Journal of Economic Perspectives—Volume 31, Number 2—Spring 2017—Pages 145–164

Underestimating the Real Growth of GDP, Personal Income, and Productivity

Martin Feldstein

Economists have long recognized that changes in the quality of existing goods and services, along with the introduction of new goods and services, can raise grave difficulties in measuring changes in the real output of the economy. Prominent economists have led and served on government commissions to analyze and report on the subject, including the Stigler Commission in 1961, the Boskin Commission in 1996, discussed in a symposium in the Winter 1998 issue of this journal, and the Schultze Commission in 2002, discussed in a symposium in the Winter 2003 issue of this journal (Stigler 1961; Boskin et al. 1996; National Research Council 2002). But despite the attention to this subject in the professional literature, there remains insufficient understanding of just how imperfect the existing official estimates actually are.

After studying the methods used by the US government statistical agencies as well as the extensive previous academic literature on this subject, I have concluded that, despite the various improvements to statistical methods that have been made through the years, the official data understate the changes of real output and productivity. The measurement problem has become increasingly difficult with the rising share of services that has grown from about 50 percent of private sector GDP in 1950 to about 70 percent of private GDP now. The official measures provide at best a lower bound on the true real growth rate with no indication of the size of the underestimation. Thus, Coyle (2014, p. 125) concludes her useful history of GDP

■ *Martin Feldstein is George F. Baker Professor of Economics, Harvard University, Cambridge, Massachusetts. His email address is mfeldstein39@gmail.com.*

† For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at
https://doi.org/10.1257/jep.31.2.145

doi=10.1257/jep.31.2.145

by saying: "Gross domestic product is a measure of the economy best suited to an earlier era."

In considering these issues, I have been struck by the difference between the official statistics about economic growth and how people judge whether their own economic condition has improved. The official figures tell us that real GDP per capita grew at an average rate of just 1.4 percent during the past 20 years. It is common to read in the press that because of this slow overall growth and changes in the distribution of income, the real income of the median household did not rise at all between 1995 and 2013 (for example, in the Council of Economic Advisers' 2015 *Economic Report of the President*, p. 30). When polls ask how the economy is doing, a majority of respondents say the country is doing badly; for example, 57 percent of respondents to a CNN–ORC poll in January 2016 said that the country is "doing poorly" (as reported in Long 2016) and in a Gallup poll in October 2016, 29 percent of respondents said the US economy is "poor" while only 29 percent said it was good or excellent (as reported in Dugan 2016). But in a Federal Reserve (2014) survey of household attitudes, two-thirds of households reported that they were doing as well or better than they had been five years earlier and that they were either "living comfortably" or "doing OK."

The contrast is revealing. People know their personal experience directly, but they depend on news media, politicians, and official statistics to judge how the economy as a whole is doing. And while the government statisticians are careful to say that GDP doesn't measure how well we are doing, there is a strong temptation on the part of the press, the politicians, and the public to think that it measures changes in the real standard of living. In this way, when the official statistics on economic growth understate real economic growth, it reduces public faith in the political and economic system. For example, the low measured growth of incomes probably exacerbates concerns about mobility, with people worrying that they and their children are "stuck" at low income levels: in a CNN/ORC poll, 56 percent of respondents said they think their children will be worse off than they are (as reported in Long 2016), and in a Pew Research Center poll, 60 percent of Americans said their children will be financially worse off than their parents (at http://www.pewglobal.org/database/indicator/74/survey/all/response/Worse+off/). Moreover, I think it creates a pessimism that contributes to political attitudes that are against free trade and critical of our market economy more generally.

I begin this essay by briefly reviewing the age-old question of why national income should not be considered a measure of well-being. I then turn to a description of what the government statisticians actually do in their attempt to measure improvements in the quality of goods and services. Next, I consider the problem of new products and the various attempts by economists to take new products into account in measuring overall price and output changes.

Although the officially measured rates of output growth have slowed substantially in recent years, the problem of understating real economic growth is not a new

one.[1] It reflects the enormous difficulty of dealing with quality change and the even greater difficulty of measuring the value created by the introduction of new goods and services. This paper is not about the recent productivity slowdown, but I return to that issue later in this paper and discuss the implications of these measurement issues for the measurement of productivity and the recent slowdown in the rate of productivity growth.

The final section of this paper discusses how the mismeasurement of real output and of prices might be taken into account in considering various questions of economic policy. Fortunately, there are important uses of nominal GDP that do not require conversion to real GDP.

Not Even Measuring Output, and Certainly Not Well-being

There is a long-running debate about the extent to which national income estimates should be designed to measure the well-being of the population or just the output of the economy. But in practice, national income concepts have been intentionally defined in ways that fall far short of measuring even *economic* well-being, let alone the broader well-being of individuals as influenced by matters like the environment and crime.

Even if we focus just on economic output, the concept of national output has been explicitly defined ever since the initial work of Kuznets (1934) and Kuznets, Epstein, and Jenks (1941) to exclude goods and services produced within the home. An earlier National Bureau of Economic Research study by Mitchell, King, and Macaulay (1921) offered a conjectural value of housewives services equal to about 30 percent of their estimate of the more narrowly defined traditional national income. Franzis and Stewart (2011) estimate that household production, under various assumptions, ranges from 31 to 47 percent of money earnings. The official statistics also exclude services that are provided outside the home but not sold. This omission has probably had a larger effect in recent years with the provision of such services as Google and Facebook and the vast expansion of publicly available videos and music, together with written commentary, stories, reports, and information, all of which are now available to web-connected users for essentially zero marginal payment.

Similarly, national income estimates focus on the positive value of the goods and services that households consume, not on the time and effort involved in earning the funds to buy those goods and services. The average workweek has declined but the number of two-earner households has increased. Working conditions have

[1] The vast literature bearing on the measurement of changes in the real output of the economy reaches back to Sidgwick (1883), Marshall (1887), Kuznets (1934), and Kuznets, Epstein, and Jenks (1941) and includes, more recently among others, Griliches (1992), Nordhaus (1997), Hausman (1996, 1999), and Gordon (2016). The NBER Conference on Research in Income and Wealth has focused work on this issue for more than 80 years (as discussed in Hulten 2015).

improved as employment has moved from factories and farms to offices. All of this affects economic well-being, but there is (by agreement) no attempt to take it into account in our measures of national income.

