What Can Be Done to Measure Productivity Better?

Chad Syverson, University of Chicago Booth School of Business

Prepared for the ESRI International Conference; July 31, 2018

Productivity is the efficiency with which producers convert inputs into outputs. It is one of the most important economic concepts because of its multiple and tight connections to economic wellbeing.

In the aggregate, productivity growth essentially acts as the speed limit on long-run economic growth. Sustained changes in productivity growth rates shift potential per capita income growth rates one-for-one. If productivity growth slows (rises) persistently, long-run average per capita income growth would be expected to drop (increase) commensurately. Aggregate productivity differences are also hugely important across countries as well, as they account for most of the variation in per capita income across economies.

Productivity also matters greatly in more disaggregated settings. Productivity differences across sectors and industries drive large resource reallocations over both short and long horizons. Enormous and persistent productivity variation exists across companies, even within narrowly defined industries. These differences determine the fortunes of businesses, their employees, and their customers. Moreover, these disaggregate productivity patterns also shape the evolution of aggregate productivity through various mechanisms.

In short, it is all but impossible to think deeply about the supply side of an economy at any level without grappling with the sources and consequences of productivity. Given this, measuring productivity as accurately as possible should be a major priority for economists.

A recent example of productivity measurement's importance was the extensive academic and policy debate about the "mismeasurement hypothesis"—the notion that the mid-2000s slowdown in productivity growth worldwide was actually an illusion of mismeasurement. The hypothesis argued that economic statistics collection and compilation methods were unable to capture the true value of the new and better products of recent years.[1] Regardless of one's assessment of the hypothesis (I believe the evidence indicates it is not able to explain the vast majority of the slowdown), the fact that measurement issues were at least a plausible explanation for a reduction in output growth that totaled trillions of dollars a year indicates that productivity measurement is far from a trivial technical matter.

1. Productivity Measures

All productivity measures, boiled down to their essence, are output-to-input ratios. They can be defined at the micro level—that is, for an individual producer (or even for a particular production process), in which case productivity is the ratio of that producer's output to its inputs. Or productivity can be measured at a more aggregate level, such as for an industry, sector, or economy. In those cases, productivity is the ratio of the combined input of producers in the category to their combined inputs.

Single-factor productivity measures reflect units of output produced per unit of a particular input. The most common of these is labor productivity (e.g., output per employee or per worker-hour). Because single-factor productivity levels are affected by the intensity of use of the excluded inputs, researchers often use total factor productivity (TFP) (sometimes called multifactor productivity, MFP). TFP measures combine inputs in the productivity denominator in a way implied by theory to create an index that is invariant to movements along a production function isoquant. Factor intensity differences do not affect TFP because TFP reflects shifts in isoquants rather than movements along them. That is, a rise in TFP indicates a "pure" increase in output per unit of (combined) input.

Because productivity metrics are all output-to-input ratios, a sensible way to parse potential improvements in their measurement is to separately consider the challenges facing the measurement of outputs and inputs and, around this, discuss areas for potential improvement. This is how I proceed below.

Before doing so, however, it is worth a reminder that all productivity measures are in essence also residuals. They are the variation in output that cannot be explained by observable

[1] See, for example, Mokyr (2014); Hatzius and Dawsey (2015); Cardarelli and Lusinyan (2015); Byrne, Fernald, and Reinsdorf (2016); Nakamura, Samuels, and Soloveichik (2017); Syverson (2017).

inputs. Thus, as with all residuals, productivity is in some ways a measure of our ignorance, and productivity metrics can reflect potential sources of output variation—market power and measurement error are prime examples—that are not tightly conceptually related to a production function.

2. Challenges in Output Measurement

There are three basic potential problems with output measurement in productivity metrics. One arises because outputs are rarely measured directly but instead are typically computed as expenditures deflated by a price. In macroeconomic data, this can create a misalignment between implied and actual output growth due to price index inaccuracies. In micro data, there is often no micro-level price index at all, making necessary using expenditure itself as the output measure. A second difficulty, present in some settings, is even more basic: defining the appropriate output to measure. A third sticking point exists for a set of outputs that, while arguably of economic importance and the outcome of an orchestrated production process, are nevertheless uncounted as output in most official economic statistics. We consider each in turn below.

