

The Boskin Commission Report and Its Aftermath

Robert J. Gordon

This paper summarizes the analysis and findings of the Boskin Commission, reviews criticisms, provides responses to criticisms, and summarizes recent changes in the CPI and recent research on price measurement.

Key words: Inflation; Price measurement; Quality change; New products

Northwestern University and National Bureau of Economic Research

.....
This research is supported by the National Science Foundation and a grant from the Institute for Monetary and Economic Studies of the Bank of Japan. The evaluations and opinions contained herein are those of the author only and should not be taken to represent the view of any other member of the Boskin Commission. Portions of the summary of the Boskin Commission findings and its reactions to its critics are adapted from Boskin *et al.* (1998).

I. Introduction

More than four years have now elapsed since the formation of the Boskin Commission, more formally the “Advisory Commission to Study the Consumer Price Index,” appointed by the Senate Finance Committee in June 1995. Almost three years have elapsed since the Boskin Commission’s final report (Boskin *et al.* [1996]) was issued on December 4, 1996. By now, the commission’s conclusion that the U.S. Consumer Price Index (CPI) overstated inflation by 1.1 percentage points per year in 1995–96 has become familiar both within the United States and in the measurement community of other countries. Since the report was issued, the U.S. Bureau of Labor Statistics (BLS) has moved rapidly to implement some of the most important of the commission’s recommendations, so that the upward bias in the CPI is substantially less today than in 1996.

The enormous amount of journalistic attention given to the commission’s findings reflects the central role of CPI measurement in almost every major economic issue. A bias in the growth rate of the CPI contaminates official estimates of growth in output and productivity, median income, and real wages; it alters the growth rates of government tax revenues and spending programs that are indexed to inflation; it has major consequences for the time path of the government budget deficit and national debt; it produces misleading estimates of inflation for monetary policy makers for whom the inflation rate is a critical target; it distorts estimates of poverty rates; and it clouds comparisons of the comparative economic performance of nations.

A sharp distinction must be drawn between the technical and scientific issues that led the commission to its bias estimate of 1.1 percentage points, and the political and redistributive implications of that conclusion. The commission’s report contained not only the technical background for its overall bias estimate but also its estimates of the vast amount by which that bias had increased the federal budget deficit looking back into the past and would increase the deficit looking forward into the future.¹ The suggestion that the bias had caused excessive growth in Social Security and other benefits evoked a sharp political reaction, as the American Association of Retired Persons (AARP) sent its lobbyists to pressure those senators and representatives who had initially been sympathetic to reducing the budget deficit by adjusting the indexation formula by some fraction of the commission’s bias estimate, the so-called “CPI minus *X*” approach to indexation. Perhaps most decisive in derailing the movement for indexation adjustment was the miraculous disappearance of the deficit itself during the first year following the release of the commission’s report, thus eliminating any sense of urgency about changing the indexation formula.²

1. The commission calculated that the 1.1 percentage point bias would contribute US\$1.07 trillion to the U.S. national debt over the period 1997–2008 as compared to an alternative scenario in which Social Security and other programs were indexed by a formula which subtracted the bias from the published CPI.

2. The federal government budget deficit on the National Income and Product Account basis declined from a peak of US\$280.9 billion in 1992 to US\$78.3 billion in 1996/IV and to US\$0.3 billion in 1997/III (all figures are at annual rates). Since 1997/IV, there has been a surplus in every quarter that has steadily grown to over US\$100 billion.

This paper on the aftermath of the Boskin Commission report says nothing further about the implications of the bias estimate for fiscal policy. Instead, it concentrates entirely on the issues of interest to the worldwide measurement community, namely the technical basis for the commission's estimates of bias, the criticism of those estimates, the commission's response, the changes that have taken place in the CPI since the report's release, and new research on quality change. In fact, the aftermath of the report has been marked by a broader consensus, a less contentious debate, and more rapid implementation of improvements by the BLS, than individual members of the commission might have predicted as they deliberated its final conclusions over that final weekend of November 1996.

II. The Boskin Commission's Approach and Method

The Boskin Commission represented the first extensive external evaluation of the nation's price statistics in more than a third of a century; the previous report was the famous Stigler (1961) Commission, an external commission of experts set up to report to the Budget Bureau (the predecessor of the Office of Management and Budget) on ways to improve the nation's price statistics. Some contrasts between the Boskin and Stigler commissions provide useful background for those evaluating the Boskin Report and subsequent reactions to that report. The reports had one thing in common; Zvi Griliches was involved in both. There were several important differences. The Boskin Report concerned only the CPI, while the Stigler Report covered the Producer Price Index (PPI) and agricultural price indexes as well. The Stigler Report did not produce any numerical estimate for biases in any of the price indexes it reviewed (although it concluded that virtually all economists would agree there was an upward bias), whereas the mandate of the Boskin Commission included the provision for a point estimate of the overall bias in the CPI. The Stigler Commission had a substantial budget to commission new research studies, including Griliches' famous paper on hedonic price indexes for automobiles (Griliches [1961]), whereas the Boskin Commission had no budget at all (except for travel expenses that enabled commission members to meet).

Boskin Commission members divided up the task of assembling all the existing research by academics, BLS insiders, and others that had been done over the previous 35 years.³ In doing so, they benefited greatly from one of the main recommendations of the Stigler Commission, which was that the BLS create a price research unit, which has produced a great deal of high-quality research on price index issues, particularly within the past decade. Nevertheless, it is important to recognize that the Boskin Commission report of necessity was a survey article based on past research, including very recent unpublished findings, rather than a producer of new research. The requirement to produce a point estimate of the overall CPI bias forced commission members to go beyond simply surveying past research to drawing out numerical

.....
3. Especially important contributions included Diewert (1976, 1995), Gordon (1990), Griliches (1961), Hausman (1997a, 1998), and Pollak (1981).

conclusions from this research. A novel aspect of the commission's method was to divide the CPI into 27 categories and develop a separate estimate of quality change bias for each. This required some extrapolation from categories for which research existed (e.g., personal computers and TV sets) to related categories where no research existed (e.g., electronic toys). While some commentators have complained of the subjective nature of the commission's assessment of quality change bias, the commission felt that to assume that the bias was exactly zero in categories which had not been the subject of previous research was just as subjective as the commission's approach, which was to extrapolate the results of related research.⁴ We return to this issue below when commenting on some of the critiques of the commission report.

A final difference between the Stigler and Boskin commissions was in the reaction to the two reports, which was low-key and virtually invisible in the case of the Stigler Report and highly visible and politically charged in the case of the Boskin Report. This difference reflects in large part the absence of indexed tax rates and government benefits in 1961 in contrast to long-established indexation in 1996, and the consequent widespread awareness of the political and redistributive implications of the Boskin conclusions. Also, the 25 years of slow productivity growth since 1972 had spurred substantial academic research on the interplay between economic growth and measurement errors (e.g., Griliches [1994]), greatly increasing the awareness in the academic community of the implications of the Boskin findings.

III. Summary of the Boskin Commission's Findings

This chapter provides a relatively brief summary of the Boskin Commission report for those who are unfamiliar with it. We begin with a summary of the commission's description of the CPI and the inherent difficulties in measuring consumer prices; then we summarize the commission's discussion of bias involving aspects of product substitution, and finally the commission's conclusions regarding bias related to quality change, new products, and general aspects of the quality of life. Subsequently, the commission's recommendations to the BLS are summarized (again, we exclude here any discussion of fiscal policy and the recommendations of the commission regarding the use of the CPI for indexation).

A. The CPI and Difficulties of Price Measurement

Why are changes in the cost of living so hard to measure? There are literally millions of goods and services available in modern market economies. New products are being introduced all the time and existing ones improved, while others leave the market. Relative prices of different goods and services change frequently, for example, in response to technological and other factors affecting costs and quality, which leads consumers to change their buying patterns. As we have become richer, demand has increasingly shifted to services away from goods, and to characteristics of goods and

4. Nevertheless, the commission did not use related research indiscriminately and concluded, in the absence of related information, that the quality change bias was zero in eight of the 27 categories.

services such as enhanced quality, more variety, and greater convenience. Hence, the very first point the Boskin Commission made in its report was that inflation is inherently difficult to measure in a complex dynamic market economy. The CPI program of the BLS is a large, complex, and impressive undertaking, staffed by dedicated professionals.

Defining what information one would need and how to use it to obtain an overall summary measure of the level and/or change in consumer prices requires various assumptions and procedures. The first question to answer is “What is the domain of consumer expenditures to be covered?” The CPI program focuses on consumer expenditures on goods and services out of household disposable income. Hence, it excludes non-market activity, broader quality of life issues, and the costs and benefits of most government programs.