I mention these issues not to criticize the official definition of national income, but to stress that it is intended by design to be a measure of national output, not a measure of well-being. The public clearly wants a description of changes in well-being and inappropriately uses the official measures of real GDP and real personal incomes for that purpose. It might be useful to develop a formal array of well-being indicators and perhaps some summary index. These indicators might include measures of health, air pollution in cities, crime, and other matters that are not measured in the official economic statistics: Coyle (2014, chap. 5) discusses some previous attempts to provide such additional indicators. Alternatively, more attention might be focused on the Federal Reserve's Survey of the Economic Well-Being of US Households and its frequency might be increased from an annual survey to quarterly to increase its public saliency.

However, in this essay I will set aside the issues concerning what economic and noneconomic factors are left out of GDP, and how a broader measure of well-being might be constructed. Instead, I will argue that the official measure of real GDP does not even achieve its stated goal of measuring real national output on its own terms.

Measuring Quality Change

The government's calculation of real GDP growth begins with the estimation of nominal GDP, which is the market value of the millions of goods and services sold in the market to households, firms, governments, and foreign buyers. The government statisticians do a remarkable and prodigious job of collecting and then updating data from a wide array of sources.[2]

But for comparisons between one time period and the next, it is necessary to convert nominal GDP to real GDP. That process requires dividing the rise in nominal quantities into a real component and an inflation component, though the use of an appropriate price index. The overall GDP price deflator uses components based on the Consumer Price Index (CPI) and the Producer Price Index (PPI), requiring estimates done by the Bureau of Labor Statistics of the US Department of Labor and by the Bureau of Economic Analysis of the US Department of Commerce.

For each good and service, there are three possibilities when one compares one year with the next: 1) it is the same good or service with the same quality as in

[2]For a detailed analysis of the sources used to estimate these sales/purchases, see "Updated Summary of NIPA Methodologies" (Bureau of Economic Analysis 2015). Boskin (2000) shows that these estimates are subject to substantial revisions, with nearly all revisions from 1959 to 1998 in the upward direction, and some of those revisions being quite large. In this journal, Landefeld, Seskin, and Fraumeni (2008) provide a very useful description of how nominal GDP and related measures are estimated from a variety of primary sources.

the previous period; 2) it is essentially the same good, but of a different quality; or 3) it is a wholly new good. Each category receives a different treatment in the official US statistics.

Fortunately, most goods and services fall in the first category of "no (significant) change in quality." For those products, it is possible to collect the number of physical units sold and the total revenue. The percentage increase in revenue in excess of the percentage increase in physical volume is pure inflation, and the rest is the rise in real output. When exactly the same good is not available in the second period, the US Bureau of Labor Statistics tries to find a very similar good that does exist in the two successive periods and compares the revenue growth and physical quantity growth for that good. The BLS calls this procedure the "matched model" method.

Although much of the growth in the real value of economic output reflects substantial quality change and the introduction of wholly new products, the official procedures do not adequately reflect these sources of increased value. For products that experience quality change, the official methods tell us more about the increase in the value of inputs, in other words about the change in the cost of production, and not much about the increased value to the consumer or other ultimate user. This is true for goods as well as for services, although measuring quality improvement for services is even more difficult than it is for goods.

The government statisticians divide the period-to-period increase in total spending on each unit of product into a part due to a pure price increase ("inflation") and a part due to an increase in quality. The part attributed to a quality increase is considered an increase in the quantity of output although, as I will explain, the method used by the BLS means that it is generally a measure of the quantity of inputs.

The Bureau of Labor Statistics is responsible for creating the Consumer Price Index and the Producer Price Indexes, as well as a number of subsidiary indexes for various categories.[3] One main difference between the CPI and the PPI is that many of the PPI indexes are used primarily to deflate the prices of intermediate products, rather than to deflate output for final demand. The Bureau of Economic Analysis uses those price indexes and other data to create measures of real output. These estimates are also used for measuring the output of the nonfarm business sector and are used by the Department of Commerce to calculate the GDP deflator and real GDP. The same underlying data are also used to calculate the Personal Consumption Expenditures price index that the Federal Reserve uses for its price stability target.[4]

[3]For a clear description of the methods of the US Bureau of Labor Statistics, see *BLS Handbook of Methods*. Chapter 14 discusses the PPI indexes (available at http://www.bls.gov/opub/hom/pdf/homch14.pdf), and chapter 17 spells out the CPI indexes (available at http://www.bls.gov/opub/hom/pdf/homch17.pdf).

[4]A list of the price indexes used to create specific output numbers is available at Bureau of Economic Analysis (2015). For details, see also Bureau of Economic Analysis (2014).

The key question is how the Bureau of Labor Statistics estimates the change in price when there is a change in the quality of the good or service. The BLS asks the producer of each good or service whether there has been a change in the product made by that producer. If there has been no change in the product, any change in its price is considered to be pure inflation as called for in the "matched model method."

If a change has occurred, one approach to estimating the quality change is the "hedonic regression" method originally developed by Griliches (1961). The basic idea, which was used extensively for computers, is to regress the prices of computers in year t on a variety of the computers' capacity and performance measures. This gives an implicit price for each of these features (if the linearity assumption of the model is correct). Applying these implicit prices to a computer model in year $t + 1$ generates a price that would apply for that computer if the values of the individual features at time t had continued to prevail.

For example, a variety of econometric studies showed that the true price of mainframe computers assessed in this way declined at an annual rate of more than 20 percent per year during the period from 1950 to 1980 (Chow 1967; Baily and Gordon 1989; Triplett 1989). For personal computers, Berndt, Griliches, and Rappaport (1995) found a 28 percent annual rate of quality-adjusted price decline during a more recent period. The lack of use of hedonic regressions in these earlier decades may be part of the explanation for Robert Solow's (1987) comment that "you can see the computer age everywhere but in the productivity statistics."

