NBER WORKING PAPER SERIES

PRODUCTIVITY MEASUREMENT:
RACING TO KEEP UP

Daniel E. Sichel

Working Paper 25558
http://www.nber.org/papers/w25558

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2019

I thank Ana Aizcorbe, Ernst Berndt, David Byrne, Stephen Oliner, and Nicholas Oulton for exceptionally helpful comments. Any omission or errors are my responsibility alone. When citing this paper, please use the following: Sichel DE. 2019. Productivity measurement: racing to keep up. Annu. Rev. Econ. 11: Submitted. DOI: 10.1146/annurev-economics-080218-030439. Sichel is grateful to the Bureau of Economic Analysis (BEA) and the National Telecommunications and Information Administration for grant support of $76,800 during 2016 to 2017 for research on price measurement. The author currently is working with BEA as a consultant as well. Finally, the author serves on the BEA's External Advisory Committee. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Productivity Measurement: Racing to Keep Up
Daniel E. Sichel
NBER Working Paper No. 25558
February 2019
JEL No. E01,E22,E24,E31

ABSTRACT

This paper provides a non-technical review of the literature and issues related to the measurement of aggregate productivity. I begin with a discussion of productivity measures, their performance in recent decades, and key measurement puzzles that emerge from the data. The remainder of the review focuses on two important questions. First, how do we make more accurate the measures of prices used to deflate nominal output so as to win (or at least not lose) the race for economic measurement to keep up with a changing economy? This section frames the issues and points to the most important and promising areas for further research. Second, what does or should GDP measure? I defend GDP as a valuable measure of production and offer suggestions for improving it. At the same time, I emphasize the importance of measuring economic welfare (well being) and highlight the value of supplementing GDP with a satellite account that measures economic welfare.

Daniel E. Sichel
Department of Economics
Wellesley College
106 Central Street
Wellesley, MA 02481
and NBER
dsichel@wellesley.edu

1. Introduction

Sluggish economic growth since the Financial Crisis and widespread criticism that GDP is no longer an adequate measure of economic activity and welfare have re-focused attention on measurement of GDP and productivity. Indeed, the *Journal of Economic Perspectives* recently included a symposium titled "Are Measures of Economic Growth Biased?" While the included articles presented a range of views, Martin Feldstein (2017) argued that "official data understate the changes of real output and productivity … [and] provide at best a lower bound on the true real growth rate with no indication of the size of the underestimation." (p. 145) Going further, Diane Coyle (2014) says, referring to GDP, that "… it is a measure of the economy best suited to an earlier era." (p. 125)

These views represent a serious criticism of conventionally-measured GDP and, by implication, measures of aggregate productivity that rely on components of GDP. Of course, questions about the quality and accuracy of economic statistics are not new. Even at the dawn of national income accounting in the 1930s, questions about how best to measure economic activity and welfare were front and center as described by Coyle (2014). And, in the decades since then, several official commissions in the United States have evaluated the price statistics that are a key ingredient in calculating productivity growth, including the Stigler Commission (1961), the Boskin Commission (United States. Congress. Senate. Advisory Commission to Study the Consumer Price Index., 1996), and the Schultze Commission (2002).

Beyond these official reports, researchers in academia and government agencies also have looked intensely at many of the same and related issues over the decades. There are far too many papers to cite, but a comprehensive view of the challenges and progress can be gleaned from the collection of conference volumes published over the years by the Conference on Research in Income and Wealth and the National Bureau of Economic Research. These volumes include such titles as *Fifty Years of Economic Measurement (Berndt and Triplett, 1990), New Developments in Productivity Analysis (Hulten et al., 2001), Price Measurements and their Uses (Foss et al., 1993), Output Measurement in the Service Sectors (Griliches, 1992), Measuring Capital in the New Economy (Corrado et al., 2005), and Price Index Concepts and Measurement (Diewert et al., 2010).* Hulten (2015) provides a brief review of these volumes. As for the specific improvements made by statistical agencies Groshen et al. (2017) and Moulton (2018) provide excellent summaries.

A reader with just a little knowledge of economic measurement looking back at this literature would find much that is familiar—even in papers that are 50 or more years old—with discussions of quality change, accounting for new goods, pricing of services, the proper design and scope of price indexes, choice of formula for combining prices, the best approach and frequency for collecting data, and a host of other issues that remain salient today. This persistence of issues over the decades highlights the difficulty of the challenges – it is relatively straightforward to describe the issues but solving them continues to be a daunting task.

As hinted at above, the literature relating to productivity measurement is vast. To limit the scope, this paper provides a non-technical review of issues related to measuring aggregate productivity and skips over industry- and firm-level productivity. Accordingly, the focus is on value-added measures of output rather than gross output as might be used in assessing disaggregated measures. Two broad and important questions facing the measurement community are emphasized in this review: First, how should we get accurate price measures to deflate nominal output and capital that adequately capture changes in the quality and variety of goods and services in an ever-changing economy? Second, what should GDP measure and how should we measure economic welfare? An important theme that runs through both of these questions is the importance of developing better measures of the knowledge economy. Without a doubt, this focus omits important topics such as measuring labor input, the role of changes in labor quality, and productivity at the industry or firm level.[1]

To set the stage, Section 2 briefly describes aggregate measures of productivity and their performance in recent decades, highlighting key measurement puzzles raised in different time periods. This section also reviews recent work investigating whether mismeasurement can explain the slowdowns in U.S. productivity growth in 2004 and 2010. In short, the story is that, although understatement of growth (especially for high-tech and knowledge-related products) likely is substantial, this mismeasurement cannot explain the productivity slowdowns. Nonetheless, mismeasurement does affect the allocation of multifactor productivity growth (*MFP*) across sectors, and the evidence indicates that the pace of *MFP* growth and innovation has been more rapid in the high-tech sector (and less rapid elsewhere) than would be inferred from official statistics. This result deepens the puzzle of the recent productivity slowdowns.

Section 3 focuses on the question of obtaining good price measures. While statistical agencies and academic researchers have made considerable progress in recent decades,

[1] Many of these other topics are covered concisely in the OECD Manual Measuring Productivity by Schreyer (2001).

the issues are difficult (and fascinating) and continued progress must be made in order to keep up in the race to maintain the quality of economic data.

Section 4 turns to the questions of what GDP measures and how to measure economic welfare. This section defends GDP (as a measure of production) against some recent criticism and highlights areas where the GDP accounts could be improved as a measure of production: more comprehensive inclusion of intangible capital, more regular updates of satellite accounts that cover household activities, and fuller capture of new developments related to the digital economy. This section also reviews different approaches to measuring economic welfare that have been proposed in the literature, and highlights the value of a supplemental measure of economic welfare or well being.

Section 5 concludes with directions for future work. Two areas are highlighted. First, the need for continued progress on measuring quality change so that the race to keep up with a changing economy can be won (or at least not lost). Second, the potential benefit of the measurement community forging a consensus on how best to develop a satellite account for economic welfare.

2. Productivity Measures, their Performance, and Mismeasurement

Measures of Productivity

In principle, measuring productivity is easy. For labor productivity, divide a measure of real output by a measure of labor input. For MFP, divide real output by an appropriately weighted average of all inputs. In practice, obtaining accurate measures is a daunting challenge. To highlight the key measurement issues, I start with a review of different measures of productivity.

Labor productivity (*LP*) is the ratio of real output (*Y*) to hours (*H*) or sometimes the number of workers when data on hours are not available. Using hours, labor productivity is:

$$LP_t = Y_t / H_t \tag{1}$$

Real output, of course, is defined as the ratio of nominal output (*YN*) to the price deflator (*P*):

$$Y_t = YN_t / P_t \tag{2}$$

This review focuses mostly on obtaining accurate measures of the denominator in Equation 2 and what should be included in the numerator.

Before moving on, a few brief comments about the nominal output measure used in official measures of labor productivity. Productivity calculations often use business or nonfarm business output rather than GDP. Relative to GDP, these measures exclude the output of general government and households and institutions. The logic of using business sector output is that the economic forces driving the productivity of non-business sectors are quite different from those driving the productivity of the business sector and likely would generate quite different productivity trends. That being said, when considering changes in living standards over time, real GDP per person often is the preferred productivity measure because it indicates, on average, the resources available to each person.

In official measures, output is based on product-side GDP (the sum of final expenditures), though some researchers—for example, see Fernald (2015)—prefer to use an output measure that is the average of product-side and income-side GDP. The logic of using an average is that both product- and income-side measures of aggregate output convey useful information about economic activity and their average, arguably, better captures all available information.

Conventional *MFP* is defined as the ratio of real output to an income share weighted average of capital (*K*) and labor inputs. Typically, *MFP* is defined and calculated in growth rate terms (where dots over variables indicate growth rates):

$$M\dot{F}P = \dot{Y} - a_t\dot{K}_t - (1 - a_t)\dot{H}_t - \dot{L}Q_t \tag{3}$$

where *a* is the income share of capital (typically calculated as the average of the current and prior year's (or quarter's) capital income shares), *K* is capital services, and *LQ* is the contribution of labor quality.[2]

Capital services (as distinct from the capital stock) captures the service flow from capital and is a share-weighted average of capital stocks for individual types of capital using marginal products of each type of capital as weights (measured as income shares). The underlying idea of the weighting is that one dollar of computer capital generates a larger service flow in a year than does one dollar of office building capital because the service-life of the computer is so much shorter than that of an office building. Capital services typically is calculated as a Törnqvist aggregate, using the average of the current and prior year's income shares for each type of capital. In many analyses, capital services is disaggregated into key components, including information technology capital, intellectual property capital (intangibles), other equipment, structures, and other capital.

