Preliminary. Please do not cite or circulate.

Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics*

Erik Brynjolfsson, MIT Sloan School of Management and NBER
Daniel Rock, MIT Sloan School of Management
Chad Syverson, University of Chicago Booth School of Business and NBER

September 2017

Abstract. We live in an age of paradox. Systems using artificial intelligence match or surpass human level performance in more and more domains, leveraging rapid advances in other technologies and driving soaring stock prices. Yet measured productivity growth has fallen in half over the past decade, and real income has stagnated since the late 1990s for a majority of Americans. We describe four potential explanations for this clash of expectations and statistics: false hopes, mismeasurement, redistribution, and implementation lags. While a case can be made for each explanation, we argue that lags are likely to be the biggest reason for paradox. The most impressive capabilities of AI, particularly those based on machine learning, have not yet diffused widely. More importantly, like other general purpose technologies, their full effects won't be realized until waves of complementary innovations are developed and implemented. The adjustment costs, organizational changes and new skills needed to for successful AI can be modeled as a kind of intangible capital. A portion of the value of this intangible capital is already reflected in the market value of firms. However, most national statistics will fail to capture the full benefits of the new technologies and some may even have the wrong sign.

*Contact information: Brynjolfsson: erik.brynjolfsson@gmail.com; MIT Sloan School, E62-414, 100 Main Street, Cambridge, MA 02142; Rock: drock@mit.edu; 100 Main Street, Cambridge, MA 02142; Syverson: chad.syverson@chicagobooth.edu; University of Chicago Booth School of Business, 5807 S. Woodlawn Ave., Chicago, IL 60637.

The discussion around the recent patterns in aggregate productivity growth highlights a seeming contradiction. On the one hand, there are astonishing examples of potentially transformative new technologies that could greatly increase productivity and economic welfare (see e.g. Brynjolfsson and McAfee, 2014). There are some early concrete signs of these technologies' promise, the recent leaps in artificial intelligence (AI) performance being the most prominent example. However, at the same time, measured productivity growth over the past decade has slowed significantly. This deceleration is large, cutting productivity growth by half or more of its level in the decade preceding the slowdown. It is also widespread, having occurred throughout the OECD and, more recently, among many large emerging economies as well (Syverson 2017).[1]

We thus appear to be facing a redux of the Solow (1987) Paradox: we see transformative new technologies everywhere but in the productivity statistics.

In this paper, we review the evidence and explanations for the modern productivity paradox and propose a resolution. Namely, that there is no inherent inconsistency between forward-looking technological optimism and backward-looking disappointment. Both can simultaneously exist. Indeed, there are good conceptual reasons to *expect* them to simultaneously exist when the economy undergoes the kind of restructuring associated with transformative technologies. In this paper we argue and present some evidence that the economy is in such a period now.

Sources of Technological Optimism

Paul Polman, Unilever's CEO, recently claimed that "The speed of innovation has never been faster." Similarly, Bill Gates, Microsoft's co-founder observes that "Innovation is moving at a scarily fast pace." Vinod Khosla of Khosla Ventures sees "the beginnings of... [a] rapid acceleration in the next 10, 15, 20 years". Eric Schmidt, Executive Chairman of Alphabet Inc., believes "we're entering… the age of abundance [and] during the age of abundance, we're going to see a new age… the age of intelligence". Ray Kurzweil famously predicts that The Singularity, when AI surpasses humans, will occur sometime around

[1] A parallel yet more pessimistically oriented debate about potential technological progress is the active discussion about robots taking jobs from more and more workers (e.g., Brynjolfsson and McAfee, 2011; Acemoglu and Restrepo, 2017; Bessen, 2017; Autor and Salomons, 2017).

2045.[2] Assertions like these especially are common among technology leaders and venture capitalists.

In part, these reflect the continuing progress of IT in many areas, from core technology advances like further doublings of basic computer power (but from ever larger bases) to successful investment in the essential complementary innovations like cloud infrastructure, and new service-based business models. But the bigger source of optimism is the wave of recent improvements in AI, especially machine learning. Machine learning represents a fundamental change from the first wave of computerization. Historically, most computer programs succeeded by meticulously codifying human knowledge, step-by-step, mapping inputs to outputs as prescribed. In contrast, machine learning systems figure out the relevant mapping on their own, typically by being fed very large data sets of examples. Using these methods, machines have made impressive gains in perception and cognition, two essential skills for most types of human work. For instance, error rates in labeling the content of photos on ImageNet, a dataset of over 10 million images, have fallen from over 30% in 2010 to less than 5% in 2016 and most recently as low as 2.2% with SE-ResNet152 in the ILSVRC2017 competition (see Figure 1).[3] Error rates in voice recognition on the Switchboard speech recording corpus, often used to measure progress in speech recognition, have improved from 8.5% to 5.5% over the past year (Saon et al. 2017). The five percent threshold is important, because that is roughly the performance of humans at each of these tasks on the same test data.

While not at professional human performance yet, Facebook's AI Research team recently improved upon the best machine language translation algorithms available using convolutional neural net sequence prediction techniques (Gehring et al. 2017). Deep learning techniques have also been combined with reinforcement learning, a powerful set

[2] http://www.khoslaventures.com/fireside-chat-with-google-co-founders-larry-page-and-sergey-brin

https://en.wikipedia.org/wiki/Predictions_made_by_Ray_Kurzweil#2045:_The_Singularity

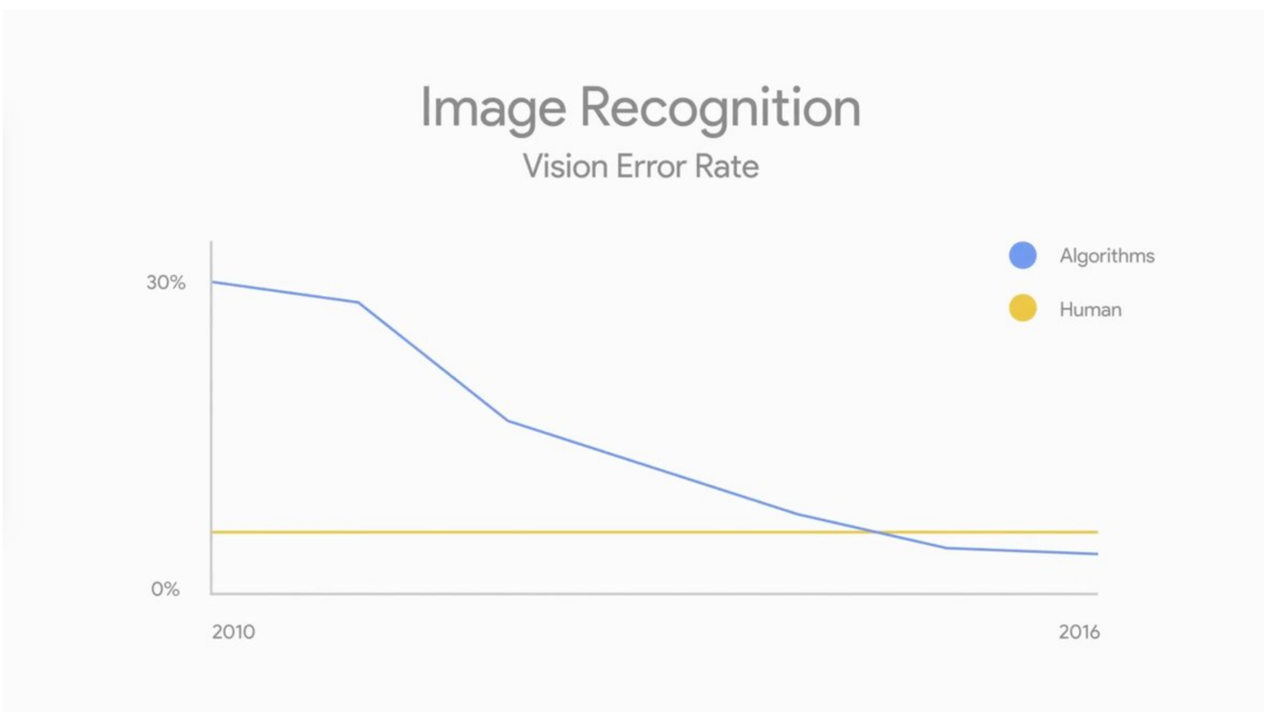
https://www.theguardian.com/small-business-network/2017/jun/22/alphabets-eric-schmidt-google-artificial-intelligence-viva-technology-mckinsey

[3] http://image-net.org/challenges/LSVRC/2017/results. ImageNet includes labels for each image, originally provided by humans. For instance, there are 339,000 labeled as flowers, 1,001,000 as food, 188,000 as fruit, 137K as fungus, etc.

of techniques used to generate control and action systems whereby autonomous agents are trained to take actions given an environment state to maximize future rewards. Though nascent, advances in this field are impressive. In addition to their victories in the game of Go, Google DeepMind has achieved superhuman performance in many Atari games (Fortunato et al. 2017).

These are notable technological milestones. But they can also change the economic landscape, creating new opportunities for business value creation and cost reduction. For example, a system using deep neural networks was tested against 21 board certified dermatologists and matched their performance in diagnosing skin cancer (Esteva et al. 2017). The neural networks at Facebook are used for over 4.5 billion translations each day.[4]

Figure 1. AI vs. Human Image Recognition Error Rates



An increasing number of companies have responded to these opportunities. Google now describes its focus as “AI first”, while Microsoft’s CEO Satya Nadella says AI is the

[4] https://code.facebook.com/posts/289921871474277/transitioning-entirely-to-neural-machine-translation/]

"ultimate breakthrough" in technology. Their optimism about AI is not just cheap talk. They are making heavy investments in AI, as are Apple, Facebook, and Amazon. As of September 2017, these companies comprise the five most valuable companies in the world. Meanwhile the tech-heavy Nasdaq composite stock index more than doubled between 2012 and 2017. According to CBInsights, global investment in private companies focused on AI has grown even faster, increasing from $589 million in 2012 to over $5 billion in 2016.[5]

The Disappointing Recent Reality

While the technologies discussed above hold great potential, there is little sign that they have yet affected aggregate productivity statistics. Labor productivity growth rates in a broad swath of developed economies fell in the mid-2000s and have stayed low since then. For example, aggregate labor productivity growth in the U.S. averaged only 1.3% per year from 2005 to 2016, less than half of the 2.8% annual growth rate sustained over 1995 to 2004. Fully 28 of 29 other countries for which the OECD has compiled productivity growth data saw similar decelerations. The unweighted average annual labor productivity growth rates across these countries was 2.3% from 1995 to 2004 but only 1.1% over 2005 to 2015.[6] What's more, real median income has stagnated since the late 1990s and non-economic measures of well-being, like life expectancy, have fallen for some groups (Case and Deaton, 2017)