Hedonic regressions are used for a variety of categories in the Consumer Price Index and the Producer Price Index. In the CPI, hedonic regressions are used in categories of goods that account for about one-third of the value in the basket of goods in the Consumer Price Index, including several categories of apparel, appliances, and electronics, but the main effect of hedonic analysis on the price index is in the analysis of housing, which by itself is more than 30 percent of the basket of goods represented in the CPI. The Bureau of Economic Analysis incorporates these estimates, and also uses hedonic price indexes to deflate nominal output, but for only about 20 percent of GDP.

The use of hedonics is no doubt very difficult to apply for many of these products and services for which, unlike computers, there is not a clear list of measured technical product attributes. There is also a problem of assuming that the attributes affect willingness to pay in a linear or log-linear way. According to the government, extensions of hedonics to even more products and services is limited by the lack of detailed data and staff resources required to build and maintain the hedonic models. In this journal, Hausman (2003) discusses the limitations of hedonic pricing.

When a producer indicates that a quality change has occurred and a hedonic regression is not used, the Bureau of Labor Statistics (2014, 2015a) typically uses the "resource cost method of quality adjustment," which is based on information about the cost of production supplied by the producer. If the producer says there has been a change in the product, the BLS asks about the "marginal cost of new input requirements that are directly tied to changes in product quality." The rationale relied on

by the BLS for this input cost as a method for defining the "quality adjustment" or, equivalently, the measure of the increased output, is described in Triplett (1983).

When the resource cost method is used, the Bureau of Labor Statistics concludes that there has been a quality improvement if and only if there is such an increase in the cost of making the product or service. The government statisticians then use the marginal cost of the product change, measured as a percentage of the previous cost of the product, to calculate a share of the price rise that is due to a quality improvement and that is therefore deemed to be an increase in the output of the product. The rest is regarded as inflation. The resource cost method can also treat a decline in production cost as evidence of a decline in quality.

This resource cost method of defining an improvement in a product or service is remarkably narrow and misleading. For the very specific case where a quality improvement is exclusively the result of adding an input, it will work. But according to this method, a pure technological innovation that makes the product or service better for the consumer doesn't count as a product improvement unless it involves an increased cost of production! In reality, product improvements generally occur because of new ideas about how to redesign or modify an existing product or service. Those changes need not involve an increased cost of production.

Government services provide an extreme version of treating costs of inputs as equivalent to the value of outputs. Government services are valued in the GDP at their cost, and so there is no possibility of reflecting changes in government productivity or the value created by the introduction of new government services.

Although the "resource cost method" may be the most common approach for quality adjustment, it and the hedonic procedure are not the only ones. The Bureau of Economic Analysis also uses what it calls the "quantity extrapolation method" and the "direct valuation method" for a few types of output. For example, the real quantity of bank services is derived from volume data on consumers' deposits and loans (for discussion, see Bureau of Economic Analysis 2015).

When government statisticians deal with quality change in services, they use a variety of different methods, but none of them attempts to capture changes in the true output of the service. For some services, like legal services provided to households, the Bureau of Labor Statistics creates a price index for a variety of specific services, like writing a will, and uses that price index and total expenditure to calculate the increase in real output.

The official GDP statistics for the healthcare industry, which accounts for more than 17 percent of US GDP, focus on costs of providing various categories of health services but do not seek to capture the effect of the health products and services on the health of the patient. For example, the "output" measure for hospitals recently shifted from a day of in-patient care to an episode of hospital treatment for a particular condition. Changes in the cost-per-episode-of-treatment is the corresponding price for the Producer Price Index, which is then used to deflate expenditure to get a measure of the quantity of output. Triplett (2012, p. 17), a careful analyst of the statistical health debate, concluded that there is a "very large error in measuring output generated in the medical care sector."

More generally, as Triplett and Bosworth (2004) note, the official data imply that productivity in the health industry, as measured by the ratio of output to the number of employee hours involved in production, declined year after year between 1987 and 2001. They conclude (p. 265) that such a decline in true productivity is unlikely, but that officially measured productivity declines because "the traditional price index procedures for handling product and service improvements do not work for most medical improvements." More recent data show that health sector productivity has continued to decline since 2001.

None of these measures of productivity attempt to value the improved patient outcomes. As one concrete example, when Triplett and Bosworth (2004, p. 335) wrote about the remarkable improvement in treating cataracts—from more than a week as an immobilized hospital inpatient to a quick outpatient procedure—they questioned whether accounting for medical improvements like that would cross over the traditional "production boundary in national accounts" and asked whether "the increased value to the patient of improvement in surgery … belongs in national accounts if no additional charges are made."

The Department of Commerce is experimenting with health sector "satellite accounts" that calculate the cost of treating a patient with a particular diagnosis for a calendar year, including the cost of hospital care, physicians, and pharmaceuticals. But these accounts also do not try to capture the value of improved health outcomes. There are some research studies that attempt to measure the effect of a certain treatment on such health outcomes as Quality Adjusted Life Years (QALYs) or Disability Adjusted Life years (DALYs).

For another example of the difficulties of adjusting for quality in a service, consider mutual fund management. The Bureau of Labor Statistics (2015b) has noted a substantial expansion over time in the types of funds that are available (including exchange-traded funds, fund-of-funds, long-short funds, a large number of emerging market funds, and more), but it ignores this increase in diversity of products and focuses only on the measuring output of mutual fund providers based on a percentage of all assets, concluding: "Under the current methodology, no special procedures are necessary for adjusting for the changes in the quality of portfolio management transactions" (p. 13).

To study the growth of output and productivity for individual industries, the Bureau of Labor Statistics sometimes measures real output at the industry level by the quantity of services provided. For passenger air travel, output of the industry is the number of passenger miles and productivity is defined as passenger miles per employee hour. The analysis of output "does not account for changes in service quality such as flight delays and route circuitry …" (Duke and Torres 2005).