The labor quality term aggregates the service flow from different types of labor, using wages as weights. This term picks up the productivity contribution of shifts in the

---

[2] The shift to growth rates in Equation 3 from levels in the Equations 1 and 2 follows from Jorgenson and Griliches (1967). They also emphasize the importance of measuring capital services.

education, experience (age), and gender mix of the workforce. It is based on the difference between the growth rate of a wage-weighted aggregate of hours (dividing the workforce into education, age, and gender cells) and the growth rate of unweighted total hours.

At the industry level, *MFP* growth can be calculated using Equation 3, with industry value added as the numerator. More typically, though, industry-level *MFP* is calculated using a KLEMS variant that includes a wider set of inputs. For output, the KLEMS variant uses gross output rather than value added in the numerator in order to include inputs beyond the primary inputs of capital and labor. For inputs, it uses capital (K), labor (H), energy (E), materials (M), and services (S). KLEMS calculations also typically include a term for labor quality (LQ). The widespread use of this variant is summarized by Jorgenson (2017) and the articles included in that special issue of the *International Productivity Monitor*. Over the years, Jorgenson has undertaken an extraordinarily successful effort to persuade statistical agencies and academics in many countries to calculate KLEMS measures.

Aggregate *MFP* often is taken—at least by macroeconomists—as a rough gauge of the pace of technological progress averaging across the entire economy. However, the *MFP* expression in Equation 3 must be interpreted cautiously as a host of other factors—such as varying utilization, non-constant returns to scale, and adjustment costs—can affect it. Elaborations on Equation 3 that account for these factors have been developed in an effort to extract a pure measure of technological progress as described in Basu, Fernald, and Kimball (2006). That being said, many estimates of aggregate *MFP*—including official estimates from the Bureau of Labor Statistics (BLS)—rely on Equation

3 rather than the more complicated formulations, and the numbers presented below also rely on Equation 3.

The measurement of each variable in these productivity definitions has been the subject of extensive research. As noted, this review focuses on getting reliable measures of prices and determining the proper scope of nominal output (what should be included in GDP) in Equation 2. Researchers have investigated several other important topics not covered in this review, including labor quality and the measurement of hours. (For example, are employees with devices enabling greater connectivity working more unrecorded hours?) Bosler et al. (2016) assess labor quality, and Eldridge and Pabilonia (2007) discuss questions around unmeasured hours. On another topic not covered here, Syverson (2011) reviews the literature on productivity at the firm level.

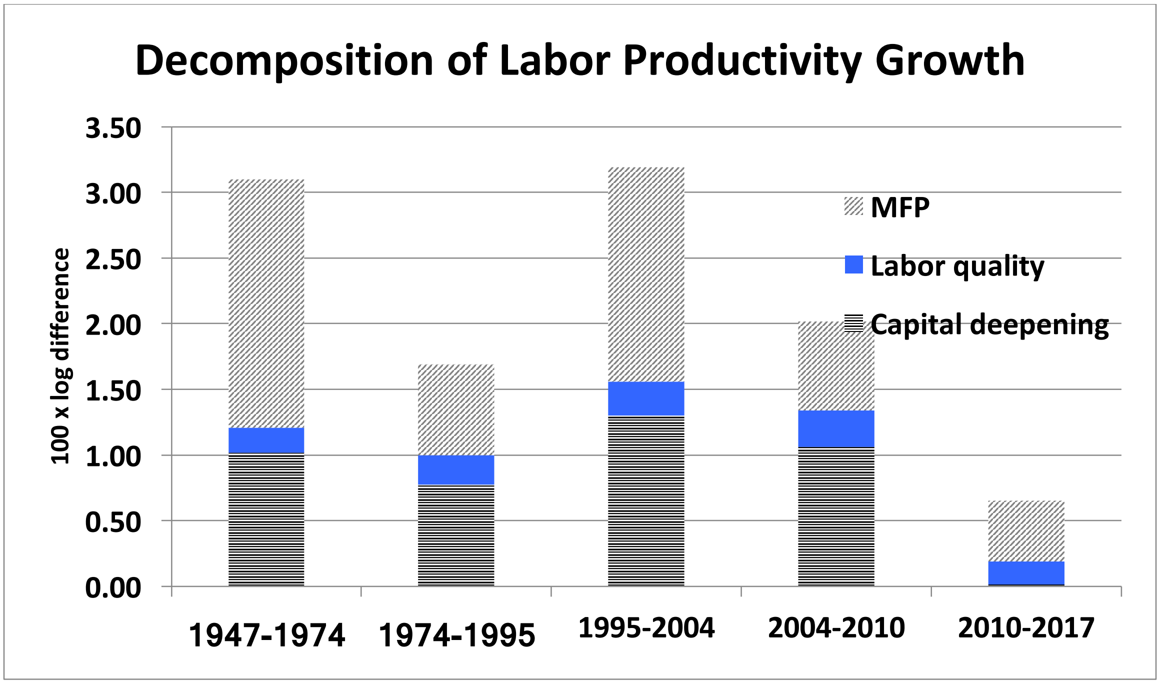
Productivity Performance in the United States

Historically, the U.S. economy has experienced alternating periods of faster and slower productivity growth, and these developments have affected the questions asked by measurement researchers. Figure 1 shows average growth rates of labor productivity and *MFP* over selected periods, relying on the decomposition:

$$\dot{LP} = \dot{Y}_t - \dot{H}_t = a_t(\dot{K}_t - \dot{H}_t) + \dot{LQ}_t + \dot{MFP}_t \qquad (4)$$

where the term with $\dot{K} - \dot{H}$ is capital deepening. Following this decomposition, Figure 1 shows contributions to labor productivity growth from *MFP,* changes in labor quality, and capital deepening. As shown, productivity growth was quite strong in the decades after the Second World War, exceeding an average annual rate of 3 percent. Then, the pace of productivity growth stepped down dramatically in the early 1970s to a pace of only about 1½ percent. This slowdown generated a great deal of attention and many papers attempting to explain it are summarized in a *Journal of Economic Perspectives* symposium [Fischer (1988)].

Figure 1



Source: Based on publicly available data from the Bureau of Labor Statistics (2018) and related data files.

Mismeasurement of output was one of the hypotheses put forward. One idea emphasized at the time was that the development of new technologies (computers) and other changes in the economy were making output more difficult to measure and, accordingly, that official measures of GDP were understating economic growth. Solow (1987) captures the puzzlement related to computers with his famous quip "You can see the computer age everywhere but in the productivity statistics." And, Griliches (1994) highlighted the increasing share of hard-to-measure sectors of the economy—including many types of services—as a possible contributor to the slowdown, though Sichel (1997) showed that the effect on productivity growth of a larger hard-to-measure sector was quantitatively small around the time of the 1970s productivity slowdown. More generally, gradual shifts in the structure of the economy—such as the rising share of services over many decades—are unlikely to be good explanations of sudden shifts in productivity growth.

Ultimately, economists never satisfactorily explained the 1970s slowdown and moved on to other questions as the mid-1990s resurgence in productivity growth took hold, with productivity growth averaging over 3 percent from 1995 to 2004. Yet, measurement questions remained in the fore, with the reports mentioned above from the Boskin and Schultze Commissions in 1996 and 2002, respectively.

Productivity growth took a big step down around 2004 and another around 2010, averaging barely over ½ percent per year from 2010 to 2017. Fernald (2015) was the first to decisively document the slowdown around 2004 (with earlier versions of his paper circulating well before it was published). As is evident from Figure 1, that slowdown reflected, in large part, a dropback in *MFP* growth. The further collapse in productivity growth around 2010 was associated with a cessation of capital deepening that reflected very weak business investment. With these developments—along with the seemingly rapid change arising from the digital revolution—researchers again focused on the degree to which measures of productivity and GDP are understating growth rates.

How Much Is Real GDP Growth Understated?

Researchers largely agree that mismeasurement of productivity and GDP growth is substantial, based mainly on concerns about price deflators—especially for products associated with the digital revolution and health care—and on the increasing prevalence of free goods (such as Facebook and Google's search and mapping tools) that are not counted as final output in GDP calculations.[3] The specifics of these issues are discussed in the next two sections. Here, I focus on selected estimates of the overall magnitude of mismeasurement and how this affects our interpretation of the productivity slowdowns after 2004.

Table 1 reports selected estimates of bias in growth rates of real GDP, based on the current definition of what is included in GDP.[4] As shown in the first column of numbers in Table 1, Moulton (2018)—the most detailed and comprehensive recent analysis—estimates that bias in the growth rate of real nonfarm business output amounted to 0.65 percentage point in 2017. (If the components of GDP outside of nonfarm business output

Table 1
Selected Estimates of Bias in Real GDP Growth

		Understatement of Real GDP Growth (pct pts)	
Study	**Coverage**	**Current estimate**	**Estimate for period before productivity slowdown**
Moulton (2018)	Nonfarm business output	.65	1.08
Goldman Sachs (2018)	GDP	⅔ to ¾	¼

Note: Moulton's estimates are for nonfarm business output. His estimate for the prior period draws from the Boskin Commission report in 1996. Goldman Sachs' estimate for the prior period is for two decades ago.

[3] Nominal GDP likely is undermeasured in some other areas as well. Several of these areas are discussed in Section 4 below.

[4] An early estimate of CPI bias is provided by Lebow, Roberts, and Stockton (1994) with an update in Lebow and Rudd (2003). Gordon (1990) provides comprehensive evidence on biases in prices of durable goods covering an earlier period.

suffer from roughly the same degree of mismeasurement as those within, this figure represents an estimate of bias in GDP growth.) Goldman Sachs' (2018) estimate of current bias is similar to Moulton's, reporting that real GDP growth currently is understated by ⅔ to ¾ percentage point a year.

