Energy Consumption Survey collected data on the average square feet of housing units over the period 1980–93. Although these data are based on smaller samples than the American Housing Survey, the trend is the same. The Energy Department data show a 3.5 percent increase in the average size of all rental units from 1980 to 1993; that is, about 0.3 percent per year.[16]

—In order for the average size of rental units to have increased as rapidly as the advisory commission claims, the average size of newly constructed apartments would need to be much larger than the average size of existing apartments. Data comparisons indicate that newly con-

13. U.S. Senate, Committee on Finance (1996, p. 30).
14. Moulton (1997).
15. U. S. Department of Commerce and U.S. Department of Housing and Urban Development (1985, 1993; table 2-3).
16. U.S. Energy Information Administration (1980, table 9; 1993, table 3.4).

structed apartments were not, in fact, much bigger than existing apartments over the period 1976–93.

After adjusting for the advisory commission's overstatement of the increase in apartment size and taking account of several other measurement issues, a small downward bias is more plausible than the upward bias claimed in the report.[17]

Appliances, including Electronic

The commission's estimate of bias for appliances—5.6 percent per year—is the largest of its estimates. It is also probably the best documented. A number of academic and government studies develop hedonic quality adjustment models and find upward quality bias for personal computers, television, video equipment, and other items in this category. Hedonic methods are already being applied in the Producer Price Index (PPI), for personal computers and peripherals. Matthew Shapiro and David Wilcox describe substitution bias as harvesting "low hanging fruit," pointing out that there are generally accepted methods for removing substitution bias.[18] We view the application of hedonic methods to the appliances category as harvesting the low hanging fruit of the quality bias problem, and the BLS is currently developing such methods and improved sampling of new products within this category.

Apparel and Upkeep

The advisory commission's estimates of bias for apparel and upkeep are based on a recent study by Robert Gordon that compares the CPI apparel indexes with indexes constructed from prices in the Sears catalog.[19] In addition to noting the obvious difference in data sources, we would point out an important methodological difference between the Gordon study and the CPI. By contrast with CPI methods, Gordon measures year-to-year price changes only for apparel items that remain identical from year to year. Many BLS studies find that price change tends to occur when new varieties or fashions are introduced.[20] Gor-

17. See Moulton (1997).
18. Shapiro and Wilcox (1997).
19. Gordon (1996).
20. See, for example, Triplett (1971), Armknecht (1984), Armknecht and Weyback (1989), Liegey (1993, 1994), and Reinsdorf, Liegey, and Stewart (1996).

don's method of analysis would exclude these price increases; in essence, it attributes any price increases associated with the introduction of new fashion lines entirely to quality improvement. In BLS terminology, the price changes would be "linked out." Empirical evidence that Gordon's study may link out too much price change is provided by the fact that his index for women's apparel, where fashion changes and implied item substitutions are most common, shows no increase from 1984 to 1993, whereas his indexes for men's, boys', and girls' apparel increase between 14 percent and 17.5 percent over the period.

Furthermore, the apparel and upkeep category seems to be particularly sensitive to an issue that may affect several of the commission's estimates of bias; because of their volatility, apparel prices may have been affected by so-called formula bias and lower-level substitution bias.[21] Part of the method that BLS adopted in 1995 and 1996 to correct the problem of formula bias has, for other reasons, been applied to many of the apparel indexes since 1989. The rate of growth of the apparel and upkeep component of the CPI has obviously been affected by these changes. In addition, simulations of possible lower level substitution bias using a geometric mean index formula indicate that apparel and upkeep may also be particularly affected by lower-level substitution bias.[22] It is unclear whether the advisory commission avoided double counting when sorting through these various sources of bias to produce its estimate of quality bias.[23]

New Vehicles

In the case of automobiles, the advisory commission makes one of its few methodological recommendations with respect to quality adjustment: it recommends that the BLS should price cars and other durables using a cost of services ("user cost" or "leasing equivalence") approach. This is an important recommendation that will require significant research to evaluate. However, the suggestion seems, in part, to be motivated by a misinterpretation of BLS quality adjustment procedures. The advisory commission report states that the CPI has not

21. The advisory commission emphasizes these two separate sources of bias (U.S. Senate, Committee on Finance (1996, sect. 4).
22. Moulton and Smedley (1995).
23. See Baker (1997).

taken into account "the increased service lifetime of the typical new car." [24] Although it is true that the CPI does not adjust directly for changes in the average service lifetime of automobiles, many adjustments have been made for improvements that are related to durability. Without a major, detailed study it is impossible to know whether the CPI adjustments related to durability fully account for the increased lifespan of the typical car, but to attribute the entire increase in automobile service life to unmeasured quality bias clearly involves some double counting.

Motor Fuel

For the motor fuel category, the advisory commission attributes "a small upward bias of 0.25 percent per year to the CPI for ignoring the convenience and time-saving contribution of automatic credit-card readers built into gasoline pumps." [25] Because the commission applies this estimate over a ten-year period, the estimate of the cumulative bias from this source amounts to 2.5 percent. [26] Our approach to measuring the consumer surplus created by pay-at-the-pump credit card technology is to attempt to value the saving in time. Suppose that paying at the pump saves two minutes per fill-up, and that the customer's time is valued at $18 per hour (average total compensation per hour for all workers in private industry was $17.49 in 1996). Then the value of paying at the pump is 60 cents per fill-up. Assuming that ten gallons are purchased, the quality bias for the customer who pays at the pump is 6 cents per gallon, or roughly 4.5 percent of the cost of a gallon of gasoline.

Since this service is of value only to the customers who use it, one must next determine the approximate percentage of gasoline purchasers

24. U.S. Senate, Committee on Finance (1996, p. 35).
25. U.S. Senate, Committee on Finance (1996, p. 36).
26. We also note that the report does not address possible unmeasured decline in retail services, such as the introduction of fees for providing air for tires at some service stations. In addition, the advisory commission incorrectly assumes that the CPI does not make quality adjustments for air pollution mandates and, agreeing with this supposed BLS practice, makes no bias adjustment for the mandates itself. Since BLS does, in fact, make cost-based adjustments for motor fuel pollution mandates, the commission presumably should have counted these as downward bias (see U.S. Bureau of Labor Statistics, "Quality Adjustment for Gasoline," *CPI Detailed Report*, January 1995, p. 8).

who use pay-at-the-pump technology. Although we have not found direct information on this percentage, the September 1996 issue of the trade journal *National Petroleum News* reports that 28 percent of the retail facilities operated by thirteen oil companies had installed pay-at-the-pump technology as of 1996.[27] Since many of the customers at these stations do not use credit cards, we attempt to find the percentage of gasoline customers who do so. We have not found published information, but an industry source has told us that roughly 35 percent of sales are made through credit cards.

A naive estimate of the proportion of sales using pay-at-the-pump technology would thus be 10 percent (28 percent × 35 percent). However, there are at least three reasons why this estimate is too low: first, pay-at-the-pump technology was doubtless first targeted at high-volume sites in areas with high credit card usage; second, the availability of the technology induces customers to make more use of credit cards; and third, the technology is spreading rapidly, so that even estimates published in September 1996 will understate current availability. Consequently we take 25 percent as our estimate of the percentage of customer sales made with pay-at-the-pump technology at the end of 1996. Under these assumptions, we calculate the cumulative index bias from neglecting the benefits of this technology as approximately 1.1 percent (4.5 percent × 25 percent), which is less than half of the advisory commission's estimate.

Medical Care

The advisory commission's estimate of bias in the medical care component of the index appears to have been largely based on just two recent empirical studies—one concerning the treatment of cataracts and the other concerning heart attacks—both of which identify large quality improvements that are missed in the calculation of the CPI.[28] Although there have been enormous improvements in medical technology over time, we are not convinced that these two examples should be considered representative of the unmeasured quality advances in the treatment of medical conditions in general.

27. "Pay-at-the-Pump Shows Solid Growth in '90s," *National Petroleum News*, September 1996, p. 22.

28. The two studies cited by the commission are Shapiro and Wilcox (1996) and Cutler and others (1996).

Without necessarily endorsing the advisory commission's estimate of bias, we agree that BLS methods are not likely to capture fully the quality improvements that have occurred in medical services. Adjustment for quality change in this component is the most challenging in the index. The BLS has recently taken steps to address some of the problems in measuring price change for medical care. In January 1995 it changed the method of pricing prescription drugs when generic alternatives become available to reflect the research of Zvi Griliches and Iain Cockburn.[29] For hospital services, in January 1997 the CPI adopted the practice, already used in the PPI, of pricing courses of treatment (as represented by bundles of services on selected patient bills) rather than individual medical inputs. These measures do not overcome the difficult measurement issues in these categories, but BLS is actively pursuing further research.

Cellular Telephone Service

The CPI does not currently include cellular telephone service in its sample—this item will be added to the market basket during the January 1998 update. Although the advisory commission's report is unclear as to the extent that the commission counts cellular phones in its estimate of bias for telephone service, a recent study by Hausman concludes that the bias from late introduction and uncounted consumer surplus is large.[30] His analysis overstates the magnitude of the bias, however, because it is based on data that include business use, which is not within the scope of the CPI.

The share of total *consumer* expenditure going to cellular telephone service in 1995, according to the Consumer Expenditure Survey, is 0.088 percent. Hausman's estimate of its price elasticity is -0.5. Since his data predominately represent business users and we think that consumers are more price sensitive than businesses, his elasticity estimate may be smaller than our ideal number. Nevertheless, applying the CEX expenditure share and his elasticity estimate to equation 1, we calculate the cumulative bias in the CPI from failure to include cellular telephones as 0.088 percent ($-0.5 \times 0.088/-0.5$) through 1995. Applied over a

29. Griliches and Cockburn (1994).
30. Hausman (1997).

five-year period, this represents a bias of a little less than 0.02 percent per year.

We would not say that 0.02 percent per year is a small number, nor that the delay in including cellular telephones is justified. To the contrary, in our experience, any bias that affects the ''all items'' index by 0.02 percent per year should be considered large. We do, however, think that showing that missing consumer surplus in one of the most important new products in recent years produces a bias of roughly this magnitude helps to explain why we consider many of the larger numbers that appear in the advisory commission's report implausible.

Summary

The above discussion does not cover all the components of bias, and we do not attempt to give our own estimate of quality bias. For two major components—medical care and appliances—we agree with the advisory commission that the evidence of upward bias due to quality and new goods is convincing. We also agree that new products contribute an upward bias to many components of the index, though we think that the commission and others have overstated the magnitude of this bias. However, there are many important components of the index— including shelter, apparel and upkeep, and new vehicles—for which we have not seen convincing evidence of upward bias and, in some circumstances, think it possible that the methods used by the BLS may overadjust for quality change.

In the case of appliances, the availability of substantial academic research together with the recently announced CPI budget initiative and research projects currently being conducted by BLS staff suggest that the quality biases can be corrected relatively soon.[31] As noted above, the BLS has also made improvements to the medical care indexes, but the quality adjustment problems in this component are difficult, and progress may come slowly.

CPI Methods for Addressing Quality Change

The methods for collecting data for the CPI were designed to identify and adjust for quality changes.[32] The BLS selects sample items by

31. On the CPI budget initiative, see Abraham (1997b).
32. New products are usually ''linked'' into the sample through the regular (cur-

probability methods, so that the items that are repriced each month are representative of consumer purchases. Each item is described in detail on a checklist, to ensure that identical items are compared for price while they remain in the sample. Each time the price data are collected, the data collector compares the item with the detailed description to see if any of the characteristics of the item have changed. If, during the monthly (or bimonthly) price collection, the exact sample item has become permanently unavailable at that outlet, the data collector selects a similar item as a replacement. This item replacement process is the focus of our analysis in the remainder of this paper.

After an item has been replaced, a BLS commodity analyst examines the descriptions of the old and new versions of the item to determine which quality adjustment procedure is to be applied. These methods were developed to be applied in an environment in which 80,000 sample prices are collected and processed each month, of which roughly 4 percent represent item replacements. These item replacements are much more important than may be suggested by the 4 percent monthly attrition rate, however. Approximately 30 percent of all sample items scheduled to remain in the sample for the full year (that is, not scheduled for regular sample rotation) need to be replaced at some time during the year. Detailed explanations of the various methods used by BLS to adjust for quality changes due to item replacements are available elsewhere; here, we give a heuristic description.[33]

COMPARABLE ITEMS. In some cases, the commodity analyst examines the differences between the two specifications and determines that the change has not resulted in a significant change in quality, so that the

---

rently, five-year) rotation of samples, although some new products may enter through the sample replacement procedures described below. Current CPI procedures do not generate a reduction in the cost of living from the consumer surplus attributable to new products.