Figure 2 replicates the Conference Board's analysis of its country-level Total Economy Database (Conference Board, 2016). It plots highly smoothed annual productivity growth rate series for the U.S., other mature economies (which combined match much of the OECD sample referred to above), emerging and developing economies, and the world overall. The aforementioned slowdowns in the U.S. and other mature economies are clear in the figure. The figure also reveals that the productivity growth acceleration in emerging

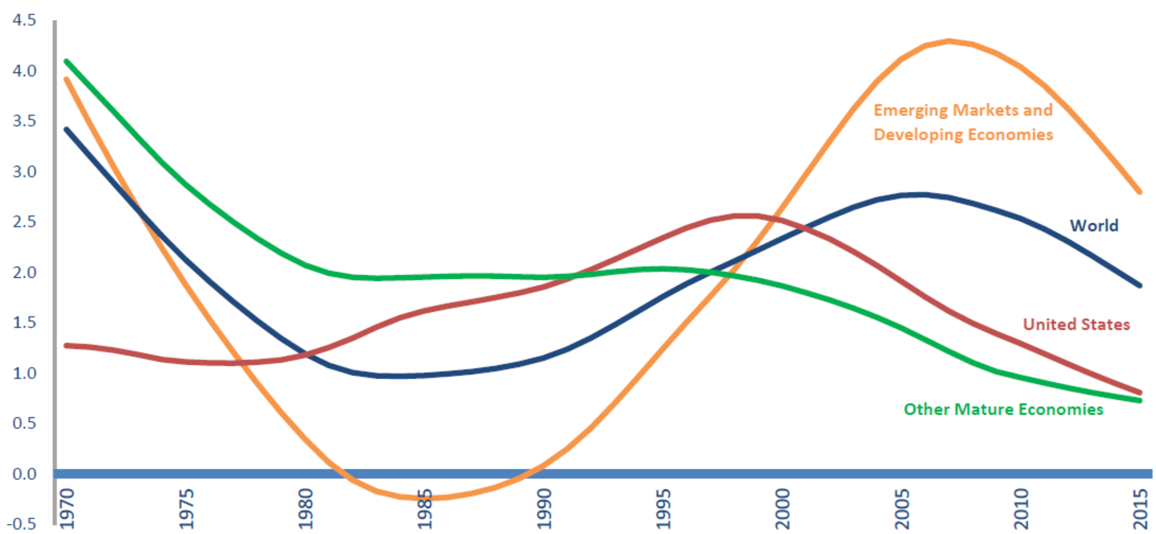
[5] And the number of deals increased from 160 to 658. See https://www.cbinsights.com/research/artificial-intelligence-startup-funding/

[6] These slowdowns are statistically significant. For the U.S., where the slowdown is measured using quarterly data, equality of the two periods' growth rates is rejected with a t-statistic of 2.9. The OECD numbers come from annual data across the 30 countries. Here, the null hypothesis of equality is rejected with a t-statistic of 7.2.

and developing economies during the 2000s ended around the time of the Great Recession, causing a recent decline in productivity growth rates in these countries too.

These slowdowns do not appear to simply reflect effects of the Great Recession. In the OECD data, 28 of the 30 countries still exhibit productivity decelerations if 2008-09 growth rates are excluded from the totals. Cette, Fernald, and Mojon (2016), using other data, also find substantial evidence that the slowdowns began before the Great Recession.

Figure 2. Smoothed Average Annual Labor Productivity Growth (Percent) by Region



Source: The Conference Board Total Economy Database™ (Adjusted version), November 2016.
Notes: Trend growth rates are obtained using HP filter, assuming a l=100.

Both capital deepening and total factor productivity (TFP) growth lead to labor productivity growth, and both seem to be playing a role in the slowdown (e.g., Fernald 2014 and OECD 2015). Disappointing technological progress can be tied to each of these components. TFP directly reflects such progress. Capital deepening is indirectly influenced by technological change because firms' investment decisions respond to improvements in current or expected capital quality.

These facts have been read by some as reasons for pessimism about the ability of new technologies like AI to greatly affect productivity and income. Gordon (2015) argues that productivity growth has been in long-run decline, with the IT-driven acceleration of 1995 to 2004 being a one-off aberration. While not claiming technological progress will be nil in the coming decades, Gordon essentially argues that we have been experiencing the new, low-growth normal and should expect to continue to do so going forward. Cowen (2011) similarly offers multiple reasons why innovation may be slow at least for the foreseeable future. Bloom et al. (2017) document that in many fields of technological progress, research productivity has been falling, while Nordhaus (2015) finds that the hypothesis of a speed up of technology-driven growth fails a variety of tests.

This pessimistic view of future technological progress has entered into long-range policy planning. The Congressional Budget Office, for instance, reduced its 10-year forecast for average U.S. annual labor productivity growth from 1.8 percent in 2016 (CBO 2016) to 1.5 percent in 2017 (CBO 2017). While perhaps modest on its surface, that drop implies U.S. GDP will be considerably smaller 10 years from now than it would in the more optimistic scenario—a difference of equivalent size to almost $600 billion in 2017.

Potential Explanations for the Paradox

There are four principal candidate explanations for the confluence of technological optimism and poor productivity performance that the world finds itself in: 1) false hopes, 2) mismeasurement, 3) concentrated distribution and rent dissipation, 4) implementation and restructuring lags.[7]

False hopes

The simplest possibility is that the optimism about potential technologies is misplaced and unfounded. Perhaps technologies won't be as transformative as many expect, and while they might have modest and noteworthy effects on specific sectors, their aggregate impact will be small. In this case, the paradox will be resolved in the future as realized productivity growth never escapes its current doldrums, ultimately forcing the optimists to mark their beliefs to market.

History and some current examples offer a quantum of credence to this possibility. Certainly one can point to many technologies that did live up to initially optimistic expectations. Nuclear power never became too cheap to meter, and fusion energy has been 20 years away for 60 years. Mars may still beckon, but it's been over 40 years since Eugene

[7] To some extent, these explanations parallel the explanations for the Solow Paradox (Brynjolfsson, 1993).

Cernan was the last person to walk on the moon. Flying cars never got off the ground[8] and passenger jets no longer fly at supersonic speeds. Even AI, perhaps the most promising technology of our era, is well behind Marvin Minsky's 1967 prediction that "Within a generation the problem of creating 'artificial intelligence' will be substantially solved".

On the other hand, there remains a compelling case for optimism. As we outline below, it is not difficult to construct back-of-the-envelope scenarios where even a modest number of currently existing technologies could combine to substantially raise productivity growth and societal welfare. Indeed, knowledgeable investors and researchers are betting their money and time on exactly such outcomes. Thus, while we recognize the potential for over-optimism—and the experience with early predictions for AI makes an especially relevant reminder for us to be somewhat circumspect in this essay—we judge that it would be highly preliminary to dismiss optimism at this point.

Mismeasurement

Another potential explanation for the paradox is output and productivity mismeasurement. In this case, it is the pessimistic reading of the empirical past, not the optimism about the future, that is mistaken. Indeed, this explanation implies that the productivity benefits of the new wave of technologies are already being enjoyed. It is just that economic statistics are not up to the task of accurately measuring these benefits, making the slowdown of the past decade illusory. This "mismeasurement hypothesis" has been forwarded in several works (e.g., Mokyr 2014; Alloway, 2015; Feldstein 2015; Hatzius and Dawsey 2015; Smith 2015).

There is a prima facie case for the mismeasurement hypothesis. Many new technologies, like smartphones, online social networks, and downloadable media, involve time-intensive consumption with little monetary cost. They might deliver substantial utility even if they account for a small share of GDP due to their low relative price. Guvenen, Mataloni, Rassier, and Ruhl (2017) also show how growing offshore profit shifting can be another source of mismeasurement.

[8] At least not yet: https://kittyhawk.aero/about/.

However, a set of recent studies have shown there is good reason to think that mismeasurement is not the entire, or even a substantial, explanation for the slowdown. Cardarelli and Lusinyan (2015); Byrne, Fernald, and Reinsdorf (2016); Nakamura and Soloveichik (2015); and Syverson (2017), each using different methodologies and data, present evidence that mismeasurement is not the primary explanation for the productivity slowdown. After all, while there is convincing evidence that many of the benefits of today's technologies are not reflected in GDP and therefore productivity statistics, the same was undoubtedly true in earlier eras as well.

Concentrated Distribution and Rent Dissipation

A third possibility is that the gains of the new technologies are already attainable, but through a combination of concentrated distribution of those gains and dissipative efforts to attain or preserve them (assuming the technologies are at least partially rivalrous), their effect on average productivity growth is modest overall, and is virtually nil for the median worker. For instance, two of the most profitable uses of AI to date have been for targeting and pricing online ads, and for automated trading of financial instruments, both applications with many zero-sum aspects.

One version of story asserts that the benefits of the new technologies are being enjoyed by a relatively small fraction of the economy, but the technologies' narrowly scoped and rivalrous nature creates wasteful "gold rush"-type activities. Both those seeking to be one of the few beneficiaries, as well as those who have attained some gains and seek to block access to others, engage in these dissipative efforts, destroying many of the benefits of the new technologies.[9]

Recent research offers some indirect support for elements of this story. Productivity differences between frontier firms and average firms in the same industry have been increasing in recent years (Andrews, Criscuolo, and Gal, 2016; Furman and Orszag, 2015). Differences in profit margins between the top and bottom performers in most industries have also grown (McAfee and Brynjolfsson, 2009). A smaller number of superstar firms are gaining market share (Autor et al, 2017, Brynjolfsson et al. 2008) while workers' earnings

[9] Stiglitz (2014) offers a different mechanism where technological progress with concentrated benefits in the presence of restructuring costs can lead to increased inequality and even, in the short run, economic downturns.

are increasingly tied to firm-level productivity differences (Song, Price, Guvenen, Bloom, and von Wachter 2015). There are concerns that industry concentration is leading to substantial aggregate welfare losses due to the distortions of market power (e.g., De Loecker and Eeckhout, 2017; Gutiérrez and Philippon, 2017). Furthermore, growing inequality can lead to stagnating median incomes and associated socio-economic costs, even when total income continues to grow.

While this evidence is important, it is not dispositive. The aggregate effects of industry concentration are still under debate, and the mere fact that a technology's gains aren't evenly distributed is no guarantee that resources will be dissipated in trying to capture them—especially that there would be enough waste to erase noticeable aggregate benefits.

Implementation and Restructuring Lags

Each of these three possibilities, especially the first two, relies on explaining away the discordance between high hopes and disappointing statistical realities. One of the two elements is somehow "wrong". In the misplaced optimism story, the expectations for technology are off base. In the mismeasurement explanation, the tools we use to gauge empirical reality aren't up to the task of accurately doing so. And in the concentrated distribution stories, the private gains for the few may be very real, but they don't translate into broader gains for the many.