2.1. Expenditure-Based Output Measures

Outputs in economic data—both at the macro and micro level—are rarely directly measured quantities (e.g., board-feet of lumber, cars, patient-nights, tax returns filed). While these might be ideal in cases with physically homogeneous goods because they reflect the "pure quantity" concept of output from a production function, direct quantities are often impractical to collect or use. This can be due to incomparability across producers (a Cirrus SR22 and a Boeing 777 are both airplanes, but not equivalent outputs of a production process) or the need to aggregate across multiple outputs.

As a result, outputs are almost invariably measured from expenditures on goods, or equivalently, producers' revenues from selling those goods. The reason for this is practical; while expenditures are measured imperfectly, they are probably one of the most accurately tracked economic values.

In macroeconomic data (e.g., at the industry- or economy-wide level), output measures are obtained by deflating these expenditures by a price index for the corresponding good or bundle of goods. A price index is defined so that when expenditures are divided by the index, the quotient reflects the effective quantity of the good that enters consumers' utility functions (or, if the good is an intermediate input, producers' production functions). A correct index should reflect changes in the effective price of a unit of the good or bundle. This means the quality of output measures in productivity metrics depends heavily on the quality of price indexes. Any error in the index creates a mirror-image error in measured output. The fidelity between price indexes as measured and this conceptual definition almost certainly varies markedly across settings.[2]

The conceptual sources of such departures are well known, among them: problems measuring quality differences, an inability to fully incorporate the benefit of new products in the bundle, and substitution bias for fixed-weight indexes like the CPI.[3] Therefore settings with an abundance of quality change or product turnover are particularly vulnerable to price index mismeasurement, and thus output and productivity mismeasurement.

Statistical agencies do often incorporate treatments of quality change into their price indexes. Numerous approaches are used; Groshen et al. (2017) summarizes those used by the U.S. Bureau of Labor Statistics. Practical limitations in funding and staffing mean that not all issues known conceptually to be problematic can be fully treated. Some products are adjusted for quality change, with varying degrees of sophistication depending on the method being used, but many are not. This disparate application of quality adjustments can result in certain product classes or sectors having considerably different trajectories in the official price and output statistics even if, on the surface, other observables about the markets look similar. For example, Houseman (2018) argues that the large implied output growth in IT-related goods over the past three decades, driven in considerable part by the large implied price cuts derived from

[2] Note this applies not just to deflating expenditures across different time periods, but also to output and productivity comparisons across countries via purchasing power parity (PPP) indexes.

[3] See the Boskin Commission report (U.S. Congress, 1996) and Groshen et al. (2017) for thorough discussions of such issues in theory and practice. Goolsbee and Klenow (2018) offer a quantitative comparison of some of these issues between the traditional CPI and a digitally collected large-scope price index.

hedonically adjusted quality models for IT goods, may have distorted aggregate output measures.

It is worth considering how statistical agencies choose which limited set of products receive more intensive quality adjustment. Perceived priorities for particular cases have often driven these choices, but the process can end up being ad hoc. A more systematic approach would be for an agency to define evaluation criteria that balance the benefits and costs of quality adjustment for a product and then prioritize for adjustment products that best fulfill those criteria. This would help allocate scarce measurement resources to where they could have the most benefit.

In micro data, there are typically no producer-specific price indexes at all. Output is instead measured as producer revenue. This means that price differences remain in output and productivity measures.[4] This can pose obvious problems in interpreting differences in measured productivity levels and growth across producers. As noted in Foster, Haltiwanger, and Syverson (2008), to the extent that prices reflect idiosyncratic demand shifts or market power variation rather than quality or production efficiency differences, producers with high productivity measures may not be especially technologically efficient. This could lead to false inferences about the connections between productivity and outcomes like survival, growth, wages, and prices, among others. Attempts to parse micro-level revenue-based productivity measures into separate implied quantity-based productivity and price effects have been a major focus of productivity researchers over the past decade.

Issues of product turnover are equally difficult to handle. The current small-sample approaches of most statistical agencies cannot adequately measure product turnover—particularly not if one defines a “product” as a good-location-of-sale, which is arguably appropriate. While the precise effect of new varieties on the effective price of final or intermediate goods depends on the particular form of (respectively) the consumer’s utility function or producer’s production function, calculations with under reasonable assumptions

[4] Aggregate-level deflators are typically still applied to adjust for inflation over time, but the cross-sectional variation in price levels across producers remains embodied in the revenue-based output measure.

and for particular samples indicate the effects could be substantial. See Feenstra (1994) and Broda and Weinstein (2010), for example.