33. Recent papers that give detailed descriptions of the methods currently used for quality adjustment include Armknecht, Lane, and Stewart (1997) and Reinsdorf, Liegey, and Stewart (1996).

The terminology used to describe the various methods of handling item substitutions is not standardized. We follow Armknecht and Weyback (1989). The advisory commission (U.S. Senate, Committee on Finance, 1996) and Gordon (1990) use different terminology: ''direct comparison'' instead of ''comparable,'' ''linking'' instead of ''overlap method,'' and ''deletion'' instead of ''link method;'' the commission's report omits the relatively new method of class-mean imputation.

Figure 1. Comparable Items
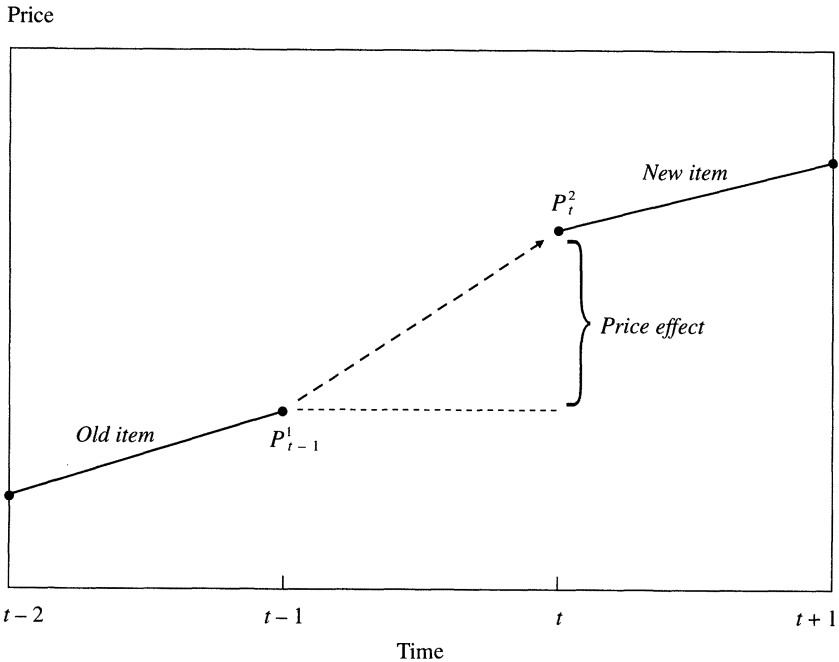
Price

prices of the old and new versions can be directly compared. Let $P^1_{t-1,i}$ denote the price in the previous period $(t - 1)$ of the old version (denoted by superscript 1) of item i, and P^2_{ti} denote the price in the current period (t) of its new version (superscript 2). As shown in figure 1, this method counts the entire price change, $P^2_{ti}/P^1_{t-1,i}$, as part of inflation; that is, no quality difference is attributed to the new version of the item. These comparable replacements would typically involve versions of an item that differ by minor changes in styling or other minor differences in characteristics that do not reflect quality.

OVERLAP METHOD. This method is used when prices of both the old and the new versions are available during an overlap period, so that the difference in price levels can be used as an estimate of the difference in quality. As shown in figure 2, the pure price change (or "price effect") before period t is measured by the price change of the old version, and the price change after period t is measured by the price change of the new version. It is relatively uncommon to have prices for

Figure 2. Overlap Method

Price

Time
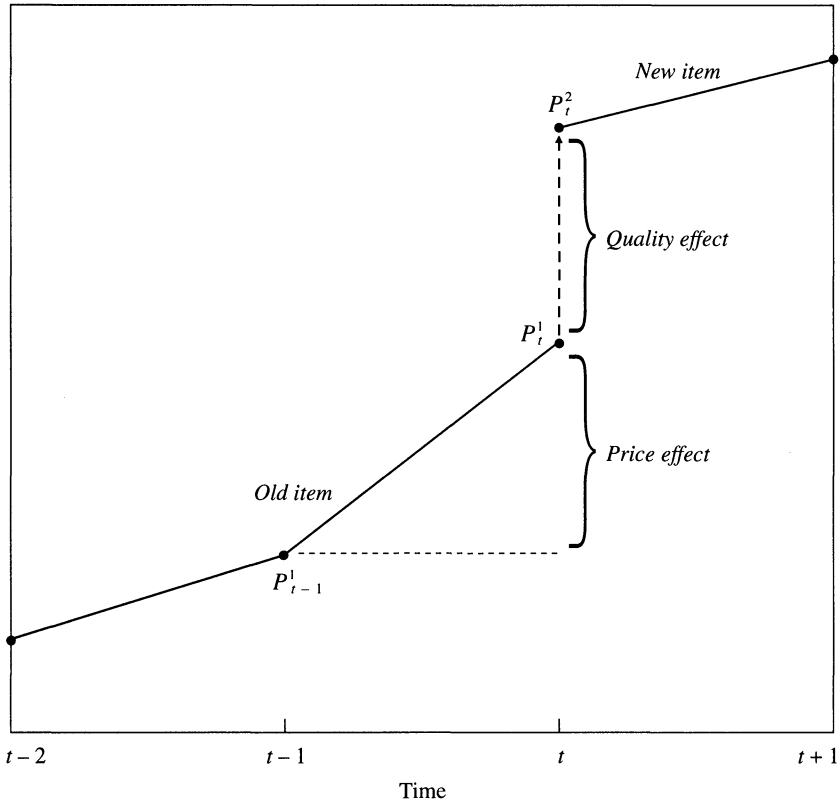
an item and its replacement during an overlap period, but an aggregate version of the overlap method is used when an entire CPI component sample is replaced during sample rotation. Both the old and new samples are collected during an overlap period, t, and the old sample is used to measure the price change from $t - 1$ to t, while the new sample is used to measure the price change from t to $t + 1$.

LINK METHOD. The disappearance of an item is typically not detected until the item is no longer available at the sample outlet, so prices of the old and new versions are not available concurrently. Consequently the overlap method cannot be used to estimate the portion of the price difference that is attributable to inflation and the portion that is attrib-

Figure 3. Link Method and Class-Mean Imputation

Price

Time
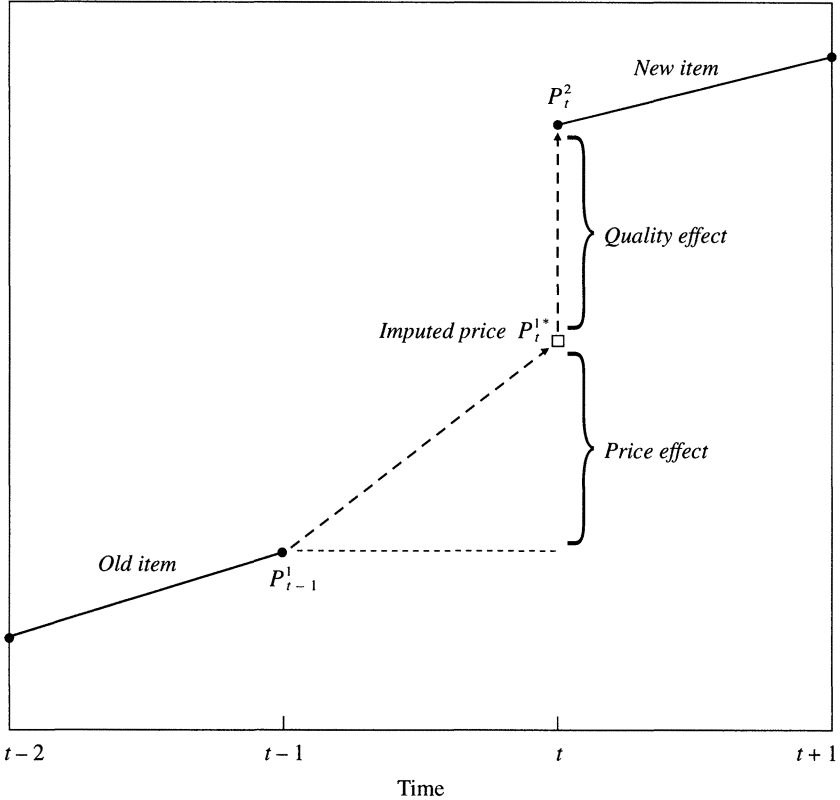
utable to quality change. Each item is part of a stratum of similar goods. When an item vanishes, the link method first calculates the rate of inflation for its stratum during that month, omitting the item from the calculation of price change. Suppose, for example, that the inflation rate for the other goods in the stratum is 2 percent, but that the replacement version of the item, when it appears, costs 5 percent more than the earlier version. As illustrated in figure 3, the link method effectively assumes that of that 5 percent, 2/5 is due to the overall rise in the price of goods and the other 3/5 is due to an improvement in quality. Note that the estimated quality change is essentially a residual in this calculation.

CLASS-MEAN IMPUTATION. This method was introduced to the CPI new cars index in 1989 and to other items in 1992.[34] Like the link method, class-mean imputation also imputes a price change and treats the quality change as a residual. In this case, however, the price change is imputed from a set of similar items that are classified as comparable replacements or are directly adjusted for quality change. This method is based on the assumption that the rate of inflation when a new model of an item replaces an earlier model is different from the inflation rate when the model does not change.

DIRECT QUALITY ADJUSTMENT. These methods are applied when there is information available with which to directly estimate the dollar value of the change in quality. Sometimes (in particular, in the cases of new and used cars and motor fuel) manufacturers provide information on the cost of the quality improvement. In other cases, the hedonic method is used to estimate the relationship between price and quality by regressing price on characteristics of the goods. The coefficients of these regressions are then used to infer the value of changes in characteristics of the goods in the sample. The CPI has used hedonic methods since 1988 for calculating the effects of depreciation and other housing characteristics on rent, and since 1991 for calculating quality changes in apparel. Figure 4 illustrates a direct quality adjustment, in which an adjustment is made to the period $t - 1$ price of the old item for the estimated value of the quality improvement embedded in the new item.

Potential Errors in Quality Adjustment

In cases when a change in characteristics is observed but the replacement is classified as comparable, it is possible that small, unmeasured quality changes are incorrectly attributed to price change. The advisory commission concludes that this leads to upward bias because "most goods tend to undergo steady improvement."[35] This conclusion does not necessarily follow from its premise, however, because replacements are not randomly classified as comparable. Suppose, for example, that the value of the relative change in quality that occurs with item replacement is uniformly distributed between -25 percent and 75 percent, and that all quality changes smaller than 25 percent are classified as

---

34. Armknecht, Lane, and Stewart (1997); Reinsdorf, Liegey, and Stewart (1996).
35. U.S. Senate, Committee on Finance (1996, p. 25).

Figure 4. Direct Quality Adjustment

Price

Time
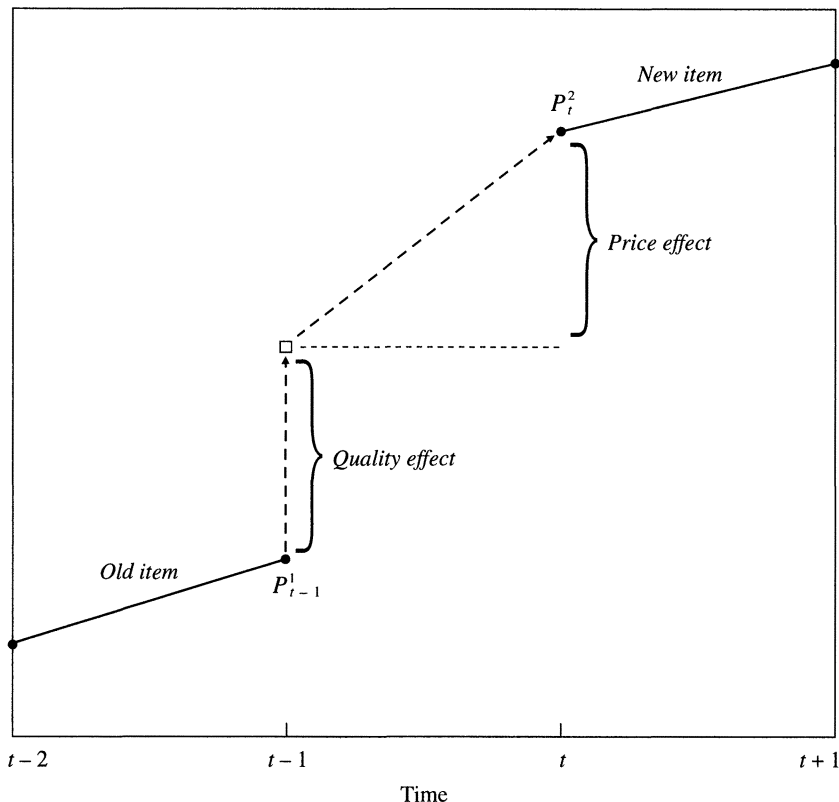
comparable. In this case, the average bias of the comparable items is zero, even though the average quality change for all replacement items is 25 percent.