But there is a fourth explanation that allows both halves of the seeming paradox to be correct. It asserts that there really is good reason to be optimistic about the future productivity growth potential of new technologies, while at the same time recognizing that recent productivity growth has been low. The core of this story is that it takes a considerable time—often more than is commonly appreciated—to be able to sufficiently harness new technologies. Ironically, this is especially true for those technologies important enough to ultimately have an important effect on aggregate statistics and welfare. That is, those with such broad potential application that they qualify as general purpose technologies (GPTs). Indeed, the more profound and far-reaching the potential restructuring, the longer the time lag between the initial invention of a technology and its full impact on the economy and society.

This explanation implies there will be a period where the technologies are developed enough that one can imagine their potentially transformative effects even though they have had no discernable effect on recent productivity growth. It isn't until the necessary build-up and implementation time has passed that the promise of the technology actually blossoms in the aggregate data.

There are two main sources of the delay between recognition of a new technology's potential and its measureable effects. One is that it takes time to build the stock of the new technology to a size sufficient enough to have an aggregate effect. The other is that complementary investments are necessary to obtain the full benefit of the new technology, and it takes time to discover what these complements are and to implement them. While the fundamental importance of the core invention and its potential for society might be clearly recognizable at the outset, the myriad necessary co-inventions, obstacles and adjustments needed along the way await discovery over time, and the required path may be lengthy and arduous. Never mistake a clear view for a short distance.

This explanation resolves the paradox by acknowledging that its two seemingly contradictory parts are not actually in conflict. Rather, they are in some sense both natural manifestations of the same underlying phenomenon of building and implementing a new technology.

While each of the first three explanations for the paradox might have a part in describing its source, they also face serious questions in their ability to describe key parts of the data. We find the fourth, the implementation and restructuring lags story, the most compelling in light of the evidence we discuss below. Thus it is the focus of our explorations in the remainder of this paper.

The Argument in Favor of the Implementation and Restructuring Lags Explanation

Implicit or explicit in the pessimistic view of the future is that the recent slowdown in productivity growth portends slower productivity growth in the future. We begin by establishing one of the most basic elements of the story: that slow productivity growth today does not rule out faster productivity growth in the future. In fact, the evidence is clear that it is barely predictive at all.

Total factor productivity growth is the component of overall output growth that cannot be explained by accounting for changes in observable labor and capital inputs. It has been called a “measure of our ignorance” (Abramovitz, 1956). It is a residual, so an econometrician should not be surprised if it is not very predictable from past levels. Labor productivity is a similar measure, but instead of accounting for capital accumulation simply divides total output by the labor hours used to produce that output.

Figure 3 and Figure 4 plot, respectively, U.S. productivity indices since 1948 and productivity growth by decade. The data include average labor productivity (LP), average total factor productivity (TFP) and Fernald’s (2014) utilization-adjusted TFP (TFPua).

Figure 3. U.S. TFP and Labor Productivity Indices, 1948-2016 (1990 = 100)

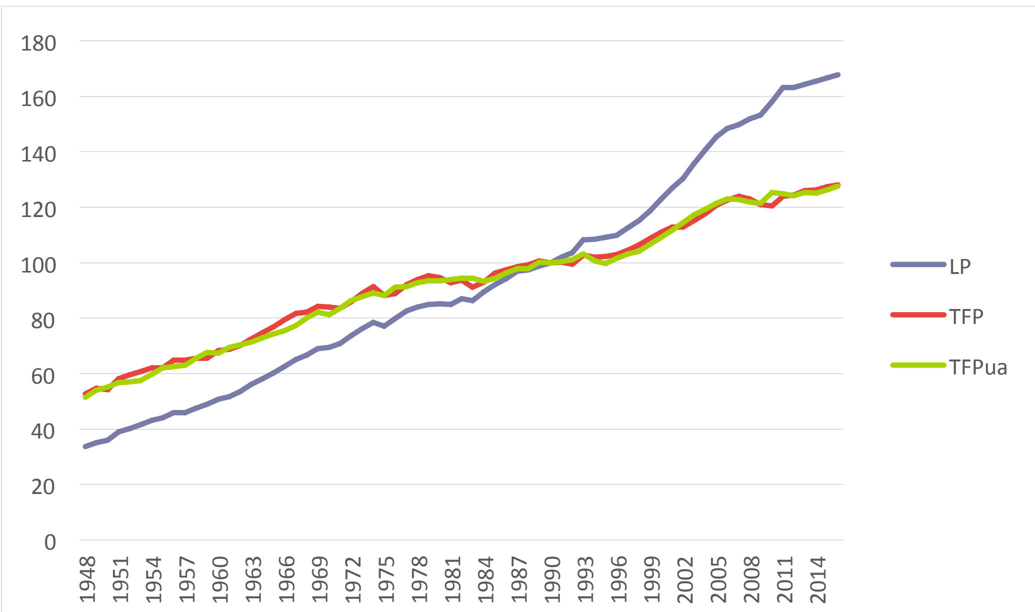
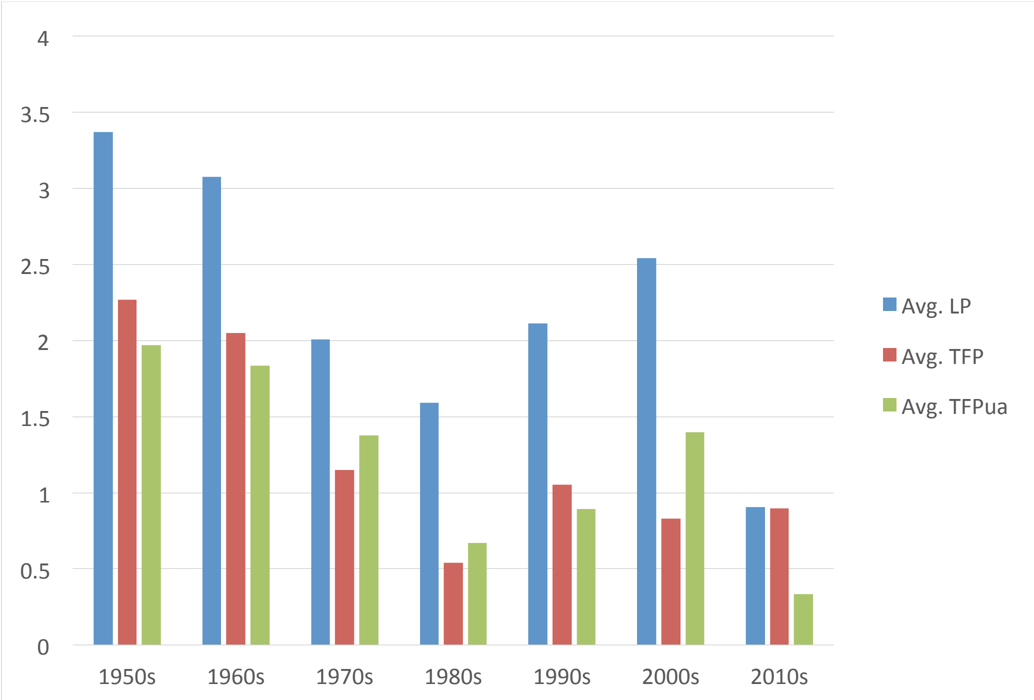


Figure 4. U.S. TFP and Labor Productivity Growth (%) by Decade



Productivity has consistently grown in the post-war era, albeit at different rates at different times. Despite the consistent growth, however, past productivity growth rates have historically been poor predictors of future productivity growth. In other words, the productivity growth of the past decade tells us little about productivity growth in for the coming decade. Looking only at productivity data, it would have been hard to predict the decrease in productivity growth at the end of the 1960s or foresee the beneficial impact of IT in the 1990s.

As it turns out, while there is some correlation in productivity growth rates over short intervals, the correlation between adjacent ten-year periods is not statistically significant. We present below the results from a regression of different measures of average productivity growth on the previous period's average productivity growth for 10-year intervals as well as scatterplots of productivity for each 10 year against the productivity in the subsequent period. The data are sourced from the Fernald (2014) TFP and utilization-adjusted TFP series and span from 1948 to 2016 (annually).[10] The R^2 of these regressions is low in all cases. The correlation coefficients in the following table

[10] Available: http://www.frbsf.org/economic-research/indicators-data/total-factor-productivity-tfp/

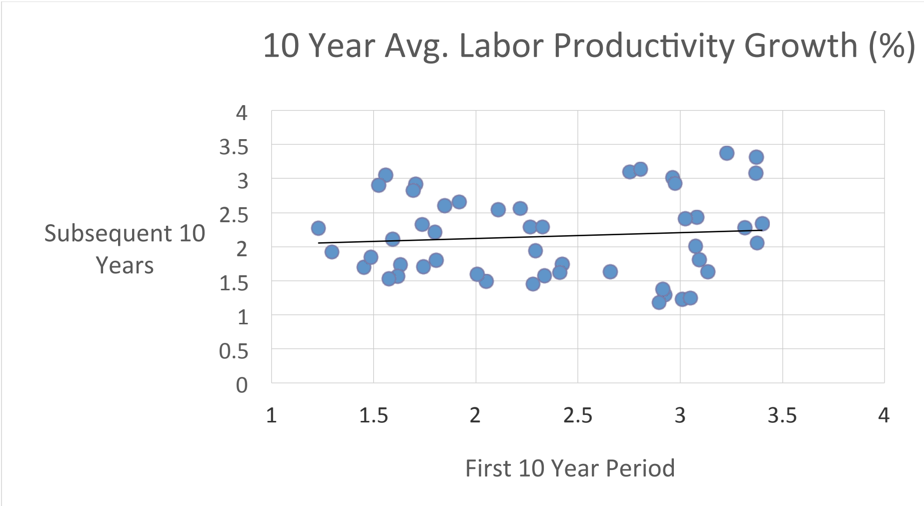
represent the correlation in productivity growth series for a given ten-year period and the subsequent ten-year period.