Another interesting estimate of bias, which is not reported in Table 1 because it is not directly comparable, is from Aghion et al. (2017). Their paper examines what happens when a product disappears and is replaced by a new product produced by a different firm. Statistical agencies often use the price change of related, surviving products to link the prices of the old and new products. (This methodology is called imputation.) As it turns out, however, the imputed price based on these related products often is higher than what the price would have been for the product that disappeared from the market if its price had still been observable and so the link from the old to the new product is made with a price that mischaracterizes the price difference. The authors estimate that this bias led growth of nonfarm business productivity to be understated by roughly ⅔ percentage point a year in 2008.

As noted, the estimates in Table 1 do not account for any mismeasurement of free goods, including the cornucopia of information available on the Internet via search engines, social media, and free apps. More pointedly, if a good is free and its use is not mediated by a market transaction, then any value created by that good beyond the cost of its production will not be counted in GDP. To be clear, the cost of producing these products is already included in GDP. Consider Google Search. Google makes this product available to users for free and funds its provision through advertising. Further, the payments that, say, an auto company, makes to Google to advertise its cars, are considered an intermediate input for the automaker and their cost is reflected in the value of cars sold. Thus, in terms of the final expenditure measure of GDP, the cost of this advertising is captured as part of household or business purchases of automobiles just as would be the case for, say, print advertising. And, in terms of the income measure of GDP, the cost of this advertising is captured in wages paid by Google or in the company's profits. However, if the free good, Google Search, provides additional value to consumers, that extra value is not captured in GDP.

But, what about this seemingly substantial value and economic welfare that is created by these goods beyond their production cost? Although calculating this value is difficult, researchers have developed a number of techniques to obtain estimates.

Nakamura, Samuels, and Soloveichik (2017) focus on the resource cost of producing free (advertising-supported) digital media and adjust the national accounts to include this value as final output consumed by households and businesses. This adjustment boosts real GDP growth by about 0.1 percentage point over 2005-2015. Another approach estimates demand curves for free goods and the attendant consumer surplus and gets significantly larger figures than did Nakamura, Samuels, and Soloveichik. For example, Brynjolfsson and Oh (2012) estimate that the incremental consumer surplus from free services obtained on the Internet averaged more than $100 billion per year in the United States from 2007 to 2011. Goolsbee and Klenow (2006) also use data on time spent on the Internet to estimate a demand curve, and they estimate that the Internet generated significant amounts of consumer surplus. Brynjolfsson, Eggers, and Gannamaneni (2018) use massive online choice experiments to gauge consumers' willingness to pay for these products. Their estimates are substantial (median valuations add up to tens of thousands of dollars per person in 2017), with the largest valuations for search engines, email, and digital maps. And, Diewert and Fox (2017) provide an analytic framework for incorporating zero-price products into the national accounts.

The profession has not yet developed a consensus on the best way to measure the value of these products or on how much of that value belongs in a production-based measure of GDP. One view is that much of the extra value is consumer surplus that does not belong in a production-based measure of GDP. Another view is that consumers' willingness-to-pay for this extra value should be included in GDP. In any case, most would agree that the extra value or consumer surplus created by these goods belongs in a measure of economic welfare. (More on that in section 4 below.) These issues are an important topic for the measurement community.

Can Mismeasurement Explain the Productivity Slowdowns After 2004?

The bulk of the evidence indicates that mismeasurement is not a good explanation for the slowdowns in productivity and real GDP growth after 2004. To sharpen focus on the relevant question, consider this simple example. Suppose real GDP growth were understated by 1 percentage point before 2004 and by 1 percentage point after 2004 as well. Then, that mismeasurement cannot explain an observed decline in real GDP growth around 2004 because any drop in growth rates must be reflecting the drop in actual GDP because the amount of mismeasurement did not change. To explain a slowdown in observed real GDP growth, the magnitude of mismeasurement must have increased around the time of the slowdown. Accordingly, the key question for the mismeasurement hypothesis is whether the amount of understatement of growth has increased.

As reported in the second column of Table 1, Moulton concludes just the opposite—that the amount of mismeasurement has fallen in recent decades. This decline reflects improvements in measurement methodology implemented by the statistical agencies as well as a shrinking GDP share of some badly measured sectors such as information technology (where a larger fraction of domestic demand is imported than in the past). In contrast, Goldman Sachs (2018) estimates that amount of mismeasurement has risen. However, Goldman Sachs' estimate of the amount of mismeasurement a couple of decades ago is at variance with Moulton’s estimate as well as most other estimates, including that from the Boskin Commission report.

Other evidence supports the view that mismeasurement cannot explain the productivity slowdowns after 2004. Byrne, Fernald, and Reinsdorf (2016a) focus on information technology products. They present alternative price indexes for high-tech products that fall more rapidly than official price measures, implying an understatement of productivity and real GDP growth. Yet, they find that the contribution of this price mismeasurement to the understatement of growth was even larger in the past (similar to Moulton’s conclusion). Using a completely different approach, Syverson (2017) also makes the case that mismeasurement cannot explain the productivity slowdown after 2004. He calculates the amount of output that must have been missing each year if the productivity slowdown were entirely attributable to mismeasurement. He then shows that none of the mismeasurement hypotheses put forward can plausibly explain that amount of missing output. For the United Kingdom, Oulton (2013) shows that mismeasurement does not provide an explanation for the U.K. productivity slowdown.

But, what about free goods? Could a surge in the value of and consumer surplus associated with free goods imply that increases in economic welfare have not slowed down as much as implied by the dropback in the growth of productivity or GDP per person? Here too, the evidence suggests that the explosion of free goods is not sufficient to explain away the puzzle of the productivity slowdowns after 2004. Syverson (2017) makes the case that even the largest estimates of consumer surplus arising from these goods are too small to offset the magnitude of the productivity slowdown. And, it stands to reason that past innovations—such as the telegraph, radio, and television—likely also generated significant amounts of consumer surplus. Thus, until researchers reach back to obtain historical estimates for earlier innovations, we should not jump to the conclusion that consumer surplus is growing more rapidly now than in the past. As for the Nakamura, Samuels, and Soloveichik (2017) approach to valuing free digital media, they find that free digital media has increased while other free media (such as print) has declined. Accordingly, they estimate that the amount of bias from this source has not increased very much.

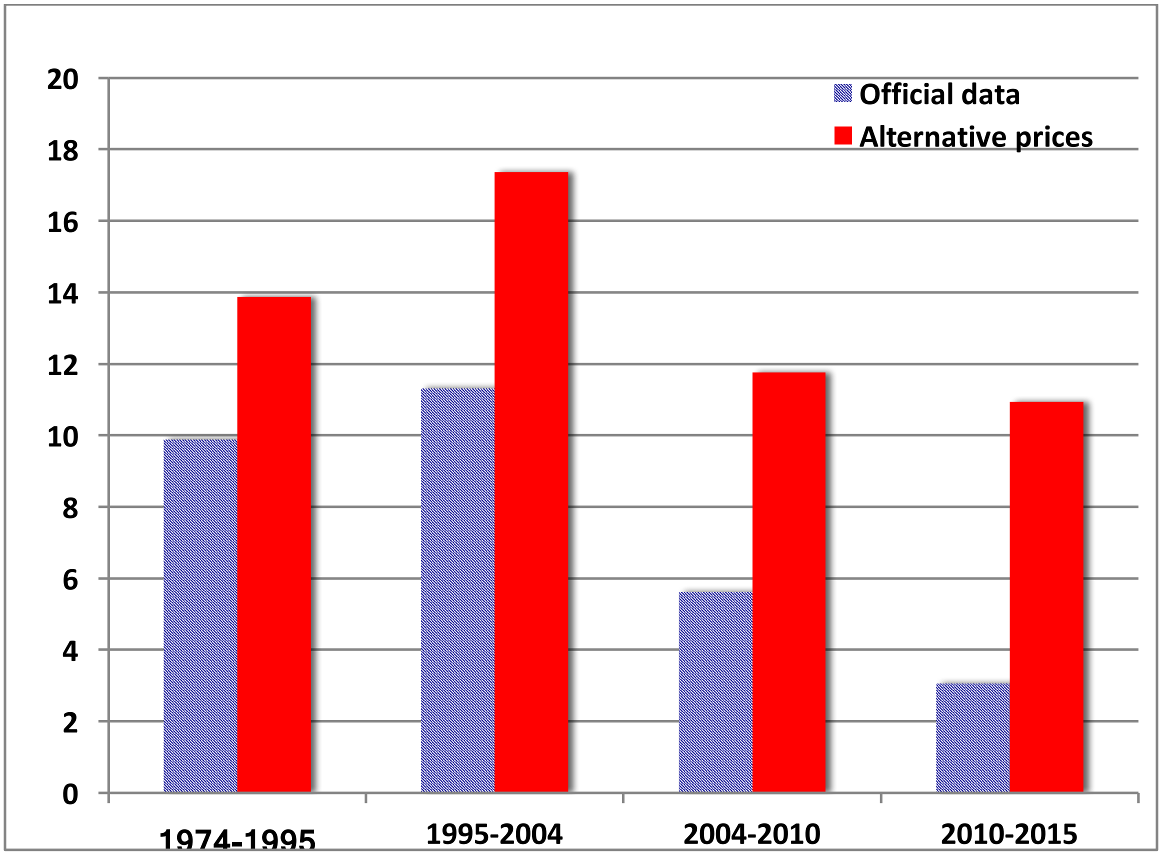
But, Biases Imply Faster Innovation in Tech and a Deeper Puzzle

As described above, the consensus of the literature is that mismeasurement, although substantial, is not a good explanation of the slowdown in productivity growth. Nonetheless, mismeasurement does have important implications for how we should think about sectoral *MFP* growth, and accounting for this mismeasurement deepens the puzzle of the productivity slowdowns after 2004.