The advisory commission's report argues that "when a new model is introduced that is more expensive than an old model, but it gains market share, we can conclude that it was superior in quality to the old model by more than the differential in price between the two." [36] This statement suggests that the bias resulting from use of the link and overlap methods is necessarily upward. We agree with this reasoning

36. U.S. Senate, Committee on Finance (1996, p. 24).

in regard to major new goods and the most significant improvements in quality.[37]

The commission does not mention, however, that in many cases the increase in price exceeds the differential in quality, so that the link and overlap methods may result in downward bias. This is especially likely to occur when a model change involves a modest improvement in quality, perhaps combined with changes in fashion or style. Triplett describes this situation:

> During a period of rising prices, if a price change and a change in the varieties in retailers' stocks are both imminent, sellers at all levels are likely to try to arrange both changes to coincide, rather than putting them into effect separately. If this occurs, prices of unchanged varieties are not moving parallel to those of the new varieties encountered. Because price increases coincide with substitution, deletion of the price quotation misses some of the true price change. Thus, when prices are rising, this method of handling quality change tends to bias the index downward.[38]

Considerable BLS research, especially in the apparel and upkeep category, has found empirical evidence of such downward bias.[39]

Direct quality adjustments are also potentially subject to bias in either direction. On the one hand, adjustments based on manufacturers' estimates of cost may underadjust for quality if, as the advisory commission says, they miss ''quality improvements achieved by those firms which supply better materials and inputs to producers of final goods.''[40] Griliches, on the other hand, has suggested a potential bias in the opposite direction: ''I am concerned that by basing such adjustments largely on data furnished by manufacturers and on 'producer costs' [the BLS] may wind up overestimating 'quality change,' accepting as 'improvements' expenditures which consumers may not interpret as such.''[41] Well-specified hedonic models probably represent the most robust method currently used by BLS, but it is well known that errors

37. Studies of price change for computers have consistently found that hedonic methods result in greater price declines than matched model (that is, linking) methods; see, for example, Cole and others (1986) and Berndt, Griliches, and Rappaport (1995).

38. Triplett (1971, p. 185). Triplett (1990) presents a mathematical analysis of this case.

39. See Armknecht (1984), Armknecht and Weyback (1989), Liegey (1993, 1994), and Reinsdorf, Liegey, and Stewart (1996).

40. U.S. Senate, Committee on Finance (1996, p. 26).

41. Griliches (1971, p. 11).

in hedonic model specification can lead to errors of quality adjustment in either direction. We conclude that the sign of any quality bias is almost never known a priori, but must be determined empirically.

Decomposing Price and Quality Changes

The present study, modifying a methodology developed by Paul Armknecht, decomposes the relative price difference between old and new items for each item replacement during 1995 into a pure price change (that is, the part of the price difference that is counted as inflation in the CPI) and the implied quality change.[42] In this study, we analyze item replacements that take place in 171 strata of nonshelter items.[43]

Methods of Decomposing Quality Change

The CPI can be written as a weighted average of price ratios for the sample items:

$$(2) \qquad I_{t,t-1} = \sum_i W_{t-1,i} R_{t,t-1,i},$$

where $I_{t,t-1}$ is the relative change of the index from $t - 1$ to t, $W_{t-1,i}$ is the weight of sample item i, which depends in a complicated way on the sampling and past price history of the item relative to other items, and $R_{t,t-1,i}$ is the relative price change of the sample item. In the case of a continuously priced item, the relative price change is simply the ratio of its price in period t to its price in period $t - 1$; that is, $R_{t,t-1,i} = P_{ti}/P_{t-1,i}$. For a replacement item, the value of the price relative depends on the method used to adjust for quality. We decompose total price change as measured in the CPI into a price change derived

---

42. Armknecht (1984) applies the methodology to CPI item substitutions that occurred during 1983, and Armknecht and Weyback (1989) analyze item substitutions that occurred during 1984. Since this method is applied solely to quality changes arising from item substitutions, any changes in the average quality of items in the sample that occur for other reasons, such as sample replacement, are omitted.

43. We exclude twenty-three unpriced strata, four health insurance strata, rent, owners' equivalent rent, household insurance, postage, babysitting, care of invalids in the home, used cars, automobile finance charges, and sports vehicles.

from replacement items and a price change derived from continuously priced items:

$$
(3) \qquad I_{t,t-1} = \underbrace{\sum}_{\substack{\text{unchanged} \\ \text{items}}} W_{t-1,i} \frac{P_{ti}}{P_{t-1,i}} + \underbrace{\sum}_{\substack{\text{replacment} \\ \text{items}}} W_{t-1,i} R_{t,t-1,i}.
$$

In the case of replacement items, we define the total (unadjusted) relative price change between the old version (denoted by superscript 1) and the new version (superscript 2) of item i as $T_{t,t-1,i} = P_{ti}^2/P_{t-1,i}^1$. Then the portion of the total relative price change that is used in the CPI calculation is $R_{t,t-1,i}$, the "pure price relative," and the remaining portion is $A_{t,t-1,i}$, the quality adjustment:[44]

$$
(4) \qquad\qquad T_{t,t-1,i} = R_{t,t-1,i} A_{t,t-1,i}.
$$

Within the category of replacement items, we further decompose price changes by method of quality adjustment used. Finally, for each method of quality adjustment, we decompose the overall price difference between the old item and the new item into a price change component and a quality difference component.

COMPARABLE ITEMS. The pure price relative for comparable replacements and continuously priced items is the entire price change:

$$
(5) \qquad\qquad R_{t,t-1,i} = T_{t,t-1,i} = P_{ti}^2/P_{t-1,i}^1.
$$

The quality is assumed not to change in these cases.

OVERLAP METHOD. For noncomparable replacements with overlap prices (that is, prices in the same time period for both varieties of the item), the assumed quality ratio is the contemporaneous price differential between varieties and the price relative is the change in the price of the old variety:

$$
(6) \qquad\qquad A_{t,t-1,i} = P_{ti}^2/P_{ti}^1,
$$

$$
(7) \qquad\qquad R_{t,t-1,i} = P_{ti}^1/P_{t-1,i}^1.
$$

A common application of the overlap method in the CPI involves creating an artificial full price for a sale-priced item to avoid recording a

---

44. Equation 4 presents the multiplicative version of the decomposition. The additive version sets $A_{t,t-1,i}^* = T_{t,t-1,i} - R_{t,t-1,i}$, and defines the "pure price effect" as $R_{t,t-1,i} - 1$ and the "quality effect" as $A_{t,t-1,i}^*$.

large quality change when the sale- priced item is replaced by a full-priced item. Consequently the price effects appear to be surprisingly large. This artificial overlap, or "return from sale," method is currently being replaced by other methods.[45]

LINK METHOD AND CLASS-MEAN IMPUTATION. Under both of these methods, the implicit quality change is the differential between the price of the new variety and the imputed price of the old variety, $P_{ti}^{1}*$, and the pure price relative is the differential between the imputed price and the previous price of the old variety:

$$(8) \qquad\qquad A_{t,t-1,i} = P_{ti}^{2}/P_{ti}^{1}*,$$

$$(9) \qquad\qquad R_{t,t-1,i} = P_{ti}^{1}*/P_{t-1,i}^{1}.$$

As described above, the two procedures differ in their methods of imputation.

DIRECT QUALITY ADJUSTMENT. Under these methods, the quality effect is the ratio of the measured value of the quality difference between the two varieties, $D_{t-1,i}$, to the price of the old item. The price effect is the differential between the price of the new variety and the price of the old variety, adjusted for the quality difference:

$$(10) \qquad\qquad A_{t,t-1,i} = 1 + (D_{t-1,i}/P_{t-1,i}^{1}),$$

$$(11) \qquad\qquad R_{t,t-1,i} = P_{ti}^{2}/(P_{t-1,i}^{1} + D_{t-1,i}).$$

Once the price effect and the quality effect for each priced item have been identified, the contribution to the total price change of the pure price changes resulting from each type of quality adjustment can be summed, using the same weights as in equations 2 and 3 for each type of replacement:

$$(12) \qquad\qquad \text{price effect} = \sum_{i} W_{t-1,i}(R_{t,t-1,i} - 1).$$

The sum of the price effects for each type of replacement and the price effect of unchanged items equals the change of the price index.

45. Return from sale processing is described by Armknecht and Weyback (1989, p. 114).

Aggregating Quality Adjustments

Aggregation of the quality adjustments to obtain an aggregate quality effect is complicated by two features of these data. First, because quality adjustment may be triggered by the attrition of items as well as changes in their characteristics, the measure of quality change must account for the effects of sample turnover as well as changes that are made to existing products. When an item is no longer available, the price collector attempts to select a similar replacement item; but when no similar item is available at the outlet, a dissimilar item may be selected—for example, the sample price of a piano lesson might be replaced by that of a violin lesson. This type of sample turnover can, in some cases, lead to quality adjustments that are very large, either up or down. Given the enormous variety of items priced by the CPI and the large number of item replacements, there is no simple way to define and screen out replacements that involve substantially dissimilar items. We attempt to eliminate all cases involving data errors or simple changes in units or quantities—for example, repackaging, such as the replacement of the price of a contract for ten piano lessons with one for twenty lessons—before aggregating the quality adjustments. Further effects of sample turnover may be alleviated by truncating some of the outlying quality adjustments.

A second issue affecting aggregation is the asymmetry (skewness) of the distribution of $A_{t,t-1,i}$. We expect the median of the distribution to be near one, but the distribution is bounded below at zero for downward quality adjustments and is unbounded for upward quality adjustments. This suggests that in estimating the ''average,'' or central, tendency of the quality changes, it may be appropriate to aggregate using the logarithms of $A_{t,t-1,i}$. The theory for aggregating measures of quality change for individual items to an overall measure of quality change has not been fully worked out. We have identified some cases for which arithmetic aggregation as used in Laspeyres indexes may be appropriate—for example, when quantity consumed is not correlated with quality—and other cases for which logarithmic aggregation is most appropriate—for example, when quantity consumed declines in proportion to quality. The data that would be needed to test these alternative assumptions, however, are not readily available.

We have one important caveat. The present analysis is an attempt to

provide some descriptive measures of the quality adjustments that occur due to item replacement, as part of the CPI data collection process. For several reasons, it would be inappropriate to interpret them as measuring the rate of improvement in the quality of goods and services in the economy. They will miss some quality changes that are captured in other phases of the CPI data collection process, in particular, those reflected in the regular sample rotations. Furthermore, the replacement item selected by the data collector may not be the same as the substitute item selected by the consumer in response to the change or disappearance of the original item. It is entirely possible that consumers systematically substitute higher or lower quality items than those selected by BLS data collectors.

We examine both an additive measure and a logarithmic measure of aggregate quality change. The additive version is exactly analogous to equation 12, using the additive quality adjustments defined above in note 44:

$$(13) \qquad \text{additive quality effect} = \sum_i W_{t-1,i} A^*_{t,t-1,i}.$$

Note that the sum of the price effect and the additive quality effect has a simple interpretation: it is the index change that would have been calculated if the unadjusted price relative, $T_{t,t-1,i}$, had been used in place of the pure price relative, $R_{t,t-1,i}$, in equation 3.