That said, there is not a lot of systematic evidence on rates of new good entry and product turnover among industries, over time, or across countries. Developing more evidence of this sort would be of great help in gauging how important and widespread of a source of mismeasurement this is. Furthermore, empirical estimations that identify more sharply the parameters of utility and production functions that are important for quantifying variety effects. Building them into price indexes should receive greater attention.

2.2. Difficulties Defining the Output

In some settings, the primary output measurement problem is even more fundamental than imperfect price indexes. Sometimes the most pressing issue properly defining what the sector's output actually is.

Consider the financial industry as an example. What is its output? What do financial firms "make"? Is it even possible to define an output metric that is comparable across, say, an insurance company and a bank? Not that narrowing the scope of analysis within the sector would solve the problem. Even just focusing on banks, measuring output by something like the total amount of credit extended would miss a host of other services that banks offer in the lending portions of their business, not to mention forgetting everything banks do on their deposit sides.

Healthcare is another example of a sector where simply defining a practical output measure is quite difficult. Conceptually, consumers' desired output from the healthcare industry is *health*, but there is no straightforward way to measure this. There is often a lag between the purchase of healthcare services and any change in health, a lot of noise in any causal relation between the two, and moreover there is typically no easily observable cardinal measure of health.

Other sectors with similarly difficult-to-define outputs include education, government, and nonprofit organizations.

Of course, activity in these sectors is often measured *somehow*. A common practice is to measure output through input use. This is a nod to practicality. While measuring outputs of financial services, health, education, and governance may be very difficult, it is more straightforward to measure employees, payroll, and tangible capital stocks in each of the sectors. Levels and changes in outputs are imputed from levels and changes in these inputs. While this may be a reasonable stop-gap solution for measures of output like GDP (output is almost surely correlated with inputs), this poses real problems for productivity measurement. Because productivity is a ratio of output to inputs, any output measure that is closely based on input use necessarily makes strong assumptions about productivity—including, at the extreme, that productivity is constant, so input growth equals output growth.

Some efforts are underway to come up with new ways to measure something closer to what might be considered the conceptually correct output measure in such sectors. An example is a recent change the U.S. Bureau of Labor Statistics (BLS) made in the measurement of output in the primary and secondary education sectors. Formerly, output was measured simply by the total number of students enrolled. While it is sensible that the education sector's conceptual output (a reasonable definition might be the addition to its students' combined stocks of human capital) is positively related to enrollment, it is also easy to recognize that enrollment is far from being a sufficient statistic for this output. The BLS has started to move in the direction of a more inclusive output measure by augmenting enrollment for quality change—namely, adjusting enrollment by attendance rates and then weighting the result based on changes in a commonly employed benchmark exam, the National Assessment of Educational Progress.[5]

Such "quantity-augmentation" approaches are one way to try to move output metrics, which are necessarily constrained by what is measureable, closer to what would ideally be measured. These approaches may be easier to implement in aggregates than in micro-level measures, however. Education, analogous to health as discussed above, often involves lags between purchase and outcomes, noisy connections between them, and limited cardinal metrics of knowledge. In aggregate data, these limitations pose less of a problem because

[5] This example makes clear how there can sometimes be a connection between the define-the-output measurement challenge and the quality-adjustment challenge. Sometimes part of the difference between these two—the measurable and the conceptual—is interpretable as an unmeasured quality difference.

person-level noise and time lags are aggregated over, creating a tighter correspondence between *average* levels of inputs and *average* educational attainment. Similarly, in health, healthcare services quantities can be augmented by, for example, quality-adjusted life expectancies.

There are reasons to believe the fundamental output definition problems discussed in this section are becoming more acute. Extending a thread first drawn out by Griliches (1994), I divide the U.S. economy into “measurable” and “unmeasurable” sectors. The former include agriculture, mining, utilities, and manufacturing. The latter is everything else.[6]

Figure 1 shows the evolution between 1947 and 2017 of the shares of each of these sectors in U.S. GDP. The share of value added taking place in the measurable sectors has been clearly shrinking throughout this time. The measurable sectors totaled 37.1% of GDP in 1947. By 2017 this had fallen by well over half, to 15.7%. The drop was driven mostly by the shrinking shares of the agriculture sector and especially manufacturing, which saw its share of GDP fall from 25.4% in 1947 to 11.6% in 2017, accounting for about two-thirds of the total drop.

