“Better Living Through Economics: Consumer Price Indexes”

By

Michael J. Boskin
T.M. Friedman Professor of Economics
and Hoover Institution Senior Fellow
Stanford University

Prepared for presentation at:

American Economics Association
Annual Meetings
New Orleans, Louisiana
January, 2008

Better Living Through Economics:
Consumer Price Indexes[1]

Michael J. Boskin

T.M. Freedman Professor of Economics and
Senior Fellow, Hoover Institution
Stanford University

Introduction

Few areas of economics research have produced as consequential societal benefits in recent decades as price index research. This research has changed how inflation, real GDP and productivity are measured, thus providing more accurate information on these key economic indicators for private investors, workers, managers and public policy-makers. That has led to more accurate indexing of government benefit programs and taxes and thus to a lower national debt. These improvements literally change (the interpretation of some) history, for example revealing that real wages and median income in recent decades have grown, albeit more slowly, than previously, not stagnated. And cumulatively the research adds greatly, not just to improved measurement of inflation but to economists' general understanding of the dynamic of economic growth. For example, it details the importance of new products, improvement in the quality of existing products and other innovations in generating rising living standards. Still, there is much more improvement to come in price measurement and in research on it.

[1] This paper draws heavily from Boskin (2005; 2008). Thanks to Jerry Hausman and John Siegfried for helpful suggestions.

Measuring prices and their rate of change accurately is central to almost every public and private economic issue and discussion, including the conduct of monetary policy, measuring economic progress over time and across countries, and the cost and structure of indexed government spending programs and taxes. Most of us are familiar with the prices of many things we purchase. We know what we paid recently for a pound of ground beef or a quart of milk. Renters know how much they pay in rent. Measuring prices, therefore, may seem simple and straightforward, but it's not.

The purpose of a price index is to summarize information on the prices of multiple goods and services over time. Consumer spending accounts for about two thirds of U.S gross domestic product (GDP). The Consumer Price Index (CPI) and the Personal Consumption Expenditure deflator (PCE) are designed to summarize the prices of goods purchased by consumers over time. In a hypothetical society with only one good, say one type of food, we would not need a price index; we would just follow the price of the one good. When there are many goods and services, however, we need a method for averaging the price changes or aggregating the information on the many different prices. The rate of change of prices—inflation—is important in both macro- and microeconomics. Estimating inflation and real growth, for example, requires measures of price changes, and in a flexible, dynamic modern market economy, obtaining accurate measures is complicated A single large superstore may contain 50,000 separately priced items. Within that individual store, new items are continually introduced and old items discontinued. The quality of many items improves in some objective way—greater energy efficiency, more durability, less maintenance, to name a

few. Of course, many more items claim to have improved. When quality really does improve but the price stays the same, the real price has fallen. How to summarize what happened to prices in just one store over a period as short as one month is complicated, even with modern scanner technology. Doing so for the entire economy is vastly more complex.

To obtain information on various prices requires us not only to measure the prices, but also to weight the various components in the index. Weighting each price change equally would be simple but not very revealing. For example, if the price of red delicious apples fell by 5% and rent rose by 5%, such an index would suggest that there had been no change in the overall price level. But that would be silly. We need to "weight" goods on which consumers spend more of their income more heavily than those on which they spend less. Hence we need data on expenditures and/or quantities of goods purchased.

The United States CPI and the Cost of Living

When economists try to measure the "true" inflation rate—the rate of change of prices—it is to answer the question, "How much more income would consumers need to be just as well off with a new set of prices as the old?" Thus, a cost-of-living concept is at the core of proper measures of prices and of changes in prices. This clearly involves tracking how consumers respond to changes in the relative prices of various goods. Their response to such changes is called "substitution." It also requires measuring quality-adjusted prices and the improvement in well being from new goods. One would

not want to count as “inflation” a major improvement in quality that resulted in a tiny price increase.

Most traditional consumer price indexes, including the CPI in the United States, measure prices with a mostly fixed-weight system, holding constant the expenditure weights from some base period. Table 1 reports the most recent weights on very broad categories of goods from 2006; the Bureau of Labor Statistics (BLS) derives these weights from expenditure surveys that report how much consumers spent on different types of goods and services. For example, at a very broad level of aggregation, those weights are 15.0% for food, 6.3% for medical care, 42.7 for housing, 17.2% for transportation, etc. Within each category, of course, are thousands of specific goods; for example, red delicious apples of a certain size and quality are a component of the apples sub-category, which is a component of fresh fruits, which, in turn, is a component of fresh fruits and vegetables, which is a component of food at home, etc.

Table 1
Relative Importance of Components in the Consumer Price Index (CPI-U)

Food and beverages	15.0
At home	7.9
Away from home	6.0
Alcoholic beverages	1.1
Housing (including utilities)	42.7
Apparel and services	3.7
Transportation	17.2
Vehicles	7.6
Gasoline	4.3
Other	5.3
Medical care	6.3
Recreation	5.6
Education and communication	6.0
Education	3.1
Communication	3.0
Other	3.5
Total	100.0

source: BLS, Relative Importance of Components in the Consumer Price Index, December 2006
n.b. individual items may not add to totals because of rounding

With these expenditure weights at hand, or alternatively quantities, since expenditures is just price times quantity, it takes a high-quality operation to track the prices accurately. And whose prices? For commodities purchased where, when and how? In the U.S., the answer is that there are several closely related consumer price indexes. One measures the change in a weighted average of consumer prices, with the base year expenditure weights, for a typical urban family, the so-called “CPI-U. A second related, but not quite identical, construct is the ”CPI-W,” the analogous measure

for urban wage and clerical workers. I focus here on the more widely cited CPI-U. Neither of these fixed-weight indexes accounts for substitution, the fact that consumers substitute away from goods whose prices increase more and towards goods whose prices increase less[2].

The CPI serves, and should serve, many purposes. For example, the CPI is used to measure consumer inflation on a monthly basis; to make cost-of-living adjustments in Social Security, income tax brackets and other government programs; to provide price data as inputs to the National Income and Product Accounts and the third price index, the Commerce Department's personal consumption expenditure (PCE) deflator, which does account for substitution.