The logarithmic version is

$$(14) \qquad \text{logarithmic quality effect} = \sum_i W^g_i \ln A_{t,t-1,i},$$

where the weights are those that are appropriate for the geometric mean index form.[46] The logarithmic version treats upward quality adjustments symmetrically with equal proportionate downward quality adjustments, so, under certain assumptions, it may better reflect the "typical" quality change.

As mentioned earlier, outliers might largely be the result of sample turnover. To deal with this problem, we consider two truncation rules, both of which are symmetric around zero in the logarithm of the quality

---

46. In particular, the weights used for the geometric mean index are not proportional to $P_{t-1,i}$, in contrast to the additive "modified Laspeyres index" form that is currently used in the U.S. CPI. For details, see Moulton and Smedley (1995).

changes.[47] Method A removes from the calculation any $A_{t,t-1,i}$ greater than 5 or less than 0.2 (ln P_t/P_{t-1} = ±1.61). Method B is more stringent and removes any $A_{t,t-1,i}$ greater than 2 or less than 0.5 (ln P_t/P_{t-1} = ±0.69).

Rates of Item Replacement in 1983, 1984, and 1995

One can examine changes in BLS quality adjustment methods by look-ing at rates of various types of sample replacement methods. Table 4 compares the percentage of each type of item replacement in the period 1983–84, as studied by Armknecht and Donald Weyback, with the percentage in 1995.[48] There have been important changes. Although the total percentage of price quotes that were replacements in 1995 (3.9 percent) is about the same as during 1983–84, the percentage of item replacements declined within food and beverages, apparel and upkeep, entertainment, and other goods and services, and increased within med-ical care. The overall rate of replacement did not fall along with the components, because of a shift in sample composition away from food and beverages and toward apparel and upkeep.

The percentage of replacements classified as comparable, relative to all quotes, has increased substantially, from 1.7 percent in 1984 to 2.5 percent in 1995. That increase can probably be attributed, first, to major efforts to redesign checklists so that BLS field staff are more likely to select directly comparable replacement items; and second, to hedonic studies which have indicated that some item characteristics have little effect on prices, so that products with different levels of those charac-teristics can be treated as comparable in quality. These changes dem-onstrate, furthermore, that the use of hedonic regression analysis within the CPI program has had effects beyond the direct quality adjustments presented in the final column of table 4.

The use of the overlap and link price methods have both declined substantially over this period, since many replacement items that were formerly thought to require these methods are now either treated as comparable or adjusted using direct quality adjustment or class-mean

47. The distribution of the logarithm of quality changes is not centered exactly at zero, but is close enough to zero that the truncation is approximately symmetric in the logarithms.

48. Armknecht and Weyback (1989).

Table 4. Product Replacements, by Item Category and Replacement Method, 1983, 1984, and 1995

Percentage of all quotes in item category

			Replacement method				
Item category[a]	*Year*	*Total*	*Comparable items*	*Overlap method*	*Link method*	*Class-mean imputation*	*Direct quality adjustment*
All items studied	1983	3.85	1.56	0.23	1.74	. . .	0.32
	1984	3.95	1.70	0.23	1.71	. . .	0.30
	1995	3.90	2.54	0.05	0.57	0.32	0.41
Food and	1983	1.81	0.52	0.04	1.25	. . .	0.00
beverages	1984	1.93	0.52	0.08	1.33	. . .	0.00
	1995	1.63	0.84	0.02	0.73	0.00	0.03
Housing	1983	4.25	2.21	0.22	1.67	. . .	0.15
	1984	4.83	2.67	0.21	1.72	. . .	0.22
	1995	4.83	3.56	0.05	0.44	0.64	0.14
Apparel and	1983	17.34	7.15	2.69	7.46	. . .	0.03
upkeep	1984	17.59	7.80	2.43	7.27	. . .	0.09
	1995	11.10	8.50	0.20	0.24	1.10	1.06
Transportation	1983	6.72	3.13	0.06	1.35	. . .	2.18
	1984	5.80	3.02	0.07	0.96	. . .	1.74
	1995	6.69	3.60	0.00	0.27	0.76	2.05
Medical care	1983	2.22	0.65	0.03	0.91	. . .	0.64
	1984	2.19	0.80	0.03	0.99	. . .	0.38
	1995	3.91	2.33	0.08	0.80	0.00	0.69
Entertainment	1983	4.61	1.92	0.23	2.28	. . .	0.18
	1984	6.08	2.85	0.26	2.70	. . .	0.27
	1995	3.47	2.05	0.03	0.45	0.61	0.33
Other goods	1983	3.30	1.44	0.06	1.64	. . .	0.17
and services	1984	3.99	1.94	0.08	1.56	. . .	0.40
	1995	2.01	1.23	0.05	0.43	0.09	0.20

Source: 1983 data are from Armknecht (1984); 1984 data are from Armknecht and Weyback (1989); and 1995 data are authors' calculations from unpublished CPI data.

a. The following price quotes are excluded from the studies: under housing, residential rent and homeowners' equivalent rent (all years) and household insurance, postage, babysitting, and care of invalids (1995); under transportation, used cars (all years) and automobile finance charges (1995); under medical care, health insurance (all years); and under entertainment, magazines, periodicals, and books (1983, 1984) and sports vehicles, including bicycles (1995).

imputation. The newer class-mean imputation method is now used nearly as much as direct quality adjustment. Within the apparel and upkeep category, the use of direct quality adjustment has increased greatly.

Effects of Item Replacements on Price Change in the CPI

Table 5 shows the contribution of each type of item replacement to the overall price change during 1995. The change of the CPI for the items in this table was 2.16 percent. Of this, 1.09 percent is attributable to replacement items and 1.07 percent is attributable to continuously priced items—that is, items for which monthly or bimonthly price comparisons involve no replacement. The total price change due to replacement items is further decomposed by type of replacement: 0.60 percent comes from comparable items, 0.10 percent from items that were adjusted for quality change using the overlap method, 0.02 percent from items adjusted using the link method, 0.18 percent from items adjusted by class-mean imputation, and 0.19 percent from items treated with direct quality adjustment.

Many readers of an earlier draft of this paper were surprised by the large aggregate price effect attributable to the 4 percent of price quotes that are replacement items. The main reasons for this appear to be the tendency of manufacturers to increase prices coincident with introduction of new models or fashions and the tendency of retailers to discount prices before dropping an item.[49] The return of the discount price of a disappearing item to the full price of the replacement item is included in the price effect.

Table 5 also provides evidence of changes in BLS procedures for handling replacements, especially in the apparel and upkeep and new

49. For example, 28 percent of the overall price change in 1995 is attributable to comparable items, although these account for only 2.5 percent of all collected prices. Such a large price effect would seem to imply an unrealistically large price change for each comparable replacement. The resolution of this puzzle lies in the fact that prices are collected monthly or bimonthly, so that after the replacement has occurred, the same item will be reclassified as an unchanged item. Suppose, for example, that the prices of new cars are collected monthly, that a replacement occurs once during the year for each car in the sample, that the price of each car increases by 3 percent when the replacement occurs, and that no other price changes occur during the year. One of twelve monthly price collections involves an item replacement, so replacement items account for about 8 percent of collected prices, but they account for all of the price change.

Table 5. Pure Price Effects of Unchanged and Replacement Items, 1983, 1984, and 1995

Percentage point contribution

				Replacement items					
Item category[a]	Year	All quotes	Unchanged items	Total	Comparable items	Overlap method	Link method	Class-mean imputation	Direct quality adjustment
All items studied	1983	2.99	1.16	1.83	0.62	0.89	0.12	...	0.20
	1984	3.40	0.14	3.26	1.37	1.38	0.16	...	0.35
	1995	2.16	1.07	1.09	0.60	0.10	0.02	0.18	0.19
Food and beverages	1983	2.64	2.41	0.23	0.14	0.06	0.03	...	0.00
	1984	3.73	3.14	0.59	0.26	0.22	0.12	...	0.00
	1995	2.08	1.94	0.14	0.04	0.05	0.04	0.00	0.01
Housing	1983	2.26	1.63	0.63	0.27	0.32	0.03	...	0.01
	1984	3.00	1.90	1.10	0.57	0.43	0.03	...	0.06
	1995	2.06	1.31	0.75	0.53	0.05	0.01	0.18	-0.02
Apparel and upkeep	1983	2.94	-21.10	24.04	8.31	14.41	1.26	...	0.06
	1984	1.95	-39.62	41.57	17.49	22.73	1.45	...	-0.10
	1995	0.21	-7.20	7.41	4.68	0.90	-0.01	0.89	0.95
Apparel commodities	1983	2.53	-25.78	28.31	9.80	17.02	1.49	...	0.00
	1984	1.42	-47.59	49.01	20.57	26.75	1.71	...	-0.02
	1995	0.15	-8.07	8.22	5.19	1.00	-0.02	0.99	1.06

Transportation	1983	1.51	0.58	0.93	0.00	0.00	0.09	. . .	0.84
	1984	2.09	0.07	2.02	0.62	0.03	0.12	. . .	1.24
	1995	1.19	0.08	1.11	0.31	0.00	0.00	0.33	0.47
New vehicles	1983	3.22	−2.01	5.23	0.31	−0.03	0.32	. . .	4.63
	1984	2.51	−6.11	8.62	1.04	0.04	0.27	. . .	7.26
	1995	1.91	−1.62	3.53	0.60	0.00	0.00	0.97	1.97
Medical care	1983	7.95	7.70	0.25	0.20	0.00	0.03	. . .	0.02
	1984	6.83	6.50	0.33	0.25	0.03	0.06	. . .	−0.02
	1995	3.73	3.70	0.03	−0.04	0.00	0.01	0.00	0.06
Entertainment	1983	3.73	2.62	1.11	0.40	0.30	0.23	. . .	0.18
	1984	4.28	2.09	2.19	0.68	0.82	0.47	. . .	0.22
	1995	3.44	2.87	0.57	0.27	0.07	0.04	0.08	0.10
Other goods and services	1983	7.95	7.14	0.81	0.58	0.06	0.12	. . .	0.05
	1984	6.07	4.51	1.56	0.92	0.13	0.03	. . .	0.47
	1995	4.18	3.69	0.49	0.40	0.02	0.01	0.01	0.04

Source: 1983 data are from Armknecht (1984); 1984 data are from Armknecht and Weyback (1989); and 1995 data are authors' calculations from unpublished CPI data.

a. Some items excluded; see table 4, note a for details.

Table 6. Percentage of Overall Pure Price Effect Due to Replacements, 1983, 1984, and 1995

Percent

Year	All items studied[a]	All items studied less apparel[a]
1983	61	20
1984	96	34
1995	50	31

Source: 1983 data are from Armknecht (1984); 1984 data are from Armknecht and Weyback (1989); and 1995 data are authors' calculations from unpublished CPI data.

a. Some items excluded; see table 4, note a for details.

vehicles categories. In both of these categories, retailers tend to discount the prices of items that are about to disappear. For apparel and upkeep, in 1995 and 1983–84 the negative pure price effect for unchanged items nearly cancels out the positive pure price effect for quotes with replacements. The magnitude of these differences, however, is much smaller in 1995 than in 1983–84. This reflects a decline in the use of the overlap method, which was often associated with large positive price effects in connection with returning sale prices to regular prices before making a quality adjustment. Likewise for new vehicles in 1995, the negative pure price effect for unchanged quotes was almost half of the positive pure price effect for replacement quotes.

Table 5 shows that the imputation of pure price change by the link method results in very small pure price effects. This is the consequence of imputing price changes from the average of price changes for all other items within the stratum. In strata where noncomparable replacement items were adjusted for quality change by the link method, the quotes that were not replaced apparently show relatively little average price change.

One noticeable difference between the results for the two periods is that in 1995 replacement items contributed only 50 percent of the pure price change for all items studied, whereas in 1984 replacement items contributed 96 percent. The source of the large difference between these ratios appears to be change in the methodology applied to replacements in apparel and upkeep. When price effects for apparel and upkeep are excluded from the study, the overall effect of replacement remains more stable across the periods studied. Table 6 shows the percentage of

overall pure price effect due to replacement items for all CPI items studied versus the percentage for all items studied less apparel.