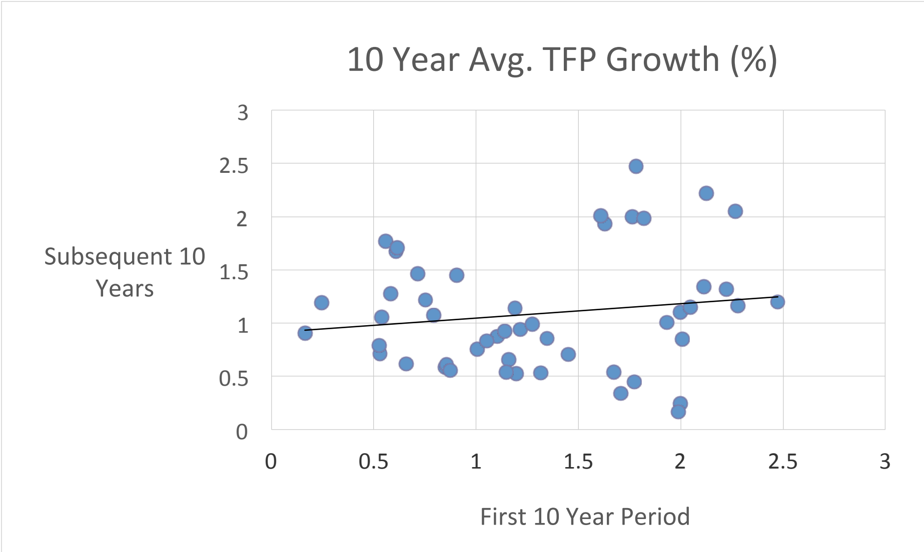
Correlation Coefficients (1^{st} vs. 2^{nd} 10-year Period)	
Labor Productivity	0.09
TFP	0.15
TFP (util. adj.)	0.17

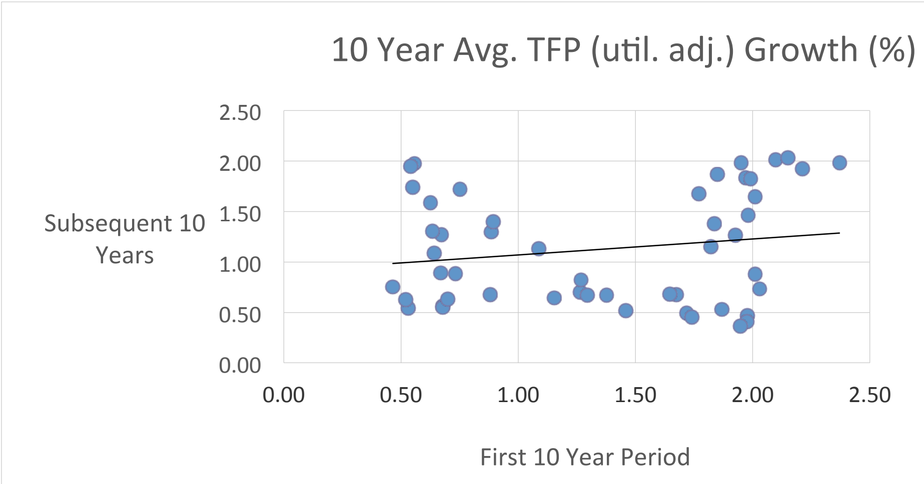
Regression results for Labor Productivity, TFP, and utilization-adjusted TFP are included below. For Labor Productivity, the R^2 is 0.009. While the intercept is significantly different from zero (productivity is positive, on average), the coefficient on the previous period is not significant. For TFP the R^2 is 0.023, and again the coefficient on the previous period is not statistically significant. Utilization-adjusted TFP is slightly higher, but still small and statistically insignificant.

Productivity Growth Regressions	(1) Labor Productivity Growth (10 Year)	(2) Total Factor Productivity Growth (10 Year)	(3) Utilization-Adjusted Productivity Growth (10 Year)
Previous 10 Year Productivity Growth	0.0857	0.136	0.158
	(0.132)	(0.121)	(0.137)
Constant	1.949***	0.911***	0.910***
	(0.297)	(0.145)	(0.189)
Observations	50	50	50
R-squared	0.009	0.023	0.030

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1


10 Year Avg. Labor Productivity Growth (%)
Subsequent 10 Years
First 10 Year Period
4
3.5
3
2.5
2
1.5
1
0.5
0
1
1.5
2
2.5
3
3.5
4


10 Year Avg. TFP Growth (%)
Subsequent 10 Years
First 10 Year Period
3
2.5
2
1.5
1
0.5
0
0
0.5
1
1.5
2
2.5
3


10 Year Avg. TFP (util. adj.) Growth (%)
Subsequent 10 Years
First 10 Year Period
2.50
2.00
1.50
1.00
0.50
0.00
0.00
0.50
1.00
1.50
2.00
2.50

The old adage that "past performance is not predictive of future results" applies well to trying to predict productivity growth in the years to come, especially in periods of a decade or longer. Historical stagnation does not justify forward-looking pessimism.

A Technology-Driven Case for Productivity Optimism

Simply extrapolating recent productivity growth rates forward is not a good way to estimate the next decade's productivity growth. Does that imply we have no hope at all of predicting productivity growth? We don't think so.

Instead relying only on past productivity statistics, we can consider the technological and innovation environment we expect to see in the near future. In particular, we need to study and understand the specific technologies that actually exist, and make an assessment of their potential.

One does not have to dig too deeply into the pool of existing technologies or assume incredibly large benefits from any one of them to make a case that existing but still nascent technologies can potentially combine to create noticeable accelerations in aggregate productivity growth. We begin by looking at a few specific examples. We will then make the case that AI is a GPT, with broader implications.

First, let's consider the productivity potential of autonomous vehicles. According to the US Bureau of Labor Statistics, in 2016 there were 3.5 million people working in private industry as "motor vehicle operators" of one sort or another (this includes truck drivers, taxi drivers, bus drivers, and other similar occupations). Suppose autonomous vehicles were to reduce, over some period, the number of drivers necessary to do the current workload to 1.5 million. We do not think this is a far-fetched scenario given the potential of the technology. Total nonfarm private employment in mid-2016 was 122 million. Therefore, autonomous vehicles would reduce the number of workers necessary to achieve the same output to 120 million. This would result in aggregate labor productivity (calculated using the standard BLS nonfarm private series) increasing by 1.7 percent (= 122/120). Supposing this transition occurred over 10 years, this single technology would provide a direct boost of 0.17 percent to annual productivity growth over that decade.

This is significant, and it doesn't include many potential productivity gains from complementary changes that could accompany the diffusion of autonomous vehicles. For instance, self-driving cars are a natural complement to transportation-as-a-service rather than individual car ownership. The typical car is currently parked 95% of the time, making it readily available for its owner or primary user (Morris, 2016). However, in locations with sufficient density, a self-driving car could be summoned on demand. This would make it possible for cars to provide useful transportation services for a larger fraction of the time, reducing capital costs per passenger-mile, even after accounting for increased wear-and-tear. Thus, in addition to the obvious improvements in labor productivity from replacing drivers, capital productivity would also be significantly improved.

A second example is call centers. As of 2015, there were about 2.2 million people working in over 6,800 call centers in the United States and hundreds of thousands more work as home-based call center agents or in smaller sites.[11] Improved voice-recognition systems coupled with intelligence question-answering tools like IBM's Watson might plausibly be able to handle 60-70% or more of the calls, especially since, in accordance with the Pareto principle, a large fraction of call volume is due to variants on a small number of basic queries. If AI reduced the number workers by 60%, it would increase US labor productivity by 1%, perhaps again spread over 10 years. Again, this would likely spur complementary innovations, from shopping recommendation and travel services, to legal advice, consulting, and real-time personal coaching.

Beyond labor savings, advances in AI have the potential to boost total factor productivity. In particular, energy efficiency and materials usage could be improved in many large-scale industrial plants. For instance, a team from Google DeepMind recently trained an ensemble of neural networks to optimize power consumption in a data center. By carefully tracking the data already collected from thousands of sensors tracking temperatures, electricity usage, pump speeds, the system learned how to make adjustments in the operating parameters. As a result, they were able to reduce the amount of energy used for cooling by 40% compared to the levels achieved by human experts. The algorithm was a general-purpose framework designed to account complex dynamics, so it is easy to see how such a system could be applied to other data centers at Google, or indeed

[11] https://info.siteselectiongroup.com/blog/how-big-is-the-us-call-center-industry-compared-to-india-and-philippines

around the world. Overall, data center electricity costs in the US are about $6 billion per year, including about $2 billion just for cooling.[12]

What's more, similar applications of machine learning could be implemented in a variety of the commercial and industrial activities. For instance, manufacturing accounts for about $2.18 trillion of value-added each year. Manufacturing companies like GE are already using AI to forecast product demand, future customer maintenance needs, and analyze performance data coming from sensors on their capital equipment. Recent work on training deep neural network models to perceive objects and achieve sensorimotor control at the same time have yielded robots that can perform a variety of hand-eye coordination tasks (e.g. unscrewing bottle caps and hanging coat hangers) (Levine et al. 2016). Liu et al. (2017) trained robots to perform a number of household chores, like sweeping and pouring almonds into a pan, using a technique called imitation learning.[13] In this approach, the robot learns to perform a task using a raw video demonstration of what it needs to do. These techniques will surely be important for automating manufacturing processes in the future. The results suggest that artificial intelligence may soon improve productivity in household production tasks as well, which in 2010 were worth as much as $2.5 trillion in nonmarket value-added (Bridgman et al. 2012).

While these examples are each suggestive of non-trivial productivity gains, they are only a fraction of the set of applications for AI and machine learning that have been identified so far. James Manyika and his colleagues analyzed 2000 tasks and estimated that about 45% of the activities that people are paid to perform in the US economy could be automated using existing levels of AI and other technologies. They stress that the pace of automation will depend on factors other than technical feasibility, including the costs of automation, regulatory barriers and social acceptance.

Artificial Intelligence is a General Purpose Technology

Important as specific applications of AI may be, we argue that the more important economics effects of AI, machine learning, and associated new technologies stem from the

[12] According to personal communication, August 24, 2017 with Jon Koomey, Arman Shehabi and Sarah Smith of Lawrence Berkeley Lab.
[13] Videos of these efforts available here: https://sites.google.com/site/imitationfromobservation/

fact that they embody the characteristics of general purpose technologies (GPTs). Bresnahan and Trajtenberg (1996) argue that a GPT should be pervasive, able to be improved upon over time, and be able to spawn complementary innovations.

The steam engine, electricity, the internal combustion engine, and computers are each examples of important general purpose technologies. Each of them not only increased productivity directly, but also by spurring important complementary innovations. For instance, the steam engine not only helped pump water from coal mines, its most important initial application, but also spurred the invention more effective factory machinery and new forms of transportation like steamships and railroads. In turn, these co-inventions helped give rise to innovations in supply chains and mass marketing, to new organizations with hundreds of thousands of employees, and even to seemingly unrelated innovations like standard time, which was needed to manage railroad schedules.

AI, and in particular machine learning, certainly has the potential to be pervasive, to be improved upon over time, and to spawn complementary innovations, making it a candidate for an important GPT.

As noted by Agrawal, Gans, and Goldfarb (2017), the current generation of machine learning systems is particularly suited for augmenting or automating tasks that involve at least some prediction aspect, broadly defined. These cover a broad range of tasks, occupations and industries, from driving a car (predicting the right way to turn the steering wheel) and diagnosing a disease (predicting its cause) to recommending a product (predicting what the customer will like) and writing a song (predicting which note sequence will be most popular). The core capabilities of perception and cognition addressed by current systems are pervasive, if not indispensable, for many tasks done by humans.