Here's the story. Byrne, Oliner, and Sichel (2017b) and others have shown that the mismeasurement of prices of high tech-products has a much smaller effect on aggregate *MFP* growth than on labor productivity growth. Equation 3 above illustrates why. Mismeasurement of high-tech prices implies that output growth is understated, but it also implies that capital services growth (which is subtracted output growth to calculate MFP growth) is understated. As it turns out, for high-tech price mismeasurement, these two effects are pretty close in magnitude so the understatement in aggregate *MFP* growth is smaller than that in labor productivity growth. However, as Byrne, Oliner, and Sichel show, the mismeasurement of the prices of high-tech goods—with actual prices falling faster than captured in official measures—implies a different pattern of *MFP* growth across sectors.[5] In particular, the dual approach to productivity analysis indicates that faster relative price declines for high-tech products reflects a faster pace of *MFP* growth in that sector, as shown in figure 2.[6] The key result is plotted in figure 2. The red bars show *MFP* growth in the tech sector using the new, or alternative prices. These growth rates are well above those for *MFP* growth base on official prices (blue bars). Moreover, their analysis indicates that *MFP* growth outside the tech sector is much slower using the alternative prices than with the official prices. This result follows given that aggregate *MFP* growth is little affected by the mismeasurement so faster *MFP* growth in tech necessarily implies slower *MFP* growth elsewhere.

[5] Oulton (2016) highlights the likely mismeasurement of *MFP* in business services and finance and its implications for the pattern of *MFP* growth across sectors.
[6] The dual approach to production analysis relates the growth rate of prices for a product to a weighted average of its input costs minus *MFP* growth. The logic is that if prices are falling for a product, either input costs must be falling or *MFP* must be rising. Jorgenson and Griliches (1967) provide one of the clearest explanations of the dual approach.

Figure 2
Tech-Sector MFP Growth Using Official and Alternative Prices
(percent)



Source: Byrne, Oliner, and Sichel (2017b) .
Note: Tech sector consists of industries producing computers, software, communications equipment, and semiconductors.

To the extent that *MFP* trends provide information about the underlying pace of innovation, this result implies faster innovation in the tech sector but slower innovation elsewhere. This finding for the tech-sector innovation runs counter to recent narratives about its disappointing performance, such as in Gordon (2017) and raises a further puzzle: Even with notably faster growth in innovation in the tech sector since

2004 than reflected in the official data, aggregate labor productivity growth still slowed substantially.

3. The Price Isn't Right: Measuring Quality Change

This review focuses on the quality change issue for two reasons. First, it is, quantitatively, one of the most important sources of bias in growth rates of real GDP. According to Moulton (2018) mismeasurement of prices coming from quality change and the introduction of new goods amounts for more than 70 percent of the overall bias in consumer price inflation. Second, I would argue that capturing quality change is the most challenging measurement issue to fix and, at the same time, one on which significant progress can be made.

In this section, I review key issues in price measurement, especially regarding quality change, and highlight important progress that has been and is being made by U.S. statistical agencies.

What are the issues?

In principle, constructing the rate of change in a price index is simple. Just collect prices of every good and service in the relevant bundle and weight them appropriately to construct an index. Then, calculate the change over time in the index. However, in practice, problems emerge at nearly every turn, both in getting weights right and in making "like-to-like" product comparisons so that measures of price change over time are meaningful.

Many of these problems were well documented in the report of the Boskin Commission (1996)—and other reports and reviews—which often have focused on the

Consumer Price Index (CPI). The Boskin report identified four sources of bias in the CPI—substitution, outlet substitution, quality change, and new goods.

Before discussing these specific biases in more detail, I want to emphasize that statistical agencies in the United States and other countries have made significant progress in improving economic statistics in recent decades. For reviews of this progress, see Moulton (2018), the symposium on the Boskin Commission in the *International Productivity Monito*r as summarized in Sharpe (2006), and Gordon (2006). Despite significant progress, many of the issues highlighted by the Boskin Commission remain important challenges, especially for knowledge-related products.

Substitution bias occurs in a fixed-weight price index, such as the CPI, when consumers adjust their purchases in response to price changes but the weights used to aggregate prices are fixed. Lower-level substitution bias arises from the aggregation of prices within an expenditure category. The BLS largely eliminated lower-level bias in 1999 through the adoption of a geometric means formula for prices within an expenditure category. Upper-level substitution bias occurs because the CPI uses fixed-weights for aggregating expenditure categories. This source of bias remains in the CPI; however, the price index for Personal Consumption Expenditures (PCE) produced by the Bureau of Economic Analysis (BEA) allows these weights to change over time and therefore eliminates this source of bias. As for the CPI, in 2002 the BLS began producing an experimental superlative index that also eliminates upper-level substitution bias. Because the statistical agencies largely have eliminated this source of bias, I do not consider substitution bias further.

The other sources of bias are more closely related to the need for like-to-like price comparisons. Outlet substitution bias arises when consumers shift purchases toward lower priced outlets (say to Walmart rather than a local store or to Amazon rather than Walmart); the CPI considers items purchased from these different outlets as different products so any drop in the price paid by consumers is missed. However, if the products really are the same—that is, just purchased through a lower-price channel—then a relevant price decline is being omitted from CPI calculations. Moulton (2018) suggests that this source of bias is relatively small, though continuing shifts to online purchases could make it more salient. A related type of bias (known as offshoring bias) occurs when a buyer shifts its purchase from a domestic producer to a foreign producer, as discussed by Houseman et al. (2011). The BLS would consider the domestic and foreign items as different products; therefore, any price decline would be omitted from official price indexes.

Quality change bias occurs when the CPI is not able to capture the change in a product's quality and so is not comparing like-to-like prices. If an item's quality improves (or diminishes), then part of any price change reflects that improvement (or reduction) in quality and that portion of price change should be removed to obtain a like-to-like price comparison. For example, if an iPhone 6s was $549 and an iPhone 7 was $649, an adjustment for improved quality must be made because the two products are not directly comparable. With this adjustment, a quality-adjusted or constant-quality price index can be constructed so that the prices compared across time periods are for comparable amounts of "quality." Significant efforts have been made to control for quality change, but the problem is difficult to solve, and in a dynamic economy with frequently changing products new challenges will continue to emerge.[7]

Bias from the introduction of new goods is closely related to that from quality change.[8] New goods bias occurs when an item is different enough from prior items to be called a new product rather than a change in quality. One way to think about this difference is that a new good includes new characteristics rather than just more of previously-present characteristics. So, a faster laptop is an example of quality change, but the first smartphone is a new product. New goods raise two issues for price measurement. First, it may take some time for a new product to be included in the official price indexes and so a price index may miss price changes early in a product's life cycle. (Famously, cell phones were not included until 1998). Second, a new good may provide significant value that far exceeds its introduction price. In principle, this value can be incorporated by estimating a reservation price for what consumers would have been willing to pay for the product in the period before its introduction as in Hausman (1996) and then including in the price index the change from that reservation price to the product's price at introduction. But, the use of these calculations and the magnitudes generated are controversial. Indeed, while the Boskin Report (and many other economists) argued that declines from the reservation price should be incorporated in the CPI, this view is not universal. The Schultze Commission (2002) report argued that declines from reservation prices should not be included. In any case, given the

---

[7] As discussed by Moulton (2018), the BLS may also miss quality change because the agency makes quality adjustments only when a product disappears from the market and must be replaced by a new product. The BLS does not make quality adjustments at the time of routine sample rotation. As Moulton highlights, when smartphones became available, basic cellphones did not disappear so no quality adjustment between smartphones and basic cell phones would have been made.

[8] Recent evidence on new goods bias is provided in Goolsbee and Klenow (2018). They also highlight possibilities for using big data from online retailers to measure consumer prices.

difficulties of obtaining estimates that would be seen as a consensus of the profession, I judge this issue largely to be moot in that there is little prospect of an adjustment for new goods to be introduced into the CPI anytime soon.

These challenges also apply to price indexes for components of GDP beyond consumption. While deflators for business investment (especially high-tech products) have received attention from researchers, price indexes for exports and imports as well as government purchases generally have been more neglected. Moulton (2018) estimates that biases affecting business investment are sizable, while those affecting other components of nonfarm business output are modest. He may be right, though with more research, I would not be surprised if additional concerns about these other indexes emerge.

Controlling for Quality Change

Economists have struggled with the issue of quality at least since Smith (1963) [1776], who said about prices of cloth in the *Wealth of Nations* that "Quality, however, is so very disputable a matter that I look upon all information of this kind as somewhat uncertain." (p. 195) Notwithstanding Smith's caution, statistical agencies must confront the issue. The work-horse method to control for quality change is the matched-model methodology. To see how this procedure works, consider tracking the price of a single product, say a personal computer (PC). The price of a particular PC model would be tracked over time for as long as that model was observed in the market. Because the characteristics of that PC were not changing, each price change recorded would reflect a like-to-like comparison. Then, when that model disappeared from the market, the price of a new replacement model—typically of higher quality—would be tracked over time. For matched-model indexes to control correctly for quality change, the difference between the price of the exiting and entering model must reflect only the difference in quality. Put another way, the market for quality must be in equilibrium so that the price-performance ratio of the exiting and entering models are the same. For example, the availability of a new, higher-quality model at a price similar to what the old model had been selling for must sufficiently drive down the price of the old model.

Of course, in practice, this procedure can break down because price-performance ratios of exiting and entering products may differ for a host of reasons such as, for example, the price of the old model exits the market before its price has dropped enough to equilibrate price-performance ratios. If so, the implicit assumption that the price difference between the exiting and entering models reflects a quality difference is not valid; in this case, a matched-model price index would inadequately adjust for quality and therefore would be biased. Byrne, Oliner, and Sichel (2018) demonstrate how this assumption broke down for microprocessors in the mid-2000s and led to significant biases in official estimates.