Second, one has to define the prices of the commodities and services one wants to measure, how to measure them, how to collect data on them, over what span of time and at what interval, where and when to collect the data, and how to aggregate them into one or several overall summary statistics. At each of these levels, various judgments and assumptions must be made in order to make practical headway. Some of the commission’s suggested improvements dealt with various of these assumptions such as the use of fixed weights for a given market basket that is held fixed for a decade or more.

The BLS collects price quotations on 71,000 goods and services, at about 22,000 retail outlets, either monthly or bimonthly. Additional information is obtained on rent and owners’ equivalent rent—that is, how much owners are paying in opportunity cost terms for housing services—from about 35,000 rental units every six months. About once per decade, the weights for different commodities are derived from the Consumer Expenditure Surveys (CES). At the time of the commission’s report in 1996, the weights were based on the average of 1982–84 and thus were on average 13 years out of date. In January, the BLS updated the weights to the period 1993–95, which are by now on average five years out of date.⁵

The CPI price collection goes through a hierarchical or pyramidal process. At the top is the all-item CPI—the overall summary measure of the change in consumer prices from month to month. Next comes a small number of broad commodity groups such as housing, food and beverages, apparel, transportation, medical care, entertainment, and other. Each of these categories is subdivided into other categories. For example, housing is divided into shelter, fuel and other utilities, and household furnishings and operations. Each of these subcategories is divided still further, e.g., household furnishings into categories such as furniture and bedding, appliances, etc. The lowest level of aggregation for which an index is calculated by aggregating price quotes is called an item stratum. There are 207 item strata collected in 44 geographical areas, leading to 9,108 strata indexes. Below the strata are entry level items (ELIs), from which specific products are sampled. The basic price quotes are on

5. A detailed description of CPI methodology is contained in the U.S. Department of Labor, BLS (1992). This section refers to the CPI as it was measured in 1996. As discussed below, numerous improvements in methodology have been introduced since 1996, including plans for much more frequent updating of the upper-level weights.

actual commodities and services, not on this lowest level of aggregate or ELI. For example, one doesn't go into a store and buy "apples," one buys a specific type and size, like large red Jonathan apples. Therefore, the price data on actual commodities purchased in various places must be aggregated to obtain the strata indexes, e.g., for apples in Los Angeles.

Outlets are chosen and rotated every five years from a point-of-purchase survey (POPS) asking consumers where they purchase goods and services, with probabilities of outlet selection proportional to expenditures. The prices are collected and compared within outlets. Within the outlet, the BLS collects prices on specific items, with probability of selection proportional to sales, and reprices the same item for five years. Often the same item is not in stock in successive months, and a judgment must be made to price a comparable item or not; if so, differences in the substituted items are noted and, as noted below, adjustments are made. At the individual item level, there is roughly a 20 percent chance each year that an item will be replaced by another item and/or another outlet. No account is explicitly taken of substitution across outlet types, as might be expected with the evolution of retailing, for example, to widespread use of discount stores in recent years.⁶

In calculating the lowest-level strata price index aggregates on the basis of price quotes, the data were aggregated using a variant of arithmetic means until the changeover in January 1999 to geometric means (this was the most important single recommendation of the Boskin Commission that was accepted and rapidly adopted by the BLS). Arithmetic aggregation was historically defended as consistent with the fixed-weight formulas used at the higher level of aggregation. In short, it was assumed that there was no substitution within these categories, even when they included items among which there is likely to be immense substitution, such as types of apples, styles and colors of men's dress shirts, and so on. To assume that the elasticity of substitution in consumption was zero, as did the BLS until 1999, was unrealistic for virtually every commodity group and dramatically so for most.⁷

Once the elementary indexes are formed in this way, higher levels of indexes are aggregated again with fixed weights up the pyramid all the way to the all-item CPI. The method used is to look at what proportion of their income consumers actually spend in each category as revealed by a past CES, which defines a fixed "basket" of goods, and then to figure what it would cost to purchase that basket in later years up to the present. This approach to measuring the changes in the cost of a fixed basket of goods is called a Laspeyres index. Notice that it assumes no consumer substitution occurs in response to changes in relative prices, an assumption that is extreme, unrealistic, and unnecessary. As noted above, the 1982–84 weights in effect at the time of the commission's report were more than a decade out of date at the time of the commission's report, and the newer 1993–95 weights in effect now are already half a decade out of date.

6. This description describes the pre-1998 rotation procedure and does not reflect the more rapid and flexible rotation scheme since adopted, as described below in the section on recent CPI improvements.

7. See Diewert (1995) for a discussion of properties of alternative methods for aggregating elementary price quotes.

There are two approaches to measuring cost of living (COL) indexes. The first uses estimated systems of demand equations to make exact utility comparison accounting for the substitution estimated in the demand system. This approach has been successfully implemented at a relatively high level of aggregation; see, e.g., Braithwait (1980) and Jorgenson and Slesnick (1983). However, the number of parameters that must be estimated in a full system of consumer demand rises with the square of (one less than) the number of commodities and quickly becomes impossible to implement at any detailed level of disaggregation. If much of the substitution goes on at the lower levels of disaggregation, this would be missed. But the estimates of substitution bias in, e.g., Jorgenson and Slesnick are quite similar to the far more detailed disaggregation used in the numerous studies by the BLS described below.

The second approach uses mathematical formulas to calculate index numbers, usually at a more, often extremely, detailed level of disaggregation. The most common index number formula is the Laspeyres index, which measures the changes in the cost of a fixed basket of goods, i.e., it assumes no substitution due to relative price changes. Thus, it usually is an overestimate of a “true” COL index. At the other extreme is the Paasche index, which weights by the current quantities, and thus likely overstates substitution and understates the COL index.

Superlative indexes (Diewert [1976]) approximate the true COL index, under certain assumptions, without having to estimate the demand equations and thus can handle the dimensionality problem inherent in disaggregation in a modern market economy with vast numbers of goods and services. The Fisher ideal index (Fisher [1922]) is the geometric mean of the Laspeyres and Paasche indexes, i.e., the square root of their product. The Törnqvist index (Törnqvist [1936]) exponentiates the sum of the weighted natural logarithms of price relatives, with the weights being the averages of the expenditure shares of the goods in the current and base periods (Aizcorbe and Jackman [1993] provide the mathematical formulas for these indexes, in two period and chained form).

When the BLS price takers confront a new product, a basic dilemma arises. If the product fits fairly reasonably into an existing category it may well be rotated in, perhaps somewhat late, as part of the sampling procedures embedded in the POPS. The use of sampling was a major recommendation of the 1961 Stigler Commission Report (Stigler [1961]) and was implemented in 1978. However, when a new product comes along that does not fit neatly into an existing category, or is “missed” for other reasons, it may be an entire decade or more before the product is introduced into the CPI. This was the case in the past with numerous products, including the automobile (not introduced until 1935), air conditioners (1964), VCRs (1987), and cellular phones (1998). The problems presented by new products are difficult and likely to be quantitatively very important. Much of the improvement in standards of living in modern economies derives from new products, especially major innovations such as in light, transportation, communications, and so on (Nordhaus [1997, 1998]).

The changes in the CPI are published monthly, and are not revised, even when subsequent information calls the previous estimate into question. Experimental

indexes are used to bring new methodological and other information to bear when the BLS decides to make improvements and are usually published alongside the traditional measure for a while before, sometimes, being incorporated into the CPI. This was done, for example, with the major improvement in the treatment of owner-occupied housing in 1983. Because the BLS methodology continually changes and the CPI is never revised backward, the CPI is inherently inconsistent methodologically over time and can incorporate very different amounts of bias at different times in the past.⁸

B. Findings Related to Substitution

The commission's findings are summarized in Table 1. In this section, we discuss the findings related to substitution, comprising the first three lines in the table. Next we discuss the category of new products and quality change.

Table 1 Boskin Commission Estimates of Bias in the Consumer Price Index in 1995–96

Percentage points per annum

Source of bias	Estimate
Upper-level substitution	0.15
Lower-level substitution	0.25
Outlet substitution	0.10
New products/quality change	0.60
Total	1.10
Plausible range	(0.80–1.60)

Despite decades of analytical and empirical research, some done by the BLS, the CPI (until 1999) relied on fixed-weight Laspeyres indexes that did not account for consumer substitution among commodities. These Laspeyres measures of inflation were inherently an upper bound, and empirical studies led the commission to conclude that this source of bias amounted to about 0.4 percentage point per year. Of this, 0.15 percentage point came from estimates of the effect of substitution among the upper-level “strata” (apples vs. bananas) and 0.25 percentage point from the effect of substitution among the lower-level ELIs (red delicious apples vs. Jonathan apples). Starting in 1999, the CPI uses geometric weights for most of the lower-level categories while retaining the Laspeyres index for the upper-level strata.