The average price change that occurs with each replacement event is also of interest. Table 7 shows the mean period-to-period pure price change, weighted by the weights that were applied in calculating the aggregate contributions to price change. These are simple percentage changes that have not been converted to an annual rate, and this calculation mixes some quotes that are priced monthly with others that are priced bimonthly.[50] The average pure price change attributed to each replacement is 2.51 percent for comparable items, 0.34 percent for the link method, 5.17 percent for class-mean imputation, and 2.66 percent for direct quality adjustment. The overlap method, which is usually applied when the item is discounted, has a much larger average pure price increase: 28.0 percent.

Measures of Quality Change

As discussed above, there are a number of problems with aggregating our data on quality adjustments by item to arrive at an aggregate quality effect. Table 8 presents the results from six alternative methods of aggregation, together with the contributions to each aggregate from the different forms of quality adjustment applied to individual items. The first column of data, presenting the full-sample arithmetic aggregation, shows an aggregate quality effect of 1.76 percent during 1995.[51] This implies that if all item replacements that were adjusted for quality change by other methods had been treated as comparable, the inflation

50. It would be inappropriate to convert these changes to an annual rate because, for a given sample item, the replacement event typically occurs only once during the year.

51. An earlier draft of this study, which has now been widely cited, reported the difference as 2.56 percent (these results are available from the authors, upon request). In that draft of the paper we noted, however, that some of the measured "quality change" might include substitutions that were done for reasons other than ordinary quality changes, in particular, changes in units of measurement and other kinds of simple repackaging. In this version, we have dropped these kinds of substitutions as well as those involving data corrections from the measured quality change, to the extent that we have been able to screen them out. Certain issues remain as to whether all of the "quality" in our revised measure is appropriately designated as quality change. We discuss some of these issues below and produce several alternative measures of quality change.

Table 7. Mean Monthly or Bimonthly Pure Price Change, by Item Category and Replacement Method, 1995[a]

Percent

Item category	All quotes	Unchanged items	Replacement items				
			Comparable items	Overlap method	Link method	Class-mean imputation	Direct quality adjustment
All items studied	0.23	0.12	2.51	28.00	0.34	5.17	2.66
Food and beverages	0.20	0.19	0.38	28.39	0.42	0.00	1.05
Housing	0.20	0.14	1.68	13.30	0.32	3.89	−1.98
Apparel and upkeep	0.03	−1.03	7.87	68.32	−0.47	12.39	17.18
Apparel commodities	0.02	−1.17	8.33	68.33	−0.61	12.40	17.81
Transportation	0.13	0.01	1.05	27.29	0.00	3.96	1.90
New vehicles	0.26	−0.25	3.52	0.00	−0.07	4.48	3.28
Medical care	0.50	0.51	−0.19	0.04	0.19	0.00	1.46
Entertainment	0.40	0.35	1.93	53.46	1.33	2.52	2.42
Other goods and services	0.38	0.34	3.57	3.75	0.60	2.79	1.82

Source: Authors' calculations from unpublished CPI data.
a. For all quotes and for unchanged items, table reports mean of monthly and bimonthly pure price changes. For replacement items, table reports mean pure price change per replacement event.

rate for these items during 1995 would have been 3.92 percent instead of 2.16 percent.[52] Although this arithmetic aggregation measure is appropriate for this calculation, the alternative measures shown in the table, which exclude outliers or use logarithmic aggregation, give substantially different estimates that, under certain cirumstances, may be more appropriate indicators of the aggregate quality effect.

For the arithmetic aggregation, the quality effect for all replacement items drops from 1.76 percent without any truncation to 1.10 percent with truncation method A and to 0.54 percent with truncation method B.[53] The drop in quality effects is concentrated in replacements that use the link method and class-mean imputation; replacements using direct quality adjustment are hardly affected. Apparently, direct quality adjustments are not applied in cases that result in outlying quality adjustments.

As discussed above, to answer the question "What is the size of the typical quality adjustment made with CPI item replacements?", logarithmic aggregation may be appropriate.[54] As shown in table 8, without truncation the aggregate logarithmic quality effect is 0.44 percent. Under truncation method A the logarithmic aggregate quality effect is 0.40 percent. Under truncation method B, it drops to 0.28 percent. Each of the methods of quality adjustment, except the overlap method, contribute to the total effect under all six aggregation formulas, although the contributions of the link method and class-mean imputation are greatly reduced by truncation and the use of logarithms.

The results from the logarithmic version suggest that the quality change accounted for by BLS item replacement procedures are especially concentrated in the transportation and apparel and upkeep categories. Without truncation, the logarithmic quality effects are 1.35

52. The quality effect that we measure for 1995 is higher than the effects reported for 1983 (1.11 percent) and 1984 (1.23 percent). As noted above, the quality effects are sensitive to outliers, changes in units, and so forth. We have not been able to determine the extent to which Armknecht (1984) and Armknecht and Weyback (1989) control for these problems, so comparisons of measured quality effects between their studies and ours may be misleading.

53. For the arithmetic aggregation, the proportion of noncomparable replacements that are truncated is 2.2 percent under method A and 11.5 percent under method B. For logarithmic aggregation, the proportion is 1.3 percent under method A and 10.4 percent under method B.

54. Note that this question is distinct from that answered by arithmetic aggregation: "What would be the effect on CPI inflation of not applying quality adjustment methods?"

Table 8. Measured Quality Effects, by Item Category and Replacement Method, 1995[a]

Percentage points[b]

Item category	Replacement method	Additive quality effect[c]			Logarithmic quality effect[c]		
		No truncation	Method A	Method B	No truncation	Method A	Method B
All items studied	Overlap method	0.00	0.00	0.00	−0.01	−0.01	−0.01
	Link method	0.99	0.55	0.23	0.21	0.18	0.10
	Class-mean imputation	0.66	0.43	0.21	0.15	0.14	0.10
	Direct quality adjustment	0.11	0.12	0.10	0.08	0.09	0.08
	Total	1.76	1.10	0.54	0.44	0.40	0.28
Food and beverages	Overlap method	−0.01	−0.01	−0.01	−0.02	−0.02	−0.01
	Link method	1.40	0.62	0.35	0.14	0.10	0.08
	Class-mean imputation	. . .	. . .	. . .	. . .	. . .	. . .
	Direct quality adjustment	0.01	0.01	0.00	0.00	0.00	0.00
	Total	1.39	0.62	0.34	0.12	0.09	0.07
Housing	Overlap method	0.05	0.03	0.00	0.01	0.01	−0.01
	Link method	0.88	0.77	0.19	0.34	0.25	0.09
	Class-mean imputation	0.57	0.47	0.28	0.06	0.02	0.07
	Direct quality adjustment	0.06	0.07	0.00	0.01	0.02	−0.01
	Total	1.57	1.33	0.47	0.42	0.31	0.15
Apparel and upkeep	Overlap method	−0.02	0.00	0.00	−0.03	−0.03	0.00
	Link method	0.50	0.48	0.37	0.26	0.21	0.16
	Class-mean imputation	2.62	2.43	0.61	0.97	0.94	0.29
	Direct quality adjustment	0.27	0.34	0.29	0.15	0.15	0.15
	Total	3.37	3.25	1.27	1.35	1.26	0.61
Apparel commodities	Overlap method	−0.02	0.00	0.00	−0.03	−0.03	0.00
	Link method	0.59	0.56	0.44	0.33	0.27	0.21
	Class-mean imputation	2.90	2.69	0.68	1.08	1.04	0.33
	Direct quality adjustment	0.30	0.37	0.32	0.16	0.16	0.17
	Total	3.77	3.64	1.44	1.54	1.44	0.70

Transportation	Overlap method	0.00	0.00	0.00	0.00	0.00	0.00
	Link method	0.33	0.35	0.22	0.16	0.25	0.17
	Class-mean imputation	0.44	0.44	0.40	0.21	0.28	0.27
	Direct quality adjustment	0.36	0.36	0.36	0.35	0.35	0.36
	Total	1.13	1.16	0.98	0.72	0.89	0.79
New vehicles	Overlap method	..	..	..	..	..	..
	Link method	0.19	0.19	0.19	0.15	0.15	0.15
	Class-mean imputation	0.97	0.96	0.87	0.71	0.71	0.64
	Direct quality adjustment	0.56	0.56	0.56	0.54	0.54	0.54
	Total	1.71	1.71	1.61	1.40	1.40	1.33
Medical care	Overlap method	0.00	0.00	0.00	0.00	0.00	0.00
	Link method	2.53	0.52	0.14	0.44	0.11	0.05
	Class-mean imputation	..	..	..	..	..	..
	Direct quality adjustment	-0.02	-0.02	-0.02	-0.03	-0.03	-0.03
	Total	2.51	0.50	0.13	0.41	0.08	0.01
Entertainment	Overlap method	-0.03	-0.03	0.00	-0.04	-0.04	-0.04
	Link method	0.67	0.55	0.16	0.10	0.22	0.07
	Class-mean imputation	4.09	0.76	0.31	0.36	0.06	0.12
	Direct quality adjustment	0.06	0.06	0.06	0.05	0.05	0.05
	Total	4.79	1.34	0.52	0.47	0.30	0.20
Other goods and services	Overlap method	0.00	0.00	0.00	0.00	0.00	0.00
	Link method	0.62	0.38	0.09	-0.01	0.04	0.06
	Class-mean imputation	0.02	0.02	-0.02	-0.01	-0.01	-0.03
	Direct quality adjustment	-0.04	-0.04	-0.02	-0.06	-0.06	-0.03
	Total	0.60	0.36	0.05	-0.09	-0.04	0.00

Source: Authors' calculations from unpublished CPI data.
a. Missing values indicate that given method is not used for given item category.
b. Entries are expressed as percentage point contributions to quality effect in item category–specific CPI.
c. Method A truncates observations when quality increases or decreases by a factor of 5 or more; method B truncates when quality changes by a factor of 2 or more. See text for further details on different methods of aggregation and truncation.

Table 9. Mean Monthly or Bimonthly Quality Change, by Item Category and Replacement Method, 1995[a]

Percent

Item category	Replacement method	Mean additive quality change[b]			Mean logarithmic quality change[b]		
		No truncation	Method A	Method B	No truncation	Method A	Method B
All items studied	Overlap method	0.92	0.73	-0.75	-1.67	-1.37	-1.65
	Link method	30.39	11.58	5.45	4.27	3.68	2.36
	Class-mean imputation	18.62	12.44	6.79	4.19	3.94	3.16
	Direct quality adjustment	1.59	1.68	1.44	1.19	1.22	1.16
	Total noncomparable replacements	14.26	7.07	3.68	2.74	2.51	1.85
Food and beverages	Overlap method	-8.77	-3.75	-4.17	-10.58	-8.09	-5.38
	Link method	28.65	6.17	3.71	1.35	0.98	0.80
	Class-mean imputation	...	...	...	...	...	...
	Direct quality adjustment	0.77	0.77	0.19	0.39	0.39	0.08
	Total noncomparable replacements	26.16	5.65	3.33	1.08	0.77	0.64
Housing	Overlap method	12.34	6.99	-1.36	2.96	2.96	-1.56
	Link method	36.99	21.82	6.15	9.24	7.16	2.95
	Class-mean imputation	12.31	10.34	7.28	1.20	0.46	1.86
	Direct quality adjustment	5.05	5.49	0.41	0.86	1.68	-0.94
	Total noncomparable replacements	20.38	13.78	5.52	4.16	3.14	1.70
Apparel and upkeep	Overlap method	-1.15	0.34	0.24	-1.57	-1.57	-0.24
	Link method	17.55	16.88	16.03	9.49	7.63	7.47
	Class-mean imputation	36.46	36.92	13.43	13.03	13.04	6.04
	Direct quality adjustment	4.84	6.15	5.35	2.52	2.52	2.69
	Total noncomparable replacements	19.96	20.07	9.42	7.59	7.23	4.18
Apparel commodities	Overlap method	-1.15	0.34	0.24	-1.57	-1.57	-0.24
	Link method	20.19	19.47	18.89	11.70	9.70	9.70
	Class-mean imputation	36.45	36.92	13.42	13.03	13.03	6.04
	Direct quality adjustment	5.01	6.39	5.55	2.61	2.61	2.79
	Total noncomparable replacements	20.64	20.79	9.90	7.97	7.61	4.50