Machine learning systems are also designed to improve over time. Indeed, what sets them apart from earlier technologies is that they are designed to improve *themselves* over time. Instead of requiring an inventor or developer to consciously codify, or code, each step of a process to be automated, a machine learning algorithm can discover on its own a function that connects a set of inputs X to a set of outputs Y as long as its given a sufficiently large set of labeled examples mapping some of the inputs to outputs (Brynjolfsson and Mitchell, 2017). The improvements reflect not only the discovery of new algorithms and techniques, particularly for deep neural networks, but also their synergies with vastly more powerful computer hardware and the availability of much larger digital datasets that can be used to train the systems (Brynjolfsson and McAfee, 2017). More and more digital data is collected as byproduct of digitizing operations, customer interactions, communications and other aspects of our lives, providing fodder for more and better machine learning applications.[14]

Most importantly, machine learning systems spur a variety of complementary innovations. For instance, machine learning has transformed the abilities of machines to perform a number of basic types of perception and these make possible a broader set of applications. Consider machine vision—the ability to see and recognize objects, to label them in photos, and to interpret video streams. As error rates in identifying pedestrians improve from one per 30 frames to about one per 30 million frames, self-driving cars become increasingly feasible (Brynjolfsson and McAfee, 2017).

Improved vision also makes a variety of factory automation tasks practical, as well as improved medical diagnoses. Gill Pratt has made an analogy to the development of vision in animals 500 million years ago, which helped ignite the Cambrian explosion and a burst of new species on earth. (Pratt, 2015). He also pointed out that machines have a new capability that no biological species has: the ability to share knowledge and skills almost instantaneously with others. Specifically, the rise of cloud computing has made it significantly easier to scale up new ideas at much lower cost than before. This is an especially important development for advancing the economic impact of machine learning because it enables cloud robotics—the sharing of knowledge among robots. Once a new skill is learned by a machine in one location, it can be replicated to other machines via digital networks. Data as well as skills can be shared, increasing the amount of data that any given machine learner can use.

This in turn increases the rate of improvement. For instance, self-driving cars that encounter an unusual situation can upload that information with a shared platform where enough examples can be aggregated to infer a pattern. Only one self-driving vehicle needs to experience an anomaly for many vehicles to learn from it. Waymo, a subsidiary of

---

[14] For example, through enterprise resource planning systems in factories, internet commerce, mobile phones, and the "Internet of Things."

Google, has cars driving 25,000 "real" autonomous and about 19 million simulated miles each week.[15] All of the Waymo cars learn from the joint experience of the others. Similarly, a robot struggling with a task can benefit from sharing data and learnings with other robots that use a compatible knowledge-representation framework.[16]

When one thinks of AI as a GPT, the implications for output and welfare gains are much larger than in our earlier analysis. For example, self-driving cars could substantially transform many non-transport industries. Retail could shift much further toward home delivery on demand, creating consumer welfare gains and further freeing up valuable high-density land now used for parking. Traffic and safety could be optimized, and insurance risks could fall. With over 30,000 deaths due to automobile crashes in the US each year, and nearly a million worldwide, there is an opportunity to save many lives.[17]

Why Future Technological Progress Is Consistent with Low Current Productivity Growth

Having made a case for technological optimism, we now turn to explaining why it is not inconsistent with—and in fact may even be naturally related to—low current productivity growth.

Like other GPTs, AI has the potential to be an important driver of productivity. However, as Jovanovic and Rousseau (2005) point out (with additional reference to David's (1991) historical example), "a GPT does not deliver productivity gains immediately upon arrival." (p. 1184). The technology can be present and developed enough to allow some notion of its transformative effects even though it is not affecting current productivity levels in any noticeable way. This is precisely the state that we argue the economy may be in now.

We discussed above that a GPT can at one moment both be present and yet not affect current productivity growth if there is a need to build a sufficiently large stock of the

[15] http://ben-evans.com/benedictevans/2017/8/20/winner-takes-all
[16] Rethink Robotics is developing exactly such a platform.
[17] These latter two consequences of autonomous vehicles, while certainly reflecting welfare improvements, would need to be capitalized in prices of goods or services to be measured in standard GDP and productivity measures. We will discuss AI-related measurement issues in greater depth below. Of course it is worth remembering autonomous vehicles also hold the potential to create new economic costs if, say, the congestion from lower marginal costs of operating a vehicle is not counteracted by sufficiently large improvements in traffic management technology or certain infrastructure investments.

new capital or if complementary types of capital, both tangible and intangible, need to be identified, produced, and put in place to fully harness the GPT's productivity benefits.

The time necessary to build a sufficient capital stock can be extensive. For example, it wasn't until the late 1980s, more than 25 years after the invention of the integrated circuit, that the computer capital stock reached its long-run plateau at about 5 percent (at historical cost) of total nonresidential equipment capital. It was only half that level 10 years prior. Thus, when Solow pointed out his now eponymous paradox, the computers were *finally just then* getting to the point where they really could be seen everywhere.

David (1991) points out a similar phenomenon in the diffusion of electrification. At least half of U.S. manufacturing establishments remained unelectrified until 1919, about 30 years after the shift to polyphase alternating current began. Initially adoption was driven by simple cost savings. The biggest benefits came later, when managers began to fundamentally re-organize work by replacing the centralized power source and giving every individual machine its own electric motor. This created much more flexibility in the location of equipment and made possible effective assembly lines materials flow.

This approach to organizing factories is obvious in retrospect, yet it took as much as 30 years for it to become widely adopted. Why? As noted by Henderson (1993; 2006), it is exactly *because* incumbents are designed around the current ways of doing things and so proficient at them that they are blind to or unable to absorb the new approaches and get trapped in the status quo—they suffer the "curse of knowledge."[18]

Similarly, Brynjolfsson and Smith (1999) document the difficulties incumbent retailers had adapting their business processes to take full advantage of the internet and electronic commerce relative to born-digital companies like Amazon. The potential of ecommerce to revolutionize retailing was widely recognized, and even hyped in the late 1990s, but actual share of retail commerce was trivial, 0.2% of all retail sales in 1999. Only in 2017, after two decades of widely predicted yet time-consuming change in the industry,

[18] Atkeson and Kehoe (2007) note manufacturers' reluctance to abandon their large knowledge stock at the beginning of the transition to electric power to adopt what was, initially, only a marginally superior technology. David and Wright (2006) are more specific, focusing on the "the need for organizational and above all for *conceptual* changes in the ways tasks and products are defined and structured" (p. 147, emphasis in original).

are companies like Amazon are having a first-order effect on more traditional retailers' sales and stock market valuations.

Another source of the time gap between a technology's emergence and its measured productivity effects is the need for complementary capital to be installed (and often, first invented). This includes both tangible and intangible investments. The timeline necessary to acquire and install these complements is typically more extensive as the time-to-build considerations just discussed.

Consider changing a specific production process to benefit large investments in IT. Brynjolfsson and Hitt (2003) examined firm level data and found that while small productivity benefits were associated with IT investments when one-year differences were considered, the benefits grew substantially as longer differences were examined, peaking after about seven years. They attributed this pattern to the need for complementary changes in business processes. For instance, when implementing large enterprise planning systems, firms almost always spend several times more on business process redesign and training than on the direct costs of hardware and software. These can be thought of as investments in organizational and human capital, and they often take years to implement.

At the firm level, additional complementary investments are required beyond those at the process level to fully harness new technologies. The organizational structure of the company often needs to be rebuilt. Hiring and other HR practices often need considerable adjustment to match the firm's human capital to the new structure of production. In fact, Bresnahan, Brynjolfsson, and Hitt (2002) find evidence of three-way complementarities between IT, human capital, and organizational changes in the investment decisions and productivity levels. Furthermore, Brynjolfsson, Hitt, and Yang (2002) show each dollar of IT capital stock is correlated with about $10 of market value. They interpret this as evidence of substantial IT-related intangible assets and show that firms that combine IT investments with a specific set of organizational practices are not just more productive, they also have disproportionately higher market values than firms that invest in only one or the other. This pattern in the data is consistent with a long stream of research on the importance of organizational and even cultural change when making IT investments and technology investments more generally (e.g. Aral et al 2012; Brynjolfsson and Hitt, 2000; Orlikowski, 1996; Henderson, 2006).

But such changes take real time and resources, contributing to organizational inertia. Firms are more complex systems than individual production lines, and this greater complexity requires a more extensive web of complementary assets to allow the GPT to fully transform the system. Transforming firms often must reevaluate and reconfigure not only their internal processes, but often their supply and distribution chains as well.

There is no assurance that the adjustments will be successful. Indeed, there is evidence that the modal transformation of GPT-level magnitude fails. Alon, Berger, Dent, and Pugsley (2017) find that cohorts of firms over five years old contribute little to aggregate productivity growth on net—that is, among established firms, there is one firm becoming less productive for each firm that increases its productivity. It is hard to teach the proverbial old dog new tricks. Moreover, the old dogs (companies) often have internal incentives to not learn them (Arrow, 1962; Holmes, Levine, and Schmitz 2012). In some ways, technology advances in industry one company death at a time.

Transforming industries and sectors requires still more adjustment and reconfiguration. As noted above, retail offers a vivid example. Despite being one of the biggest innovations to come out of the 1990s dot-com boom, the largest change in retail in the two decades that followed was not e-commerce but instead the expansion of warehouse stores and supercenters (Hortaçsu and Syverson, 2015). It is only very recently that e-commerce has become a force for general retailers to reckon with. Why did it take so long? Many complementary investments were required. An entire distribution infrastructure had to be built. Customers had to be "retrained." None of this could happen quickly. Ecommerce may have been readily foreseeable once the Internet began to reach most homes, but has taken over 20 years for ecommerce sales to rise to its current share of 9 percent of total retail sales.

Viewing Today's Paradox through Previous General Purpose Technologies

We have indicated in the discussion above that we see parallels between the current paradox and those that have happened in the past. It is closely related to the Solow paradox era circa 1990, certainly, but it is also tied closely to the experience during the diffusion of portable power (we prefer this to "electrification" so as to also reflect the parallel growth and transformative effects of the internal combustion engine).

Comparing the productivity growth patterns of the two eras is instructive.