In addition to substitution bias among commodities, there is an outlet substitution bias, the third line in Table 1, which refers to the practice of the BLS in ignoring differences in prices for the same item across outlets. However, there has been a fundamental change in the nature of retailing, perhaps most pronounced in the United States, but spreading virtually everywhere, with the advent of superstores and discount chains. The fraction of department store sales occurring at discount stores increased from roughly 45 percent in 1988 to roughly 68 percent in 1998. The same VCR available for US\$200 in a local appliance store may be only US\$160 at a superstore. The ratio for ordinary products like toothpaste varies more, perhaps

8. A recent and extremely helpful attempt to produce a methodologically consistent CPI for the period 1978–98 is contained in Stewart and Reed (1999).

in the range of US\$2.49 to US\$1.39. In the past two years, I have observed the same exact type of Kodak film selling for anything from US\$7.49 in a convenience store or airport to US\$1.75 with a promotional coupon in a nearby discount drugstore chain.⁹ Since price data are collected *within* outlets, the shift of consumers to purchases from discounters does not show up in the CPI as a price decline even though consumers reveal by their purchases, measured by the steady shift in market share to discount outlets, that the price decline more than compensates for the potential loss of personal services. In the United States, we estimated this adds another 0.1 percentage point of upward bias.¹⁰

C. Findings and Recommendations Related to New Products and Quality Change

The three types of substitution bias listed in Table 1 account for just a little under half (0.5 percentage point) of the 1.1 percentage point bias identified by the Boskin Commission. Slightly over half (0.6 percentage point) results from the difficulty of adjusting fully for quality change and the introduction of new products. Economists have known since Hicks (1940) that the introduction of a new product should be dealt with in a COL index by estimating its reservation price and including the consumer surplus attributable to the introduction of the product. While accepting this framework, the Boskin Commission did not make any of its own estimates of the consumer-surplus value of new products. Instead, it took the more cautious view of primarily including estimates of explicit dimensions of quality change and the late introduction of major new products into the index. As stated above, VCRs, cellular phones, and other products were included in the CPI a decade or more after they had penetrated the market and their price had fallen by 80 percent or more.¹¹ Previous research allowed an estimate to be made of the bias in the CPI that occurred as a result of the late introduction of these new productions.

In addition to taking into account the substitution issues raised above, and the bias attributable to the late introduction of new products, the commission attempted to identify commodity categories in which the CPI was biased for failing to take account of improvements in quality. To carry out this task, the commission examined separately 27 subcomponents of the CPI to determine the extent of quality bias, if any, and established an estimate for quality bias within each of the 27 categories.

-
9. The Walgreen promotional coupon comes with every roll of film processed at the same chain (the processing price is lower than any local independent film developer and the quality is comparable). The coupon provides a US\$4.00 discount on a four-pack roll of Kodak ASA-200 24-exposure 135-millimeter film, reducing the price from US\$10.99 to US\$6.99, or US\$1.75 per roll. It is worth noting that in many U.S. metropolitan areas Walgreen drug stores are ubiquitous and often require, as in my case, a drive of less than a mile from home.
10. Alternatively, the bias could be expressed as 1 percent per decade, which seems minor in view of the large shifts in market share and large price differentials between old and new format stores. The BLS has argued that service levels differ between high- and low-price outlets. While this may be partially correct, much outlet substitution occurs between inefficient self-service outlets and more efficient self-service outlets that provide identical levels of service. We note that numerous inefficient outlets, e.g., Woolworth and Montgomery Ward, have lost market share and ultimately gone out of business in the face of market-share gains by such outlets as Wal-Mart and Home Depot.
11. Hausman (1997a) estimated that the quality adjusted price of cellular services declined by 90 percent from its reservation price between 1989 and 1996. Cellular phones were introduced into the CPI in 1998. As discussed in the chapter below on recent research, Hausman's more recent paper (1998) comes up with a smaller 51 percent figure for the decline in cell phone prices from 1985 to 1997.

Since no separate table was created to deal with bias due to late introduction of new products, the calculation for quality change implicitly included any additional bias due to the late introduction of new products. In eight of the categories, the commission discovered no research evidence or other factor that would indicate a bias other than zero. In the other 19 categories, bias estimates were assigned that ranged as high as 3.0 percent per year for medical and hospital services and 5.6 percent for appliances and electronic goods.¹² When these 27 bias estimates were weighted by the relative importance of each category in 1995 (based on 1982–84 expenditure weights), the overall quality and new product bias was determined to be 0.6 percentage point per year.¹³ While the commission formulated a point estimate of CPI bias related to quality change and new products, it did not attempt any quantitative estimate of changes in the quality of life. Nevertheless, it did present an informal discussion of changes in the quality of life in an attempt to determine whether there were additional factors that might make its overall bias estimate tend to be too high or too low. Interestingly from today's perspective, this section was almost apologetic and was not phrased in a "maybe too high, maybe too low" symmetry. The section was written for those who might find the bias estimates too high and divided the offsetting negatives into three categories. The first of these was the possible loss of welfare by users of old goods when new goods drive them out of the marketplace (as a perpetual user of WordPerfect 6.0 for DOS, on which this paper was written, I can understand at least one small part of this issue). The second was that an existing good could deteriorate in quality, as in the report's example of the effect of health maintenance organizations (HMOs) in restricting access to medical care. The third contained a survey of externalities ("physical, social, and economic") that might "impose on us higher expenditures necessary to keep up with our previously achieved utility levels" (Boskin *et al.* [1996]).

The commission's quality-of-life assessment concluded that the "good" outweighed the "bad." It cited a reduction in air and water pollution, a decline in crime by various measures, a decline in the suicide and infant mortality rates, and an increase in life expectancy.¹⁴ It also cited a wide range of improvements in goods related to quality and new products for which the commission's estimates made no allowance, namely "the faster speed and reduced vibration of jet planes, improved reliability of appliances and automobiles, improved sound quality of audio equipment in homes and automobiles, improved safety devices on home power tools and power lawn mowers, reduction in the noise, weight, and installation cost of room air conditioners, and immeasurably better picture quality of color TV sets." As other improvements for which no allowance was made, it cited the spread of cable and satellite TV and the new availability of the

12. Table 2 of the Boskin Report attempted to discriminate between different periods in assigning the estimates for quality change and the late introduction of new products. Thus, for instance, the bias estimate for prescription drugs was 3.0 percent for 1970–95 and 2.0 percent for 1995–96, reflecting the change in CPI methodology in 1995 which recognized the introduction of a generic version of a given drug as a price decline.

13. The only case in which alternative weights were used applied to consumer appliances and electronic goods, where the 1982–84 CPI weights were deemed to be too low, and weights were taken instead from the National Income and Product Accounts. See Boskin *et al.* (1996), Table 2.

14. The one serious negative that was cited was the increase in births to unmarried women. Other "intangible negatives" cited were "increased job insecurity, possible increased inequality, and decreased job opportunities for workers with only a high school education."

World Wide Web to owners of personal computers (another product that was introduced into the CPI long after it started becoming common in homes). The commission's section on quality-of-life issues concluded that "the major increase in longevity . . . perhaps swamps everything else. Accordingly, our estimate of the current bias in the CPI is, if anything, probably understated."

D. The Commission's Recommendations

Rather than dividing its recommendations by the type of bias (substitution, new products, quality change), it divided them instead into three time horizons. The first group comprised short-run recommendations that could be implemented immediately with few additional resources or new data collection initiatives. These centered on changing the current CPI computation, primarily to make it more current, and second, on computing an alternative, annually updated, and continuously revised COL index. The second group comprised intermediate-run recommendations that incorporate reforms which are feasible in the current state of the art, but would require new data collection, reorganization of activities, and/or changes in the detail of the various subindexes produced by the CPI. The third group consisted of longer-run recommendations, emphasizing topics in areas that require additional research and attention.

Abstracting from the time frame for implementation, the commission's first and overarching recommendation was that the BLS should establish the COL index as its objective in measuring consumer prices. All of the other, more specific recommendations were aimed at achieving this goal. The emphasis on this first recommendation may seem strange to economists, but the BLS in its publications for decades had explicitly stated that the CPI is not a COL index. Still, its most common and pervasive use and interpretation is as a COL index. The commission stated that a fundamentally sound COL index could and should be developed.¹⁵

As part of the short-run set of recommendations, the commission recommended that the single current CPI, which can never be revised (due to its use in legal contracts and for indexation) be supplemented by a second "research-based" index that is published annually rather than monthly and is continuously revised to incorporate new research results. The timely monthly index would continue to be called the CPI and would move toward a COL concept by adopting a "superlative" index formula to account for changing market baskets, abandoning the Laspeyres formula at both the upper level and lower level of aggregation.