From time to time the Bureau of Labor Statistics re-examines its approach to a particular industry. When the productivity program re-examined its measure of the commercial banking industry in 2012, it revised the activities of commercial banks and raised the estimated annual output growth from 1987 to 2010 by 58 percent, from 2.4 percent a year to 3.9 percent a year (Royster 2012, p. 8).

My own judgment is that, for most goods and services, the official estimate of quality change contains very little information about the value of the output to consumers and other final purchasers. As a result, the corresponding official measures of total real output growth are underestimates, and there is a substantial but unknown upward bias in the measure of price inflation. We don't know what the true values are, and we don't know how wide a margin of error there is around the official estimates.

Dealing with New Products

Although the sales of new products become immediately a part of *nominal* GDP, the extent to which they increase the real incomes of consumers is underestimated. Similarly, the effects of new products are not well reflected in the measures of real output and in price indexes. Moreover, the resource cost method and other government procedures for valuing changes in quality do not provide an approach to dealing with the value to consumers of new goods and services.

Instead, new products and services are not even reflected in the price indexes used to calculate real incomes and output until they represent a significant level of expenditures. They are then rotated into the sample of products used for price index calculations, and subsequent changes in their price are taken into account in the usual way. It is only at that secondary stage, sometime long after the new product has been introduced, that it affects officially measured changes in real output.

As an example to clarify how this works in practice, consider statins, the remarkable class of drugs that lowers cholesterol and reduces deaths from heart attacks and strokes. By 2003, statins were the best-selling pharmaceutical product in history and had become part of the basket of goods and services measured for the Consumer Price Index. When patents on early versions of statins then expired and generic forms became available, their prices fell. The Bureau of Labor Statistics recorded those price declines, implying a rise in real incomes. But the official statistics never estimated any value for the improvement in health that came about as a result of the introduction of statins.

To understand the magnitude of the effect of omitting the value of that single healthcare innovation, here is a quick history of the impact of statins. In 1994, researchers published a five-year study of 4,000-plus patients. They found that taking a statin caused a 35 percent reduction in cholesterol and a 42 percent reduction in the probability of dying of a heart attack. It didn't take long for statins to become a best-selling product with dramatic effects on cholesterol and heart attacks. According to the US Department of Health and Human Services (2011, pg. 26, fig. 17), between 1999–2002 and 2005–2008, the percentage of men aged 65–74 taking a statin doubled to about 50 percent. High cholesterol levels declined by more than half among men and women over age 75, and the death rate from heart disease among those over 65 fell by one-third. Grabowski et al. (2012) calculated that the combination of reduced mortality and lower hospital costs associated with heart

attacks and strokes in the year 2008 alone was some $400 billion, which was almost 3 percent of GDP in that year. None of this value produced by statins is included in the government's estimate of increased real income or real GDP.

This example of how statins have been treated in the national income statistics is representative of how all new products and services are treated. The value to consumers of a new good or service is ignored when the new product is at first introduced. Its price level becomes part of the Consumer Price Index when spending on that good or service is large enough to warrant inclusion. Subsequent declines in the price of the product are treated as real income gains, while price increases are part of inflationary real income losses. In short, the basic value to the consumer of the new good is completely ignored.

Ignoring what happens at the time of introduction of new products is therefore a serious further source of understating the real growth of output, incomes, and productivity. In addition, new products and services are not only valuable in themselves but are also valued by consumers because they add to the variety of available options. In an economy in which new goods and services are continually created, their omission in the current method of valuing aggregate real output makes the existing measure of real output even more of a continually increasing underestimate of true output. Hulten (2015, p. 2) summarizes decades of research on dealing with new products done by the Conference on Research in Income and Wealth with the conclusion that "the current practice for incorporating new goods are complicated but may miss much of the value of these innovations."

The introduction of new products into the official price indexes has historically also been subject to remarkably long delays. The Boskin Commission (Boskin et al. 1996) noted that at the time of their report in 1996 there were 36 million cellular phones in the United States, but their existence had not yet been included in the Consumer Price Index. The earlier Stigler Commission (Stigler 1961) found that decade-long delays were also noted for things like room air conditioners. Autos were only introduced to the Consumer Price Index in 1940 and refrigerators in 1934. More recently, the Bureau of Labor Statistics has introduced procedures that cause new products to be rotated into the analysis more quickly, but only after they have achieved substantial scale in spending. These delays cause the price index to miss the gains from introducing the product in the first place as well as the declines in prices that often happen early in product cycles.

But these delays in the introduction of new products to the price indexes are not the key problem. Much more important is the fact that the official statistics ignore the very substantial direct benefit to consumers of new products per se, causing an underestimate of the rate of increase in real output and an overestimate of the corresponding rate of increase of the price index.

There is great uncertainty about the size of these potential biases. For example, the Boskin Commission (Boskin et al. 1996) was charged by the US Senate with calculating the bias in the Consumer Price Index that is used for adjusting Social Security for changes in retirees' cost of living. The Commission considered several sources of bias in the existing Consumer Price Index, including the bias caused by

changes in quality and by the omission of new products and provided estimates of each type of bias in the CPI (see also the discussion of the report in the Winter 1998 issue of this journal).

But because the Boskin Commission was not able to do new research on the issue of quality change and innovation bias, it drew on existing research and on personal perceptions. For example, for "food and beverage," which accounts for 15 percent of the CPI, the commission members asked themselves how much a consumer would be willing to pay "for the privilege of choosing from the variety of items available in today's supermarket instead of being constrained to the much more limited variety available 30 years ago." They concluded, based on pure introspection, that "a conservative estimate … might be 10 percent for food consumed at home other than produce, 20 percent for produce where the increased variety in winter (as well as summer farmers' markets) has been so notable, and 5 percent for alcoholic beverages …" They used these numbers for 30 years and converted them to annual average rates of change for the 30-year period. This may be plausible, or not, but there is no real basis for believing that any of these estimates is even vaguely accurate.