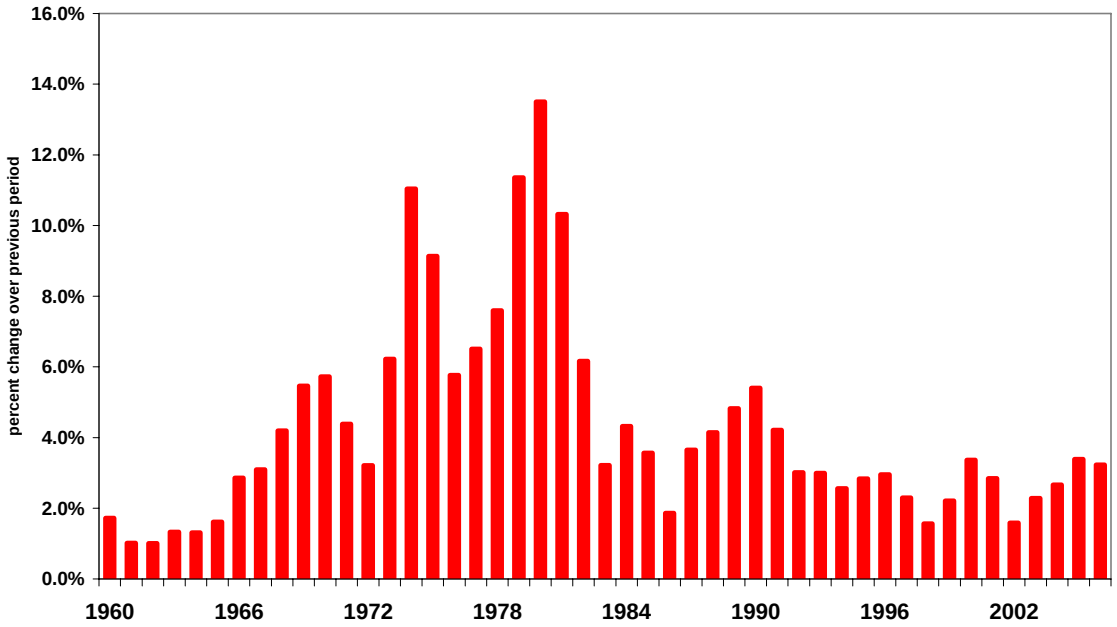
Figure 1 provides recent data on the U.S. CPI-U. The CPI-U sets the index=100 for the years 1982-84. As can be seen, the pace of measured consumer inflation has slowed considerably relative to the 1970s and 80s, has recently been running in low single digits, and has had considerably less variation than in the high-inflation 1970s and early 1980s.

People change their spending patterns over time, and do so specifically in response to changes in relative prices. When the price of chicken increases, for example, people may buy more fish, and conversely. Hence the weights change, and a

[2] A recent improvement by the BLS substitutes geometric for arithmetic mean formulas for aggregating at the lower levels for about 60% of items, thus allowing for some partial substitution.

price index such as the fixed-weight base period CPI, that fails to account for that, overstates the true change in the cost of living.

Figure 1
Percentage Change in U.S. CPI-U



There are two obvious approaches to weighting the prices. The first uses a fixed-base period weighting: quantity or expenditure weights remain fixed at their base period levels, and then we see what happens to the weighted average of prices as prices subsequently change. An alternative possibility, of course, is to use the expenditure weights or quantities in the second period, after the substitution. Economic theory strongly supports the idea of taking an average of these two numbers, a point originally made by the great American economist Irving Fisher[3]. For several years now, the BLS has computed a closely related measure called the chained-CPI, or C-CPI-U

[3] Fisher, Irving, *The Making of Index Numbers*, 1922

(Cage, Greenlees and Stewart); it has been rising much less rapidly than the traditional CPI-U, suggesting that the failure to account for consumer substitution explicitly is a serious weakness of the official CPI. ***[Figure 2]***

Similarly, *where* people make their purchases changes over time. Discount stores and on-line sales have become more important relative to traditional small retailers. Because price data are collected within outlets, the shift of consumer purchasing from discounters does not show up as a price decline, even though consumers reveal by their purchases that the price decline more than compensates for the potential loss of personal services. Thus, in addition to substitution bias among commodities, there is an outlet substitution bias. (See discussion in Boskin, et al. and Hausman and Leidtag, 2004).

Even *when* purchases are made can become important. We typically measure prices monthly, during a particular week. But if, for example, consumers get wise to post-Christmas discounts and start buying a lot more holiday items after Christmas, surveys that look solely at prices in the second week of December will miss this change in buying patterns and misstate inflation.

Another problem is that price data tend to be collected during the week. In the United States, about 1% of price quotes are collected on weekends, despite the fact that an increasing share of purchases is made on weekends and holidays (probably reflecting the increase in prevalence of two-earner couples). Because some outlets

emphasize weekend sales, there may be another “when” bias, as well as a “what” and a “where” bias. This phenomenon may explain, in part, recent research suggesting that prices rise less rapidly in data collected by scanners on actual transactions rather than by the BLS employees gathering data on prices on shelves and racks. (See Silver and Heravi)

Finally, additional biases result from the difficulty of adjusting fully for quality change and the introduction of new products (Bresnahan and Gordon). In the U.S. CPI, for example, VCRs, microwave ovens, and personal computers were included a decade or more after they had penetrated the market, by which time their prices had already fallen 80% or more. Cellular telephones were not included in the U.S. CPI until 1998, although they were introduced in the United States in 1984 (Hausman 1999).

The CPI currently overstates inflation by about 0.8 percentage point: 0.3 to 0.4 is attributable to failing to account for substitution among goods; 0.1 for failing to account for substitution among retail outlets; 0.3-0.4 for failing to account adequately for new products and quality improvements. (Lebow and Rudd). Thus, the first 0.8 percentage point or so of measured CPI inflation is not really inflation at all. This may seem small but the bias, if left uncorrected for, say, twenty years, would cause the change in the cost of living to be overstated by about 20 percent.

The United States CPI is one of the few economic statistics that is never revised, even if subsequent data reveal that the published statistic is wrong. This was done

historically because many contracts and other government programs are expressly indexed or adjusted to the CPI, and revisions would cause practical and legal complexities.