At the upper level, the commission recommended that the BLS should move away from the assumption that consumers do not respond at all to price changes in close substitutes, moving instead to a "trailing Törnqvist" index (weighted geometric mean of price relatives) or another approximation to a superlative index, at the stratum and ELI level, and also, concurrently, to geometric means of price relatives at the elementary aggregation level. These changes would eliminate the problem of the growing irrelevancy of market baskets based on decade-old consumption patterns,

15. The BLS has stated (1997a) that it has embraced the commission's overarching recommendation concerning the objective of the CPI.

reduce significantly the substitution and (any remaining) formula bias, and facilitate the speedier introduction of new goods and services into the index.

Because of the lag in collecting up-to-date information on consumer spending patterns, the weights would have to be based on a trailing two- or three-year average of past expenditures, e.g., 1993–94 weights for the 1996 price changes. These weights should be changed every year. Thus, the commission recommended that the BLS should organize itself to implement ongoing rather than decadal revisions in the CPI. Also, wherever possible, scanner data and other “outside” data should be used to expand the assortment of goods and services priced concurrently, to provide current item weights, to track actual transaction prices more accurately, and to introduce new items as quickly as possible after they enter the market.

The distinction between the “timely” CPI and the new annual research-based index rests on the fundamental proposition that the basic monthly CPI can never be revised. The commission recommended that as subsequent data became available, the weights were updated, and new goods were introduced and the history of their price changes was extended backward, the information incorporated in the published CPI could undergo retroactive revision of a new annual COL index using a compatible “superlative-index” formula that would no longer be affected by the lag in the availability of the required expenditure weight data. This alternative COL index would be published annually, with a lag of a year or two, and would be subject to additional revisions after new information emerged and new methodology was introduced.

Several additional remarks on the recommendation for a second index are warranted here. First, the United States suffers not only from having (in contrast to Canada) a divided statistical system but also two partly overlapping but partly independent measures of consumer prices—namely, the CPI published by the BLS and the deflator for personal consumption expenditures (PCE) published by the Bureau of Economic Analysis (BEA) as part of the National Income and Product Accounts (NIPA). The PCE deflator comes closer to representing the second type of index that the commission had in mind, primarily because it is frequently revised to reflect new research results. For instance, the NIPA deflators incorporated hedonic price indexes for computers more than a decade before these were introduced into the CPI, and the NIPA deflators switched in 1993–94 to deflate a major portion of medical care expenditures by the PPI instead of the CPI when the former were deemed to represent more adequate indexes. Nevertheless, the PCE deflator is still largely reliant on the CPI for its underlying commodity detail, and the BEA does not have the research resources or inclination to revise the PCE deflator to reflect research results that have not yet been incorporated into the CPI. Because of its prohibition against revision, improvements in the CPI are always forward-looking, whereas the academic and statistical community has a strong interest in an index that is revised frequently, incorporates new research results retrospectively, and to the maximum possible extent is based on a consistent methodology.¹⁶

16. Since the commission referred to its desired second index as a “research-based index,” it was not helpful that Stewart and Reed (1999) refer to their new experimental index as the “CPI Research Series.” The Stewart-Dunn index attempts to apply current CPI methodology in a consistent manner backward to 1978 but does not incorporate any academic research on quality change or new product bias as the commission had in mind.

Continuing with recommendations that would require a somewhat longer term for implementation, the commission felt that the BLS should revise its approach to sampling. To put its reaction in a nutshell, the commission was astonished at the number of price quotations that the BLS collected on ordinary products like bananas that were not subject to quality-change or new-product bias, relative to the effort devoted to collecting price data on new products like cellular phones, personal computers, and computer peripherals.

To determine which commodities to emphasize and deemphasize in its sampling procedure, the commission urged that the BLS study the behavior of the individual components of the index to ascertain which components provided the most information on the movements in the index and which items had fluctuations that were largely unrelated to the total. In addition to reducing the data collection effort devoted to apples and bananas, the commission believed that the data collection effort should be divided between national and local goods. In this revised scheme, there would no longer be any collection of data on appliances, other consumer durables, and imported produce (like tomatoes and bananas) in each city—data for these national goods could be collected in a much smaller sample applied to the entire country. This would free up resources to do a better job of collecting prices for local goods with components that might vary across cities, e.g., fuel costs, rent, household services, and non-imported produce.

The commission's intermediate-term recommendations would imply a radical change in the distribution of CPI price quotations collected by city and by product category. Moving to a new "basket" each year would allow for a faster introduction of new items and new outlets. Moving to a national sample for most of such items would allow expansion of the number of specific items (models, varieties, types) sampled within a particular ELI and reduce the number of forced substitutions. Also, this would allow for the use of new sources of data, such as scanner data on prices, and industry-wide information on sales of specific items (for more detailed weights), leading to a quicker identification of new goods and their faster incorporation into the index. This is also the level at which more extensive hedonic quality adjustments and "comparable" substitutions could be made, recognizing the appearance of new outlets and new versions of services that provide consumers, effectively, with cheaper sources for the same or similar items consumed previously.

Longer-run considerations in the commission's recommendations involved examining the ramifications of the assumptions of the law of one price; developing (in conjunction with other statistical agencies) research programs to look beyond the current "market basket" framework; and developing a number of new data collection initiatives, in particular, expanding the CES, health status surveys to obtain more information on various quality of life issues, and detailed time use from a large sample of consumers to deal with numerous issues, including commuting time and the value of new products.¹⁷

17. For instance, the value of the Internet as a consumer product depends in part on the extent to which web-surfing or game-playing are substitutes for television and other types of entertainment. The rapid decline in television viewership for the traditional national networks suggests that a substantial part of cable viewership and web-surfing may represent substitution within forms of electronic entertainment.

IV. Criticism of the Commission's Findings and Its Reaction

Volumes could be written about reactions to the commission's findings and recommendations. Here we attempt to maintain a feasible scope for the paper by concentrating only on issues that involve the validity of the commission's bias estimates, and we skip entirely the firestorm of political controversy generated by the implications of the commission's bias estimate for the indexation of Social Security and other government programs.¹⁸

The report of the Boskin Commission received much attention. Most of the findings and recommendations have generally been accepted by the economics profession, if we are to judge by various symposia in which prominent academic economists have been invited to comment on the commission's findings.¹⁹ Other analyses come to quite similar conclusions. External corroboration comes from several sources, including the Federal Reserve's study of productivity (Slifman and Corrado [1996]), Nordhaus' (1998) analysis of survey data on changes in economic well-being, and Diewert's (1998) thoughtful complementary analyses of bias by type.

A variety of technical criticisms of our report and findings have appeared in the media, press conferences, Congressional testimony, professional meetings, discussions, and technical papers. Perhaps the most extensive critique of the commission's finding is by Moulton and Moses (1997), who are in places generous to our report, calling it "the most influential critique of the CPI in decades." It is noteworthy that few if any criticisms addressed the commission's basic recommendations that the CPI should become a COL index and that substitution at both the upper and lower level should be addressed within the framework of superlative index numbers. Instead, most of the criticisms of the commission's findings centered on its treatment of quality change and new products.

- (1) The commission did not give adequate attention to quality deterioration (Abraham [1997b], BLS [1997a], and Abraham *et al.* [1998]).
- (2) The BLS already makes lots of quality adjustments, which the commission did not adequately credit (Abraham [1997b], BLS [1997a], Moulton [1996], and Moulton and Moses [1997]).
- (3) The commission made too many back-of-the-envelope calculations and was too willing to generalize from research on one item to research on related

18. Perhaps the most interesting intellectual issue involves the interplay between Social Security indexation and improvements in the quality of life taking the form of increased life expectancy. Should quality adjustments in the CPI include those improvements in medical technology that lead to improved life expectancy (see our survey below of research on medical care price indexes)? If such quality adjustments are warranted, should Social Security benefits be reduced when such adjustments make the CPI rise more slowly? In citing evidence on quality change in medical care that included an adjustment for life expectancy, and then applying its full bias estimate to calculate potential savings in Social Security expenditures over the 1997–2008 period, the commission implicitly seemed to recommend a link between life expectancy and Social Security benefits that none of its members intended or would endorse today.

19. The most important published symposium appears in the Winter 1998 issue of the *Journal of Economic Perspectives*. In the autumn of 1997, the Conference Board convened a group to review the issue of CPI bias, consisting entirely of prominent economists who were not members of the Boskin Commission. While the results of the Conference Board panel have not yet been released, I have read their preliminary report and find that it differs in no significant way from the conclusions of the Boskin Commission.

items, both with respect to quality change and with respect to the desirability of geometric means to deal with lower-level substitution bias (Abraham *et al.* [1998] and Nordhaus [1998]).