There has been a clear shift in economic activity toward sectors where defining and measuring output is more difficult. In other words, the modal unit of output in the economy is becoming decreasingly measurable. This trend portends an increasingly difficult task for economic measurement practitioners as time rolls forward.

2.3. Uncounted Outputs

A further difficulty in output measurement exists for those cases where outputs are completely unmeasured. These cases involve economic goods produced by the application of scarce resources but that, for one reason or another, are not captured in current standard economic statistics.

The value of home production is a prominent example. Meal preparation, house cleaning, DIY repair, childcare, and so on would be measured as output if they occurred as the result of a market transaction, but are ignored if done by a household member. Another

[6] Note that these are gross discretizations. The measurable sectors are only very imperfectly so, and the unmeasurable sectors hold some ability to be measured. The difference between the two categories is one of degree.

example includes the gains (or losses) in environmental goods such as clean air and water or species diversity. This uncounted output problem also applies to the omission of a more direct measure of health in the healthcare sector.

There is reason to believe missing outputs of these types are large. The Bureau of Economic Analysis estimates that household production in the U.S. totaled $3.8 trillion in 2010, 26% of that year's $14.6 trillion GDP. This is a lot of economic activity—much of it a very close substitute to market activity—that is not systematically tracked due to convention and measurement difficulties. Changes in this value over time raise interesting issues as well. The BEA calculated that household production in 1965 was a notably larger share of measured GDP, 39%. This implies that the shift from household production to market activity over the past several decades has caused GDP growth to overstate the true change in economic activity.

For environmental goods, Eurostat estimates the EU-28 countries annually spend an average of 2.2% of GDP on environmental protection. (This includes, for example, moving and treating solid waste and abating emissions of waterborne and airborne pollutants.) By revealed preference, this indicates nontrivial valuations for environmental goods. Environmental quality has broadly improved over the last four decades in many economies but is not considered in traditional national accounts measures.

Regarding health, life expectancy in the U.S. in 1970 was 70.8 years; by 2015, it had grown to 78.8 years. This increase of 0.178 years per year is absent from national income accounts, though it arguably creates a considerable increase in economic wellbeing. Using the modest estimate of $100,000 to roughly approximate the dollar-equivalent welfare value of a quality-adjusted-life-year, this implies the average American saw an additional $17,800 in effective wealth (in 2015 dollars) per year over this period. By way of comparison, real GDP per capita growth from 2014 to 2015 was about $1600, an order of magnitude smaller.

An additional category of missing output receiving more attention lately is the value of "free" digital goods such as Google, Facebook, and Snapchat. Consumers do not engage in monetary transactions when they use these products, so their searches, posts, and views do not enter GDP calculations. This measurement shortcoming, while salient in recent digital goods, is not new; for over 60 years, consumers have enjoyed considerable utility from watching television without paying for it, as they did still decades earlier for radio and many other kinds of free media.

However, it is important to recognize that there *are* measured transactions associated with these goods that enter output. As Nakamura, Samuels, and Soloveichik (2017) explain, the provision of these products is supported by advertising revenues, all measured in national statistics. Further, as I point out in Syverson (2017), consuming these goods requires the purchase of complementary goods like smart phones, iPads, broadband access, and mobile telephony. Sellers of those complements should be pricing consumers' values of "free" goods into their own prices—which again are counted as outputs.

Nakamura, Samuels, and Soloveichik (2017) propose a method for imputing consumer-side shadow values for such goods. It uses advertising revenues as a starting point, in effect treating use of such goods as a barter of ad-watching for content. If the user is a consumer, the imputed value of the content is personal consumption. If the user is a business, it is an intermediate input. Using this approach, they find rather modest implied changes to national accounts, with an estimated change in real GDP growth of -0.002% per year over 1929-1998 and +0.009% per year for 1998-2012. While these are small changes, the logic of the approach is economically sensible and implementation costs do not appear to be overwhelming. Additional research into the use and effects of similar approaches for these kinds of goods seems worthwhile.