Category	Method						
Transportation	Overlap method	. . .	. . .	. . .	. . .	. . .	. . .
	Link method	15.36	14.16	9.17	6.05	9.82	6.83
	Class-mean imputation	5.34	5.37	4.97	2.47	3.36	3.24
	Direct quality adjustment	1.46	1.46	1.46	1.43	1.43	1.43
	Total noncomparable replacements	3.35	3.25	2.78	2.00	2.48	2.22
New vehicles	Overlap method	. . .	. . .	. . .	. . .	. . .	. . .
	Link method	4.39	4.39	4.39	3.41	3.41	3.41
	Class-mean imputation	4.47	4.45	4.06	3.23	3.23	2.96
	Direct quality adjustment	0.94	0.94	0.94	0.90	0.90	0.90
	Total noncomparable replacements	2.00	1.99	1.89	1.62	1.62	1.54
Medical care	Overlap method	. . .	. . .	. . .	. . .	. . .	. . .
	Link method	61.82	13.46	4.24	10.57	2.76	1.37
	Class-mean imputation	. . .	. . .	. . .	. . .	. . .	. . .
	Direct quality adjustment	−0.44	−0.44	−0.44	−0.86	−0.86	−0.86
	Total noncomparable replacements	28.31	5.82	1.54	4.57	0.87	0.16
Entertainment	Overlap method	. . .	. . .	. . .	. . .	. . .	. . .
	Link method	19.94	16.39	5.63	3.10	6.62	2.58
	Class-mean imputation	124.76	24.19	12.59	10.28	1.88	4.64
	Direct quality adjustment	1.45	1.45	1.45	1.24	1.24	1.24
	Total noncomparable replacements	43.83	12.43	5.55	4.28	2.72	2.08
Other goods and services	Overlap method	. . .	. . .	. . .	. . .	. . .	. . .
	Link method	27.25	17.31	6.07	−0.58	1.86	4.15
	Class-mean imputation	4.70	4.70	−6.52	−3.81	−3.81	−9.08
	Direct quality adjustment	−1.84	−1.84	−1.00	−3.06	−3.06	−1.65
	Total noncomparable replacements	11.48	6.97	1.08	−1.75	−0.75	−0.06

Source: Authors' calculations from unpublished CPI data.

a. Table reports mean quality change per replacement event. Missing values indicate fewer than twenty replacement events.

b. Method A truncates observations when quality increases or decreases by a factor of 5 or more; method B truncates when quality changes by a factor of 2 or more. See text for further details on different methods of aggregation and truncation.

percent for apparel and upkeep and 0.72 percent for transportation, as compared to 0.12 percent for food and beverages, 0.42 percent for housing, 0.41 percent for medical care, 0.47 percent for entertainment, and −0.09 percent for other goods and services. Under truncation method B, the quality effects drop to 0.61 percent for apparel and upkeep, 0.15 percent for housing, 0.01 percent for medical care, 0.20 percent for entertainment, and 0.00 percent for other goods and services.

The mean quality change is presented in table 9. The typical quality adjustment is larger for the link method and class-mean imputation than for direct quality adjustment, and is negative for the overlap method. Again, the largest average quality adjustments tend to occur in the apparel and upkeep category.

Conclusions

The advisory commission and others who have analyzed the measurement of quality change and new products in the CPI have identified a number of areas in which the methods used by the BLS still require attention. The BLS has made a series of modifications to take better account of quality change, but further improvements clearly are needed in some components of the index. For certain important categories of items considered by the advisory commission, it would be difficult to argue that the CPI does not overstate the rate of price change. In other cases, however, any bias seems likely to be considerably smaller than the advisory commission has estimated and, in certain cases, it could even be negative.

The results of our decomposition of price differences between old and new items into quality and price changes are somewhat ambiguous. The arithmetic method of calculating these changes demonstrates that the quality adjustment methods have a profound effect on measured price change; they cause the measured price change of the CPI to be reduced by 1.76 percentage points for the items studied. The logarithmic method indicates a range of 0.28 to 0.44 percent for quality change, which we consider to be a significant amount of quality adjustment.

Our measurements of quality effects do not provide direct evidence of quality bias in a particular direction. They do show that any quality

bias could go in either direction, either through inadequate quality adjustment (as emphasized by the advisory commission) or through excessive quality adjustment by the application of the link method to items with rising prices. Avoiding both downward and upward bias should be of concern to the Bureau of Labor Statistics.

Comments and Discussion

Robert J. Gordon: In some past eras, there was a military atmosphere to debates between the Bureau of Labor Statistics and those academics who were doing research on price measurement. The BLS was harrassed and besieged, as if in a medieval fortress, throwing hot oil from the parapets at the academics who were attacking with bows and arrows, battering rams, and other tools of their trade.

The dialogue over the past two years between the CPI commission (of which I was a member) and the BLS has been very different. The BLS generously hosted several meetings at its offices, in which its procedures were explained to us at a fine level of detail, and we subsequently met a number of times off-site, as our findings and recommendations began to come together. BLS officials were open in their recognition that there are many problems in the CPI; they worked with us to identify the nature of the problems and potential solutions and they have generally welcomed the attention that the commission and its final report has directed to the need for more investment in the quality of government statistics, both the CPI and more broadly.

Our report acknowledges that some of the best research on the issues that we address has been produced within the BLS, and indeed, the whole issue of formula bias—now relabeled lower-level substitution bias—has come to light because of pathbreaking research at the BLS by Marshall Reinsdorf, by himself initially and more recently with Brent Moulton. Our report has a long reference list, and it is striking how many of the citations are to studies by current or former employees of the BLS.

350

In turn, this paper by Moulton and Karin Moses is generous to our report, calling it "the most influential critique of the CPI in decades." Indeed, comparisons are inevitable between ours and the famous Stigler commission report of 1961. Some contrasts between our commission and the Stigler commission provide useful background for those evaluating our report and the subsequent criticisms of it, in this paper and elsewhere. The reports have one thing in common: Zvi Griliches was involved in both. They have several important differences. Our report concerns only the CPI, while the Stigler report covered the PPI and agricultural price indexes as well. The Stigler commission did not produce any numerical estimate for bias in any of the price indexes reviewed, whereas our mandate included providing a point estimate of the overall bias in the CPI. A final and important difference, in the context of the present discussion, is that the Stigler commission had a substantial budget to commission new research studies, including Griliches's famous paper on hedonic price indexes for automobiles, whereas ours had no budget at all (except for travel expenses).[1] We did our best to assemble all the relevant existing research by academics, BLS insiders, and others, but we did not have any research support for conducting new studies.

The CPI commission's report arrives at an overall bias estimate of 1.1 percent, consisting of two main parts. Almost half (0.5 percent) involves upper- and lower-level substitution bias and outlet substitution bias. This estimate seems to be uncontroversial and to be widely accepted; indeed, it may be too conservative, for reasons mentioned below. The other portion (0.6 percent) is our estimate of the bias due to inadequate adjustment for quality change and inadequate allowance for the introduction of new products—we did not separate the quality change and new product sources of bias, but treated them together.

As Moulton and Moses emphasize, our approach to the problem of quality change differs from past evaluations. Previously, it had been common to take available research on a limited number of CPI commodities, multiply the bias figures for those commodities by their weight in the CPI, and take the product as the estimate of the overall quality change bias in the CPI, assuming that all categories for which research was not available had a bias of exactly zero. We rejected that

1. Griliches (1961).

approach: "The evaluation that the rest of the CPI is unbiased represents an extreme one-sided answer to the question as to whether the components of the CPI subject to relatively little research are biased. They may be as likely to be subject to the average rate of bias of those components which have been subject to careful research as to no bias at all." Further, "because the magnitude of quality change bias differs so much across product categories, any overall evaluation of the magnitude of quality change bias must be conducted 'down in the trenches,' taking individual categories of consumer expenditure, assessing quality change bias for each category, and then aggregating using appropriate weights."[2]

Moulton and Moses do not object to our "down in the trenches" approach to the problem. Indeed, they state that "this is the first systematic analysis, category by category, of quality bias in the CPI, and it is a noteworthy accomplishment." However, as indicated, the analysis was conducted without research staff or a research budget. Inevitably, there is a back-of-the-envelope quality to some of the estimates. But Moulton and Moses recognize that "the commission's approach to the problem of producing an overall assessment of bias seems sensible, and this type of structure will likely prove to be useful in the future." They concur with us that the quality change problem is usefully viewed as a matrix—our report plugged numbers into many slots of the matrix, and some of those numbers are better than others. As some of the weaker numbers come under scrutiny, more solid numbers will emerge, and the matrix will be continuously updated. Some of Moulton and Moses's findings suggest that some of our bias estimates may be too high, and I suggest, for other reasons, that some are too low.

Before discussing the substance of this paper, I want to emphasize some broader issues of relevance to the debate over bias in the CPI. To start with, there is the question of whether estimates of quality change bias are inevitably too "subjective" and "judgmental" to be taken seriously. One response is that quoted above from our report: it is just as subjective to assume that every CPI category not subject to careful research has a zero bias as to extrapolate research-based estimates from one category to another. Another response is that it is better to be imprecisely right than precisely wrong. Even though one can never

2. U.S. Senate, Committee on Finance (1996, p. 22).

precisely measure the value of the invention of the video-cassette re-
corder or the jet airplane, one can use the economist's toolbox to esti-
mate the size of consumer surplus triangles. We know that the size of
these triangles is positive, or, in the language of this paper, the bias
due to the failure to value the invention of new products inherently
must be positive.

In addition, there are three reasons to suspect that the commission's
overall estimate of a 1.1 percent CPI bias is too low for the period
1995–96 and one reason to suspect that at some point in the past the
bias may have been much smaller, or even negative.

The first reason that the bias may be too low is that we did not
quantify any of the numerous intangible aspects of quality change, such
as the improved safety and lighter weight of home power tools and the
improved quality of stereo sound and television pictures, among many
others. Second, with the exception of food and beverages, motor fuel,
and personal banking, we did not attempt to make any estimate of the
consumers surplus triangles created by the invention of new goods and
services.

The third reason concerns the growing divergence between the infla-
tion rate as measured by the CPI and by the chain-weighted personal
consumption expenditures (PCE) deflator. In the four quarters ending
in 1995:4, and again in 1996:4, these measures differed by 0.7 per-
centage point, with the CPI growing faster. In an unpublished study
conducted by the BLS and Bureau of Economic Analysis, fully 0.35
percentage point of the difference over the period from mid-1994 to
mid-1996 was attributable to weighting differences; the PCE deflator
uses much more current weights and also places much higher weights
on medical care and personal computers than does the CPI. This number
is much higher than our estimate of 0.15 for upper-level substitution
bias, an important reason why our overall bias estimate may be too
low. Much of the remaining difference between the CPI and PCE de-
flator arises because the latter uses the PPI rather than CPI for medical
care (the PPI apparently goes part but not all of the way to an outcome-
based, rather than an input-based, approach to pricing medical care)
and because it registers a much faster decline in the price of personal
computers.

If one were to accept the PCE deflator treatment of upper-level
weighting, medical care prices, and computer prices, one would start

with a 0.7 percent upward bias in the CPI over the two years 1995–96. But the PCE deflator still incorporates the lower-level substitution bias, the outlet substitution bias, and at least 0.4 percent of our 0.6 percent estimate of quality change bias. The sum of 0.7, 0.4, and 0.4 provides an alternative estimate of an upward bias of 1.5 percent in the CPI.

Could the bias in the CPI have been smaller in the past, and perhaps in the opposite direction? This possibility emerges from the "Nordhaus paradox." If one takes the bias in the CPI as 1 or 1.5 percent and extrapolates this rate back two centuries, the implied level of per capita income in today's prices is so low that it would not be sufficient to keep a person alive on a diet consisting solely of potatoes, without leaving anything over for clothing and shelter. At some point in the past, the bias must have been lower. One will never detect bias, or lack of it, for data from the nineteenth and early twentieth centuries, because it is based on academic research by pioneers such as Dorothy Brady, Ethel Hoover, and Albert Rees, who essentially used all the information available. However, as the commission reports, the back-of-the-envelope technique that I used to quantify the shelter bias in the CPI from 1976 to 1993 suggests a severe downward bias for shelter between 1920 and 1976, and this may have been important enough to create a downward bias in the overall CPI for much of that period.[3]

Turning to the present paper, it is really two papers. The first part provides a critique of some of the numbers in the commission's quality change matrix. The second part provides valuable information quantifying the frequency and extent of quality change adjustments in the current CPI. Let me discuss the second part first.