- (4) The Commission's estimates, whether right or wrong, are of limited use because they cannot be generated from a mechanical procedure implementable in a month-to-month price program (BLS [1997b] and Abraham [1997a]).
- (5) We should have been more aggressive in our estimates of the value of new products (Hausman [1997b] and Nordhaus [1998]).

V. The Commission's Response to the Critiques²⁰

Because there was so little criticism of the commission's estimates of substitution bias, it is perhaps worth noting that research by Shapiro and Wilcox (1996) suggests that the bias could have been a tenth of a percent or so higher than the estimate in our report, which was based heavily on BLS research.²¹ The BLS response to our recommendations about lower-level substitution bias pointed out that, while the degree of substitution between some goods like white shirts and blue shirts may be very high, that between other goods like particular types of pharmaceuticals may be very low (Abraham *et al.* [1998]). However, pharmaceuticals are the exception, not the rule, and within the vast majority of the 207 strata there is ample latitude for substitution, whether we are talking about types of food, shelter, apparel, cars, tires, appliances, entertainment, or personal care items. After all, medical care goods and services other than pharmaceuticals comprise only 6.3 percent of the upper-level weight in the CPI, and pharmaceuticals only 1.2 percent. If anything, the unitary elasticity of substitution embodied in the commission's recommendation of geometric weights at the lower level probably understates the extent of substitutability within the majority of the strata.

Before turning to quality change, there is another issue that may cause the commission's estimate to understate the overall bias yet was omitted from its quantitative bias estimates. In addition to outlet substitution bias, there are additional types of discounts that may be missed with current BLS collection techniques. As the report pointed out, virtually all of BLS price quotations are collected on weekdays, thus missing frequent weekend sales. From discussions with BLS officials, I emerged unconvinced that the CPI makes adequate allowance for discounts at food stores and other outlets enjoyed by shoppers possessing frequent shopper discount cards.²²

20. This chapter combines sections of the jointly authored paper by commission members (Boskin *et al.* [1998]) with my own further observations.

21. Many of the best studies of substitution bias have been done by BLS researchers (see, e.g., Aizcorbe and Jackman [1993]). The studies from systems of demand equations at high levels of aggregation generally also come to an estimate of 0.2 to 0.25 percentage point on both U.S. data and data from other countries. See Greenlees (1997) for further elaboration of the BLS research.

22. During the period of the commission's deliberations, I saved receipts from my local supermarket and found that each receipt reflected frequent shopper discounts of between 2 and 5 percent.

Other more subtle forms of discounting made available to frequent customers may be invisible to CPI field agents.²³ Finally, it is clear that the CPI for air travel is overstated by the failure to place a value on frequent-flyer award travel, which amounts to a discount on the tickets that were originally paid for in order to earn the miles for the subsequent free ticket.

The potential bias because of the failure to collect price data on weekends and holidays, unmeasured discounts of various types, and outlet bias, combined with the discussion above, might well lead us to conclude that substitution bias among commodities at the upper and lower level, across outlets and at different times, might be somewhat larger than the 0.5 percentage point we indicated in our report.

VI. The Debate over Quality Change and New Product Bias

Most of the criticism focused on our extensive analysis of quality change and new product bias. On the question as to whether estimates of quality change bias are inevitably too “subjective” and “judgmental” to be taken seriously, it is, of course, at least 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. The notion that assuming zero bias is scientific, whereas attempting to generalize cautiously from related goods or practical reasoning is not precise enough, seems to throw out information that, even if not precise, is available. For instance, even though we will never precisely measure the value of the invention of the jet airplane, as economists we *know* that consumer surplus triangles have an area that is positive rather than zero. Likewise, we know that PCs with Pentium processors are objectively higher quality (faster) than the 386 and 486 machines they replaced.

Hence, the commission examined 27 subcomponents of the CPI, and most of our estimates of quality change were based on the collection of price data from independent sources and the careful quality adjustment of those independent data. Independent sources of price data were employed in our bias estimates for shelter, appliances, radio-TV, personal computers, apparel, public transportation, prescription drugs, and medical care. Estimates derived from these categories were extrapolated, sometimes partially rather than fully, to other house furnishings, nonprescription drugs, entertainment, commodities, and personal care. This left only a few remaining categories where we added a bias estimate to the CPI category in which there were already quality adjustments, rather than computing the bias estimate indirectly by subtracting an independent estimate from the CPI estimate for the same category. These categories were food and beverages, other utilities, new and used cars, motor fuel, and personal expenses. The BLS did not object to our “down in the trenches” approach to the problem. Indeed, Moulton and Moses (1997) stated, “This is the first time that a systematic analysis of quality bias has been done

23. At my local dry cleaner, for example, an electronic calculator is used to add up the list price of the items, then the calculator screen is shown to certain clients and a discount is “announced.” In my case, a calculator total of, say, US\$45.40 translates into “That will be US\$40, please.” The same technique of rounding down for good customers also occurs at a local independently owned photo finishing shop.

category by category, which we consider to be a noteworthy accomplishment of the commission . . . [the] overall approach seems to us to be a sensible and useful way to approach the problem of coming up with an overall assessment of bias, and we expect this type of structure will prove to be useful in the future.”

Some outside critics of the commission argued that the BLS already does a great deal of quality adjustment, and that the commission’s report is flawed for ignoring the extent of the BLS adjustments.²⁴ However, for most categories, the extent of current BLS quality adjustments is irrelevant to an assessment of the commission’s treatment of quality change. We were comparing our own evidence to the corresponding CPI indexes—however they are quality adjusted, in a major or minor way—and thus our estimates of quality change bias are a residual that remains after the BLS has completed its efforts.

However, it is still instructive to discuss what the BLS calls quality adjustment, since it illustrates the substantive and communication difficulties in this field.²⁵ There is very little explicit adjustment for quality change (Nordhaus [1998]). Most of the reported “quality adjustment” by the BLS—1.65 out of the 1.76 percentage points in Moulton and Moses (1997)—comes from “linking” procedures, where a missing item is replaced by another (excluding outliers, commodity pairs where the implicit price-quality differential exceeds 100 percent, the quality adjustment number shrinks to 0.3 percent). 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 of other commodities in the particular class. Roughly one out of three items disappears sometime during the year and has to be replaced by a different item in the same general class, such as a larger versus a smaller package of yogurt, a blue raincoat versus black, or a 12-cubic-foot refrigerator with its 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, greater speed, durability, variety, convenience, safety, energy efficiency, etc., for example, the increased variety and freshness of vegetables and fish due to improving transport facilities and the globalization of trade, or the substitution of laparoscopic procedures for gallstone operations, and so on.

Yes, the BLS does lots of “price adjustments.” It is forced to by its sampling framework and the product turmoil in the markets. However, the BLS is not looking for the “quality change” that we were worried about. And it does not adjust explicitly for quality *change*, as we were defining it (except in the case of automobiles, apparel, and possibly rental apartment units and the occasional truly new goods caught by their substitution procedures). Moreover, such explicit quality adjustments account for only about 6 percent of the total “treatment of substitutions” effect, and amount to only 0.08 percent per year in the “outlier-cleaned” recomputations. While some of the commission’s estimates can be questioned—in both directions—there is very little overlap between them and the recent numbers produced by the BLS.

24. In fact, most of the Moulton-Moses paper (1997) is devoted not to a direct critique of the commission’s estimates but to an explanation of how the CPI is adjusted for quality change and to an attempt to estimate the quantitative significance of those adjustments.

25. Some of the material in this chapter comes from my published discussion of Moulton and Moses (1997).

The helpful Moulton-Moses (1997) discussion of several categories would probably lead us to reduce our overall quality change bias estimate by perhaps 0.1 of the total 0.6, if that were the only new information since the report; but other new research information goes in the opposite direction.²⁶ However, even though it may turn out that some of our estimates of quality change were too high, others are likely to be too low.²⁷ Further, with the exception of a few cases having low overall weight in the CPI, we did not explicitly estimate the additional welfare gain of the numerous new commodities in the economy.

The area in which the commission's bias estimates are surely likely to be too low, rather than too high, is in the area of new products. We did not make explicit allowance for the late introduction into the CPI of numerous new products. We indicated that the appropriate way to deal with new products is to value the consumer surplus from their introduction, as first demonstrated by Hicks (1940), and recently elaborated and applied by Hausman (1997a, 1998) to the case of cellular telephones. We chose to deal with this by being deliberately cautious, but indicating that there was an asymmetrical bias with more potential bias on the upside than the downside because of the likely future new product introductions which were unlikely to be captured in the CPI program.