3. Challenges in Input Measurement

The most commonly computed productivity measures are for labor productivity, where the denominator of the productivity metric is some measure of labor inputs (employees, hours, or either of these adjusted for some labor quality factor). I therefore start discussing input measurement issues with those tied to labor.

3.1. Labor

One of the trickiest issues in accurately measuring labor inputs is variations in worker quality. Workers (or worker-hours, as the case may be) may well be heterogeneous, and treating them as homogenous in productivity measures to production function shifts what would instead be increases in quality-adjusted labor inputs. An enormous research literature has documented multiple potential dimensions of labor heterogeneity. These include, among others, education, training, overall experience, and tenure at a firm.

For more aggregate labor inputs measures, one might merge employment quantities with similarly aggregate data on workers' average levels of education, training, experience, etc. This builds a quality-adjusted labor series. Of course this procedure requires one to model the weights that each human capital component has on overall human capital of the relevant labor force. Examples of this process are in Acheson and Franklin (2012), Nomura and Akashi (2017), and Bosler et al. (forthcoming).

Another common approach to capture differences in labor quality in productivity measures, especially in microdata, is to measure labor inputs using the wage bill—the product of worker-hours and the average wage, i.e., the entire expenditure on labor. The idea is that market wages reflect variations in workers' marginal products, so the wage bill serves as a quality-adjusted labor input. This approach is not foolproof; wage variation might reflect the competitive structure of local labor markets, or causation could be in the other direction if more productive firms share rents with employees. Hence, more direct labor quality measures are needed to definitively pin down labor quality's productivity contribution. Interestingly, however, the results in Fox and Smeets (2011) suggest that the wage bill captures much of the labor quality variation that exists, even as measurable in highly detailed matched employer–employee data.

Total factor productivity measures require measurement of non-labor inputs. In that case, potential measurement problems for intermediate inputs and capital come into play, as do imperfections of estimating the factor elasticities used to combine separate inputs into a composite input, a necessary input into constructing TFP.

3.2. Intermediate Inputs

As with outputs, quantities of intermediate inputs are usually measured as expenditures divided by a price. This raises a need for accurate input price indexes and can fall short in the same ways as described for output price indexes above, like product (input) turnover and quality differences. In fact it may be that such problems may be more acute for input price indexes than output price indexes, given the frequency of input turnover and the large amount of customization in some industries.

Groshen et al. (2017), along with its overview of final-goods quality adjustment practices, also details what the U.S. BLS does about quality adjustment in producer price indexes, which tend to be heavy with input prices. As with quality adjustment for final goods, a rationalization of these practices given resource constraints may be useful.

3.3. Capital

Capital could be the most mismeasurement-prone component of any productivity measure. It holds multiple sources of potential measurement problems, each of plausibly substantial size.

First, capital can exhibit considerable unmeasured quality variation. For example, capital vintages might vary in the extent to which they embody the latest in technological progress. Capital quality measurement poses conceptually similar problems to the labor quality measurement issues discussed above. As a practical matter, however, capital measurement can be much more difficult to tackle. For labor, there are often measurable auxiliary correlates of quality, such as education and experience, that can be used to adjust measured labor units (employees or hours) for quality differences as discussed above. Capital, on the other hand, has a dearth of analogous proxies for quality.

Second, practitioners are often forced to use a capital *stock* to measure capital input. But the true capital input into production is the *flow* of services that capital provides. This means differences in capital utilization rates (whether across firms, industries, or time) create measurement error. A given stock of capital can be providing a high flow of capital inputs when used intensively, or a low flow when not. But a typical productivity measure that uses the stock to measure inputs would deem capital inputs in both cases as equal.

Third, capital stocks are often built up using the perpetual inventory method. (That is, the future period’s capital stock equals this period’s stock, minus depreciation, plus current investment.) Both the depreciation rate and current investment are potential conduits for mismeasurement here. The depreciation rate depends on the mix of capital the producer uses and the intensity with which it is used. Both of these are typically unobserved, creating measurement error in the constructed capital stock. Furthermore, capturing the correct level of investment requires obtaining a price deflator that ensures an effectively equivalent dollars’ worth of investment today adds an equivalent amount of capital to the stock as in any other year. Such investment deflators are probably more noise laden than output deflators.