We know that different sets of consumers have different expenditure weights because they spend different fractions of their income on the various commodities: renters vs. homeowners; the middle aged vs. the elderly, etc. Interestingly, most analyses find only modest differences in inflation rates across groups with different expenditure weights, although even modest differences can be consequential when compounded over time.

What about differences across groups in prices and rates of change of prices? For example, do the prices paid by the elderly differ from those paid by the general population, and if they do differ, have the differences changed over time? Economic theory suggests the prices will not differ much for most items, but we do not have serious empirical evidence on this score.

Thus inflation—the rate of change of prices—is hard to measure accurately. Government statisticians in all countries, especially those at the U.S. Bureau of Labor Statistics, have made numerous important improvements over the years. (Stewart and Reed). Yet new products are introduced all the time, existing ones are improved, and other products leave the market. Relative prices of various goods and services change frequently, for example, in response to technological and other factors affecting cost

and quality, causing consumers to change their buying patterns. Literally hundreds of thousands of goods and services are available in rich, industrialized economies. As we have become richer, our demands have shifted to services away from goods, and to characteristics of goods and services such as enhanced quality, more variety, and greater convenience. Technology and entrepreneurship provide them. But all these factors mean that a larger fraction of what is produced and consumed in an economy today is harder to measure than decades ago, when a larger fraction of economic activity consisted of easy-to-measure items such as tons of steel and bushels of wheat (although even these items have important quality dimensions). Thus, how to obtain information on who is buying what, where, when, why and how, in an economy, and then to aggregate it into one or a few measures of price change raises a host of complex analytical and practical problems.

Economics Research and Improving Price Measurement

Once considered very important for practicing economists, price index research declined for decades in relative importance in academic curricula. Fortunately, important research continued: Erwin Diewert and Bob Pollak on index number theory; Dale Jorgenson's theoretically consistent set of production, income, accumulation and wealth accounts; Zvi Griliches and Sherwin Rosen on hedonics; Jerry Hausman on new products, to name the most important contributions that have transformed the way economists think about these issues and, with a lag, how government statisticians measure prices[4].

[4] See Diewert (1993) for a discussion of the early history of price index research.

Important research was also done at the Bureau of Labor Statistics (BLS) – by Marshall Reinsdorff, Jack Triplett, Brent Moulton, Paul Armkeknecht and Patrick Jackman, to name a few – and some practical improvements (within the fixed-weight, called Laspeyres, context) were made to the CPI. Related important work occurred at the Bureau of Economic Analysis (BEA) and other statistical agencies worldwide. For example, sampling was introduced to the CPI in 1978 and the treatment of owner-occupied housing was changed to a rental equivalence basis in 1983. Public policy decisions were made to index major programs that comprised large fractions of both sides of the budget: Social Security benefits in the 1970s and income tax brackets in the 1980s. While I would argue these were valuable policy improvements, they had the side effect of costing government statisticians their innocence. Not surprisingly, occasional errors were made, both by statisticians and politicians, and then corrected. For example, the introduction of sampling – an enormous improvement – was accompanied by an unexpected upward bias of about 0.2 percent per year, subsequently discovered and corrected by BLS. Initially, Social Security benefits were doubly indexed by the political system; this too was corrected after a brief period.

While my own research and policy interests were in Public Economics and Macroeconomics, exposure to this and related research on these issues piqued my curiosity in economic statistics. From a cost-benefit analysis point of view, weren't we, as a profession and a society, underinvesting in economic statistics? From curricula, training, and professional rewards to the public funds devoted to government statistics, couldn't some quite valuable improvements be made at pretty low cost? But in the

1980s, growing pressure on the public fisc led to increasing constraints on financial resources for the statistical agencies.

When I became CEA chairman in 1989, I decided to see what could be done to improve economic statistics. Together with representatives of the statistical agencies, I developed an initiative to improve the quality of statistics. We looked at fundamental issues such as whether to centralize the disparate statistical agencies, whether some government activities should be performed in the private sector, etc. But our main product was a list of priority improvements in the generation of the statistics. President Bush agreed to make our proposals an initiative in his budget, and we sought funds from Congress. About 70% of the requested funds were appropriated, but different agencies were treated quite disparately because of separate appropriations subcommittees. Funding to improve the National Income and Product Accounts (NIPAs) was increased substantially for the first time in a decade.

My next involvement in economic statistics (other than helping get budgets passed) was in 1995-1996 as chairman of a commission on the CPI (henceforth BC, shorthand for the proclivity to name it after me). My colleagues—Ellen Dulberger, Bob Gordon, Zvi Griliches and Dale Jorgenson – and I were asked both to estimate the likely biases in the CPI and to recommend improvements to the BLS and Congress. We examined everything from data collection (we were each deputized and went on price collection site visits) to index number theory to practical issues (aided by extensive interaction with BLS personnel and others). Based on our evaluations of the research,

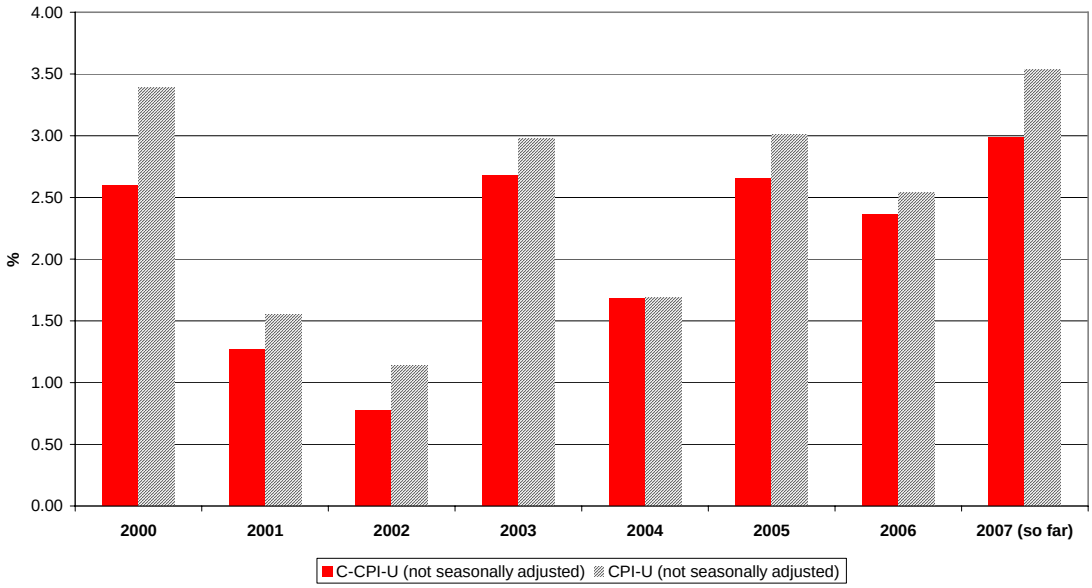
we made a dozen recommendations to the BLS, some of which have been adopted; four to our elected officials, some of which have been adopted; and one to the economics and statistics profession, which is making some but not enough progress. Remarkably, statistical agencies around the world have been using the Commission report as a major input to their own agendas for improvement. The real social value added certainly exceeded our opportunity costs by a large amount[5].