Hedonics

Hedonic techniques provide a solution to this problem by explicitly controlling for quality change. Drawing directly from Byrne, Oliner, and Sichel (2018) consider a simple dummy-variable hedonic specification:

$$\ln(P_{it}) = a + \sum_{k} b_k X_{k,i,t} + \sum_{t} d_t D_{i,t} + e_{i,t} \tag{5}$$

where $P_{i,t}$ is the price of model i in period t, $X_{k,i,t}$ is the value of characteristic or performance metric k for model i in period t (measured in logs or levels, as appropriate), $D_{i,t}$ is a time fixed effect that equals 1 if model i is observed in period t and zero otherwise, and $e_{i,t}$ is an error term. In this equation, the X variables directly control for quality change, and the coefficients on the time fixed effect estimate quality-adjusted price changes.

A potential shortcoming of Equation 1, highlighted by Pakes (2003) and Erickson and Pakes (2011), is that the coefficients on the characteristic or performance variables are constrained to remain constant over the full sample period. One response to that concern, as described by Aizcorbe (2014) and Triplett (2006) is to run a cross-section regression for every time period and then to use predicted values from those regressions to build up a price index. Such an approach is appealing because it provides maximum flexibility for estimated coefficients to change over time and allows the results to be used in any price index formula. For further background on hedonic price indexes, see the *Handbook on Hedonic Indexes* prepared for the OECD by Jack Triplett (2006).

Court (1939), Griliches (1961) , and Chow (1967) were among the first to empirically implement hedonic methods, applying these techniques to automobiles (Court and Griliches) and mainframe computers (Chow). These efforts picked up steam with the further application of hedonic methods to computers in the 1970s and 1980s. These efforts ultimately led the BEA to introduce a hedonic index for computers into the national income and product accounts in 1985. For reviews of early research on hedonic indexes for computers see Gordon (1987) and Triplett (1989). For early work on PC prices, see Berndt and Rappaport (2001) . For more recent work on PC prices, see Byrne, Oliner, and Sichel (2016b) .

Hedonic indexes have been developed for many other products as well, and, over the past two decades the BLS introduced hedonic indexes or made other quality-related improvements to price indexes for many products (especially electronics-related products). Interestingly, as reported by Moulton (2018) , the hedonic adjustments outside of computers had relatively little effect on price trends. This result is somewhat puzzling. On the one hand, it certainly is possible that the old-style matched-model indexes were correctly capturing quality change. On the other hand, it also seems possible that the hedonic indexes, at least as implemented, still are not fully capturing quality change.

The efforts to improve price measures of high-tech products quieted down for a time after 2000 or so, but were revived in the 2010s, including work by various combinations of Byrne, Corrado, Oliner, and Sichel, as well as many other authors. In particular, Byrne and Corrado (2017) develop new price indexes for high-tech products and pull together other work to provide an overview of bias for business investment in information and communications technology products. Their price indexes are significantly different from those in official measures and were the basis for the alternative MFP calculations in Section 2 and for the Byrne, Fernald, and Reinsdorf paper on why mismeasurement cannot explain the productivity slowdown.

Despite this progress, more needs to be done. Byrne (2015) highlights the need for additional measurement attention to special-purpose equipment containing significant electronics components, including medical equipment, military gear, aerospace equipment, and a host of other products. Price indexes of these products generally have moved higher in recent years—despite their significant electronics content—in contrast to sharp price declines for computers and other high-tech products whose price and quality changes have been extensively researched. This discrepancy raises the possibility of significant bias in official measures of these prices.

In addition, recent advances in robotics and artificial intelligence raise a new set of challenges. On robotics, recent papers assessing its economic impact often rely on simple counts of industrial robots, as in Acemoglu and Restrepo (2017). The capability, size, and mobility of robots surely have improved over time, yet, very little is known about their prices, let alone prices on a quality-adjusted basis. For artificial intelligence, the challenge of measuring prices and quantities is the latest incarnation of earlier struggles to develop price indexes for software. Price indexes for software have been developed for narrowly-defined products—such as prepackaged PC software—but it has remained difficult to obtain reliable quality-adjusted indexes for larger, more idiosyncratic software projects (including artificial intelligence). Brynjolfsson and Kemerer (1996) , Gandal (1994) , and Oliner and Sichel (1994) provide early estimates of prices for prepackaged PC software. Abel, Berndt, and White (2003) and Copeland (2013) provide somewhat more recent estimates. Software has become an increasingly large share of total information technology investment so making further progress here is essential.

More generally, the growing recognition of the role of intangible capital in advanced-market economies as described in Haskel and Westlake (2017) highlights the need for better price measures for intellectual property products such as for research and development, organizational capital, and, as noted above, software.

Services

Hedonics also have been used for prices of services. Byrne, Corrado, and Sichel (2017a) develop hedonic price indexes for cloud computing services available from Amazon Web Services. Hedonics are readily applicable to this application because cloud providers rent virtual machines to users, and these virtual machines are configured to have characteristics much like those of PCs so the same types of characteristics used for PC hedonics can be used for cloud computing services. Their paper finds double-digit price declines for cloud services for computing, database, and storage in recent years.

Developing hedonic price indexes is often more challenging for a variety of other services than even for dynamic high-tech products. Defining the good to be priced can be difficult. For example, the output of health care is the improvement in health outcomes so that, in principle, it is simply necessary to measure the amount of that improvement and the price of achieving it. Yet, even in cases where the improvement in health outcomes and associated prices can be measured, how should the observed price change be allocated between quality change and pure price change? What is the role of the provider's skill, the patient's prior health status, the patient's willingness or ability to comply with recommended treatment, and a host of other difficult-to-measure factors? Moreover, the nature of markets for health care—with third-party payments, asymmetric information, changes in channels of provision, and other features—raises many additional challenges about how much of a price change reflects quality change.

Despite these challenges, significant progress has been made. For example, on health care measurement, Berndt et al. (2001) and Aizcorbe et al. (2018) summarize key issues and progress. One central idea is to price disease episodes (such as a heart attack) rather than specific inputs into (such as an aspirin, a doctor's visit, or a medical device) into medical care. This idea allows changes in treatment—such as a shift from a surgical to a pharmaceutical intervention—to be captured. And, the BEA has begun publication of a satellite health care account that tracks spending for treatment of diseases (rather than individual inputs). Dunn et al. (2015) provide an overview of these new accounts. (The original idea for pricing disease episodes appeared in Scitovsky (1964)).

Yet, this methodological advance does not solve the quality adjustment problem. Dauda et al. (2018) provide a recent assessment of alternative approaches to measuring quality change (and cites to earlier papers). They consider four alternative approaches to quality change and show that, perhaps not surprisingly, quality change is a big deal. Indeed, on their preferred estimates, the annual average growth rate of TFP for hospitals and nursing homes from 2001 to 2014 is 2.9 percent, a far more credible figure than the -0.3 percent figure reported in official statistics. However, the alternative approaches deliver different answers and the profession has not yet reached a consensus on the best approach.

Other services—such as education and financial services—pose their own set of unique measurement challenges, both in defining the service that is provided and in developing an appropriate technique for quality adjustment. Just as in the case of health care, the quality-adjustment techniques used for other products—such as electronics or apparel—generally are not directly applicable. The measurement of education and financial services are vast topics that I will not cover in any detail. For an overview of the issues related to education, see Schreyer (2010). For a discussion of issues related to financial services, see Diewert et al. (2016).

4. Production or Welfare: What does/should GDP Measure?

What does GDP measure? The short answer: Production. Although real GDP is highly correlated with measures of welfare and components of GDP should be key ingredients of any welfare calculation, real GDP is designed to be a measure of production: that is, a measure of the quantity or volume of goods and services produced within a country in a given period of time. Although many commentators have criticized GDP and suggested that economic measurement should shift its emphasis from production to economic welfare or well being (for example, Stiglitz et al. (2009)), measuring production remains an important task for statistical agencies. It is a crucial metric for monetary and fiscal policymakers to track, it is more highly correlated with employment than are many proposed measures of welfare, and it is useful in times of national emergency to gauge an economy's productive capacity. (See Coyle (2014) for a discussion of the original motivations for measuring production.)

If real GDP measures production, what about measuring economic welfare? Later in this section, I highlight the potential value of supplementing GDP with a satellite account measure of economic welfare. But first, I focus on how the GDP accounts could be improved as a tool for measuring production.

Where could Nominal GDP be improved as a measure of production?

GDP could be improved as a measure of production in three particularly important areas. First, including a more complete set of intangible capital—known as intellectual property products in the GDP accounts—as investment. Second, regularly updating and expanding a satellite GDP account for to better account for the full range of non-market household activity. Third, by sorting out the issues surrounding a host of new goods arising from the digital economy.

Including a Fuller Set of Intangible Capital as Investment

Business investment occurs when a firm devotes resources to acquiring or building an asset that is expected to generate a revenue stream in a future period. Traditionally, statistical agencies included only investment in tangible capital—such as machines and buildings—as business investment in GDP. Purchases of intangibles—such as software as well as R&D—were considered intermediate inputs that were used up in the current period and whose value would be incorporated into other products, on par with the purchase of pencils and paper. Thus, intangible capital did not get counted as investment in GDP.