Fourth, some of the capital used in production may not just be measured with error; it may not be measured at all. Intangible capital (brand value among a customer base, production know-how, organizational culture, relationships with suppliers or distributors, and so on) plays an important role in explaining producers’ real outputs in many markets. Yet, by nature, these kinds of capital are typically unmeasured. Their production contributions are therefore attributed to productivity. While this can be managed in part by interpreting productivity as including the effect of intangibles, it can cloud important questions about the mechanisms through which intangibles add to production, how they vary across producers or over time, and the extent to which measured productivity is affected by factors *other than* intangible capital.

3.4. Estimating Output Elasticities

Total factor productivity measures require disparate inputs (labor, capital, and intermediates) to be combined into a single, composite input that forms the denominator of the TFP measure.

The precise way to combine the inputs depends on the production function. The most common method is to weight each input by the elasticity of output with respect to that input, so the (logged) composite input is equal to the output-elasticity weighted sum of the logged individual inputs. This summation is exactly correct for a Cobb-Douglas production function. It is also, conveniently, the correct first-order approximation for any general production function.

Thus I couch the discussion in terms of this practice, but the discussion here applies more generally when other production functions or higher-order approximations are used.

None of the output elasticities are directly observable; they are properties of the production function. Hence they must somehow be estimated, either of which can introduce measurement error.

One approach uses fact that cost minimization implies producers should equate an input's output elasticity with the product of that input's cost share and the scale elasticity. This of course requires measuring inputs' cost shares, posing its own measurement problems. In some cases, measuring cost shares is less measurement prone than measuring the quantities of inputs, because the former are based on expenditures, which tend to be directly measured and not rely on separate accurate measurement of price indexes. The more difficult exception is often in obtaining capital costs. It is rare to have direct measures of expenditures on capital inputs, especially if that capital is owned rather than rented (necessitating the estimation of implicit rental rates).

The cost-minimization approach to output elasticity measurement is not assumption-free. The cost minimization it assumes is based on the static first-order condition. If there are factor adjustment costs, the link between observed cost shares and the needed output elasticities will not hold at any given moment because firms will be operating at input levels away from their long-run desired level. This misspecification error is mitigated in part by using cost shares that have been averaged over time (or in micro settings, producers as well). Such averaging smooths out idiosyncratic adjustment-cost-driven misalignments between actual and optimal input levels. (The notion being that input levels are optimal on average, though individual producers might be operating with idiosyncratically high or low inputs.) That said, some mismeasurement is likely to remain.

A second approach to obtain output elasticities is to estimate the production function by regressing output on inputs. The estimated coefficients are (depending on the functional form assumptions) either direct estimates of the elasticities or invertible functions thereof, and total factor productivity is the estimated residual from the production function. There are econometric challenges involved in this approach. Marschak and Andrews (1944) first pointed out the so-called "transmission bias": producers' input choices are likely to be correlated with their productivity levels. In settings where productivity is being measured at the firm-level, an additional potential selection bias is present, because less efficient producers are more likely to exit from the sample. A considerable literature has attempted to address these concerns (e.g., Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg, Caves, and Frazer, 2015; Gandhi, Navarro, and Rivers, forthcoming). None of these approaches, of course, is perfect. Each introduces potential measurement error through both model misspecification (i.e., the assumptions of the model used to avoid transmission and selection biases do not hold in the data) and sampling error. More recent work in production function estimation has tried to incorporate this treatment of input endogeneity while also tackling additional estimation problems like capital mismeasurement (Collard-Wexler and De Loecker, 2016) and multi-product outputs (Goldberg et al., 2016; Orr, 2018).

4. Another Measurement Issue: Market Power

There is an additional complication in productivity measurement that, while known about for some time (e.g., Hall 1988), has received increasing attention in the research literature (though less in official statistics): market power.

When firms have market power in their product markets, it creates a wedge between changes in inputs and changes in outputs. In a perfectly competitive output market, an input is paid a share of output equal to its output elasticity. When markets are imperfectly competitive, inputs instead receive income shares equal to their elasticity divided by the markup (the producer keeps the excess income as a rent). This means TFP measures from regressions of output on inputs can embody not just the "technology" concept of TFP (that is, as a production function shifter), but also market power variation.[7] As a result, demand-side market forces enter the measurement of what is conceptually a supply-side primitive.[8]

[7] The cost-minimization index-number approach to TFP measurement does account for these market power effects when inputs' shares of total *costs*, rather than shares of *revenues*, are used as the estimated elasticities. This is because the cost-minimization condition underlying the use of such shares accounts for potential market power.