The Bureau of Labor Statistics (BLS), the agency of the Labor Department responsible for producing and disseminating the Consumer Price Index, has made considerable progress improving the CPI. Some of these improvements have been ongoing, others enacted or accelerated subsequent to the publication of the *Report of the Advisory Commission on the Consumer Price Index* (Boskin et al., 1996). In that report, my colleagues and I estimated that, taken as a measure of change in a true cost of living index[6], the change in the Consumer Price Index at that time was overstating inflation by slightly over one percentage point per annum., and had been doing so for some time. Of course, there was a range to our estimates, perhaps it was a little bit smaller or larger. Since that time, BLS has made several important improvements that reduced some measures of the bias, but, as it turns out, our estimates, based in part on Aizcorbe and Jackman, of one important component of the bias appears now to have been substantially underestimated according to the BLS' own subsequent calculations. The new BLS C-CPI-U series implies upper level substitution bias for the last five years

[5] Also see the papers in the symposium edited by Triplett, 2006.
[6] It is widely accepted that a cost-of-living index is the theoretically correct conceptual foundation of measures of inflation for input into monetary policy, making cost-of-living adjustments to government programs and tax rules and measuring economic progress. Indeed, it is difficult to conceive of any other conceptual basis.

was much larger than previously assumed (0.4% or more per year vs. 0.2% or less). Over the last year, (through October 2007) the C-CPI-U rose 3.0% vs. 3.5% for the CPI-U. Since its inception in December, 1999, the C-CPI-U has risen on average about 0.4% per year less rapidly than the regular CPI-U. Similar differences occur for the core C-CPI and core CPI, which remove volatile food and energy components. The data for 2000-2007 are shown in Figure 2.

Figure 2
A Comparison of C-CPI & CPI
(annual % change)

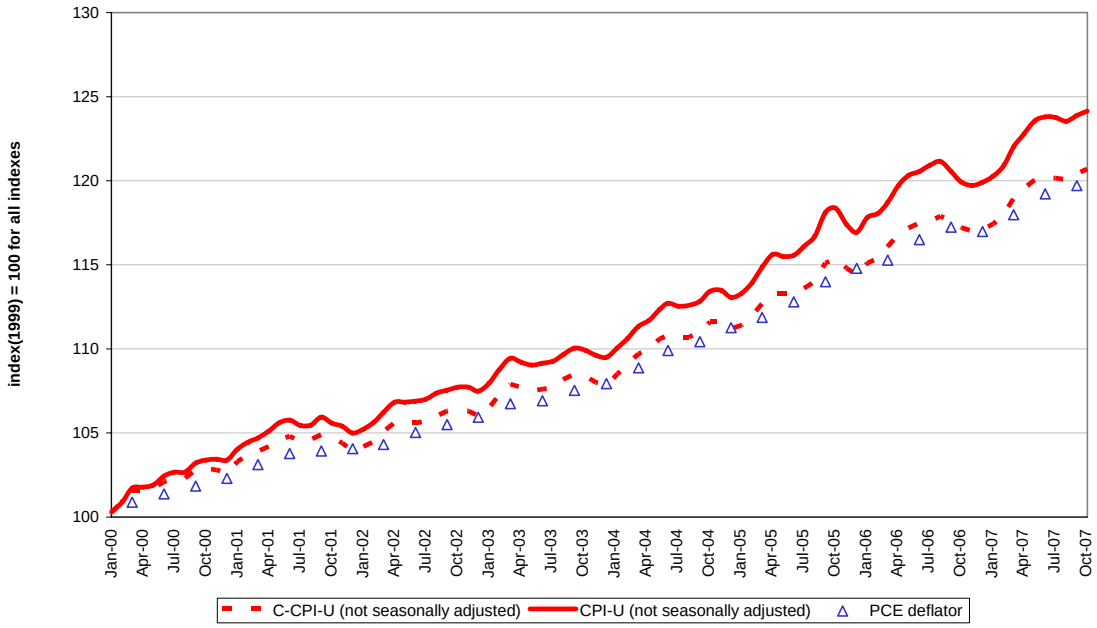


In the construct of the CPI, data on some 70+ thousand prices for separate items are collected and aggregated. How that is done is a complex and instructive, important subject in itself, but I refer the reader to Boskin et al. (1996) for a summary and to Hausman (2002) for important suggestions on improving the process. Upper-level substitution bias refers to the bias in the estimate of the true cost-of-living index created by the use of a fixed base-period weight price index, called “Laspeyres”, that ignores the

likelihood of consumer substitution in response to relative price changes. (There is also a lower level substitution bias before the CPI is aggregated to the upper level; but the BLS has made a substantial improvement by moving to geometric means for about 60% of the goods and services at the lower level (Dalton, Greenlees and Stewart)). So, for about 200 broader categories of goods and services, the BLS calculates the fixed-weight Laspeyres Index using base period expenditure weights obtained from consumer expenditure surveys.

That a fixed-weight index misses the likely consumer substitution is well known. What is startling is that, in part a response to our Commission's recommendation to calculate a chained Tornqvist Index – a formula that accounts for consumer substitution and falls into a category of indexes called "superlative indexes" because of their desirable properties [Diewert, 1976], as noted above – the BLS now calculates a chained CPI, called the C-CPI-U. This chained Tornqvist has been published for several years and has risen on annual average 43 basis points less than the regular CPI-U per year. Our original estimate of substitution bias was 0.2% or less, based on econometric studies at a much higher level of aggregation and some very preliminary BLS unpublished research. A difference of 43 basis points for years cumulates to a substantively important issue, to a total of 3.4% since 2000 [**Figure 3**]. Even if the larger estimate of upper-level substitution bias proved to be temporary, the ramifications for over-indexing government programs and tax brackets amount to "excess" debt of many billions of dollars annually.