Nor did we try to quantify all of the intangible aspects of quality change, such as the improved safety of home power tools or the improved quality of stereo sound and TV pictures. But we did try to do so in some cases, e.g., the increased freshness and seasonal availability of fruits and vegetables.

VII. Problems of Implementation

The BLS response to some of the commission's recommendations has been that they are very difficult to implement in real time in a monthly CPI program using mechanical rules which are straightforward to implement. We have some sympathy for this position, which is why we proposed that the BLS calculate a second index, published annually, that would constantly be updated and continuously revised, and which need never be final, as new information, research, etc., became available both inside the BLS and externally. This second index addresses many—though not all—of the BLS concerns about practicality. For example, the BLS correctly notes

26. In particular, the Moulton-Moses discussion would lead us to reduce our estimates of quality bias in the following CPI categories: fruits and vegetables, shelter, apparel, and motor fuel. However, there is no criticism in the Moulton-Moses paper that would cause us to reduce our estimates for those categories where our bias estimates were the largest—consumer appliances and electronics, medical care, and pharmaceuticals.

27. Recent evidence that we may have underestimated the biases in some of the areas we did examine comes from an alternative measure of consumer prices, the personal consumption expenditures (PCE) deflator, which rose at 0.94 percentage point less per annum than the CPI over the period 1992–98 (the respective growth rates were 1.56 and 2.50 percent). An unpublished examination of this difference by the BLS indicates that most of it arises from the use by the BEA of alternative price indexes for hospital expenditures, personal computers, and airfares. There is some overlap between the commission's bias estimates and the PCE-CPI difference, since the BEA uses measures for hospitals and personal computers that eliminate part of the quality bias estimate, perhaps 0.2 percentage point.

that while it is undoubtedly true that much quality change in health care has gone uncorrected in the CPI, it is difficult to quantify “what’s new in surgery this month.” But numerous studies have appeared (some are surveyed below) on a wide variety of medical areas from cataracts to generic drugs to depression to heart attack treatments. While it may be some time before procedures can be developed for implementing various quality adjustments and the consumer surplus from new products into a monthly CPI and perhaps never for some categories, there is no reason why the BLS cannot incorporate such information into a continually revised annual index.

Such an annual index has a number of benefits since it would be substantially more accurate than an uncorrected monthly index. Indeed, the only advantage of a monthly CPI is that it is timely, and to the extent it is desirable for contractual purposes, never revised. Those who are using the CPI for indexing purposes could decide if they wanted a more accurate measure that was not quite as up to date, or a far less accurate but quicker one.

The BLS has been quite critical of generalizing research results from one category of goods to other, even quite similar categories. A serious issue arises as to whether quality adjustment can be done individual commodity by individual commodity, and at what level of aggregation. The BLS takes the position that it is only for very specific commodities, when extensive research is done with a procedure that can be implemented mechanically in a monthly index, that they will make explicit quality adjustments. This is unfortunate. The major problem is that even with substantially augmented resources, by the time the BLS does a substantial amount of new research on quality change and decides whether to do something for a particular commodity, many of the ones they studied several years ago will have changed and perhaps been replaced by new products. It would be preferable to develop reasonable rules of thumb for when it is permissible to generalize from closely related categories, for example, those for which many of the inputs are common (e.g., microprocessors).

The BEA of the Commerce Department has moved to a Fisher ideal index for its measures of real output and prices, as have most other statistical agencies around the world. Since quantity information is not usually available contemporaneously, this immediately raises the issue of the need to have lags in the system. In our final report, we suggested that one approach would be to use a trailing Törnqvist index with the data continually updated. This was meant to be illustrative, not definitive, and we remain open to other possibilities. In fact, Shapiro and Wilcox (1997) have demonstrated that a very good approximation can be obtained with a one-parameter CES function constructed in the same real time frame as the current CPI.

VIII. Recent Changes in the CPI

Despite the initial set of critical comments, the most important of which were published in the previously cited set of articles authored by BLS officials, the BLS has moved with surprising speed to implement some of the commission’s recommendations. At the same time, the BLS has introduced other changes that were planned previously to the release of the commission’s report and for which the commission

can take no credit. Here we list the eight changes that the CPI has introduced since the December 1996 release of the commission's report.²⁸

A. Geometric Weighting

Effective with data for January 1999, the BLS introduced the geometric mean estimator for index categories that comprise approximately 61 percent of total consumer spending in the CPI, and this was expected to reduce the rate of increase of the CPI by about 0.2 percentage point (Dalton, Greenlees, and Stewart [1998]). This was the most important change in the CPI that could be interpreted as a response to the commission's recommendations.

B. More Rapid Change in Upper-Level Weights

In another major change that appears to be a response to the commission's recommendations, the BLS plans to change upper-level weights much more rapidly than in the past (see U.S. BLS [1999]). Eleven years elapsed between the initial use of 1982–84 weights in 1987 and the switch to 1993–95 weights in 1998. Now the plan is to switch to 1999–2000 weights in January 2002, and to change weights every two years thereafter, e.g., to 2001–2002 weights in January 2004. Not only is the shift in weights greatly speeded up, but there is a shorter lag in implementing the weights. This shift should largely eliminate upper-level substitution bias.

C. Change from Area- to Item-Based Sample Rotation Procedures

In a change planned before release of the commission's report, the CPI switched its sample rotation methodology in 1998 (Cage [1996]). It shifted its POPS from time-consuming in-person visits to computer-assisted telephone surveys. This allows the BLS to break out of the five-year rotation sample approach and to rotate items and outlets more rapidly and flexibly; types of outlets that tend to go out of business frequently can be sampled much more frequently than types that are very stable, such as electric utilities. Telephone surveying allows for an increase in sample size and focus on specific item categories where products turn over rapidly and where new products are frequently introduced. "Ultimately, this allows the bureau to rotate items and outlets that make up the CPI sample on a rolling, quarterly basis in every area, as opposed to rotating all items and all outlets on an annual basis in only a select group of areas. Without question, this will give the bureau the capability to introduce new products into the CPI sample more rapidly" (Cage [1996]). This promising methodological improvement directly addresses the commission's concern over the long lag in the introduction of new products even though by its timing it cannot be interpreted as a response to the commission's recommendations.

28. The list of changes comes from background material supplied to commission members with a recent General Accounting Office (GAO) survey regarding recent CPI measurement changes. I omit one change as unimportant (a change in the procedure for sampling housing) and combine two into one, regarding the more rapid updating of expenditure weights.

D. Changes in the Methods for Pricing of Hospital Services

In another change planned before the release of the commission's report and implemented in January 1997, the BLS has improved its procedures for pricing hospital services (Cardenas [1996]). Instead of the old approach, which was a straightforward input cost index that did not reflect shifts in the use of inputs (e.g., shorter hospital stays for a given procedure or a shift from inpatient to outpatient treatment), the new methodology obtains prices for a sample of specified treatments for particular diseases, rather than for a day in the hospital. This approach was introduced into the PPI in 1992 and extended to the CPI in 1997.²⁹ During the interim, the PPI for hospital services increased at 2.0 to 2.5 percentage points less than the equivalent CPI index, not far from the commission's estimate of a 3.0 percentage point upward bias in medical care services (the commission also made an allowance for the effects of improved technology that are not taken into account in the new CPI methodology).

E. Adjustment of Personal Computer Prices

Again following in the footsteps of the PPI, which had used the hedonic regression technique to adjust personal computer prices as long ago as 1991, the CPI adopted the hedonic regression approach for personal computers in 1998. This will eliminate part of the commission's quality bias estimate.

F. Adjustment of Television Prices

Similarly, the CPI introduced a hedonic regression approach to adjusting television set prices, beginning in January 1999. This technique was introduced following the research of Moulton, LaFleur, and Moses (1998).

G. Treatment of Utility Refunds

Starting in January 1999, the CPI no longer makes an adjustment for utility refunds that occur when utility companies or regulators determine that customers have been overcharged in the past. The elimination of refund adjustments is designed to make this component of the CPI less volatile but will have the undesirable effect of causing the CPI to overstate the secular rise in utility prices. This change constitutes an excellent example of the need for a second, research-based annual index that can be revised. Ideally, utility refunds based on past overcharges should be credited back to the period when the overcharges originally occurred.