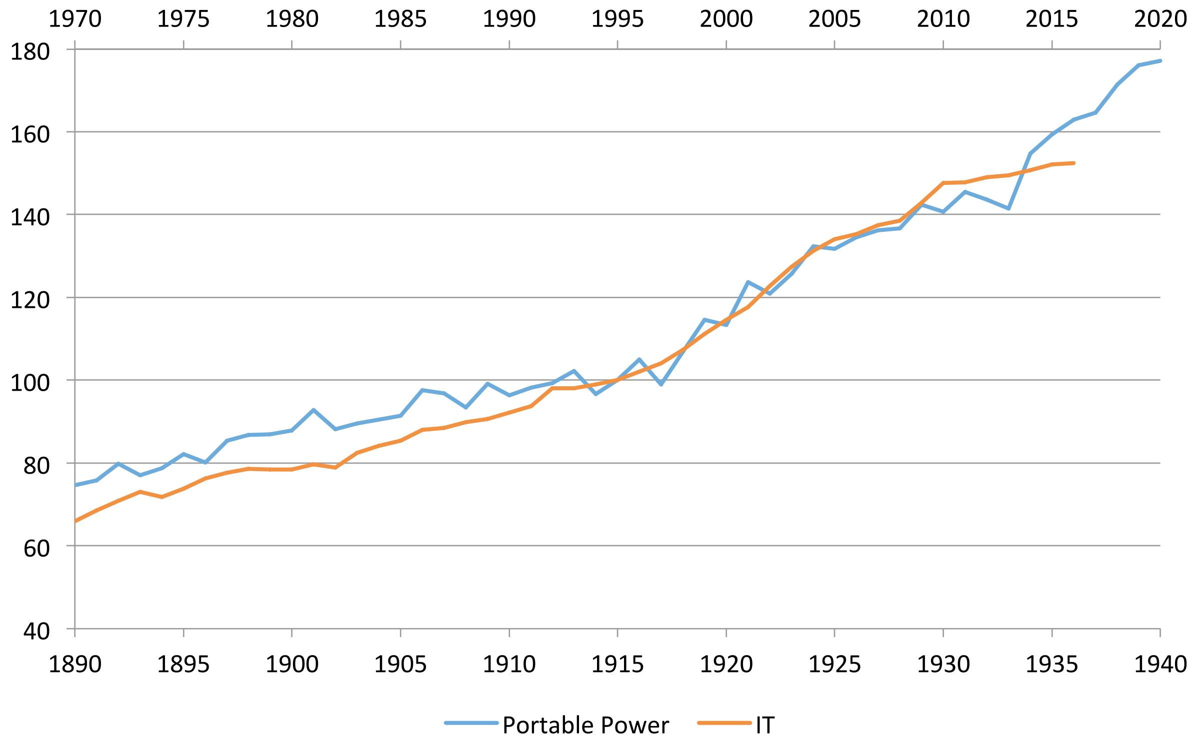
Figure 5 is an updated version of an analysis from Syverson (2013). It overlays U.S. labor productivity since 1970 with that from 1890 to 1940, the period after portable power technologies had been invented and were starting to be placed into production. (The historical series values are from Kendrick 1961.) The modern series timeline is indexed to a value of 100 in 1995 and is labeled on the upper horizontal axis. The portable power era index has a value of 100 in 1915, and its years are shown on the lower horizontal axis.

Labor productivity during the portable power era shared remarkably common patterns with current series. In both eras, there was an initial period of roughly a quarter century of relatively slow productivity growth. Then both eras saw decade-long accelerations in productivity growth, spanning 1915 to 1924 in the portable power era and 1995-2004 more recently.

The late-1990s acceleration was the (at least partial) resolution of the Solow Paradox. We imagine the late 1910s acceleration could have similarly answered some economist's query in 1910 as to why one sees electric motors and internal combustion engines everywhere but in the productivity statistics.[19]

Figure 5. Labor Productivity Growth in the Portable Power and IT Eras

[19] We aren't aware of anyone who actually said this, and of course today's system of national economic statistics did not exist at that time, but we find the scenario amusing, instructive, and in some ways plausible.



Very interestingly, and quite relevant to the current situation, the productivity growth slowdown we have experienced after 2004 also has a parallel in the historical data, a slowdown from 1924 to 1932. As can be seen in the figure, and instructive to the point of whether a new wave of AI and associated technologies (or if one prefers, a second wave of IT-based technology) could re-accelerate productivity growth, labor productivity growth at the end of the portable power era rose again, averaging 2.7 percent per year between 1933 and 1940.

Of course this past breakout growth is no guarantee that productivity must speed up again today. However, it does raise two relevant points. First, it is another example of a period of sluggish productivity growth followed by an acceleration. Second, it demonstrates that productivity growth driven by a core GPT can arrive in multiple waves.

Expected Productivity Effects of an AI-Driven Acceleration

To understand the likely productivity effects of AI, it is useful to think of AI is a type of capital, specifically a type of intangible capital. It can be accumulated through investment; it is a durable factor of production; and it can depreciate. Treating AI as a type of capital clarifies how its development and installation as a productive factor will affect productivity.

As with any capital deepening, increasing AI will raise labor productivity. This would be true regardless of how well AI capital is measured (which we might expect it won't be for several reasons discussed below) though there may be lags.

AI's effects on total factor productivity (TFP) are more complex and the impact *will* depend on its measurement. If AI (and its output elasticity) were to be measured perfectly and included in the both the input bundle in the denominator of TFP and the output bundle in the numerator, then measured TFP will accurately reflect true TFP. In this case, AI is treated just like any other measurable capital input. Its effect on output will be properly accounted for and "removed" by the TFP input measure, leading to no change in TFP. This isn't to say that there wouldn't be productive benefits from diffusion of AI; it is just that it would be valued like any other type of capital input.

There are reasons why economists and national statistical agencies might face measurement problems when dealing with AI. Some are instances of more general capital measurement issues, but others are likely to be idiosyncratic to AI. We discuss this next.

Measuring AI Capital

Regardless of the effects of AI and AI-related technologies on actual output and productivity, it is clear from the productivity outlook above that the ways AI's effects will be *measured* are dependent on how well countries' statistics programs measure AI capital.

The primary difficulty in AI capital measurement is, as mentioned above, that it will largely be intangible. This will present itself as a problem for both AI capital itself as well as its outputs. This potential issue is exacerbated by the likelihood that AI will primarily be used as an input in making other capital, including new types of software, human and organizational capital, rather than final consumption goods. Human capital per worker is rising throughout the world, compounding the measurement issue (Jones and Romer, 2010). Moreover, this other capital will, like AI itself, be mostly intangible.

Effective use of AI requires developing datasets, building firm-specific human capital, and implementing new business processes. These all require substantial capital outlays and maintenance. The tangible counterparts to these intangible expenditures, including purchases of computing resources, servers, and real estate, are easily measured in the standard neo-classical growth accounting model (Solow, 1957). On the other hand, the value of capital goods production for complementary intangible investments is difficult to quantify. Both tangible and intangible capital stocks generate a capital service flow yield that accrues over time. Realizing these yields requires more than simply renting capital stock as well. After purchasing capital assets, firms incur additional adjustment costs (e.g. business process redesigns and installation costs). These adjustment costs make capital less flexible than frictionless rental markets would imply. Much of the market value of AI capital in specific and IT capital more generally may be derived from the capitalized short-term quasi-rents earned by firms that have already reorganized to extract service flows from new investment.

Yet while the stock of tangible AI assets is booked on corporate balance sheets, expenditures on the intangible complements and adjustment costs to AI investment largely are not. Without including the production of intangible AI capital, the usual growth accounting decompositions of changes in value added can misattribute AI intangible capital deepening to growth in TFP. As discussed in Hall (2000) and Yang and Brynjolfsson (2001) this constitutes an omission of a potentially important component of capital goods production in the calculation of final output. Estimates of TFP will therefore be inaccurate, though possibly in either direction. Nevertheless, in the case that claims on the assets of the firm are publicly traded, the financial market will properly value the firm as the present value of its risk-adjusted discounted cash flows.

We can combine q-theory of investment with the neoclassical growth accounting framework to improve estimates of TFP. In particular, we show in the appendix that in the case that the shadow price of AI investment is close to the purchase price of investment, there is no missing growth in output. But if the intangible AI capital stock is growing faster than the accumulation of ordinary capital, then TFP growth will be underestimated. The intuition for this result is that in any given period t, the output of (unmeasured) AI capital stock in period t+1 is a function the input (unmeasured) existing AI capital stock in period t. When AI stock is growing rapidly, the unmeasured outputs will be greater than the unmeasured inputs. (Of course, in steady state there is no longer a mismeasurement problem as further investment serves precisely to replenish depreciated capital. In this case, the unmeasured inputs and outputs cancel out.) Furthermore, suppose the relevant costs needed to create intangible assets, in terms of labor and other resources, are measured, but the resulting increases in intangible assets are not measured as contributions to output. In this case, not only will total GDP be undercounted but so will productivity, which uses GDP as its numerator. Thus periods of rapid intangible capital accumulation may be associated with *lower* measured productivity growth, even if true productivity is increasing.

These problems may be particularly stark for AI capital, as its accumulation will almost surely outstrip the pace of ordinary capital accumulation in the short-run. AI capital is a new category of capital—new in economic statistics, certainly, but we would argue practically so as well. While the concept of AI is decades old, very little actual AI capital was accumulated in prior decades. Thus the current stock is close to zero.

This also means that capital quantity indexes that are computed from within-type capital growth might have problems benchmarking size and effect of AI early on. National statistics agencies do not really focus on measuring capital types that aren't already ubiquitous. New capital categories will tend to either be rolled into existing types, possibly with lower inferred marginal products (leading to an understatement of the productive effect of the new capital), or missed altogether. This problem is akin to the new goods problem in price indexes.

A related issue is—once AI is measured separately—how closely its units of measurement will capture AI's marginal product relative to other capital stock. That is, if a dollar of AI stock has a marginal product that is 10 percent higher than the modal unit of non-AI capital in the economy, will the quantity indexes of AI reflect this? This requires measured relative prices of AI and non-AI capital to capture differences in marginal product. Measuring levels right is less important than having proportional differences (whether intertemporally or in the cross section) correct. What is needed in the end is that a unit of AI capital twice as productive as another should be twice as large in the capital stock.

It is worth noting that these are all classic problems in capital measurement and not new to AI. Perhaps these problems will be systematically worse for AI, but this is not obvious ex ante. What it does mean is that economists and national statistical agencies at least have experience in, if not quite a full solution for, dealing with these sorts of limitations.

Some measurement issues are likely to be specifically prevalent for AI. One is the likelihood that a substantial part of the value of AI output will be firm-specific. Imagine a program that figures out individual consumers' price elasticities and matches pricing to these elasticities. This has different value to different companies depending on their customer bases, and knowledge may not be transferrable across firms. The value also depends on companies' abilities to implement price discrimination. Such limits could come from characteristics of company's market, like resale opportunities, which are not always under firms' control, or from the existence in the firm of complementary implementation assets and/or abilities. Likewise, each firm will likely have a different skill mix that it seeks in its employees, unique needs in its production process and a particular set of supply constraints. As noted by Brynjolfsson and McAfee (2017), firm-specific data sets and applications of those data can differentiate the machine learning capabilities of one firm from another.

Conclusion

In 2017, there are plenty of both optimists and pessimists about technology and growth. The optimists tend to be technologists and venture capitalists, and many are clustered in technology hubs. The pessimists tend to be economists, sociologists, statisticians and government officials. Many of them are clustered in major state and national capitals. There is much less interaction between the two groups than within them, and it often seems as though they are talking past each other. In this paper, we argue that in an important a sense, they are.

When we talk with the optimists, we are convinced that the recent breakthroughs in AI and machine learning are real and significant. We also would argue that they form the core of a new, economically-important GPT. When we speak with the pessimists, we are convinced that productivity growth has slowed down recently and what gains there have been are unevenly distributed, leaving many people with stagnating incomes, declining metrics of health and well-being, and good cause for concern. People are uncertain about the future, and many of the industrial titans that once dominated the employment and market value leaderboard have fallen on harder times.