Figure 3
CPI-U, C-CPI-U, and PCE Deflator



Just this year, that bias has implications of many billions of dollars of increased future outlays and decreased revenue for the federal government. The cumulative bias just from 2000-2007 will amount to many hundreds of billions of dollars over the next couple of decades, since the overstatement of inflation gets built into the baseline amounts of spending and revenue. The likely substantial future bias for years beyond 2007[7] will make this problem dramatically worse.

It is important to understand how pervasive the use of the Consumer Price Index is in our economy and, indeed, in academic economic research. As mentioned above,

[7] Of course, as mentioned above, there is no guarantee that the bias will remain this large; it could change in either direction, in response to relative price dispersion and the response of expenditures to that dispersion in the future.

other price indexes are available, some of which make important improvements in this dimension relative to the CPI. For example, the BEA's personal consumption expenditures deflator uses a Fisher Ideal Index, an alternative method of adjusting for consumer substitution. Unfortunately, but not to detract from that important effort, it is available only quarterly. Therefore, many people who need monthly data, and especially people who want to measure things on literally a calendar-year basis and therefore want to measure, for example, real stock returns over the year, need monthly data, especially December over December. Thus many researchers continue to rely on a quite faulty CPI, faulty from the standpoint of accurately measuring changes in the cost of living for the purposes they have in mind.

While many aspects of the CPI imply a likely upward bias, it is important to realize downward biases can and also do occur; for example, it appears that the shelter component of the CPI may well have been overstated historically (Gordon, 2003; Gordon and Van Goethen). Likewise, it is important that researchers are not just looking for upward biases when attempting to evaluate and improve the CPI, but for any bias, positive or negative. We want the most accurate price index, not the slowest growing one.

Likewise, 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, discount chains, and online retailing. The same VCR available for $150 in a local appliance store may be only $120 at Circuit City, or $115

online. Since price data are collected *within* outlets, the shift of consumers to purchasing from discounters does not show up as a price decline, even though consumers reveal by their purchases that the price decline more than compensates for the potential loss of personal services. Thus, in addition to substitution bias among commodities, there is an outlet substitution bias (see Hausman, [2004] for one example).

Understanding Economic Progress

As noted above, the original Boskin Commission 1996 estimate of the bias using changes in the CPI as a measure of the cost of living was about 1.1 percentage point per year. A technical problem called formula bias, introduced into the CPI in 1978 and corrected by the BLS early in 1996, added about 0.2 percentage points per year. Other improvements, net of the larger upper-level substitution bias, reduce the total bias to about 0.8-0.9% per year. To gain some appreciation for the potential importance and ramifications of the overstatements, suppose that the changes in the CPI have overstated changes in the cost of living by 1.1 percentage points from 1973-1995 (ignoring, to be conservative, the 0.2 percentage points attributed to formula bias from 1979-1995) and by 0.8 percentage points since. Note that it is not possible that these rates of bias in recent decades have been “permanent” for time spans of centuries, as back-casting such rates would lead to implausibly low standards of living in the distant past.

Table 2 presents estimates (updated from Boskin and Jorgenson [1997]) of commonly used measures of economic progress: real average hourly and weekly earnings and real median family income, as deflated by the official CPI and by an adjusted cost-of-living index that grows more slowly than the official CPI by the amount of the estimated bias. Instead of falling by about 8 % (hourly) or 15% (weekly), real average earnings have risen about 20% or more. Instead of growing a mere 10%, real median family income increased 48%, from 1973-2005. Clearly, the pace of improvement in living standards has slowed relative to the previous quarter century, but it has neither virtually stagnated nor declined, even relative to the early 1970s. These revised estimates accord much more closely with descriptions of living standards in terms of products and services consumed, including by those of modest means[8].

Table 2

Historical Implications Of Upward Bias In The CPI Measure of the Cost of Living, Selected Measures of Economic Progress*

Measure	CPI-U/CPI-U-RS	Adjusted COLI
Real Average Earnings (1973 - 2006)		
- Hourly	**-8.4%**	**30.5%**
- Weekly	**-15.9%**	**23.0%**
Real Median Income (1973 - 2006)	**15.7%**	**36.3%**

source: author's estimates

* Bias 1.1% 1973-95; 0.8%, 1996-2006, in CPI-U measure. Median income already deflated by CPI-U-RS, which grew more slowly than CPI-U by about 0.43% per year.

[8] See the related analysis over a longer period by Nordhaus and the analysis by Bils and Klenow.

Consumption accounts for approximately two-thirds of GDP. Taking 0.4% as a reasonable estimate of the bias flowing into the price data in the National Income and Product Accounts (NIPA), (after subtracting substitution bias, since the NIPA PCE deflator accounts for it) would imply an understatement of the growth of real GDP, as a result of understating the growth of real consumption expenditures of perhaps a quarter of a percentage point per year. There undoubtedly are additional and analogous problems raised in the investment-goods price measures, as well as those for government purchases and net exports. But even just relying on the consumption numbers, real GDP would be understated by perhaps a quarter percentage point per year. Even a quarter percentage point per year compounded over a long span of time accumulates to a sizable number. Real GDP, total and per capita, would be about 8 percent higher than the official statistics if such a bias held on average since 1973.[9]

There are numerous other conceptual and measurement issues involved in all of the measures of economic progress. Fringe benefits and bonus payments become more important through time, rendering nominal average hourly earnings less relevant than a broader measure of compensation. Compositional effects change the size, age, and other characteristics of the median family, etc.

[9] There are numerous other conceptual and measurement issues involved in all of the measures of economic progress. Fringe benefits and bonus payments become more important through time, rendering nominal average hourly earnings less relevant than a broader measure of compensation. Compositional effects change the size, age, and other characteristics of the median family, etc. Economists are also conducting research that will be helpful to improve, and interpret these statistics.