For most categories in the CPI, the extent of current quality adjustments is irrelevant to an assessment of the treatment of quality change in the commission's report, simply because most of our estimates of quality change bias are valid independent of how the BLS arrives at its estimates of price change or the extent to which its adjustments for quality change are large or small. Most of our bias estimates are based on the collection of price data from independent sources and the careful quality adjustment of those data. The difference between these quality-adjusted independent price indexes and the corresponding CPI indexes (however they are adjusted for quality change) forms the basis of our

---

3. U.S. Senate, Committee on Finance (1996, p. 31).

estimates of bias. We use such independent sources of price data in our bias estimates for shelter, appliances, radio-television, personal computers, apparel, public transportation, prescription drugs, and medical care. Estimates derived from these categories are extrapolated, sometimes partially rather than fully, to other housefurnishings, nonprescription drugs, entertainment commodities, and personal care.

Only in the few remaining cases do we add a bias estimate to a CPI category in which there are already quality adjustments, rather than computing the bias estimate indirectly by subtracting an independent estimate from the CPI estimate for that category. These categories are food and beverages, other utilities, new and used cars, motor fuel, and personal expenses.

Moulton and Moses previously estimated that the BLS made adjustments that reduced the rate of inflation in the CPI by about 2.6 percent in 1995. In the present revised version of their paper, this number has shrunk to 1.76 percent. Using logarithmic aggregation and excluding ''outliers''—commodity pairs for which the implicit price-quality change is two-fold or greater (which are simply noncomparable pairs and are not likely to reflect what we had in mind by the concept of quality change)—the number shrinks to a mere 0.3 percent. At this point, the argument loses any quantitative significance.

But it is still instructive to discuss this argument, since it illustrates the substantive difficulties and the problems of communication in this field. Most of the reported ''quality adjustment'' by the BLS (1.65 percent out of the 1.76 percent that includes outliers) comes from ''linking'' procedures or class-mean imputation, in which a missing item is replaced by another. No judgment at all is made about the quality differential between the new and old items. The price change during the link period is imputed, by using either the inflation rate in the overall CPI or that of other commodities in the particular class. These adjustments are the consequence of the BLS sampling procedures, which focus on pricing a very specific item in a particular store and city. There are thousands upon thousands of such commodities in the market, but only a small fraction of them is in a particular store at any time. The pricing agent has to deal with rapid turnover and high probability of stockout. Roughly one out of three items disappear sometime during the year and have to be replaced by a different item in the same general class: a larger versus a smaller package of yogurt, a blue raincoat versus

a black one, a refrigerator with a freezer at the bottom rather than at the top. But this churning is not what we had in mind by "quality change," which rather involves the appearance of new and improved goods, such as the increased variety and freshness of vegetables and fish due to improved transport facilities and the globalization of trade, or the substitution of laporascopic procedures for gallstone operations, and so on.

This brings me to the first part of Moulton and Moses's paper. Much of this is solid, and I accept some of it. To quantify what is at stake, recall that the commission's estimate of quality change and new product bias is 0.61 percent annually for the period 1995–96. Of this, 0.31 percent comes from appliances, radio-television, personal computers, drugs, and medical care, all of which Moulton and Moses accept at face value. Their analysis of fruits, vegetables, shelter, and motor fuel is very helpful. But these categories together only account for 0.08 point of our 0.61 percent.

On apparel, Moulton and Moses do not actually come up with a quantitative reason to doubt the commission's estimate of a 1.0 percent bias. This estimate is based on the difference between my Sears catalogue estimate of apparel prices for the four major subdivisions of apparel and the corresponding four categories of the CPI. That difference was 1.92 percent, and we cut it down to 1.0 percent just to be conservative. But recall that during the last decade of the Sears catalogue, the company was losing market share to new competitors like Wal-Mart, so the Sears apparel index probably overstates the rate of inflation actually experienced by consumers who are free to choose where to shop.

On motor vehicles, our bias estimate is based solely on the increase in durability. I agree that some unknown part of the increase in durability has already been taken into account in the CPI. But going in the other direction is an enormous improvement in the quality of new vehicles, as measured by the J. D. Power survey of initial defects, as well as by the *Consumer Reports* questionnaire, which registers a decline in defects by a factor of 3 over the past two decades. Decreased frequency of repairs is not adequately taken into account in the CPI, so the paper's discussion does not convince me that our bias estimate for motor vehicles is too high.

Moreover, while Moulton and Moses's helpful discussion of certain

categories does reduce our overall quality change bias estimate slightly, not all new research goes in this direction. For the television category, the commission took my estimate of a 3.3 percent bias over the period 1973–83, the final decade in my earlier study.[4] I have recently updated that work with a new study of the prices of three different sizes of television sets over the period 1984–97, using annual evaluations in *Consumer Reports* that allow one to hold constant not only picture size, but also such features as picture-in-picture, stereo capability, and the number of input-output channels, among others. The annual rate of price decline for television sets in the CPI over these years is 3.0 percent and in my index is 11.1 percent, for an upward bias in the CPI of 8.1 percent, in contrast to the 3.3 percent that we assumed. Yet even this much larger rate of bias is understated. Some features have improved in ways that could not be taken into account, such as picture and sound quality and the number of cable channels that could be received. Perhaps more important, my study fails to take into account a significant increase in reliability: the median percentage of sets repaired in their first five years decreased from 19 percent in 1986 to 6 percent in 1997.

To conclude, Moulton and Moses have contributed a very useful discussion suggesting that several of the commission's estimates of bias are too high. I contribute four reasons why they are too low: failure to include intangible quality improvements, failure to include the consumer surplus contribution of new products for most categories, the divergence between the CPI and the personal consumption expenditures deflator, and new research on television sets that suggests the possibility of similarly large biases in other products involving electronics.

The road to solid estimates of quality change bias is long and arduous. I am pleased that the journey down that road has become such a fruitful partnership between academic research and research within the government agencies producing numbers that are crucial ingredients in virtually every measure of American economic performance and well-being.

Barry P. Bosworth: This an interesting and obviously timely paper. It has two major parts; the first responds to portions of the CPI advisory commission's report that deal with the issue of quality change; and the

4. See Gordon (1990).

second provides a new empirical analysis of the frequency and extent of product substitutions in the CPI for the years 1983, 1984, and 1995.

Very little is said about the portions of the commision's report that focus on the issue of substitution bias. I think that is because the authors basically agree with the commission's suggestion that the methodology of the CPI should be changed from the current Laspeyres index to some superlative index that would yield a closer measure of the cost of living—in effect, following the recent methodological changes to the national accounts. The only real question is the speed with which the change will be implemented.

The role of quality change is far more controversial, however. The commission alleges that current BLS procedures fail adequately to capture improvements in quality, and that the result is an overestimate of annual inflation of about 0.6 percentage point. Moulton and Moses challenge the notion of a pervasive bias in the treatment of quality change. The first portion of their paper provides a useful response to the report—the commission's analysis itself seems somewhat biased in ignoring the potential for overstating the quality change. In combination, the two statements provide a more comprehensive discussion of a difficult issue, but they are largely confined to a war of anecdotes. I am skeptical of alternative indexes based on catalogues or *Consumer Reports*, when there is no way of determining the relevant quantity weights. The amount of research that actually bears on the issue is very limited.

Thus the really significant contribution comes in the second part of the paper, where the authors present some data on the decomposition of the total price change for items in the CPI sample into a quality change component and a pure price change component. While some of the public debate presents the issue as an argument over whether there are or are not quality improvements, the question is, more accurately, whether quality change is greater or less than the BLS estimates. Since it has never really been known how much quality change is captured in the BLS procedures, it difficult to determine if it is too much or too little. A study of the actual BLS adjustments is a crucial starting point.

It is evident that this paper is very much a study in progress, and to draw any conclusions from the analysis at this point may be hazardous. First, the data are derived from situations in which it was necessary to substitute for one of the items in the sample. These substitutions occur

for a wide variety of reasons other than a change in the quality char-
acteristics of the surveyed item. The authors have tried to exclude those
substitutions that result from simple changes in the units in which an
item is sold, but there remain a large number of substitutions that do
not involve the introduction of a new model. However, one could
possibly argue that differences in price levels for those substitutions
that result from the simple disappearance of an item might be randomly
distributed. In the event of a substitution, the data collector is instructed
to search for a close substitute, but there may be no particular tendency
to chose an item that is more or less costly. Thus the mean change
would be dominated by situations involving the introduction of a new
model.

The paper classifies substitutions according to the different tech-
niques that the BLS uses to factor out the quality change, ranging from
comparable substitutions (when there is no difference in quality be-
tween old and new items) to direct adjustment for the quality compo-
nent. The major issues arise with respect to the link method. Consider
two extreme views of the process by which firms introduce changes in
price and quality. On the one hand, if quality changes are typically
introduced in a continuous process that is uncorrelated with price
changes, it seems reasonable to believe that the BLS procedures may
miss some part of the quality change: small differences would be over-
looked. On the other hand, if firms tend to save up price and quality
changes and incorporate them in a new model, the timing of price and
quality changes are correlated and the BLS link procedures would tend
to eliminate nearly all of the price difference between the old and new
items and overstate the amount of quality change. Concern about just
such a practice was a primary reason why the BLS adopted direct
adjustment for the automotive market; and Moulton and Moses provide
a nice example of the same phenomenon for women's clothing. The
use of the link method is probably responsible for some of the CPI's
failure to capture price declines fully in areas of rapid innovation, such
as consumer electronics.

This paper finds that substitutions occur very frequently: if one as-
sumes that all items are priced monthly, the typical item will require a
substitution within a two-year period. However, in 1995 two-thirds of
the substitutions were deemed comparable and hence were not adjusted
for quality change. This is a sharp increase from 1983–84 and supports

the commission's view that small changes in quality are overlooked. One should also worry that in two-thirds of those cases with a quality adjustment, the adjustment was derived from the link method or its close equivalent, class-mean imputation.

The data arrayed in table 5 also highlight the important role of subjective judgements in the classification of substitutions as comparable or noncomparable. If the data collector cannot find the item and reports the price of a substitute item, the analyst has to decide if it is comparable or noncomparable. The distinction is critical. If the substitute item is ruled to be noncomparable, nearly all of the price change is linked out. If it is ruled comparable, there is a relatively large contribution to the overall estimate of inflation. Yet if the items are so comparable, why are the price increases so large?

Moreover, the large category of items with no substitutions (96 percent of the total) represents a surprisingly small part of the overall price increase: only about one-half. If the BLS counts improvements as price increases, why does the category of nonsubstitutions not account for a larger portion?

At first glance, the data indicate that current BLS procedures account for a surprisingly large amount of quality improvement: 1.76 percent in 1995, according to table 8. The authors go on, however, to raise some serious questions about their estimate of the mean quality change; in particular, they show that it is very sensitive to extreme values. As shown in table 8, the mild truncation of method A—a loss of 2 to 3 precent of the sample—reduces the mean estimate of the quality change component from 1.76 to 1.10. The tighter distribution of method B— representing an 11 percent loss—reduces it to 0.54. Why do these restrictions have the greatest effect in the category of food and beverages, where it should be possible to find very close substitutes? One might expect a large role for extreme values in the cases of medical care and entertainment, because of the heterogeneous nature of those categories. The logarithmic aggregation probably provides a better adjustment for extreme values in the ratio of new to old prices: it reduces the estimated mean quality change to 0.4 percent.

If the data from item substitutions are to provide useful insights into the issue of quality change in the CPI, more work needs to be done to understand the nature of the substitutions and the reasons for some of the large changes. I would hope that the BLS would expand on this

type of analysis and make it a regular part of an annual review of the CPI. One needs to know more about the factors that influence the decision to classify an item as comparable or noncomparable; and some further classification is needed for those substititutions for which the link method is used to adjust for quality change.