With the rise of information technology and the digital economy, intangible capital has become increasingly important. Indeed, as documented in Corrado, Hulten, and Sichel (2009) and Haskel and Westlake (2017), intangible capital is now more important than tangible capital in many advanced-market economies. And, statistical agencies have, over time, adjusted the asset boundary to include more types of intangible capital as investment. In the United States, the BEA began counting software as business investment in 1999 and began counting R&D and the creation of entertainment and artistic originals as investment in 2013. However, the current asset boundary still excludes several important types of intangible capital that were identified in the framework developed by Corrado, Hulten, and Sichel (2004) and (2009). The excluded categories include nonscientific product development, brand equity, training, and organizational capital. Investment and capital stocks for these excluded assets are quite substantial—as highlighted in Corrado et al. (2012) and the databases referenced in that paper—and counting these assets as investment would significantly boost the level of GDP. Moreover, because these assets often are associated with innovative activity and the knowledge economy, including them would facilitate tracking and understanding these dynamic contributors to economic growth and welfare.

There is significant agreement in the measurement community that a fuller range of intangibles should, in principle, be counted as investment. In practice, however, the statistical agencies face both conceptual and data challenges. Indeed, the investment and capital stock estimates in Corrado, Hulten, and Sichel (2009) for the types of intangible capital currently outside the asset boundary for GDP—such as organizational capital—were generated with some strong assumptions to overcome conceptual and data issues. Moreover, getting deflators to translate nominal to real values for these intangibles is particularly challenging. Accordingly, there is less of a consensus on how to measure these types of capital than was the case for, say, R&D when it was brought inside GDP's asset boundary.

Household Activities

By design, GDP focuses on market-mediated activities. While this makes sense, it excludes important economic activity that occurs within households. These activities include, among other things, meal preparation, childcare, elder care, the service flow from household durables, time spent shopping, and human capital accumulation. Another type of household activity that has received attention recently is household production of a particular intangible asset—household R&D; that is, the dedication of household resources to create a product or idea that will generate a service flow to the household (or other households) in the future. For example, as documented by von Hippel (2017), a parent hacked a simple device for monitoring blood sugar levels of a diabetic child to make readings trackable remotely via the Internet (and posted the hack on the Internet for others to use).

Household production has long been of interest to economists with mentions as early as Gilman (1898). Estimates of human capital accumulation were pioneered by Jorgenson and Fraumeni (1989). More recently, BEA and others have periodically developed satellite accounts for household production and human capital, with recent versions in Bridgman (2016) for household production and Fraumeni, Christian, and Samuels (2017) for human capital formation. And, the BEA currently is exploring the feasibility of updating these accounts on an annual basis. For household R&D, Sichel and von Hippel (2018) present estimates—following the methodology used for business investments in intangible assets—and make the case that it is large enough to be consequential.

Although it makes sense to measure these non-market household activities in satellite accounts outside the scope of official GDP, tracking them is important for understanding the full range of economic activity and sources of economic welfare. In particular, regularly tracking these activities would help to elucidate the full welfare effect of changes over time in the structure of the economy. For example, the dramatic increase in women's labor market participation in the United States was accompanied by a shift of many activities from non-market to market-mediated, including, for example, childcare and meal preparation. GDP as currently defined will be affected by these shifts from non-market to market activity even if the overall amount of activity is not changing. A full understanding of the effect of such shifts on economic welfare requires measures of non-market household activity.

The Digital Economy

The ongoing evolution of the economy will always present new measurement challenges. Three examples of that are mentioned here. First, as discussed above, the increasing prevalence of free goods raises the question of how much of the value created by these goods should be included in GDP. Second, tax shifting by multinational firms likely has led to an undercount of value added located in the United States and, accordingly, has been a source of bias in nominal GDP as discussed by Guvenen et al. (2017). Third, high-tech firms—such as Google, Facebook, Amazon, and Microsoft—have been assembling a significant quantity of investment goods from purchased electronic components, that is, own-account investment just as when a business constructs a building on its own account. Because these capital goods are not mediated through market transactions, they likely are not fully counted as business investment in GDP, as discussed in Byrne, Corrado, and Sichel (2017a). Fortunately, the BEA partially accounted for this own-account investment in the latest comprehensive revision.

If GDP measures production, what about economic welfare?

Measuring economic welfare is vitally important. After all, providing economic welfare or well-being is the ultimate objective of economic systems, and recent criticisms of GDP highlight the demand for such measures. As noted, however, measuring production remains a key objective of statistical agencies. Moreover, many measures of welfare rely on components of GDP as key ingredients. Thus, a powerful argument can be made that GDP should not be scrapped or replaced, but rather should be supplemented with a measure of economic welfare. In this section, I first discuss how GDP relates to economic welfare and then review alternative possible approaches to measuring economic welfare that have been proposed in the literature.[9]

How does real GDP relate to economic welfare?

In popular accounts, real GDP or real GDP per person, often is taken as a measure of economic welfare. Understandably, that approach has generated considerable criticism. However, a combination of GDP components is pretty closely related to many measures of economic welfare. After all, it is consumption of goods and services—whether purchased or produced by a household or provided by the government or other

[9] Even broader measures of well-being that incorporate factors such as environmental quality and climate change can be constructed. Those considerations are beyond the scope of economic welfare measures considered here.

institutions—that generates utility. Dynan and Sheiner (2018) highlight that a broad measure of consumption—they label it aggregate economic well-being—can be built up from components of GDP—including personal consumption expenditures and some other pieces of GDP—along with some measures of non-market activity. Dynan and Sheiner also show that real consumption and real GDP have tracked pretty closely during the past several decades, indicating that, even though GDP measures production, it appears to be well correlated with, at least, the consumption piece of economic welfare. Oulton (2012) makes an even stronger case that GDP is an important component of welfare and a useful indicator of welfare.

These connections highlight the importance of GDP and associated measures, for measuring both production and economic welfare. They also reify why it is important to supplement, rather than to replace, GDP. More generally, Jorgenson (2018) provides a detailed evaluation of the relationship between measured GDP and economic welfare, highlighting both linkages of the sort described by Dynan & Sheiner and Oulton as well as how ingredients beyond GDP can be used to create other measures of economic welfare.

Measuring Economic Welfare

One approach for measuring economic welfare is to use a “dashboard” of economic and social indicators. Examples include the Human Development Index published by the United Nations Development Programme, a proposal from Coyle and Mitra-Kahn (2017), and the OECD’s Better Life Initiative. Each of these measures is a weighted average of various social and economic indicators. Although this approach has the virtue of incorporating any indicators of interest beyond GDP—such as the distribution of income, social and institutional capital, environmental quality, and life expectancy—they suffer from two problems as highlighted in the essay I co-wrote with Carol Corrado, Kevin Fox, Peter Goodridge, Jonathan Haskel, Cecelia Jona-Lasinio, and Stian Westlake [Corrado et al. (2017)]. In particular, the weighting across components can seem arbitrary, and they may double count some components. On double counting, the OECD Better Life index, for example, includes measures of housing and also includes household income (which itself would include measures of rent and an imputation for the income stream from owner-occupied housing).

Other approaches rely more on economic theory to construct measures of welfare and explicitly avoid double counting. Nordhaus and Tobin (1972) proposed a measure of economic welfare building on components of GDP. More recently, Jorgenson and Slesnick (2014) use individual and social welfare functions to create a measure of welfare that is based on per capita consumption and its distribution across the population, thereby very neatly incorporating distributional factors into a welfare measure. Jones and Klenow (2016) create a welfare measure using the standard economics of expected utility to combine data on consumption, leisure, inequality, and mortality. Interestingly, while they find that welfare and per capita GDP are highly correlated, there are differences, with countries in Western Europe looking more similar to the United States on a welfare basis than on a GDP per capita basis. Corrado et al. (2017) propose a measure of welfare that (using the insight of Weitzman (1976) that net national product is a proxy for the discounted value of future consumption) combines net product with measures of leisure and security along the lines of Jones and Klenow. Corrado et al. also propose more completely incorporating the value of "free" goods, relying on online experiments of the type implemented by Brynjolfsson et al. (2018) and coauthors to assess consumers' willingness to pay for free goods. Finally, Hulten and Nakamura (2017) propose an approach to measuring consumer welfare that draws on Lancaster's (1966) consumer theory and accounts for how consumer utility is affected not only by how much we consume but also by the technology that transforms goods and services into utility. And, new digital products appear to be dramatically changing these consumption technologies.[10]

These latter approaches have considerable merit. They incorporate important information beyond GDP into a measure of economic welfare in a way consistent with generally accepted principles of economic measurement and welfare analysis. Moreover, the pieces of these measures related to GDP (or its components) avoid the problems of the dashboard approach because GDP is designed to avoid double counting and, rather than use arbitrary weights, they use prices to aggregate different components. Prices are very appealing weights given that they represent valuations based on interactions of many buyers and sellers.

I believe that a powerful case can be made for the measurement community to forge a consensus about the best way to develop satellite accounts that provide a welfare measure, considering proposals along the lines of Jorgenson & Slesnick (2014), Jones and Klenow (2016), or Corrado et al. (2017). There is tremendous demand for such a measure and if the measurement community does not come to together to take this on, the

[10] One other approach is to ask people how happy they are as described in Stevenson and Wolfers (2008). This approach, however, is far removed from the topic of this essay.

demand for welfare measures may create its own supply, perhaps leading to inferior measures of welfare gaining traction among researchers and policymakers.

5. Conclusion

This paper has selectively reviewed the literature on measuring aggregate productivity growth. Through this review, I argue that productivity growth is understated by a significant amount, especially for rapidly changing high-tech products, other knowledge products (including intangible capital and free goods), and for hard-to-measure services such as health care and education. However, biases in productivity measures do not explain the slowdowns in productivity growth after 2004 and again after 2010 because, according to the best available evidence, there was considerable mismeasurement of productivity growth in the past as well. Although not an explanation for the productivity slowdown, mismeasurement—for which the evidence is particularly strong for high-tech products—deepens the puzzle of the slowdown as the mismeasurement implies that growth in innovation in the tech sector has been even faster than would be inferred from official data and that growth of innovation elsewhere in the economy has been even slower.