It is not just across time that economic progress or conditions are compared, but also across countries. It is important to understand that basic index number problems affect the statistics of every country. While some countries have made progress in ameliorating some of these problems (e.g., Statistics Canada with the movement to geometric means at the lowest level of aggregation removes some of the substitution bias), in other dimensions the American statistical system is far ahead of most other countries. For example, the widespread sampling done in the United States to collect price statistics undoubtedly gets some quality-change and new-product bias that is not picked up in other countries. Thus, the rates of real GDP growth, inflation, productivity growth, real wage growth, real family-income growth, and analogous measures are all misstated in virtually every country, to varying degrees, and these measurement issues and problems should be borne in mind when such comparisons are made. Fortunately, analogous economic research in many counties, in academia and central banks, will over time improve these statistics as well. It should also be borne in mind that there are numerous other important differences between the U.S. NIPAs and those of other countries.

Sometimes the mechanical use of flawed statistics can be inadvertently misleading when used as input to other data. For example, Table 3 presents historical data on real returns to stocks, bonds, and other financial instruments, as well as measured inflation, over several sub-periods for the last two centuries , taken from the invaluable work of Siegel (2002) and Morningstar (2006).

Table 3
Compound Annual Real Returns by Type of Instrument (1802-2006)

Period	Stock (arithmetic mean)	Stocks	Bonds	Bills	Gold	Inflation[1]	Equity premium
1802-2001*	8.4	6.9	3.5	2.9	0.0	1.4	3.4
1802-1870*		7.0	4.8	5.1	0.2	0.1	2.2
1871-1925*		6.6	3.7	3.2	-0.8	0.6	2.9
1926-2001*		6.9	2.2	0.7	0.4	3.1	4.7
1946-2001*		7.1	1.3	0.6	-0.3	4.1	5.8
1926-2006**		7.4	2.4	0.7	--	3.0	5.0

Sources: *Siegel 2002; **Morningstar, 2007

1 N.B. Inflation measured as change in the official CPI, and therefore overstated considerably, and real returns understated considerably, in recent decades

The long-run average of about 7% real returns to stock is impressively stable over two centuries of economic and demographic change – several depressions, numerous recessions, an almost-doubling of life expectancy, world wars, etc. Of course, these averages suppress the substantial short-run variation in real returns – the standard deviation of real stock returns is more than twice the mean. The long-run real return to government bonds averaged about 2.4%. The invaluable Siegel and Ibbotson measures are careful estimates of nominal returns; unfortunately, they overstate historical inflation and understate real returns because the change in the official consumer price index (CPI) is generally used as the measure of inflation. In contrast to the care in measuring nominal returns, the change in the official CPI is just assumed to measure inflation correctly. In recent decades, the change in the CPI has overstated inflation by about 1.0% per year (Boskin, et al., 1996; 1997; 1998; Lebow and Rudd)

because of the several types of bias discussed above. A more accurate inflation measure would thus substantially increase all the measures of real returns (for stocks and bonds), by about 100 basis points in recent decades[10]. Just using a CPI series consistent with current, as opposed to mid-1990s, BLS procedures would raise the estimate of real returns for recent decades by 40-50bp.[11]

Finance theory implies that, when investment comes out of retained earnings which produce a rate of return equal to the discount rate, the reciprocal of the P:E ratio equals the rate of return. The historical long-term P:E ratio of just under 15 fits well with the 7% real return estimates. However, if real returns in fact have been perhaps 8% for several decades; this raises several important issues for finance economists[12].

Priorities Moving Forward

A useful way to organize one's thoughts on the CPI is to ask the basic questions: What price and quantity of/or expenditure data should be collected? How? Where? How Often? Whose prices and quantities? What methodology should be used to aggregate them? How many indexes? For what purposes? Answering each of these questions involves a series of conceptual, methodological and practical considerations.

First, the CPI ought to approximate a true cost of living index (COLI), a position the BLS has endorsed. For most of the uses of the CPI, from making monetary policy to cost-of-living adjustments in government programs and tax rules to measuring

[10] Of course, it would correspondingly decrease the inflation estimate.
[11] See Stewart and Reed (1999)
[12] See the discussion of the use of historical data in Shoven (2001) and the other papers in that volume.

economic progress in many dimensions, it is hard to imagine not at least starting with a cost-of-living concept. Further, my own long interaction with the Executive and Legislative branch policy makers in this area suggests they did indeed have and still have a cost-of-living concept in mind in indexing the tax brackets and benefit programs. It would be appropriate to adopt a superlative index such as the chained CPI as the official CPI and to live with the need for revision. This would move the official CPI much closer to a true cost-of-living index. If our elected officials do not like the implications for indexing government programs, the indexing can easily be changed. Government statisticians should not be making policy judgments, and politicians should not be politicizing the production of government statistics.

Better measures of the user cost of housing and consumer durables should be a high priority. Leaving out real capital gains (and losses) in the user cost may avoid some practical oddities, but for a full evaluation of economic well-being, real capital gains need to be included. If my house goes up in real value, it will indeed cost me more (implicitly) to rent it to myself, but I may well be better off. If we leave the real capital gains out of the price index, we ought to think twice about that component of a COLA to a wealthy home-owner financed by payroll taxes on young renters! The recent paper by Bajari, et al. (2002) suggests that a conceptually proper treatment of housing can be quite consequential, and that current methods appear to include a substantial upward bias[13].

[13] Recall the discussion above about likely downward bias for other reasons historically in the shelter component of the CPI.

Next, it would be useful to greatly expand the use of hedonics to account for explicit quality change (as in Berndt, Griliches and Rapport; Berndt, Cockburn and Griliches; Chwelos, Berndt and Cockburn; White, Jarmin, Berndt and Monroe; and the analysis of Pakes; Diewert; and in Triplett). The BLS is in fact making some progress here.

Closely related to quality change is the issue of new goods. Historically, BLS has at times entered new products into the index very late, after their prices have fallen appreciably. An important recent example is cellular telephony. Since Hicks (1940), there has been a standard analytical economics treatment of new products in evaluating real income. Hausman (1997) has developed econometric techniques to implement the virtual price concept and demonstrated the failure to do so can lead to quite large biases. While the more frequent BLS updating of the market basket should help reduce the incidence of major long-lived omissions, it is not sufficient to deal with new products (Griliches and Cockburn; Hausman 1999; Petrin).