H. Treating Mandated Pollution Control Measures as Price Increases

As discussed in Fixler (1998), the BLS switched in January 1999 to treat changes in vehicle or motor fuel characteristics arising from air pollution mandates as a change in price rather than quality. This reverses a BLS policy in effect since 1971 and follows from a recommendation in the commission's report. Since most of the changes in automobile technology introduced to reduce air pollution occurred during the 1970s and 1980s, this change will have little effect moving forward into the

.....
29. This shift in methodology is discussed in Triplett (1999).

future. This change provides another example of the need for a second, annual research-based index that can be revised into the past.³⁰

The overall impact of these changes is to reduce significantly the upward bias in the CPI below the commission's estimate of 1.1 percentage points that applied to the CPI during the 1995–96 period, before any of these changes were put into effect. By shifting upper-level weights much more rapidly, by shifting to geometric means at the lower level, and by improving its quality adjustments for medical care, personal computers, and TV sets, and by introducing a new procedure for sample rotation that allows the earlier introduction of new products, the commission's estimate of bias may be cut in half. Gone would be most of the 0.4 percentage point upper- and lower-level substitution bias, and perhaps 0.2 to 0.3 percentage point of the quality change bias. Remaining would be possibly a small residual of upper-level substitution bias, say 0.05 percentage point, the 0.1 percentage point for outlet substitution bias, and perhaps 0.3 to 0.4 percentage point of the quality change and new product bias, for a total of 0.45 to 0.55 of the original 1.1 percentage points.

IX. Recent Research on Quality Change

Most of the recent research on quality change has focused on medical care, including pharmaceuticals, and hence most of this brief survey is devoted to those categories. Subsequently we will cite scattered studies on other topics.

A. Medical Care and Pharmaceuticals

Most of the research on medical care and pharmaceuticals has been conducted by academic economists affiliated with the National Bureau of Economic Research. Some of this research has recently been brought together in a new book edited by Triplett (1999), whose comprehensive and thoughtful introduction provides the best available introduction into the issues, problems, techniques, and results in this area. It is important to note that earlier versions of some of the same research were available to the Boskin Commission and were the basis of its judgment that the pre-1997 CPI contained an upward bias of 3.0 percent per year in medical care services and 2.0 percent in prescription drugs.

In a newer version of research that the commission had reviewed, Cutler, McClellan, and Newhouse (1999), the authors use information from hospital records to price heart attack treatments using the CPI's traditional input-cost approach and find an upward bias of 2.0 percent. Adding in the value of improved life expectancy from better treatment procedures, the bias rises to the range of 3.1–3.5 percent.³¹

30. Gordon (1990) exhibits a table showing the time series of BLS quality adjustments for new automobiles divided among safety, environmental mandates, and other factors. The environmental adjustments had the effect of reducing the rate of inflation of new auto prices by 1.22 percentage points per year over the period 1967–85.

31. See also the other references under Cutler, McClellan, and Newhouse (1999) in the list of references. Cutler, McClellan, Newhouse, and Remler (1999) contains larger bias estimates, with a synthetic index based on pre-1997 BLS methods rising at an annual rate of 3.3 percent for 1984–94 contrasted with a COL index measuring changes in the quality of life that declines at 1.7 percent per year, for a total difference of 5.0 percent per year.

A study that implies a much larger bias was carried out by Frank, Berndt, and Busch (1999) on treatments for mental depression. There is no component of the CPI that is directly comparable, but the medical services CPI rose more than 25 percent over the 1991–95 interval for which the authors provide estimates of price declines in the treatment of depression ranging from 22 to 30 percent per year, implying a possible bias of more than 10 percent per year. In a new study of pharmaceuticals, Ellison and Hellerstein (1999) analyze a large data set on the prices of the cephalosporin class of antibiotics and find a price increase for 1988–96 of 0.76 percent per year compared to the PPI cephalosporin component that rose at 4.54 percent a year, for a bias estimate of 3.78 percent per year.³²

The number of aspects of medical care subject to research has been growing, and it seems reasonable to extrapolate from areas where studies have occurred to those that may have similar characteristics. Shapiro, Shapiro, and Wilcox (1999) point to the numerous dimensions of cataract surgery. An operation that once required a week in the hospital is now a brief outpatient procedure. Subsequent recovery is much faster; complication rates have declined; intraocular lenses have replaced cumbersome cataract spectacles. As a result of improved technology and a drastic decline in prices, the rate of cataract surgery among individuals in the United States aged 65 years has increased by almost a factor of four. Because the operation now occurs earlier in the disease, the period of steadily obscured vision has been eliminated, with enormous benefits in welfare. The authors point to similar benefits in angioplasty, joint replacement, and laparoscopic removal of gall bladders.

In looking back to the commission's estimate of a 3.0 percentage point per year upward bias in the CPI for medical care services, fully 2.0 percentage points are accounted for by the change from input costs to a treatment basis, introduced into the PPI in 1992 and the CPI in 1997. This leaves only 1.0 percentage point for all remaining improvements in medical care technology, and if anything, the recent studies suggest that medical care technology improvements may be reducing the true price of medical care by more than 1.0 percentage point per year.

B. Other Products

Scattered pieces of new research have emerged for other products. Moulton *et al.* (1998) concluded that there was an upward bias in the CPI for television sets of 3 to 5 percent over the period 1993–97. Ohashi (1999) developed hedonic price indexes for VCRs during the first decade of their introduction (1978–87) prior to their introduction into the CPI and found an average rate of price decline of 12 percent per year over that period. Hausman (1998) criticizes the Boskin Commission for understating the CPI bias related to cellular phones, which only appeared in the CPI in 1998, by which time there were 55 million cellular phone subscribers in the United States. Hausman estimates that the price of cellular phone service declined by 51 percent between 1985 and 1997, when only the straight cost of phone service and phone instruments is included. When he constructs a broader COL index that takes account of the welfare benefits from cellular phones, he emerges with an estimate

.....
32. Related research available to the commission was Griliches-Cockburn (1994) and Berndt *et al.* (1996).

that the CPI for telephone service (taking account of the gradually increasing weight on cell phones) is biased upward for 1985–97 by between 0.8 and 1.9 percentage points per year.

X. Conclusion

Partly as a result of its prompt response to the Boskin Commission recommendations and partly as a result of plans already in place at the time of the commission's deliberations, the BLS has implemented or announced plans to implement improvements in CPI measurement that eliminate roughly half of the 1.1 percentage point per annum upward bias estimated by the commission for the time period 1995–96. The greatest amount of progress has been made in eliminating substitution bias, both at the upper level and lower level, although the thorny issue of outlet substitution bias remains untouched. Several specific improvements address part of the commission's estimated upward bias involving quality change and new products, including new hedonic indexes for television sets and personal computers as well as an improved treatment-based methodology for measuring medical care prices.

The commission's report received a heavy dose of initial criticism. But in retrospect, much of that was directed at the political and redistributive implications for tax and Social Security benefit indexation than the scientific basis for its estimate of the bias in the CPI. Considering the fact that the commission operated without a budget for new research studies and was a part-time activity for its members, its report and conclusions have survived relatively well since its issuance in December 1996. Some criticism of its quality change estimates in several categories are valid and might shave 0.1 percentage point off its overall bias estimate, but other research emerges that points in the other direction. Its bias estimates for medical care, pharmaceuticals, and telephone service may have been too small. Further, there is the nagging discrepancy between the CPI and the PCE deflator, which grew at a rate 0.9 percentage point slower than the CPI between 1992 and 1998, indicating the possibility of a large potential bias in categories that the PCE deflator and CPI treated differently during that interval.

The debate surrounding the Boskin Commission report and its aftermath may seem quite insular in the view of measurement experts based in other countries. Virtually none of the research cited by the commission or in this paper has made any reference at all to price behavior in other countries or attempted any kind of comparative exercise in the size of measurement bias across countries. The Boskin Commission has stimulated related investigations of CPI bias in Germany, Japan, Switzerland, and perhaps other countries, but a survey of those results is beyond the scope of this paper. It is unfortunate but perhaps inevitable that foreign experts know far more about the measurement debate within the United States than we know about measurement issues relevant for other countries.

One fortunate side effect of the Boskin Commission report has been not only the rapid pace of change at the BLS but also a new attention to research issues raised by measurement bias. The BLS may benefit from the commission's aftermath by

receiving increased resources to launch a more ambitious research program. Both the BEA and BLS have advertised the availability of fellowships to graduate students and young scholars willing to do new research on measurement issues, and this hopefully will bring fresh blood into the now lamentably underpopulated research area of price and productivity measurement.