Improving measures of productivity is difficult work. Indeed, the key issues were documented decades ago. Much work has been done since then, and statistical agencies in the United States and in other countries have made considerable progress in improving economic statistics. That being said, the structure of the economy and the variety and quality of products continues to rapidly change. As that happens, new measurement challenges will emerge even as old ones are resolved.

As for directions for research, reports of official commissions, even though dated, still provide a useful roadmap of specific problems to be solved. In addition, Berndt (2006), Moulton (2018), Diewert, Greenlees, and Hulten (2010), and many other papers have provided sensible recommendations, and I will not reiterate those. Rather, I emphasize here two broad areas where progress can be made. The first area is continuing to develop the best adjustments for quality change. This cuts across almost all goods and services in the economy and requires, in the words of Shapiro and Wilcox (1996), "house-to-house combat"; that is, careful work for every product that undergoes consequential quality change and for every new product. That is a big task. Nonetheless, making progress is imperative to keep up in the race to ensure that the GDP accounts retain a high degree of accuracy.

The second area is in developing measures of economic welfare. Although real GDP is very valuable as a measure of production, there is a growing recognition among the public and policymakers that GDP does not measure welfare and, accordingly, considerable demand for measures that do. As noted, I believe that the measurement community should continue to build expertise on measures of welfare and to forge a consensus about the best way to develop a satellite account that includes measures of economic welfare.

All of this work requires resources. At a time when government resources are likely to be limited (at least in the United States), it is, perhaps, especially important for academic and other researchers to contribute to this effort. Such contributions seem to be increasingly occurring with a wider range of economists becoming interested in these issues, as evidenced by the Brookings Institution's recently launched Productivity

Measurement Initiative as well as ongoing efforts by the Conference Board and a host of other organizations. These are welcome developments that have the potential to significantly improve measures of productivity and welfare.

Literature Cited

ABEL, J. R., BERNDT, E. R. & WHITE, A. G. 2003. Price Indexes for Microsoft's Personal Computer Software Products. National Bureau of Economic Research.

ACEMOGLU, D. & RESTREPO, P. 2017. Robots and Jobs: Evidence from US Labor Markets. National Bureau of Economic Research.

AGHION, P., BERGEAUD, A., BOPPART, T., KLENOW, P. J. & LI, H. 2017. Missing growth from creative destruction. National Bureau of Economic Research.

AIZCORBE, A., BAKER, C., BERNDT, E. R. & CUTLER, D. M. 2018. *Measuring and Modeling Health Care Costs*, University of Chicago Press.

AIZCORBE, A. M. 2014. *A practical guide to price index and hedonic techniques*, Practical Econometrics.

BASU, S., FERNALD, J. & KIMBALL, M. 2006. Are Technology Improvements Contractionary? *American Economic Review,* 96**,** 1418-1448.

BERNDT, E. R. 2006. The Boskin Commission Report After a Decade: After-life or Requiem? *International Productivity Monitor,* 12**,** 61.

BERNDT, E. R., CUTLER, D. M., FRANK, R., GRILICHES, Z., NEWHOUSE, J. P. & TRIPLETT, J. E. 2001. Price Indexes for Medical Care Goods and Services--An Overview of Measurement Issues. *Medical care output and productivity.* University of Chicago Press.

BERNDT, E. R. & RAPPAPORT, N. J. 2001. Price and quality of desktop and mobile personal computers: A quarter-century historical overview. *American Economic Review,* 91**,** 268-273.

BERNDT, E. R. & TRIPLETT, J. E. 1990. *Fifty years of economic measurement: the jubilee of the Conference on Research in Income and Wealth*, University of Chicago press.

BOSLER, C., DALY, M. C., FERNALD, J. G., HOBIJN, B. & NATIONAL BUREAU OF ECONOMIC RESEARCH. 2016. The Outlook for U.S. Labor-Quality Growth. Cambridge, Mass.: National Bureau of Economic Research.

BRIDGMAN, B. 2016. Accounting for Household Production in the National Accounts: An Update, 1965–2014. *Survey of Current Business,* 96**,** 1-5.

BRYNJOLFSSON, E., EGGERS, F. & GANNAMANENI, A. 2018. Using massive online choice experiments to measure changes in well-being. National Bureau of Economic Research.

BRYNJOLFSSON, E. & KEMERER, C. F. 1996. Network externalities in microcomputer software: An econometric analysis of the spreadsheet market. *Management Science,* 42**,** 1627-1647.

BRYNJOLFSSON, E. & OH, J. 2012. The attention economy: measuring the value of free digital services on the Internet.

BUREAU OF LABOR STATISTICS 2018. Multifactor productivity trends, 2017. Washington, DC.

BYRNE, D. & CORRADO, C. 2017. ICT Services and Their Prices: What Do They Tell Us About Productivity and Technology? *International Productivity Monitor***,** 150-181.
BYRNE, D., CORRADO, C. & SICHEL, D. 2017a. The Rise of Cloud Computing: Minding your P's and Q's. *NBER Chapters*.
BYRNE, D., FERNALD, J. & REINSDORF, M. 2016a. Does the United States have a productivity slowdown or a measurement problem? *Brookings Papers on Economic Activity,* 2017**,** 109-182.
BYRNE, D., OLINER, S. D. & SICHEL, D. E. 2017b. Prices of high-tech products, mismeasurement, and the pace of innovation. *Business Economics,* 52**,** 103-113.
BYRNE, D. M. 2015. Domestic Electronics Manufacturing: Medical, Military, and Aerospace Equipment and What We Don't Know about High-Tech Productivity. Board of Governors of the Federal Reserve System (US).
BYRNE, D. M., OLINER, S. D. & SICHEL, D. E. 2016b. A New Look at Prices of Personal Computers: Desktops, Laptops, and Tablets.
BYRNE, D. M., OLINER, S. D. & SICHEL, D. E. 2018. How fast are semiconductor prices falling? *Review of Income and Wealth,* 64**,** 679-702.
CHOW, G. C. 1967. Technological change and the demand for computers. *The American Economic Review***,** 1117-1130.
COPELAND, A. 2013. Seasonality, consumer heterogeneity and price indexes: the case of prepackaged software. *Journal of Productivity Analysis,* 39**,** 47-59.
CORRADO, C., FOX, K., GOODRIDGE, P., HASKEL, J., JONA-LASINIO, C., SICHEL, D. & WESTLAKE, S. 2017. Improving GDP: Demolishing, repointing or extending. *Indigo prize paper, Global Perspectives*.
CORRADO, C., HALTIWANGER, J. & SICHEL, D. 2005. *Measuring Capital in the New Economy*, NBER and University of Chicago Press.
CORRADO, C., HULTEN, C. & SICHEL, D. 2004. Measuring capital and technology: an expanded framework. *Measuring capital in the new economy.* University of Chicago Press.
CORRADO, C., HULTEN, C. & SICHEL, D. 2009. Intangible capital and US economic growth. *Review of income and wealth,* 55**,** 661-685.
CORRADO, C. A., HASKEL, J., IOMMI, M. & JONA LASINIO, C. 2012. Intangible capital and growth in advanced economies: Measurement and comparative results.
COURT, A. 1939. Hedonic price indexes with automotive examples. *The dynamics of automobile demand.* General Motors Corporation.
COYLE, D. 2014. *GDP : a brief but affectionate history,* Princeton ; Oxford, Princeton University Press.
COYLE, D., MITRA-KAHN, B. & AUSTRALIA, I. P. 2017. Making the future count. *Indigo Prize winning entry*.
DAUDA, S., DUNN, A. & HALL, A. 2018. Are Medical Care Prices Still Declining? A Systematic Examination of Quality-Adjusted Price Index Alternatives for Medical Care? *In:* PRESENTED AT CRIW/NBER SUMMER INSTITUTE, J. (ed.).

DIEWERT, E., FIXLER, D. & ZIESCHANG, K. 2016. Problems with the Measurement of Banking Services in a National Accounting Framework. *In:* HARTWICK, J. (ed.) *National Accounting and Economic Growth.* Cheltham, UK
Northampton, MA: Edward Elger Publishing.
DIEWERT, E. & FOX, K. 2017. The Digital Economy, GDP, and Consumer Welfare. Presented at the NBER Summer Institute CRIW Workshop, July 17-18.
DIEWERT, W. E., GREENLEES, J. & HULTEN, C. R. 2010. *Price index concepts and measurement*, University of Chicago Press.
DUNN, A., RITTMUELLER, L. & WHITMIRE, B. 2015. Introducing the new BEA health care satellite account. *Survey of Current Business,* 95**,** 1-21.
DYNAN, K. & SHEINER, L. 2018. GDP as a Measure of Economic Well-Being.
ELDRIDGE, L. P., PABILONIA, S. W. & UNITED STATES. BUREAU OF LABOR STATISTICS. OFFICE OF PRODUCTIVITY AND TECHNOLOGY. 2007. Are those who bring work home really working longer hours? implications for BLS productivity measures. [Washington, D.C.]: U.S. Dept. of Labor, U.S. Bureau of Labor Statistics, Office of Productivity and Technology.
ERICKSON, T. & PAKES, A. 2011. An experimental component index for the CPI: from annual computer data to monthly data on other goods. *American Economic Review,* 101**,** 1707-38.
FELDSTEIN, M. 2017. Underestimating the Real Growth of GDP, Personal Income, and Productivity. *Journal of Economic Perspectives,* 31**,** 145-64.
FERNALD, J. G. 2015. Productivity and Potential Output before, during, and after the Great Recession. *NBER macroeconomics annual,* 29**,** 1-51.
FISCHER, S. 1988. Symposium on the slowdown in productivity growth. *Journal of Economic Perspectives,* 2**,** 3-7.
FOSS, M., MANSER, M. & YOUNG, A. 1993. *Price Measurements and Their Uses*, University of Chicago Press.
FRAUMENI, B. M., CHRISTIAN, M. S. & SAMUELS, J. D. 2017. The Accumulation of Human and Nonhuman Capital, Revisited. *Review of Income and Wealth,* 63**,** S381-S410.
GANDAL, N. 1994. Hedonic price indexes for spreadsheets and an empirical test for network externalities. *the RAND Journal of Economics***,** 160-170.
GILMAN, C. P. 1898. *Women and economics: A Study of the Economic Relation between Men and Women as a Factor in Social Condition*, Small, Maynard & Company.
GOLDMAN SACHS 2018. iPhones, eBay, and the Fate of Productivity Growth. *US Economics Analyst.*
GOOLSBEE, A. & KLENOW, P. J. 2006. Valuing consumer products by the time spent using them: An application to the Internet. *American Economic Review,* 96**,** 108-113.
GOOLSBEE, A. D. & KLENOW, P. J. 2018. Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce. *American Economic Review, Papers and Proceedings,* 108**,** 488-92.