These two stories are not contradictory. In fact, any many ways, they are consistent and symptomatic of an economy in transition. Our analysis suggests that while the recent past has been difficult, it is not destiny. Although it is always dangerous to make predictions, and we are humble about our ability to foretell the future, our reading of the evidence does provide some cause for optimism. The breakthroughs of AI technologies already demonstrated are not yet affecting much of the economy, but they portend bigger effects as they diffuse. More importantly, they will enable complementary innovations that will multiply their impact. Entrepreneurs, managers and end-users will find powerful new applications for machines that can now learn how to recognize objects, understand human language, speak, make accurate predictions, solve problems, and interact with the world with increasing dexterity and mobility.

Further advances in the core technologies of machine learning are likely to yield large benefits. However, our perspective suggests that an underrated area of research is understanding better the complements to the new ML technologies, not only in areas of human capital and skills, but also new processes and business models. The intangible assets associated with the last wave of computerization were about ten times as large as the direct investments in computer hardware itself. We think it is plausible that ML-associated intangibles can be of a comparable or greater magnitude. Given the big changes in coordination and production possibilities made possible by ML, the ways that we organized work and education in the past are unlikely to remain optimal in the future.

Relatedly, we need to update our measurement toolkits. As AI and its complements more rapidly add to our (intangible) capital stock, the traditional metrics like GDP and productivity can be increasingly misleading. Successful companies don't need large investments in factories or even computer hardware, but they do have intangible assets that are costly to replicate. The large market values associated with companies developing and/or implementing AI suggest that investors believe there is real value in those companies. What's more, the effects on living standards may be even larger than the benefits that investors hope to capture, though it's also possible, even likely, that many people will not share in those benefits. Economists are well positioned to contribute to a research agenda of documenting and understanding the often-intangible changes associated with AI and its broader economic implications.

Realizing the benefits of AI is far from automatic. It will require effort and entrepreneurship to develop the needed complements, and adaptability at the individual, organizational, and societal levels to undertake the associated restructuring. Theory predicts that the winners will be those with the lowest adjustment costs and the as many of the right complements in place as possible. This is partly a matter of good fortune, but with the right roadmap, it is also something for which they, and all of us, can prepare.

References

Abramovitz, Moses. "Resource and Output Trends in the U.S. since 1870." *American Economic Review*, May 1956 (Papers and Proceedings), 46(2), pp. 5-23.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2017. "What to Expect from Artificial Intelligence." *Sloan Management Review* Feb 7, 2017.

Alloway, Tracy. 2015. "Goldman: How 'Grand Theft Auto' Explains One of the Biggest Mysteries of the U.S. Economy." *Bloomberg Business*, May 26, 2015. http://www.bloomberg.com/news/articles/2015-05-26/goldman-how-grand-theft-auto-explains-one-of-the-biggest-mysteries-of-the-u-s-economy

Alon, Titan, David Berger, Robert Dent, and Benjamin Pugsley. 2017. "Older and Slower: The Startup Deficit's Lasting Effects on Aggregate Productivity Growth." Mimeo.

Andrews, Dan, Chiara Criscuolo, and Peter Gal. 2016." The Best *versus* the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy." OECD Productivity Working Papers, No. 5, OECD Publishing, Paris.

Aral, Sinan, Brynjolfsson, Erik and Wu, Lynn (2012) "Three-Way Complementarities: Performance Pay, HR Analytics and Information Technology." *Management Science*, 58(5):913-931.

Arrow, Kenneth. 1962. "Economic Welfare and the Allocation of Resources for Invention." in *The Rate and Direction of Inventive Activity: Economic and Social Factors*, Richard R. Nelson (ed.), Princeton: Princeton University Press: 609–26.

Atkeson, Andrew and Patrick J. Kehoe. 2007. "Modeling the Transition to a New Economy: Lessons from Two Technological Revolutions." *American Economic Review*, 97(1): 64-88.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson and John Van Reenen. 2017. "Concentrating on the Fall of the Labor Share," *American Economic Review Papers and Proceedings,* American Economic Association, vol. 107(5), pages 180-185, May.

Autor, David and Anna Salomons, 2017. "Robocalypse Now–Does Productivity Growth Threaten Employment?" MIT working paper.

Barter, Paul "'Cars are parked 95% of the time.' Let's check!" Reinventing Parking blog. http://www.reinventingparking.org/2013/02/cars-are-parked-95-of-time-lets-check.html

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb. 2017. "Are Ideas Getting Harder to Find?" Mimeo.

Bresnahan, T, E. Brynjolfsson, and L. M. Hitt (2002). "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence," *Quarterly Journal of Economics*, Vol. 117, Issue 1, Feb 2002 pp 339-376.

Bresnahan, Timothy F., Manuel Trajtenberg. 1996. "General Purpose Technologies: 'Engines of Growth'?" *Journal of Econometrics, Annals of Econometrics*, 65(1): 83–108.

Bridgman, B., Dugan, A., Lal, M., Osborne, M., & Villones, S. (2012). Accounting for household production in the national accounts, 1965–2010. *Survey of Current Business*, *92*(5), 23-36.

Brynjolfsson, Erik. 1993. "The Productivity Paradox of Information Technology", *Communications of the ACM*.

Brynjolfsson, E., & Hitt, L. M. (2000). Beyond Computation: Information Technology, Organizational Transformation and Business Performance. *Journal of Economic Perspectives*, 14(4), 23-48.

Brynjolfsson, Erik and Lorin Hitt. 2003. "Computing Productivity: Firm-level Evidence." *Review of Economics and Statistics*, 85(4): 793-808.

Brynjolfsson, Erik, Lorin Hitt, and Shinkyu Yang. 2002. "Intangible Assets: Computers and Organizational Capital." *Brookings Papers on Economic Activity*, 2002(1), Brookings Institution.

Brynjolfsson, E., and McAfee, A. 2011. *Race against the machine*. Digital Frontier, Lexington, MA.

Brynjolfsson, E., and McAfee, A. 2014. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Brynjolfsson, Erik and Andrew McAfee 2017 "What's Driving the Machine Learning Explosion?" *Harvard Business Review*, July 18, 2017.

Brynjolfsson, E., A. McAfee, M. Sorell, and F. Zhu. 2008. "Scale Without Mass: Business Process Replication and Industry Dynamics" *Harvard Business School Technology & Operations Mgt*." Unit Research Paper 07-016.

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 WP/15/116.

Cette, Gilbert, John G. Fernald, and Benoit Mojon. 2016. "The Pre-Great Recession Slowdown in Productivity." *European Economic Review*, 88, 3-20.

Congressional Budget Office. 2016. *The Budget and Economic Outlook: 2016 to 2026.*

Congressional Budget Office. 2017. *The 2017 Long-Term Budget Outlook.*

Cowen, Tyler. 2011. *The Great Stagnation: How America Ate All the Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better*. (New York: Dutton)

David, Paul. 1991. "Computer and Dynamo: The Modern Productivity Paradox in a Not-Too-Distant Mirror." In: *Technology and Productivity: The Challenge for Economic Policy*, OECD, Paris: 315–47.

David, Paul A. and Gavin Wright. 2006. "General Purpose Technologies and Surges in Productivity: Historical Reflections on the Future of the ICT Revolution." in *The Economic Future in Historical Perspective, Volume 13*, Paul A. David and Mark Thomas (eds.). Oxford: Oxford University Press/British Academy.

Deng, Jia, et al. 2009. "ImageNet: A Large-Scale Hierarchical Image Database." CVPR. *IEEE Conference Computer Vision and Pattern Recognition* IEEE, 2009.

De Loecker, Jan and Jan Eeckhout. "The Rise of Market Power and the Macroeconomic Implications." NBER Working Paper 23687.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. 2017. "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks." *Nature*, 542(7639): 115-18.

Feldstein, Martin. "The U.S. Underestimates Growth." *Wall Street Journal*, May 18, 2015.

Fernald, John G. 2014. "A Quarterly, Utilization-Adjusted Series on Total Factor Productivity." FRBSF Working Paper 2012-19 (updated March 2014).

Fortunato, Meire, Mohammad Gheshlaghi Azar, Bilal Piot, Jacob Menick, Ian Osband, Alex Graves, Vlad Mnih et al. 2017. "Noisy Networks for Exploration." *arXiv preprint arXiv:1706.10295*.

Furman, Jason, and Peter Orszag. 2015. “A firm-level perspective on the role of rents in the rise in inequality.” Presentation at “A Just Society” Centennial Event in Honor of Joseph Stiglitz Columbia University.

Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. N. (2017). Convolutional Sequence to Sequence Learning. *arXiv preprint arXiv:1705.03122*.

Gordon, R. J. 2014. *The demise of US economic growth: restatement, rebuttal, and reflections* (No. w19895). National Bureau of Economic Research.

Gordon, Robert J. 2015. *The Rise and Fall of American Growth: The U.S. Standard of Living since the Civil War*. Princeton, NJ: Princeton University Press.

Gutiérrez, Germán and Thomas Philippon. 2017. “Declining Competition and Investment in the U.S.” NBER Working Paper No. 23583.

Guvenen, Fatih, Raymond J. Mataloni, Jr., Dylan G. Rassier, and Kim J. Ruhl. 2017. “Offshore Profit Shifting and Domestic Productivity Measurement.” NBER Working Paper No. 23324.

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

Hayashi, Fumio. 1982. “Tobin’s Marginal q and Average q: A Neoclassical Interpretation.” *Econometrica*, 50(1): 213-24.

Hayashi, Fumio and Tohru Inoue. 1991. “The Relation between Firm Growth and Q with Multiple Capital Goods: Theory and Evidence from Panel Data on Japanese Firms.” *Econometrica*, 59(3): 731-54.

Henderson, Rebecca. 1993. “Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Industry.” *Rand Journal of Economics*, 24(2): 248-70.

Henderson, Rebecca. 2006. “The Innovator’s Dilemma as a Problem of Organizational Competence.” *Journal of Product Innovation Management*, 23: 5-11.

Holmes, Thomas J., David K. Levine, and James A. Schmitz. 2012. "Monopoly and the Incentive to Innovate When Adoption Involves Switchover Disruptions." *American Economic Journal: Microeconomics*, 4(3): 1-33.

Hortaçsu Ali and Chad Syverson. 2015. "The Ongoing Evolution of US Retail: A Format Tug-of-War." *Journal of Economic Perspectives*, 29(4): 89-112.

Jones, C. I., & Romer, P. M. 2010. The new Kaldor facts: ideas, institutions, population, and human capital. *American Economic Journal: Macroeconomics*, *2*(1), 224-245.

Jovanovic, Boyan and Peter L. Rousseau. 2005. "General Purpose Technologies." in *Handbook of Economic Growth, Volume 1B*. Philippe Aghion and Steven N. Durlauf, eds. Elsevier B.V.: 1181-1224.