Housing is the most heavily weighted component of the Consumer Price Index with a weight of nearly one-third. The Boskin Commission (Boskin et al. 1996) concluded that "a conservative estimate is that the total increase in apartment quality per square foot, including the rental value of all appliances, central air conditioning, and improved bathroom plumbing, and other amenities amounted to 10 percent over the past 40 years, or 0.25 percent per year." Maybe that is right, or maybe a better estimate would be 1 percent per year. There is nothing in the commission's report that helps to choose between differences of this magnitude.

In the end, the Boskin Commission concluded that the weighted average of these individual biases implied a total bias from product innovation and quality change in the annual CPI inflation rate for 1996 of 0.6 percentage points. I have no idea how much margin of error should be attached to that estimate. It served to satisfy the background political purpose for the Boskin Commission of providing a politically acceptable basis for reducing the rate of increase of Social Security benefits.

A formal analytic approach to the problem of valuing new products was developed by Hausman (1996, 2003). He showed how the value to consumers of a single new product could be measured by estimating the value of introducing a new brand of breakfast cereal—specifically Apple-Cinnamon Cheerios. His approach, following the theory presented by Hicks (1940), was to estimate the "virtual price," that is the price that would prevail when the good is just introduced at zero quantity. The consumer gains an amount of real income when the good is introduced implied by the decline in its price from the virtual price to the actual market price. He concluded that the Consumer Price Index component for cereals may be overstated by about 20 percent because of its neglect of new cereal brands. The Hausman estimates were controversial, but if the magnitude is even roughly indicative of the overstatement of the Consumer Price Index from a failure to reflect the

introduction of new varieties of cereal brands, then surely the overstatement of the Consumer Price Index and the understatement of real income that result from failing to take into account new products like statins and new anti-cancer drugs must be substantially larger.

Broda and Weinstein (2010) and Redding and Weinstein (2016) extend the Hausman (1996) approach and present a new method for valuing new products as well as the value to consumers of changes in product quality. They analyze a very large set of data on bar-coded package goods for which prices and quantities are available over time. By studying these data in the framework of a demand system based on constant-elasticity-of-substitution utility functions, they find that conventional price indexes overstate inflation for this set of goods by as much as 5 percentage points because the conventional measure ignores quality and new goods biases. Of course, this method is limited to goods and services for which the bar-coded price and quantity data are available and requires accepting a specific theoretical demand specification for these products. But as the availability of data on prices and quantity grows, it provides a starting point for improving the overall measurement of consumer prices and the corresponding estimates of real income.

The creation of new products also means an increased variety of choice, a form of quality improvement in itself, as Hausman (1996) noted. The value to consumers of access to an increased variety of options, which allows individuals to make choices that conform to their personal taste, can be substantial. Coyle (2014) noted that in the 30 years after 1970, the number of commonly available television channels rose from five to 185, and the number of soft drink brands climbed from 20 to 87.

The failure to take new products into account in a way that reflects their value to consumers may be an even greater distortion in the estimate of real growth than the failure to reflect changes in the quality of individual goods and services. At present, there is no way to know.

Productivity Change and Its Recent Slowdown

Labor productivity is defined as the ratio of real output to the number of hours worked by all employed persons. The Bureau of Labor Statistics estimates labor productivity for the nonfarm business sector, as well as for some parts of that sector, using output estimates provided by the Bureau of Economic Analysis.[5]

The key problem in measuring labor productivity is in the numerator—that is, in measuring output. The failure to measure quality changes adequately and to incorporate the value of new products means that true output has grown faster than

[5] In contrast, multifactor productivity is the ratio of real output to a combination of labor and capital input services. It is intended to measure the increase in output that is not attributable to either labor inputs or capital inputs. A good deal of research has been devoted to the very difficult problem of measuring the input of capital services and to the correct way to combine labor and capital inputs. Here, I will sidestep these issues by focusing on labor productivity.

measured output and therefore that the pace of productivity growth has been underestimated. This problem is particularly difficult in service industries. Bosworth and Triplett (2000, p. 6; Triplett and Bosworth 2004, p. 331) note that the official data imply that productivity has *declined* in several major service industries—including health care, hotels, education, entertainment, and recreation—and concluded that this apparent decline was "unlikely" and probably reflected measurement problems.

While the understatement of productivity growth is a chronic problem, there has been a sharp decline in the officially measured rate of productivity growth in the last decade. That sharp decline remains a puzzle that is yet to be resolved, as Syverson discusses in this issue. His work, along with papers by Fernald (2014) and Byrne, Fernald, and Reinsdorf (2016) show that the recent productivity slowdown cannot be attributed to the effects of the recession of 2008–2009, to changes in the labor force demographics in recent years, or to the growth of unmeasured internet services. One possible explanation of the recent downturn in productivity growth may be that the unusually rapid *increase* in the productivity growth in the prior few years was an anomaly and the recent decline is just a return to earlier productivity patterns.

A further hypothesis for explaining the recent downturn in productivity growth that has not yet been fully explored involves the mismeasurement of official estimates of output and productivity. Any attempt to explain the recent decline in the estimated productivity growth rate must attempt to understand not just the aggregate behavior for the nonfarm business sector as a whole, but also what happened at the disaggregated level. (Official estimates of productivity by industry, are available from the Bureau of Labor Statistics ("Industry Productivity" 1987–2015), although it should be noted that the overall productivity measure is not calculated by combining the individual industry numbers but is estimated separately based on a measure of real value added.)

The recent decline in the official measure of overall labor productivity growth in the nonfarm business sector reflects an enormous diversity of changes of productivity in specific industry groups. For the nonfarm business sector as a whole, the rate of productivity growth fell from 3.2 percent a year in the decade from 1995 to 2004 to just 1.5 percent in the decade from 2004 to 2013. The decline of 1.7 percentage points in the overall productivity change reflects an enormous range of changes in various industries. Even if attention is limited to the relatively aggregate three-digit level, the official productivity data show that productivity in apparel manufacturing went from annual growth at 1 percent in the earlier decade to an annual productivity decline of 5 percent in the later period, a drop of 6 percentage points. For manufacturing of computers and electronic products, productivity growth fell from a 15 percent annual rate to a 4 percent annual rate, a fall of 11 percentage points. Some industries experienced faster productivity growth, with productivity in the manufacturing of wood products increasing from a 2 percent annual rise in the early period to a 2.4 percent rise in the later period.