GORDON, R. J. 1987. The postwar evolution of computer prices. National Bureau of Economic Research Cambridge, Mass., USA.
GORDON, R. J. 1990. *The measurement of durable goods prices*, University of Chicago Press.
GORDON, R. J. 2006. The Boskin Commission report: A retrospective one decade later. National Bureau of Economic Research.
GORDON, R. J. 2017. *The rise and fall of American growth: The US standard of living since the civil war*, Princeton University Press.
GRILICHES, Z. 1961. Hedonic price indexes for automobiles: An econometric of quality change. *The Price Statistics of the Federal Goverment.* NBER.
GRILICHES, Z. 1992. *Output measurement in the service sectors*, University of Chicago Press.
GRILICHES, Z. 1994. Productivity, R&D, and the Data Constraint. *The American Economic Review,* 84**,** 1-23.
GROSHEN, E. L., MOYER, B. C., AIZCORBE, A. M., BRADLEY, R. & FRIEDMAN, D. M. 2017. How Government Statistics Adjust for Potential Biases from Quality Change and New Goods in an Age of Digital Technologies: A View from the Trenches. *Journal of Economic Perspectives,* 31**,** 187-210.
GUVENEN, F., MATALONI JR, R. J., RASSIER, D. G. & RUHL, K. J. 2017. Offshore profit shifting and domestic productivity measurement. National Bureau of Economic Research.
HASKEL, J. & WESTLAKE, S. 2017. *Capitalism without capital: the rise of the intangible economy*, Princeton University Press.
HAUSMAN, J. A. 1996. Valuation of new goods under perfect and imperfect competition. *The economics of new goods.* University of Chicago Press.
HOUSEMAN, S., KURZ, C., LENGERMANN, P. & MANDEL, B. 2011. Offshoring bias in US manufacturing. *Journal of Economic Perspectives,* 25**,** 111-32.
HULTEN, C. R. 2015. Measuring the Economy of the 21st Century. *NBER Reporter***,** 1-7.
HULTEN, C. R., DEAN, E. R. & HARPER, M. 2001. *New developments in productivity analysis*, University of Chicago Press.
HULTEN, C. R. & NAKAMURA, L. 2017. Measuring GDP and Welfare in the 21st Century. Presented at CRIW/NBER conference Measuring and Accounting for Innovation in the 21st Century, March 2017.
JONES, C. I. & KLENOW, P. J. 2016. Beyond GDP? Welfare across countries and time. *American Economic Review,* 106**,** 2426-57.
JORGENSON, D. 2017. Productivity and Economic Growth in the World Economy: An Introduction. *International Productivity Monitor,* Fall**,** 1-7.
JORGENSON, D. W. 2018. Production and Welfare: Progress in Economic Measurement. *Journal of Economic Literature,* 56.
JORGENSON, D. W. & FRAUMENI, B. M. 1989. The accumulation of human and nonhuman capital, in: Lipsey, RE, Tice, HS,(Eds.), The Measurement of Savings, Investment and Wealth. The University of Chicago Press, Chicago, IL.

JORGENSON, D. W. & GRILICHES, Z. 1967. The explanation of productivity change. *The review of economic studies,* 34, 249-283.
JORGENSON, D. W. & SLESNICK, D. T. 2014. Measuring social welfare in the US national accounts. *Measuring Economic Sustainability and Progress.* University of Chicago Press.
LANCASTER, K. J. 1966. A new approach to consumer theory. *Journal of political economy,* 74, 132-157.
LEBOW, D. E., ROBERTS, J. M. & STOCKTON, D. J. 1994. Monetary Policy and the Price Level. Board of Governors of the Federal Reserve System.
LEBOW, D. E. & RUDD, J. B. 2003. Measurement error in the consumer price index: where do we stand? *Journal of economic literature,* 41, 159-201.
MOULTON, B. 2018. The Measurement of Output, Prices, and Productivity: What's Changed Since the Boskin Commission.
NAKAMURA, L. I., SAMUELS, J. D. & SOLOVEICHIK, R. H. 2017. Measuring the "free" digital economy within the GDP and productivity accounts. *In:* PHILADELPHIA, F. R. B. O. (ed.) *Research Department Working Papers.*
NATIONAL BUREAU OF ECONOMIC RESEARCH. PRICE STATISTICS REVIEW COMMITTEE., UNITED STATES. CONGRESS. JOINT ECONOMIC COMMITTEE. & UNITED STATES. BUREAU OF THE BUDGET. OFFICE OF STATISTICAL STANDARDS. 1961. *The price statistics of the Federal Government: review, appraisal, and recommendations; a report to the Office of Statistical Standards, Bureau of the Budget, together with twelve staff papers,* [New York].
NORDHAUS, W. D. & TOBIN, J. 1972. Is growth obsolete? *Economic Research: Retrospect and prospect, Volume 5, Economic growth.* Nber.
OLINER, S. D. & SICHEL, D. E. 1994. Computers and output growth revisited: How big is the puzzle? *Brookings Papers on Economic Activity,* 1994, 273-334.
OULTON, N. 2012. Hooray for GDP! GDP as a measure of wellbeing. *VoxEU. org,* 22.
OULTON, N. 2013. Has the growth of real GDP in the UK been overstated because of mismeasurement of banking output? *National Institute Economic Review,* 224, R59-R65.
OULTON, N. 2016. The mystery of TFP. *International Productivity Monitor*.
PAKES, A. 2003. A Reconsideration of Hedonic Price Indexes with an Application to PC's. *American Economic Review,* 93, 1578-1596.
SCHREYER, P. 2001. *OECD productivity manual: A guide to the measurement of industry-level and aggregate productivity growth*, Organisation for Economic Co-operation and Development.
SCHREYER, P. 2010. Towards measuring the volume output of education and health services.
SCHULTZE, C. L., MACKIE, C. D. & NATIONAL RESEARCH COUNCIL (U.S.). PANEL ON CONCEPTUAL MEASUREMENT AND OTHER STATISTICAL ISSUES IN DEVELOPING COST-OF-LIVING INDEXES. 2002. *At what price? : conceptualizing and measuring cost-of-living and price indexes,* Washington, DC, National Academy Press.

SCITOVSKY, A. A. 1964. An index of the cost of medical care-a proposed new approach. *The Economics of Health and Medical Care. Ann Arbor: Bureau of Public Health Economics, University of Michigan*.
SHAPIRO, M. D. & WILCOX, D. W. 1996. Mismeasurement in the consumer price index: An evaluation. *NBER macroeconomics annual,* 11**,** 93-142.
SHARPE, A. 2006. Symposium on the Boskin Commission Report After a Decade: Editor's Overview. *International Productivity Monitor***,** 1-4.
SICHEL, D. & VON HIPPEL, E. 2018. Household Innovation, R&D, and New Measures of Intangible Capital.
SICHEL, D. E. 1997. The productivity Slowdown: is a growing unmeasurable sector the culprit? *Review of Economics and Statistics,* 79**,** 367-370.
SMITH, A. 1963. *The Wealth of Nations, volume 1, 1776*, Richard D. Irwin.
SOLOW, R. 1987. We'd Better Watch Out. *New York Review of Books.*
STEVENSON, B. & WOLFERS, J. 2008. Economic growth and happiness: Reassessing the Easterlin paradox. *Brookings Papers on Economic Activity*.
STIGLITZ, J., SEN, A. & FITOUSSI, J.-P. 2009. The measurement of economic performance and social progress revisited. *Reflections and overview. Commission on the Measurement of Economic Performance and Social Progress, Paris*.
SYVERSON, C. 2011. What Determines Productivity? *Journal of Economic Literature,* 49**,** 326-65.
SYVERSON, C. 2017. Challenges to mismeasurement explanations for the US productivity slowdown. *Journal of Economic Perspectives,* 31**,** 165-86.
TRIPLETT, J. 1989. Price and Technological Change in a Capital Good: A Survey of Research on Computers. *In:* JORGENSON, D. & LANDAU, R. (eds.) *Technology and Capital Formation.* Cambridge, MA: MIT Press.
TRIPLETT, J. 2006. Handbook on hedonic indexes and quality adjustments in price indexes.
UNITED STATES. CONGRESS. SENATE. ADVISORY COMMISSION TO STUDY THE CONSUMER PRICE INDEX. 1996. *Toward a more accurate measure of the cost of living : final report to the Senate Finance Committee,* [Washington, D.C.?], Advisory Commission to Study the Consumer Price Index.
VON HIPPEL, E. 2017. *Free Innovation* MIT Press.
WEITZMAN, M. L. 1976. On the welfare significance of national product in a dynamic economy. *The Quarterly Journal of Economics,* 90**,** 156-162.