Greater use should be made of scanner data (Silver and Herari, plus early work by Reinsdorf), which despite their problems have the advantage of not only sample size but actual transaction prices as opposed to prices of goods on shelves, racks or other displays. For example, the widely reported growing tendency of consumers to shop after rather than before Christmas to take advantage of discounts is difficult to measure accurately with current procedures. Finally, more work on outlets is badly needed. Casual empiricism certainly suggests a growing role of discount outlets, and for

example, Hausman and Leibtag, 2004 and 2005, estimate the failure to account just for the emergence of Walmart as the largest retail chain has caused a measurable bias in the CPI. And, of course, the rapid growth of online sales raises new challenges for improving price measurement.

Most importantly to social welfare, delays in improving our price indexes are enormously consequential. Irving Fisher, perhaps the greatest American economist of the first half of the 20th century, in his classic *The Making of Index Numbers* in 1922, said his entire life's work would have been worth while if it but led to the replacement of arithmetic with geometric index number formulas! Three-quarters of a century later, we have finally started using geometric means in about three-fifths of the lower-level data and have an experimental superlative index at the upper level, but have not yet accounted for upper-level substitution in the official CPI. While I would be the first to applaud any statistical agency for care in evaluating, experimenting with and adopting new techniques, the consequences of having adopted such formulas on the 50th rather than the 75th anniversary of Fisher's work would have been enormous, amounting to over a trillion dollars of lower federal debt as a result of reduced overindexing of outlays and income taxes from the inaccurate CPI. Explicitly, adopting a superlative index[14] would have immense desirable effects on the accuracy of our data.

These consequences are the most important reason to reiterate the need for more attention to be paid by economists and statisticians to training in data collection,

[14] See Diewert (1976) and Shapiro and Wilcox (1997), for discussions of theory and practical alternatives, respectively.

analysis and interpretation, with far more attention to these matters in the standard curriculum and more emphasis on measurement and sampling issues in the training of economists and statisticians. The economics research on price indexes, some of which is described above, has made immense contributions to better understanding of price movements. It has, in a very real sense, changed (the interpretation of) history. If economic research can continue to help the statistical agencies to improve our economic statistics, there will be a large social value added: more accurate COLAs, better information for government policy makers and for the private economic decisions of firms, workers, consumers and investors and, most importantly, better assessments of economic conditions and progress by our citizens, not to mention economists.

Perhaps it is just how consequential the (mis)-measurement of inflation could be that sparked the renaissance of research interest in these very practical issues. Economists and statisticians in and out of government will continue to contribute as much or more social value-added here as in virtually any other area of economics. While perhaps regretting their lost innocence, those working on such government statistics hopefully will be more than adequately compensated by their increased relevance.

References

1. Aizcorbe, Ana M. and Patrick C. Jackman (1993). “The Commodity Substitution Effect in CPI Data,” 1982-1991, *Monthly Labor Review,* 116 (12), December, 25-33.
2. Bajari, Patrick, C. Lanier Benkard and John Krainer (2002). “House Prices and Consumer Welfare”. *Unpublished mimeo*, Stanford University.
3. Berndt, E.R.Z. Griliches, and N.J. Rapport (1995), “Econometric Estimates of Price Indexes for Personal Computers in the 1990s”, *Journal of Econometrics,* 68, 243-268.
4. Berndt, Ernst R., Iain Cockburn and Zvi Griliches (1996). “Pharmaceutical innovations and market dynamics: Tracking effects of price indexes for anti-depressant drugs,” *Brookings Papers on Economic Activity,* 133-99.
5. Bils M. and P. Klenow (2001). “Quantifying Quality Growth”, *American Economic Review,* 91, 1006-1030.

6 Boskin, Michael J. (2005). “Causes and Consequences of Bias in the Consumer Price Index as a Measure of the Cost of Living”, *Atlantic Economic Journal,* 33(1), March.

7. Boskin, Michael J. (2008). “Consumer Price Indexes”, *Concise Encyclopedia of Economics*, D. Henderson, Ed., (Liberty Fund, Inc., Library of Economics and Liberty)
8. Boskin, Michael J., E. Dulberger, R. Gordon, Z. Griliches, and D. Jorgenson (1998). “Consumer Prices, the Consumer Price Index, and the Cost of Living,” *Journal of Economic Perspectives*, 12(1), Winter, 3-26.

9. Boskin, M., Dulberger, R., Gordon, Z. Griliches and D. Jorgenson (1997). “The CPI Commission: Findings and Recommendations,” *American Economic Review,* May.
10. Boskin, M., E. Dulberger, R. Gordon, Z. Griliches and D. Jorgenson (1996). “Toward A More Accurate Measure of the Cost of Living”. Report of the Advisory Commission on the Consumer Price Index. Washington, D.C., United States Senate, December.
11. Boskin, M. and D. Jorgenson, (1997). “Implications of Overstating Inflation for Indexing Government Programs and Understanding Economic Progress”, *American Economic Review*, May.
12. Cage, Robert, John S. Greenlees and Patrick Jackman (2003). “Introducing the Chained Consumer Price Indices,” in Thierry Lacroix, ed., International Working Group on Price Indices (Ottawa Group)*: Proceedings of the Seventh Meeting.* Paris: INSEE, 213-246.
13. Chwelos, Paul D., Ernst R. Berndt, and Iain M Cockburn (2004). “Faster, Smaller Cheaper: An Hedonic Price Analysis of PDAs,” NBER Working Paper 10746, September.
14. Chwelos, Paul D.(2005). “Apparel Prices 1914-93 and the Hulten/Brueghel Paradox,” in C. Corrado, E. Diewert, and C. Hulten, eds, *Price Index Concepts and Measurement,* Conference on Research in Income and Wealth (Chicago: University of Chicago Press for NBER, forthcoming). Also NBER Working Paper 11548, August.