The differences are even greater at a more disaggregated level. At the four-digit level, for example, productivity growth increased by 5 percentage points annually

for radio and TV broadcasting but declined by 18 percent for semiconductors and electronic components. The deflation of output for disaggregated industries is even harder than for the economy as a whole because nominal outputs must be deflated by quality-adjusted prices for the disaggregated industries (Dennison 1989).

It would be intriguing, although difficult, to explore how or whether productivity differences across industries might be correlated with the problems of dealing with product change and the introduction of new goods and services in those industries.

Using Our Imperfect Data

What can be learned from the imperfect measures of real output and from the corresponding overstatement of price inflation? How should our understanding of the mismeasurement affect the making of monetary and fiscal policies?

Assessing Cyclical Economic Conditions

Consider first the assessment of short-run business cycle conditions. Policymakers and financial markets often focus on short-term fluctuations of real GDP as an indication of the state of the business cycle. Although measuring the size of fluctuations of real GDP is flawed by the difficulty of dealing with new products and quality changes, the official measure of real GDP fluctuations can in principle capture the short-term up or down changes in the pace of economic activity. Of course, it is important to recognize the substantial uncertainty about the estimated short-run fluctuations in GDP and the subsequent revisions.[6]

But it is interesting to note that when the Business Cycle Dating Committee of the National Bureau of Economic Research meets to consider appropriate dates for the start and end of a recession, it places relatively little emphasis on GDP. Contrary to popular belief, the NBER Committee has never used two quarters of decline in real GDP as its definition of a recession. Instead, it has traditionally looked at employment, industrial production, wholesale-retail sales, as well as real income. In recent years, the NBER Committee has also looked at monthly GDP when Macro Advisers began creating monthly estimates of GDP.

All data involve problems of interpretation in judging the state of economic activity, but employment, industrial production, and nominal sales are relatively free from the problem of quality adjustment and price measurement that affect measures of real GDP. Employment data are available monthly with substantial detail based on a large survey of employers. Industrial production is estimated by

[6]The Federal Reserve Banks of New York and Atlanta have recently begun using official data to produce preliminary estimates of changes in real GDP even before the corresponding quarter is over, but with some variability in results. In April 2016, the New York Federal Reserve estimated that real GDP increased by 1.1 percent in the recently completed first quarter of 2016, while the Atlanta Federal Reserve estimated that the increase in the same quarter was only 0.1 percent.

the Federal Reserve based primarily on data on physical production (such as tons of steel and barrels of oil) obtained from trade associations and government agencies, supplemented when necessary with data on production-worker hours and for some high-tech products by using nominal output and a price index (for details, see the Federal Reserve Board data https://www.federalreserve.gov/releases/G17/). These measures of industrial production as well as wholesale-retail sales deal with economic activity without having to impute value in large amounts, as must be done for the services of owner-occupied homes that are involved in the estimate of GDP.

Assessing Longer-Term Growth and Inflation

For the longer term, the official measures of changes in real output are misleading because they essentially ignore the value created by the introduction of new goods and services and underestimate changes in the quality of these products. It follows therefore that "true" real output is growing faster than the official estimates imply and that the corresponding "true" GDP price index is rising more slowly than the official one—or is actually declining.

The economics profession should educate the general public and the policy officials that "true" real incomes are rising faster than the official data imply. We can reassure people that it is very unlikely that the real incomes of future generations will be lower than real incomes today. Even if the future will not see the "epochal innovations" of the type that Kuznets (1971) referred to or such fundamental changes as electricity and indoor plumbing that caused jumps in living standards (as emphasized by Gordon 2016), current and future generations can continue to experience rising real incomes due to technological changes, improvements in education, and increases in healthcare technology.

One can only speculate about whether the bias in the officially measured pace of real output change is greater now than in the past. One reason to think that the gap between true output growth and measured growth is greater now than in the past is that services now represent about 70 percent of private value added, up from about 50 percent of private value added back in 1950, and the degree of underestimation of quality change and product innovation may be greater for services. Within services, health occupies a larger share of output—and quality improvements there may be greater than in other parts of the service sector. The internet and services through the internet have become much more important, and are also harder to measure.

Poverty and Distribution

Trends in the overestimation of inflation and therefore in the underestimation of real incomes may vary among demographic groups and income groups because of differences in the mix of goods and services consumed by these different groups. For example, are the goods and services bought by older people improving relatively faster than the goods and services bought by younger households? Health care is an obvious example, although most of the consumption of health care by the elderly is financed by government transfers.

Implications for Fiscal and Monetary Policy

Policy issues that depend on nominal measures of output are unaffected by the problems discussed in this essay. The most obvious of these is the ratio of debt to GDP, since both the numerator and the denominator are nominal values. Similarly, the rate of change of the debt-to-GDP ratio depends only on the nominal value of the annual deficit and the annual rate of nominal GDP growth. If the debt-to-GDP ratio is not on an explosive path, its long-run equilibrium value is equal to the annual nominal deficit ratio divided by the rate of nominal GDP growth.

The evidence that the true inflation rate is less than the measured inflation rate may imply that the true inflation rate is now less than zero. Fortunately, this does not imply that the US economy is experiencing the traditional problem of debt deflation (Fisher 1933) that occurs when a declining price level reduces aggregate demand by increasing the value of household debt relative to current incomes. The traditional problem of debt deflation does not arise under current conditions because the nominal value of wage income is not declining and the real monthly wage is rising more rapidly.

Overestimating the true rate of inflation does imply that the real rate of interest is higher than the conventionally measured rate. If households recognize that their dollars will buy relatively more in the future, this could alter the household saving rate—either increasing saving in response to the greater reward for saving or decreasing saving because a given volume of assets will buy more in the future, depending on whether substitution or income effects dominate. Because many factors affect the household saving rate, it is not clear which of these effects now dominates.