[8] This is related to, though distinct from, the issue of demand variations entering productivity measures when revenues are used as a measure of output, typically due to lack of producer-specific price information.

National statistical agencies do not currently make any adjustments to productivity measures for market power effects. Users must therefore be mindful of the potential inference issue that these measures present. There is a practicality constraint; adjusting published measures for market power is not simple. However, further exploration of potential adjustments makeable within limited time and resource budgets would be worthwhile.

A useful aspect of this measurement issue is that the relationship can be turned on its head to use production data to obtain measures of the magnitude of market power. Hall (1988) first implemented this strategy in industry data; a more recent example using microdata is De Loecker and Warzynski (2012). Researchers have used these metrics to explore both the causes of market power and their effects on producers' and industries' behavior and trajectories.

5. Priority Areas to Address

This essay has considered a multitude of potential sources of productivity mismeasurement. In this section, I outline three areas that I suggest would be worthy of priority attention from both researchers and practitioners going forward.

Improve and Expand Price Data. High quality price data are critical for accurate productivity measurement. This is true at both the macro and micro levels.

For aggregates, improving price indexes would offer considerable benefits in terms of measuring productivity at the economy-wide and industry levels. A rationalization of the type described above regarding which products should receive priority when it comes to quality adjustment would be helpful, as would programs to measure product turnover more systematically. Leveraging existing digital pricing information (e.g., Goolsbee and Klenow, 2018) may allow current price measurement programs to be augmented in these ways without imposing considerable extra costs on already budget-strapped government agencies.

At the micro level, having producer-specific prices opens up an entirely new set of issues that can be studied and understood about not just how productivity varies and evolves at the micro level, but also how firms', workers', and consumers' fortunes are affected by both supply- and demand-side phenomena. Detailed price information are available in some limited cases (heavily used by researchers because of the importance of the underlying issues), but these data are the exception rather than the rule.

Expand the Set of Measured Outputs. Efforts to expand the set of outputs that are collected would pay substantial dividends in terms of our understanding of productivity at both the macro and micro levels. There are two areas that arguably deserve particular priority: health outputs and household production.

Given the enormous amount of resources that the economy applies toward producing healthcare, it is an easy case to make that we should improve how we measure what the sector actually producers. At the micro- (e.g., hospital-) level, sensible and feasible outcomes for measurement might be survival and/or readmission rates. At more aggregate levels, this might be quality-adjusted life expectancy.

Household production is enormous in aggregate (see the estimates above) and plays an important role in understanding changes in income over time as well as, likely, differences in income across households. It was not included in national income accounts on the basis of an early judgment call tied in part to practical measurement limitations. Economically, the case for their inclusion is strong. Moreover, some basic building blocks of a household production measurement system are already in place. Time-use surveys offer important information that can be used to build aggregate household production accounts (depending on the size of the sample, geographic breakouts may be possible as well). This is already been done on an experimental basis; see for example the use of the American Time Use Survey in Bridgman (2016).

Improve Capital Measurement, Especially Intangibles. As noted above, capital is probably the most difficult input to measure, and as such considerable improvements to productivity measures can be obtained through better capital measurement.

Perhaps the most pressing of the capital measurement issues is the treatment of intangibles. Intangible capital has become a highly important factor (in both the economically literal and figurative senses of that word) on the economy's supply side, and continues to grow in importance. An impressive array of productivity-related phenomena are closely tied to the rise of intangibles (see Haskel and Westlake, 2017, for a review).

None of these proposed expansions of the scope and quality of productivity measurement would be trivial to implement. The fact that they have still been left unfinished after decades of careful and important measurement work indicates that they are inherently among the more difficult issues to address. Progress will require considerable research and organization building. However, these challenges are matched by a higher expected return to collecting and analyzing such data—a return that should pay for itself many times over in terms of our understanding of productivity and the economy.

References

Acheson, Jean and Mark Franklin. 2012. “Quality-Adjusted Labour Input: Estimates to 2010.” Office of National Statistics (UK) report.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. “Identification Properties of Recent Production Function Estimators.” *Econometrica*, 83(6): 2411–2451.