15. Dalton, Kenneth V., John S. Greenlees and Kenneth J. Stewart (1998). "Incorporating a Geometric Mean Formula into the CPI," *Monthly Labor Review, 121* (10), October, 3-7.
16. Diewert, W.E. (1976). "Exact and Superlative Index Numbers, " *Journal of Econometrics,* 4, 114-145.
17. Diewert, W.E. (1993). "The Early History of Price Index Research," in W.E. Diewert and A.O. Nakamura, eds., *Essays in Index Number Theory,* vol. I, Amsterdam: Elsevier.
18. Diewert, W.E. (1998). "Index Number Issues in the Consumer Price Index," *Journal of Economic Perspectives* 12, 47-58.
19. Diewert, W.E. (2001). "Hedonic Regressions: A Consumer Theory Approach", forthcoming in *Scanner Data and Price Indexes,* R. Feenstra and M. Shapiro (eds.), Studies in Income and Wealth, Volume 61, Chicago: University of Chicago Press.
20. Fisher, I. (1922). *The Making of Index Numbers,* Boston: Houghton Mifflin.
21. Gordon, Robert J. (2003). "A Century of Housing Shelter Prices: How Big Is the CPI Bias?". Conference on Research in Income and Wealth, *mimeo*.
22. Gordon, Robert J. (2006). "The Boskin Commission Report: A Retrospective One Decade Later," in Triplett, *op cit.*
23. Gordon, Robert J. and van Goethem, Todd (2005). "A Century of Downward Bias in the Biggest CPI Component: The Case of Rental Shelter, 1914-2003," in E. Berndt and C. Hulten, eds., *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches,* Conference on Research in Income and Wealth (Chicago: University of

Chicago Press for NBER, forthcoming). Also NBER Working Paper 11776, November.

24. Griliches, Zvi, and Iain Cockburn (1994). “Generics and New Goods in Pharmaceutical Price Indexes”, *American Economic Review, 84* (5), December, 1213-1232.

25. Hausman, J. (1997). “Valuation of New Goods under Perfect and Imperfect Competition,” in Bresnahan, T. and Gordon, Robert J., eds., *The Economics of New Goods* (Chicago: University of Chicago Press for NBER), 209-37.

26. Hausman, J. (1999). “Cellular Telephone, New Products and the CPI,” *Journal of Business and Economic Statistics*.

27 Hausman, J. (2003). “Sources of Bias and Solutions to Bias in the CPI”, *Journal of Economic Perspectives*, Vol. 17 (1), 23-44.

28. Hausman, J., and Leibtag, Ephraim (2004) “CPI Bias from Supercenters: Does the BLS Know that Wal-Mart Exists?” Cambridge, MA, National Bureau of Economic Research, NBER Working Paper 10712. Forthcoming in W.E. Diewert, J.P. Greenlees, and C.R. Hulten, eds., *Price Index Concepts and Measurement*, 2008.

29. Hausman, J., and Leibtag, Ephraim (2005). “Consumer Benefits from Increased Competition in Shopping Outlets: Measuring the Effect of Wal-Mart. NBER Working Paper 11809, December; forthcoming, *Journal of Applied Econometrics*, 2007.

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

31. Ibbotson Associates (2002). *Stocks, Bonds, Bills and Inflation – 2002 Yearbook*. Chicago, IL, Ibbotson Associates.

32. Konus, A. (1924), “The Problem of the True Index of the Cost of Living,” *Econometrica*, 7, 1939, 10-29.
33. Lebow, D., and J. Rudd (2003). “Measurement Error in the Consumer Price Index: Where Do We Stand?” *Journal of Economic Literature*, March.
34. McClellan, Mark, Newhouse, Joseph P., and Remler, Dahlia (1996). “Are Medical Prices Declining?“ NBER Working Paper 5750, September.
35. Moses, Karin E. (1997), “Addressing the Quality Change Issue in the Consumer Price Index, *Brookings Paper on Economic Activity, 28* (1), 305-49.
36. National Academy of Sciences (2002). “At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes”. Panel on Conceptual, Measurement, and Other Statistical Issues in Developing Cost-of-Living Indexes. Charles Schultze and Christopher Mackie, eds. Committee on National Statistics, National Research Council.
37. Nordhaus, William D. (1997). “Do Real-Output and Real-Wage Measures Capture Reality? The History of Light Suggests Not,” in Bresnahan T., and R.J. Gordon, eds., *The Economics of New Goods* (Chicago: University of Chicago Press for NBER), 29-66.
38. Pakes, A. (2001), “Some Notes on Hedonic Price Indices, with an Application to PC’s”, paper presented at NBER Productivity Program Meeting, March 16, Cambridge MA.
39. Petrin, A. (2002), “Quantifying the Benefits of New Products: The Case of the Minivan,” Journal of Political Economy, 110, 705-729.
40. Pollak, R. (1989), The Theory of the Cost-of-Living Index, Oxford University Press.

41. Shapiro, M. and D. Wilcox (1997). “Alternative Strategies for Aggregating Prices in the CPI”. Federal Reserve Bank of St. Louis, *Review,* 79 (May/June), 113-125.
42. Shapiro, Irving, Matthew D. Shapiro, and David W. Wilcox (1999), “Quality Improvement in Health Care: A Framework for Price and Output Measurement,“ NBER Working Paper 6971, February.
43. Siegel, J. (2002). *Stocks for the Long Run,* 2d ed. New York, McGraw-Hill.
44. Silver, M.S. and Heravi S. (2001a), “Scanner Data and the Measurement of Inflation”, *The Economic Journal,* 111, F384-F405.
45. Silver, M.S. and Heravi S. (2001b), “Why the CPI Matched Models Method may Fail Us: Results from an Hedonic and Matched Experiment Using Scanner Data, “Univ. of Cardiff, *mimeo*.
46. Shoven, J. (2001). “What Are Reasonable Long-Run Rates of Return to Expect in Equities?” Washington, D.C.: Social Security Advisory Board, August.
47. Stewart, K. and S. Reed (1999). “Consumer Price Index Research Series using Current Methods, 1978-1998”. *Monthly Labor Review*, June. (An update is available on the BLS website.)
48. Triplett, Jack, ed. (1999). *Measuring the Price of Medical Treatments,* Washington, Brookings.
49. Triplett, Jack, et al. (2006). “The Boskin Commission Report after a Decade”, *International Productivity Monitor*, Spring.
50. White, Alan G., Jaison R. Abel, Ernst R. Berndt, and Cory W. Monroe (2004). “Hedonic Price Indexes for Personal Computer Operating Systems and Productivity Suites,” NBER Working Paper 10427, April.