Uncertainty about the true rate of inflation should affect the optimal monetary policy. There seems little point in having a precise inflation target when the true rate of inflation is measured with a great deal of uncertainty. The goal of price stability also takes on a new meaning if true inflation is substantially negative while measured inflation is low but positive. Would it be better to have a target range for measured inflation as the Federal Reserve does now? Or to have a target range for measured inflation that is higher and further from the zero bound, thus leaving more room for larger changes in nominal interest rates while recognizing that the actual inflation rate is lower than the officially measured one? Or to restate the inflation goal of monetary policy as reacting when there is a rapid movement in measured inflation either up or down?

The underestimation of real growth has affected Federal Reserve decision-making in the past. Back in 1996, Fed chairman Alan Greenspan persuaded members of the Federal Open Market Committee that the official data underestimated productivity growth, so that maintaining strong demand would not cause a rise in inflation and there was no reason to raise interest rates (Mallaby 2016). In the last few years, the perception of slow real growth is often mentioned in support of a Federal Reserve policy of exceptionally low interest rates, but if real growth rates are actually higher (or if real growth rates have not dipped as much as the official statistics seem to show), then the Fed's policy of ultra-low interest rates has been providing little gain while contributing to certain risks of potential financial instability.

A great deal of effort and talent has been applied over past decades to the measurement of real income and inflation. These problems are extremely difficult. In my judgement, they are far from being resolved, and as a result, substantial errors of unknown size remain in our ability to measure both real output and inflation. It is important for economists to recognize the limits of our knowledge and to adjust public statements and policies to what we can know.

■ *I am grateful for extensive help with this paper to Anna Stansbury and for comments on earlier drafts from Katherine Abraham, Graham Allison, Michael Boskin, Erica Groshen, Jim Stock, David Weinstein, members of the staffs of the Council of Economic Advisers, the Bureau of Labor Statistics (especially David Friedman), and the Bureau of Economic Analysis, as well as participants at a meeting of the Group of Thirty and at a Harvard seminar.*

References

Baily, Martin, and Robert J. Gordon. 1989. "Measurement Issues, the Productivity Slowdown and the Explosion of Computer Power." CEPR Discussion Paper 305.

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

Boskin, Michael J., Ellen R. Dulberger, Zvi Griliches, Robert J. Gordon, and Dale Jorgensen. 1996. *Toward a More Accurate Measure of the Cost of Living.* Final report to the Senate Finance Committee for the Advisory Commission to Study the Consumer Price Index.

Boskin, Michael J. 2000. "Economic Measurement: Progress and Challenges." *American Economic Review* 90(2): 247–52.

Bosworth, Barry, and Jack E. Triplett. 2000. "Numbers Matter: The US Statistical System and a Rapidly Changing Economy." *Brookings Policy Brief 63.*

Broda, Christian, and David E. Weinstein. 2010. "Product Creation and Destruction: Evidence and Price Implications." *American Economic Review* 100(3): 691–723.

Bureau of Economic Analysis. 2014. *Concepts and Methods of the U.S. National Income and Product Accounts.* February. https://www.bea.gov/national/pdf/allchapters.pdf.

Bureau of Economic Analysis. 2015. "Updated Summary of NIPA Methodologies." *Survey of Current Business,* November, 95(11).

Bureau of Labor Statistics, US Department of Labor. 1987–2015. "Industry Productivity." Database, available at https://www.bls.gov/lpc/data.htm (accessed April 3, 2016).

Bureau of Labor Statistics, US Department of Labor. 1987–2015. "Major Sector Productivity and Costs." Database, available at https://www.bls.gov/lpc/data.htm (accessed April 3, 2016).

Bureau of Labor Statistics. No date. "Producer Prices." Chap. 14 in *Handbook of Methods.* http://www.bls.gov/opub/hom/pdf/homch14.pdf.

Bureau of Labor Statistics. No date. "The Consumer Price Index." Chap. 17 In *Handbook of Methods.* http://www.bls.gov/opub/hom/pdf/homch17.pdf.

Bureau of Labor Statistics. 2014. "Quality Adjustment in the Producer Price Index." Last modified August 2014. http://www.bls.gov/ppi/qualityadjust.pdf.

Bureau of Labor Statistics. 2015a. "Hedonic Models in the Producer Price Index (PPI)." Last modified October 21, 2015. http://www.bls.gov/ppi/ppicomqa.htm.

Bureau of Labor Statistics. 2015b. "Synopsis: NAICS 523920. Portfolio Management." Producer Price Index Program. Revised March 2, 2015.

Byrne, David M., John G. Fernald, and Marshall B. Reinsdorf. 2016. "Does the United States have a Productivity Slowdown or a Measurement Problem?" Brookings Papers on Economic Activity Conference Draft, March 10–11.

Chow, Gregory C. 1967. "Technological Change and the Demand for Computers." *American Economic Review* 57(5): 1117–30.

Council of Economic Advisors, The. 2015. *Economic Report of the President.* Transmitted to the Congress, February 2015. Washington, DC: US Government Printing Office.

Coyle, Diane. 2014. *GDP: A Brief But Affectionate History.* Princeton University Press.

Dennison, Edward F. 1989. *Estimates of Productivity Change by Industry: An Evaluation and an Alternative.* Washington, DC: Brookings Institution Press.

Dugan, Andrew. 2016. "U.S. Economic Confidence Index Flat at -10." Gallup, October 18. http://www.gallup.com/poll/196526/economic-confidence-index-flat.aspx.

Duke, John, and Victor Torres. 2005. "Multifactor Productivity Change in the Air Transportation Industry." *Bureau of Labor Statistics Monthly Labor Review,* March, pp. 32–45.

Federal Reserve, Board of Governors. 2014. *Report on the Economic Well-Being of U.S. Households in 2013.* July.

Federal Reserve, Board of Governors. "Industrial Production and Capacity Utilization – G.17." http://www.federalreserve.gov/releases/g17/current/.

Federal Reserve Bank of Atlanta. No date. "GDP Now". https://www.frbatlanta.org/cqer/research/gdpnow.aspx?panel=1 (accessed April 2016).