Kendrick, John W. 1961. *Productivity Trends in the United States*. National Bureau of Economic Research. Princeton University Press.

Levine, S., Finn, C., Darrell, T., & Abbeel, P. 2016. End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, *17*(39), 1-40.

Liu, Y., Gupta, A., Abbeel, P., & Levine, S. 2017. Imitation from Observation: Learning to Imitate Behaviors from Raw Video via Context Translation. *arXiv preprint arXiv:1707.03374*.

Lucas, Jr., Robert E. 1967. "Adjustment Costs and the Theory of Supply." *Journal of Political Economy*, 75(4, Part 1): 321-334.

Manyika, James, Michael Chui, Mehdi Miremadi, Jacques Bughin, Katy George, Paul Willmott, and Martin Dewhurst. "Harnessing automation for a future that works." *McKinsey Global Institute* (2017).

McAfee, Andrew, and Erik Brynjolfsson. 2008. "Investing in the IT that makes a competitive difference." *Harvard Business Review*: 98.

Brynjolfsson, Erik and Tom Mitchell, 2017. "Automating Automation: How and Where Machine Learning Affects the Workforce" Draft.

Morris, David Z. 2016. "Today's Cars Are Parked 95% of the Time." *Fortune*, March 13, 2016

Nakamura, Leonard and Rachel Soloveichik. 2015. "Capturing the Productivity Impact of the 'Free' Apps and Other Online Media." Federal Reserve Bank of Philadelphia Working Paper 15-25.

Nordhaus, W. D. (2015). *Are We Approaching an Economic Singularity? Information Technology and the Future of Economic Growth* (No. w21547). National Bureau of Economic Research.

Orlikowski, W. J. (1996). Improvising organizational transformation over time: A situated change perspective*. Information systems research*, 7(1), 63-92.

Organization for Economic Cooperation and Development. 2015. *The Future of Productivity*.

Pratt, Gill A. "Is a Cambrian Explosion Coming for Robotics?" *Journal of Economic Perspectives* 29, no. 3 (2015): 51-60.

Saon, G., Kurata, G., Sercu, T., Audhkhasi, K., Thomas, S., Dimitriadis, D., ... & Roomi, B. 2017. English conversational telephone speech recognition by humans and machines. *arXiv preprint arXiv:1703.02136*.

Smith, Noah. 2015. "The Internet's Hidden Wealth." *Bloomberg View*, June 6, 2015. http://www.bloombergview.com/articles/2015-06-10/wealth-created-by-the-internet-may-not-appear-in-gdp.

Solow, Robert M. 1957. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics*, 39(3): 312-20.

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

Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. 2015. "Firming Up Inequality." NBER Working Paper No. 21199.

Stiglitz, Joseph E. 2014. "Unemployment and Innovation." NBER Working Paper No. 20670.

Syverson, Chad. 2013. "Will History Repeat Itself? Comments on 'Is the Information Technology Revolution Over?'" *International Productivity Monitor*, 25: 37-40.

Syverson, Chad. 2017. "Challenges to Mismeasurement Explanations for the US Productivity Slowdown." *Journal of Economic Perspectives*, 31(2): 165-86.

Wildasin, David E. 1984. "On Public Good Provision with Distortionary Taxation." *Economic Inquiry*, 22(2): 227–43.

Yang, S., & Brynjolfsson, E. 2001. Intangible assets and growth accounting: Evidence from computer investments. *Unpublished paper. MIT*, *85*(61), 28.

Appendix: Derivation of the Productivity Bias from Unmeasured AI Capital

Our setup adopts the approach of Yang and Brynjolfsson (2001) as follows. Take a constant returns to scale production function

$$Y = pF(K, N, t) \quad (1)$$

where Y is the final goods output of the firm, p is the price of final goods output, K is the vector of capital goods, N is the vector of variable inputs (e.g. labor), and t represents the level of total factor productivity at time t. With flexible capital and input prices (*r, w)*, we have the following, with g representing a growth rate:

$$g_Y = \frac{\dot{Y}}{Y} = \frac{p\left(F_K\dot{K} + F_N\dot{N} + F_t\right)}{Y} = \left(\frac{rK}{Y}\right)g_K + \left(\frac{wN}{Y}\right)g_N + g_T \quad (2)$$

The values with an upper dot represent the total derivative with respect to time.

In words, the growth in output over time can be decomposed into the growth in capital stock multiplied by capital's share of output plus the growth in flexible input quantity multiplied by the expenditure share of flexible inputs and a final total factor productivity growth term. This is the familiar Solow Residual. As mentioned above, it represents a kind of "measure of our ignorance" in how a firm converts inputs to outputs, but growth in TFP indicates improvement in productive efficiency.

To incorporate adjustment costs, we modify (1) following Lucas (1967):

$$Y = pF(K, N, I, t) \quad (3)$$

Now the production function incorporates an investment term I with market price z such that the total cost of investment in one unit of capital goods is $(z - pF_I)$. F is assumed non-increasing and convex in I to represent the idea that adjustment costs grow increasingly costly for larger I. This helps model why firms cannot, for example, instantaneously replicate the capital stocks of their competitors without incurring larger costs.

We can relate firm investment behavior to market value using this production function.[20] For the price-taking firm, market value is equal to the sum of the capitalized adjustment costs. The firm must solve:

[20] See for example Hayashi (1982), Wildasin (1984), and Hayashi and Inoue (1991).

$$\max_{I,N}\left[\int_0^{\infty}\pi(t)u(t)dt = V(0)\right]$$

$$\text{where} \quad \pi(t) = \ pF(K,I,N,t) - w'N - z'I$$

$$\text{and} \quad \frac{dK_i}{dt} = I_i - \delta_i K_i \ \forall i = 1,2,\dots,J. \quad (4)$$

That is, K_i is the capital stock of type i (indexes capital variety), N is a vector of flexible goods, $u(t)$ denotes the discount rate at time t, and δ_i is the depreciation rate of capital of type i. F is assumed non-decreasing and concave in K and N, and with homogeneity of degree one for F we get the solution to the maximization of the Hamiltonian in (5) at time 0:

$$H(K,N,I,t) = (pF(K,N,I,t) - w'N - z'I)u(t) + \textstyle\sum_{i=1}^{J}\lambda_i(I_i - \delta_i Ki) \qquad (5)$$

with first order conditions:

$$\frac{\partial H}{\partial \lambda_j} = \dot{K}_j = I_j - \delta_j K_j \quad \forall j \in \{1,2,\dots,J\}, \forall t \ \in [0,\infty]$$

$$\frac{\partial H}{\partial K_j} = -\dot{\lambda}_j = pF_{K_j}u - \lambda_j\delta_j \quad \forall j, \forall t$$

$$\frac{\partial H}{\partial I_j} = 0 = \left(pF_{I_j} - z_j\right)u + \lambda_j \quad \forall j, \forall t$$

$$\frac{\partial H}{\partial N_i} = 0 = \left(pF_{N_i} - w_i\right)u \quad \forall i \ \in \{1,2,\dots,L\}, \forall t$$

$$\lambda(\infty)K(\infty) = 0$$

leading to an equation for the value of the firm:

$$V(0) = \sum_{j=1}^{J}\lambda_j(0)K_j(0) \qquad (6)$$

The value of the firm at t = 0 is the sum over all varieties of the capital stock quantities multiplied by the "shadow price" of investment of the respective varieties. This shadow price, representative of adjustment costs in the original formulation, corresponds directly to intangible AI capital in our context. Assume that market prices correctly represent the value of claims on publicly traded firms. Equation (6) suggests that a regression of firm value on dollar quantities of asset varieties will yield a coefficient vector that represents the value of one unit of each type of capital. In a frictionless efficient market, that vector would be equal to unity for all assets. In the presence of adjustment costs, the coefficient is equal to unity plus the marginal adjustment cost for all asset varieties. This of course assumes that all asset stocks are measured perfectly.

We can extend this logic to intangible AI investments that are correlated or complementary to tangible assets and imperfectly measured. Suppose an AI-intensive firm must invest in two assets: data centers and firm-specific AI specialist training. If a firm owns a measurable quantity of tangible capital in data centers and has invested in firm-specific training of AI specialists, the estimated shadow price coefficient for the data center investment will exceed the "true" data center coefficient by the amount necessary to represent the training as well. The specialist training is not capitalized on the firm's balance sheet, yet the financial market adequately values the training service flow if no arbitrage conditions are to hold. The market value premium over book value implies a value greater than unity for Tobin's Q; the value of the firm is higher than the simple replacement cost of its *observed* assets. Technology firms have considerably higher values of Q, suggesting that they have higher levels of adjustment costs, intangible correlate investments to the booked assets, or both.

MARKET

In the growth accounting framework, the value of final goods in any given year can be divided into the value of consumption goods and the value of capital goods as follows:

$$p_c C + zI = Y = p_y F(K, N, I) = p_y F_n N + p_y F_k K + p_y F_I I = wN + rK + (z - \lambda)I \quad (7)$$

This is the growth accounting identity. The value of consumption goods plus the value of capital investment is equal to total output *Y.* This, in turn, is equal to the total income of flexible inputs, capital rental costs, and investment (both measured and unmeasured).

If $(\lambda – z)I$ value of capital goods production goes unmeasured, then part of the expenditure on capital goods is missing when the growth decomposition is performed. In the context of AI, this means that much of the training, the investment in implementing data-driven decision processes, the reorganization costs, and the incentive designs necessary to generate capital service flow from AI capital are left out.

Furthermore, if the economy is accumulating AI capital faster than it accumulates measurable capital, then TFP will be underestimated. To see why, we can update the growth decomposition equation as follows:

$$g_Y = \frac{\dot{Y}}{Y} = \frac{p\left(F_K\dot{K} + F_N\dot{N} + F_I\dot{I} + F_t\right)}{Y} \quad (8)$$

following the first order conditions for the Hamiltonian above, we have

$$\lambda_j(0) = (z_j - pF_{I_j}) \quad \text{and}$$

$$g_Y = \left(\frac{pF_KK}{Y}\right)\left(\frac{\dot{K}}{K}\right) + \left(\frac{pF_NN}{Y}\right)\left(\frac{\dot{N}}{N}\right) + \left(1 - \frac{\lambda}{z}\right)\left(\frac{zI}{Y}\right)\left(\frac{\dot{I}}{I}\right) + \left(\frac{F_t}{F}\right) \quad (9)$$

The growth decomposition now clearly shows the missing component of investment in the second to last term. Because the growth of productivity in the last term is a residual, it will also subsume the missing investment.

Thus, we have shown that in the case that the shadow price of AI investment is close to zero, there is no missing growth in output. But when there are extensive unmeasured AI investments that correlate the tangible capital goods production, as is likely to be the case, then estimates of TFP growth will be biased downward. The extent of this bias will depend on the magnitude of the unmeasured capital.