References

- Abraham, Katharine G., "The CPI Commission: Discussion," *American Economic Review*, 87 (1), May 1997a, pp. 94–95.
- , "Testimony on the CPI," Testimony before the House Budget Committee, March 7, 1997b.
- , John S. Greenlees, and Brent R. Moulton, "Working to Improve the Consumer Price Index," *Journal of Economic Perspectives*, 12 (1), Winter 1998, pp. 27–36.
- Aizcorbe, Ana M., and Patrick C. Jackman, "The Commodity Substitution Effect in CPI Data, 1982–1991," *Monthly Labor Review*, 116 (12), December 1993, pp. 25–33.
- Berndt, Ernst R., Iain Cockburn, and Zvi Griliches, "Pharmaceutical Innovations and Market Dynamics: Tracking Effects of Price Indexes for Anti-Depressant Drugs," *Brookings Papers on Economic Activity*, 1996.
- Boskin, Michael J., E. Dulberger, R. Gordon, Z. Griliches, and D. Jorgenson, "Toward a More Accurate Measure of the Cost of Living," Final Report to the Senate Finance Committee, December 4, 1996.
- , ———, ———, ———, and ———, "Consumer Prices, the Consumer Price Index, and the Cost of Living," *Journal of Economic Perspectives*, 12 (1), Winter 1998, pp. 3–26.
- Braithwait, Steven D., "Substitution Bias of the Laspeyres Price Index: An Analysis Using Estimated Cost-of-Living Indexes," *American Economic Review*, 70 (1), March 1980, pp. 64–77.
- Cage, Robert, "New Methodology for Selecting CPI Outlet Samples," *Monthly Labor Review*, 119 (12), December 1996, pp. 49–61.
- Cardenas, Elaine M., "Revision of the CPI Hospital Services Component," *Monthly Labor Review*, 119 (12), December 1996, pp. 40–48.
- Cutler, David, and Mark McClellan, "The Determinants of Technological Change in Heart Attack Treatment," *NBER Working Paper* No. 5751, National Bureau of Economic Research, September 1996.
- , ———, and Joseph P. Newhouse, "The Costs and Benefits of Intensive Treatment for Cardiovascular Disease," in Jack Triplett, ed. *Measuring the Prices of Medical Treatments*, Washington, D.C.: Brookings, 1999, pp. 34–71.
- , ———, ———, and Dahlia Remler, "Are Medical Prices Declining?" *NBER Working Paper* No. 5750, National Bureau of Economic Research, September 1996.
- , ———, ———, and ———, "Pricing Heart Attack Treatments," *NBER Working Paper* No. 7089, National Bureau of Economic Research, April 1999.
- Dalton, Kenneth V., John S. Greenlees, and Kenneth J. Stewart, "Incorporating a Geometric Mean Formula into the CPI," *Monthly Labor Review*, 121 (10), October 1998, pp. 3–7.
- Diewert, W. Erwin, "Exact and Superlative Index Numbers," *Journal of Econometrics*, 4 (2), May 1976, pp. 115–145.
- , "Axiomatic and Economic Approaches to Elementary Price Indexes," *NBER Working Paper* No. 5104, National Bureau of Economic Research, May 1995.
- , "Index Number Issues in the Consumer Price Index," *Journal of Economic Perspectives*, 12 (1), Winter 1998, pp. 47–58.
- Ellison, Sara Fisher, and Judith K. Hellerstein, "The Economics of Antibiotics: An Exploratory Study," in Jack Triplett, ed. *Measuring the Prices of Medical Treatments*, Washington, D.C.: Brookings, 1999, pp. 118–143.
- Fisher, Irving, *The Making of Index Numbers*, Boston: Houghton Mifflin, 1922.
- Fixler, Dennis, "Treatment of Mandated Pollution Control Measures in the CPI," *CPI Detailed Report*, September 1998, pp. 4–7.
- Frank, Richard G., Ernst R. Berndt, and Susan H. Busch, "Price Indexes for the Treatment of Depression," in Jack Triplett, ed. *Measuring the Prices of Medical Treatments*, Washington, D.C.: Brookings, 1999, pp. 72–117.
- Gordon, Robert J., *The Measurement of Durable Goods Prices*, Chicago: University of Chicago Press for National Bureau of Economic Research, 1990.

- Greenlees, John S., "Expenditure Weight Updates and Measured Inflation," unpublished, U.S. Bureau of Labor Statistics, 1997.
- Griliches, Zvi, "Hedonic Price Indexes and Quality Change," in George Stigler, ed. *The Price Statistics of the Federal Government: Report to the Office of Statistical Standards*, Bureau of the Budget. New York, National Bureau of Economic Research, 1961.
- , "Productivity, R&D, and the Data Constraint," *American Economic Review*, 84 (1), March 1994, pp. 1–20.
- , and Iain Cockburn, "Generics and New Goods in Pharmaceutical Price Indexes," *American Economic Review*, 84 (5), December 1994, pp. 1213–1232.
- Hausman, Jerry, "Valuation of New Goods Under Perfect and Imperfect Competition," in T. Bresnahan and Robert J. Gordon, eds. *The Economics of New Goods*, Chicago: University of Chicago Press, 1997a.
- , "Comment on CPI Commission Report," Papers and Proceedings, *American Economic Review*, 87 (2), May 1997b, pp. 96–98.
- , "Cellular Telephone, New Products and the CPI," revised draft, May 1998, *Journal of Business and Economic Statistics* (forthcoming).
- Hicks, John R., "The Valuation of the Social Income," *Economica*, 7 (26), May 1940, pp. 105–124.
- Jorgenson, Dale W., and Daniel T. Slesnick, "Individual and Social Cost-of-Living Indexes," in W. E. Diewert and C. Montmarquette, eds. *Price Level Measurement*, Ottawa: Statistics Canada, 1983, pp. 241–336.
- Moulton, Brent R., "Bias in the Consumer Price Index: What is the Evidence?" *Journal of Economic Perspectives*, 10 (4), Fall 1996, pp. 159–177.
- , and Karin E. Moses, "Addressing the Quality Change Issue in the Consumer Price Index," *Brookings Papers on Economic Activity*, 28 (1), 1997, pp. 305–349.
- , Timothy J. LaFleur, and Karin E. Moses, "Research on Improved Quality Adjustment in the CPI: The Case of Televisions," presented to the Conference of the Ottawa Group, April 1998.
- Nordhaus, William D., "Do Real-Output and Real-Wage Measures Capture Reality? The History of Light Suggests Not," in T. Bresnahan and R. J. Gordon, eds. *The Economics of New Goods*, Chicago: University of Chicago Press, 1997, pp. 29–66.
- , "Quality Change in Price Indexes," *Journal of Economic Perspectives*, 12 (1), Winter 1998, pp. 59–68.
- Ohashi, Hiroshi, "Quality-Adjusted Price Indexes for Home Video Cassette Records in the US, 1978–87," presented at National Bureau of Economic Research, Summer Institute, July 21, 1999.
- Pollak, Robert A., "The Social Cost of Living Index," *Journal of Public Economics*, 15 (3), June 1981, pp. 311–336.
- Shapiro, Irving, Matthew D. Shapiro, and David W. Wilcox, "Quality Improvement in Health Care: A Framework for Price and Output Measurement," *NBER Working Paper No. 6971*, National Bureau of Economic Research, February 1999.
- Shapiro, Matthew D., and David W. Wilcox, "Mismeasurement in the Consumer Price Index: An Evaluation," *NBER Macroeconomics Annual 1996*.
- , and ———, "Alternative Strategies for Aggregating Prices in the CPI," *NBER Working Paper No. 5980*, National Bureau of Economic Research, March 1997.
- Slifman, Leonard, and Carol Corrado, "Decomposition of Productivity and Limit Costs," *Occasional Staff Studies*, Federal Reserve Board, November 18, 1996.
- Stewart, Kenneth J., and S. B. Reed, "CPI Research Series Using Current Methods, 1978–98," *Monthly Labor Review*, 122 (6), June 1999, pp. 29–38.
- Stigler, George, ed. *The Price Statistics of the Federal Government: Report to the Office of Statistical Standards*, Bureau of the Budget. New York, National Bureau of Economic Research, 1961.
- Törnqvist, Leo, "The Bank of Finland's Consumption Price Index," *Bank of Finland Monthly Bulletin*, 10, 1936, pp. 1–8.
- Triplett, Jack, ed. *Measuring the Prices of Medical Treatments*, Washington, D.C.: Brookings, 1999.

- U.S. Bureau of Labor Statistics, "Measurement Issues in the Consumer Price Index," *June 1997, Response to the U.S. Congress*, Joint Economic Committee, 1997a.
- , "Using a Hedonic Model in the Consumer Price Index to Adjust Personal Computer Prices for Changes in Quality," *CPI Detailed Report*, June 1997, p. 18, 1997b.
- , "Scheduled Updates for Expenditure Weights in the Consumer Price Index," *CPI Detailed Report*, February 1999, pp. 5–6.
- U.S. Department of Labor, Bureau of Labor Statistics, *BLS Handbook of Methods*, Bulletin 2414, Washington, D.C.: U.S. Government Printing Office, 1992.