Federal Reserve Bank of New York. 2016. "Nowcasting Report: April 8, 2016." https://www.newyorkfed.org/medialibrary/media/research/policy/nowcast/nowcast_2016_0408.pdf?la=en.

Fernald, John G. 2014. "Productivity and Potential Output Before, During and After the Great Recession." NBER Working Paper 20248.

Fisher, Irving. 1933. "The Debt-Deflation Theory of Great Depressions." *Econometrica* 1(4): 337–57.

Franzis, Henry, and Jay Stewart. 2011. "How Does Nonmarket Production Affect Measured Earnings Inequality." *Journal of Population Economics* 24(1): 3–22.

Gordon, Robert J. 2016. *The Rise and Fall of American Growth: The US Standard of Living since the Civil War.* Princeton University Press.

Grabowski, David, Darius N. Lakdawalla, Dana P. Goldman, Michael Eber, Larry Z. Liu, Tamer Abdelgawad, Andreas Kuznik, Michael E. Chernew, and Tomas Philipson. 2012. "The Large Social Value Resulting from Use of Statins Warrants Steps to Improve Adherence and Broader Treatment." *Health Affairs* 31(10): 2276–85.

Griliches, Zvi. 1961. "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change." In *The Price Statistics of the Federal Government: Review, Appraisal, and Recommendations,* General Series 73, pp. 173–96. A Report to the Offices of Statistical Standards Bureau of the Budget prepared by the Price Statistics Review Committee of the National Bureau of Economic Research. New York: National Bureau of Economic Research.

Griliches, Zvi, ed. 1992. *Output Measurement in the Service Sectors.* National Bureau of Economic Research Studies in Income and Wealth. University of Chicago Press.

Hausman, Jerry A. 1996. "Valuation of New Goods under Perfect and Imperfect Competition." In *The Economics of New Goods,* edited by Timothy F. Bresnahan and Robert J. Gordon, 207–248. University of Chicago Press.

Hausman, Jerry A. 1999. "Cellular Telephone, New Products, and the CPI." *Journal of Business and Economic Statistics* 17(2): 188–94.

Hausman, Jerry A. 2003. "Sources of Bias and Solutions to Bias in the Consumer Price Index." *Journal of Economic Perspectives* 17(1): 23–44.

Hicks, John R. 1940. "The Valuation of the Social Income." *Economica* 7(26): 105–24.

Hulten, Charles R. 2015. "Measuring the Economy of the 21st Century." *NBER Reporter* no. 4, pp. 1–7.

Kuznets, Simon. 1934. "National Income, 1929–1932." Bulletin 49, National Bureau of Economic Research. July 7.

Kuznets, Simon, Lillian Epstein, and Elizabeth Jenks. 1941. *National Income and Its Composition, 1919–1938,* vol. 1. National Bureau of Economic Research.

Kuznets, Simon. 1971. "Modern Economic Growth: Findings and Reflections". Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel, Prize Lecture, December 11, 1971.

Landefeld, J. Steven, Eugene P. Seskin, and Barbara M. Fraumeni. 2008. "Taking the Pulse of the Economy: Measuring GDP." *Journal of Economic Perspectives* 22(2): 193–216.

Long, Heather. 2016. "56% of Americans Think Their Kids Will Be Worse Off." *CNN Money,* January 28. http://money.cnn.com/2016/01/28/news/economy/donald-trump-bernie-sanders-us-economy.

Mallaby, Sebastian. 2016. *The Man Who Knew: The Life and Times of Alan Greenspan.* Bloomsbury.

Marshall, Alfred. 1887 [1925]. "Remedies for Fluctuations of General Prices" (1887) in *Memorials of Alfred Marshall,* edited by A. C. Pigou (1925). London: MacMillan.

Mitchell, Wesley Clair, Wilford Isbell King, and Frederick R. Macaulay. 1921. *Income in the United States: Its Amount and Distribution, 1909–1919,* Vol. 1: *Summary.* New York: National Bureau of Economic Research.

National Research Council. 2002. *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes,* edited by Charles Schultze and Christopher Mackie. Washington, DC: National Academies Press

Nordhaus, William D. 1997. "Traditional Productivity Estimates Are Asleep at the (Technological) Switch." *Economic Journal* 107(444): 1548–59.

Redding, Stephen J., and David E. Weinstein. 2016. "A Unified Approach to Estimating Demand and Welfare." http://www.columbia.edu/~dew35/Papers/Estimating-Demand-and-Welfare.pdf.

Royster, Sara E. 2012. "Improved Measures of Commercial Banking Output and Productivity." *Monthly Labor Review,* July, 135(7): 3–12.

Sidgwick, Henry. 1883. *The Principles of Political Economy.* Macmillan and Co.

Solow, Robert. 1987. "We'd Better Watch Out." *New York Times Book Review*, July 12. p. 36.

Stigler, George, ed. 1961. "The Price Statistics of the Federal Government: Review, Appraisal, and Recommendations." Report to the Office of Statistical Standards, Bureau of the Budget. New York: National Bureau of Economic Research.

Triplett, Jack E. 1983. "Concepts of Quality in Input and Output Price Measures: A Resolution of the User-Value Resource-Cost Debate." Chap. 5 in *The U.S. National Income and Product Accounts: Selected Topics,* edited by Murray F. Foss. University of Chicago Press.

Triplett, Jack E. 1989. "Price and Technological Change in a Capital Good: A Survey of Research on Computers." In *Technology and Capital Formation,* edited by Dale W. Jorgenson and Ralph Landau, 127–213. Cambridge, MA: MIT Press.

Triplett, Jack E. 2012. "Health System Productivity." Chap. 30 in *The Oxford Handbook of Health Economics*, edited by Sherry Glied and Peter C. Smith. Oxford University Press.

Triplett, Jack E., and Barry P. Bosworth. 2004. *Productivity in the US Services Sector: New Sources of Economic Growth.* Washington, DC: Brookings Institution Press.

US Department of Health and Human Services. 2011. "Health, United States, 2010: With Special Feature on Death and Dying." https://www.cdc.gov/nchs/data/hus/hus10.pdf.