General discussion: William Nordhaus raised an issue about the purpose of indexing and how that relates to the construction of the CPI. He noted that improved life expectancy was a quality improvement discussed by the commission's report in connection with pricing medical costs. Valuing this improvement with numbers conventionally used by environmentalists for the price of life would attribute a substanatial increase in living standards over the past thirty years to health care (market and nonmarket). While it may be appropriate to take that into account in indexing the tax system, he questioned whether the annuity benefits of social security recipients indexed to the CPI should be reduced simply because they are living longer. Katharine Abraham expanded on Nordhaus's point, noting similar quandaries with medical procedures that deliver better outcomes, even without increased longevity. Should the government be saying to social security recipients that they will not be compensated for the full price of operations just because those operations are now better? She believed that such a policy might be difficult for the average citizen to accept. Some other suggested adjustments, such as for the utility coming from greater variety of products available to consumers raise similar questions. How can it be explained to voters that the price level has really dropped because of increased variety, even if no price has fallen?

Abraham further observed that while substitution bias and outlet substitution bias are important problems, there is room for disagreement about how much to change present procedures. Although switching to geometric mean aggregation makes sense in many categories where price changes lead to item substitution, she argued that it should not become universal. In prescription drugs, for instance, although drugs for treating ulcers have declined in price relative to those for treating heart conditions, one should not expect significant substitution between them, and geometric mean aggregation would not be appropriate. She was skeptical about outlet bias, which is difficult to measure because buying at different stores involves different services and amenities. She

added that the bias arising from the emergence of new outlets could
even go in the opposite direction than usually assumed: consider a case
where discount stores move into an area, people shift away from tra-
ditional stores, and, in order to compete, the traditional stores cut back
on staffing and other services. The CPI could then pick up the fall in
prices due to the entry of discount stores, without picking up any of the
deterioration in the quality of service at the traditional stores. Gordon
replied that much of the actual outlet substitution has not been from
high-service to low-service outlets, but from inefficient self-service
stores, such as Sears, to more efficient ones, such as Wal-Mart.

James Duesenberry noted the great difference between the historical
BLS approach, which focused on pricing a list of items over time, and
the commission's approach, which looks for ways to measure how
people are better off as a consequence of new products and quality
changes. The ambitious approach taken by the commission calls atten-
tion to the problems of using one index for a diverse population, because
it highlights the fact that whether people are made better off depends
on which group is being referenced. From the perspective of upper-
middle income people like those on the commission, the prices of
computers and televisions that can get 150 channels are important. But
people relying on social security, who tend not to own computers or
get cable television, are more interested in items like rent and food.

Robert Hall questioned Duesenberry's assertion that television sets
were relatively unimportant for those lower in the income distribution
and was convinced that Gordon's television data identified a large CPI
bias in this category. Television sets are constantly being upgraded, in
which case the replacement is noncomparable, and so almost all the
price decline has been taken out by BLS methodology. Moulton replied
that a hedonic study of televisions and some other items is underway.
While he would quibble over some points in Gordon's analysis, he
agreed with the direction of the bias for television sets. But he noted
that since televisions account for only 0.3 to 0.4 percent of the consumer
market basket, their relative importance is small.

Several participants offered other specific instances of bias in the
CPI data. Hall explained that the CPI prices long distance telephone
calls by tabulating the regular tariff rates that the long distance com-
panies report, even though many customers are switching to lower-
priced plans. The problem here is not with quality changes, but rather,

is just the failure to check what prices people actually pay. He calculated that according to the CPI, everyone is paying about 30 cents a minute, whereas the aggressive consumer nowadays pays only 10 cents a minute and the average customer pays 14 cents a minute. Although automobiles have undoubtedly become more rust-resistant and more reliable, Jack Triplett expressed reservations about the assertions by the CPI commission and Gordon that these quality improvements are missed by the CPI. Automobile manufacturers submit costs of quality changes to the BLS, and from Triplett's work at the BLS, he reported that they seldom overlooked quality changes. In fact, manufacturers tried to attribute too much price change to quality improvements—he recalled arguing with one car company about whether taking the ''90'' and ''100'' off of the speedometer was an improvement—leading to a downward bias in the CPI. He reasoned that whether the CPI, on balance, is biased upward or downward can only be determined by looking at hundreds of procedural details. Finally, reflecting on these examples of specific items in the index, Nordhaus reasoned that real progress could not be made via ''war by anecdotes.'' As an alternative approach, he recommended that the BLS and outside experts draw a probability sample of detailed CPI components and, on the basis of in-depth study, develop an estimate of the bias (positive or negative) in each of those components. Then, using the weights that guided the initial sample selection, it would be a straightforward matter to estimate the bias in the overall CPI, as well as a confidence interval for same. David Wilcox agreed that Nordhaus's suggestion was a good one, but urged that it be taken as a model for how the BLS should be organized permanently: the BLS should review all of the components of the CPI on a rotating basis (say, once every ten years). Achieving this objective might require both administrative reorganization and the commitment of substantial additional resources, but it would pay significant dividends in terms of improvement in the index.

References

Abraham, Katharine G. 1997a. "The CPI Commission: Discussion." *American Economic Review, Papers and Proceedings* 87(2): 94–95.

———. 1997b. "Testimony before the Senate Finance Committee." 105 Cong., 1 sess., February 11.

Armknecht, Paul A. 1984. "Quality Adjustment in the CPI and Methods to Improve It." In *American Statistical Association, Proceedings of the Business and Economic Statistics Section*.

Armknecht, Paul A., Walter F. Lane, and Kenneth J. Stewart. 1997. "New Products and the U.S. CPI." In *The Economics of New Goods*, edited by Timothy F. Bresnahan and Robert J. Gordon. University of Chicago Press.

Armknecht, Paul A., and Donald Weyback. 1989. "Adjustments for Quality Change in the U.S. Consumer Price Index." *Journal of Official Statistics* 5(2): 107–23.

Baker, Dean. 1996. "The Overstated CPI—Can It Really Be True?" *Challenge* (September–October): 26–33.

———. 1997. "Does the CPI Overstate Inflation? An Analysis of the Boskin Commission's Report." Unpublished paper. Washington: Economic Policy Institute (March).

Berndt, Ernst R., Zvi Griliches, and Neal J. Rappaport. 1995. "Econometric Estimates of Price Indexes for Personal Computers in the 1990's." *Journal of Econometrics* 68(1): 243–68.

Berndt, Ernst R., Zvi Griliches, and Joshua G. Rosett. 1993. "Auditing the Producer Price Index: Micro Evidence from Prescription Pharmaceutical Preparations." *Journal of Business and Economic Statistics* 11(3): 251–64.

Bosworth, Barry P. 1997. "Testimony before the Senate Finance Committee." 105 Cong., 1 sess., February 11.

Cole, Roseanne, and others. 1986. "Quality-Adjusted Price Indexes for Computer Processors and Selected Peripheral Equipment." *Survey of Current Business* 66(1): 41–50.

Congressional Budget Office. 1994. "Is the Growth of the CPI a Biased Measure of Changes in the Cost of Living?" CBO Paper. Congressional Budget Office (October).

Cutler, David M., and others. 1996. "Are Medical Prices Declining?" Working Paper 5750. Cambridge, Mass.: National Bureau of Economic Research (September).

Diewert, W. Erwin. 1983. "The Treatment of Seasonality in a Cost of Living Index." In *Price Level Measurement: Proceedings from a Conference Sponsored by Statistics Canada*, edited by W. E. Diewert and C. Montmarquette. Ottawa: Statistics Canada.

———. 1996. "Comment on CPI Biases." *Business Economics* 31(2): 30–35.

———. 1997. "The CPI Commission: Discussion." *American Economic Review, Papers and Proceedings* 87(2): 95–96.

Gordon, Robert J. 1990. *The Measurement of Durable Goods Prices.* University of Chicago Press.

———. 1996. "The Sears Catalog Revisited: Apparel and Durable Goods." Unpublished paper. Northwestern University (April).

Griliches, Zvi. 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.

———. 1971. "Introduction: Hedonic Price Indexes Revisited." In *Price Indexes and Quality Change: Studies in New Methods of Measurement,* edited by Zvi Griliches. Harvard University Press.

Griliches, Zvi, and Iain Cockburn. 1994. "Generics and New Goods in Pharmaceutical Price Indexes." *American Economic Review* 84(5): 1213–32.

Hausman, Jerry. 1997. "Cellular Telephone, New Products, and the CPI." Unpublished paper. Massachuetts Institute of Technology (February).

Hicks, J. R. 1940. "The Valuation of the Social Income." *Economica* 7(May): 105–24.

Hulten, Charles R. 1997. "Quality Change in the CPI." Federal Reserve Bank of St. Louis *Review* (forthcoming).

Klumpner, Jim. 1996. "Fact and Fancy: CPI Biases and the Federal Budget." *Business Economics* 31(2): 22–29.

Lebow, David E., John M. Roberts, and David J. Stockton. 1994. "Monetary Policy and 'The Price Level.'" Unpublished paper. Board of Governors of the Federal Reserve System (July).

Liegey, Paul R. Jr. 1993. "Adjusting Apparel Indexes in the Consumer Price Index for Quality Differences." In *Price Measurements and Their Uses,* edited by Murray F. Foss, Marilyn E. Manser, and Allan H. Young. University of Chicago Press.

———. 1994. "Apparel Price Indexes: Effects of Hedonic Adjustment." *Monthly Labor Review* 117(5): 38–45.

Moulton, Brent R. 1996. "Bias in the Consumer Price Index: What Is the Evidence?" *Journal of Economic Perspectives* 10(4): 159–77.

———. 1997. "Issues in Measuring Price Changes for Rent of Shelter." Paper prepared for Conference on Service Sector Productivity and the Productivity Paradox. U.S. Bureau of Labor Statistics, April 11.

Moulton, Brent R., and Karin E. Smedley. 1995. "A Comparison of Estimators for Elementary Aggregates of the CPI." Unpublished paper. U.S. Bureau of Labor Statistics (June).

Reinsdorf, Marshall B., Paul Liegey, and Kenneth Stewart. 1996. "New Ways of Handling Quality Change in the U.S. Consumer Price Index." Working Paper 276. U.S. Bureau of Labor Statistics (November).

Rothbarth, E. 1941. "The Measurement of Changes in Real Income under Conditions of Rationing." *Review of Economic Studies* 8: 100–07.

Shapiro, Matthew D., and David W. Wilcox. 1996. "Mismeasurement in the Consumer Price Index: An Evaluation." In *NBER Macroeconomics Annual 1996*, edited by Ben S. Bernanke and Julio J. Rotemberg. MIT Press.

———. 1997. "Alternative Strategies for Aggregating Prices in the CPI." Federal Reserve Bank of St. Louis *Review* (forthcoming).

Triplett, Jack E. 1971. "Quality Bias in Price Indexes and New Methods of Quality Measurement." In *Price Indexes and Quality Change: Studies in New Methods of Measurement*, edited by Zvi Griliches. Harvard University Press.

———. 1990. "Hedonic Methods in Statistical Agency Environments: An Intellectual Biopsy." In *Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth*, edited by Ernst R. Berndt and Jack E. Triplett. University of Chicago Press.

———. 1997. "Measuring Consumption: The Post-1973 Slowdown and the Research Issues." Federal Reserve Bank of St. Louis *Review* (forthcoming).

U.S. Department of Agriculture. 1996a. *Fruit and Tree Nuts: Situation and Outlook Report*. FTS-278. Department of Agriculture, Economic Research Service.

———. 1996b. *Vegetables and Specialties: Situation and Outlook Yearbook*. VGS-269. Department of Agriculture, Economic Research Service.

U.S. Department of Commerce and U.S. Department of Housing and Urban Development. Various years. *American Housing Survey for the United States*. Current Housing Report H150. Department of Commerce.

U.S. Energy Information Administration. Various years. *Housing Characteristics*. Department of Energy.

U.S. Senate. Committee on Finance. 1995. *Consumer Price Index*. Hearings before the Committee on Finance. S. Hrg. 104-69, 104 Cong., 1 sess. Government Printing Office.

U.S. Senate. Committee on Finance. 1996. *Final Report of the Advisory Commission to Study the Consumer Price Index*. S. Prt. 104-72, 104 Cong., 2 sess. Government Printing Office.

Wynne, Mark A., and Fiona D. Sigalla. 1994. "The Consumer Price Index." Federal Reserve Bank of Dallas *Economic Review* (2d quarter): 1–22.