Bosler, Canyon, Mary C. Daly, John G. Fernald, and Bart Hobijn. Forthcoming. “The Outlook for U.S. Labor-Quality Growth.” In *Education, Skills, and Technical Change: Implications for Future US GDP Growth*, Charles R. Hulten and Valerie A. Ramey (eds.). Chicago: University of Chicago Press.

Bridgman, Benjamin. 2016. “Accounting for Household Production in the National Accounts: An Update, 1965–2014.” *Survey of Current Business*, 96(2): 23–36.

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

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*, *Spring*, 109-182.

Cardarelli, Roberto, and Lusine Lusinyan. 2015. “U.S. Total Factor Productivity Slowdown: Evidence from the U.S. States.” IMF Working Paper 15/116.

Collard-Wexler, Allan and Jan De Loecker. 2016. “Production Function Estimation with Measurement Error in Inputs.” NBER Working Paper No. 22437.

De Loecker, Jan, Penelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik. 2016. “Prices, Markups, and Trade Reform.” *Econometrica*, 84(2): 445-510.

De Loecker, Jan and Frederic Warzynski. 2012. “Markups and Firm-Level Export Status.” *American Economic Review*, 102(6): 2437-71.

Feenstra, Robert C. 1994. “New product varieties and the measurement of international prices.” *American Economic Review*, 84(1): 157–77.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*, 98(1): 394-425.

Fox, Jeremy T. and Valérie Smeets. 2011. "Does Input Quality Drive Measured Differences in Firm Productivity?" *International Economic Review*, 52(4): 961-89.

Gandhi, Amit, Salvador Navarro, David Rivers. Forthcoming. "On the Identification of Gross Output Production Functions." *Journal of Political Economy*.

Goolsbee, Austan and Peter J. Klenow. 2018. "Internet Rising, Prices Falling: Measuring Inflation in a World of E-Commerce." NBER Working Paper 24649.

Griliches, Zvi. 1994. "Productivity, R&D, and the Data Constraint." *American Economic Review*, 84(1): 1–23.

Groshen, Erica L., Brian C. Moyer, Ana M. Aizcorbe, Ralph Bradley, and David M. Friedman. 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(2): 187–210.

Hall, Robert E. 1988. "The Relation between Price and Marginal Cost in U.S. Industry." *Journal of Political Economy*, 96(5): 921-47.

Haskel, Jonathan and Stian Westlake. 2017. *Capitalism without Capital: The Rise of the Intangible Economy*. Princeton: Princeton University Press.

Hatzius, Jan, and Kris Dawsey. 2015. "Doing the Sums on Productivity Paradox v2.0." *Goldman Sachs U.S. Economics Analyst*, No. 15/30.

Houseman, Susan N. 2018. "Understanding the Decline of U.S. Manufacturing Employment." Upjohn Institute for Employment Research Working Paper.

Levinsohn, James and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies*, 70(2): 317–342.

Marschak, Jacob, and William H. Andrews. 1944. "Random Simultaneous Equations and the Theory of Production." *Econometrica*, 12(3–4): 143–205.

Mokyr, Joel. 2014. "Secular Stagnation? Not in Your Life." Chap. 6 in *Secular Stagnation: Facts, Causes and Cures*, Coen Teulings and Richard Baldwin (eds.). London: CEPR Press.

Nakamura, Leonard I., Jon Samuels, and Rachel H. Soloveichik. 2017. "Measuring the 'Free' Digital Economy within the GDP and Productivity Accounts." FRB of Philadelphia Working Paper No. 17-37.

Nomura, Koji and Naoyuki Akashi. 2017. "Measuring Quality-adjusted Labor Inputs in South Asia, 1970–2015." KEO Discussion Paper No. 143.

Olley, G. Steven and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64(6): 1263–1297.

Orr, Scott. 2018. "Productivity Dispersion, Import Competition, and Specialization in Multi-product Plants." Working Paper.

Syverson, Chad. 2017. "Challenges to Mismeasurement Explanations for the US Productivity Slowdown." *Journal of Economic Perspectives*, 31(2): 165-186.

U.S. Congress. 1996. "Toward a More Accurate Measure of the Cost of Living: Final Report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index." 104th Congress, 2nd sess., S.Prt. 104-072. http://www.ssa.gov/history/reports/boskinrpt.html

Figure 1. Share of U.S. GDP by Sector ("Measurable" Sectors Broken Out)