Economic Measurement of AI

(Preliminary- comments welcome)

Kristina McElheran*

September 9, 2018

This paper provides a partial review of empirical research on how firms use new information technologies, arguing for early and careful measurement of recent advances in machine learning and artificial intelligence (AI). It summarizes novel findings on important precursors to AI such as big data analytics and cloud computing, distilling key implications for researchers and policy makers interested in the diffusion and economic impact of AI. It further highlights a new data collection effort by the U.S. Census Bureau that promises to make progress with representative micro data on both the use and organizational context of new business technologies across the U.S. economy.

Keywords: artificial intelligence, cloud computing, analytics, information technology, diffusion, productivity, annual business survey

Acknowledgments and Disclaimer: Thanks are due to Chris Forman, Avi Goldfarb, Kathryn Shaw, and Prasanna Tambe for helpful early feedback. **Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. All results excerpted here have been reviewed to ensure that no confidential information is disclosed.** All errors or misconceptions remain my own.

* University of Toronto, UTSC and Rotman School of Management, k.mcelheran@utoronto.ca, 1095 Military Trail, Toronto, ON M1C 1A4.

I. Introduction

A group of technologies commonly referred to as "artificial intelligence" have been developing rapidly over the past few years. While the popular media has been captivated by the triumph of Google's AlphaGo against grandmaster Lee Sedol,[1] under the radar, its AI reduced the cooling costs for Google data centers by 40%. [2] JP Morgan's AI-powered COIN system has reduced the work of 360,000 lawyer-hours to mere seconds.[3] Starbucks uses AI to help you figure out your ideal beverage based on past preferences, the weather, and your location.[4]

Many believe that these technologies have the potential to transform economic and social outcomes such as productivity, innovation, inequality, and the nature of work (Brynjolfsson and McAfee 2014; Agrawal, Gans, and Goldfarb 2018a & b). Given the potential – and some would argue, the threat– of these technologies, it is crucial that the research community proactively consider the implications of these new technological developments.

To that end, the NBER held the first conference on the Economics of Artificial Intelligence in 2017, involving a group of leading economists in setting a research agenda for research into AI. A range of topics were introduced, from the impact of AI on the economy, to the nature of "intelligence" (human or otherwise), to the impact of AI on economic research. The emphasis of the conference was largely theoretical.[5]

The hypothetical and even speculative nature of this first foray was natural, as we lack systematic evidence on the use of AI in the economy. The technology is still relatively new, and

---

[1] https://en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol
[2] https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/
[3] https://futurism.com/an-ai-completed-360000-hours-of-finance-work-in-just-seconds/
[4] https://www.forbes.com/sites/bernardmarr/2018/05/28/starbucks-using-big-data-analytics-and-artificial-intelligence-to-boost-performance/#512d7dfd65cd
[5] See http://papers.nber.org/books/agra-1. One exception was Brynjolfsson, Rock and Syverson (2018), which provided some preliminary statistics on productivity and the gap between expectations and reality in the early days of this technological development.

"hype" far outstrips actual practice at the moment.[6] However, a notable gap in the agenda-setting process was a systematic discussion of how the profession should pursue empirical measurement moving forward. In the published volume that followed the conference, Raj and Seamans (2018) outline important issues for the measurement of robotics in the economy, some of which apply here, as well. However, a number of important considerations – particularly as they go beyond physical automation to encompass the full spectrum of AI and related technology – warrant early and serious attention. Moreover, many key insights arise out of very recent – much of it unpublished – research that has had less time to inform the conversation.

Why is worrying about measurement of AI so important at this early stage? Recent work at the frontier of economics, management, and information systems suggests that we already risk being late. Research on the economic impact of the internet was significantly hampered by data constraints, and many important dimensions of that technological sea change remain poorly understood to this day. Recent work has advanced our measurement and understanding of technologies that emerged from the Internet Age such as "big data" analytics and cloud computing (Saunders and Tambe 2013; Brynjolfsson and McElheran 2016 & 2017; Jin and McElheran 2018) that may inform our expectations concerning AI. However, key distinctions among these digital technologies, the specific ways in which they are applied, and the types of organizations that benefit from them suggest that we have very far yet to go.

This paper outlines general issues in how to measure the use of new technologies by firms at early stages of their diffusion and highlights recently-discovered methodological concerns that should play a role in AI-focused efforts. It proceeds as follows. First, I motivate why we cannot necessarily rely on prior intuitions about how firms engage with these new

[6] http://www.cityam.com/270451/gartner-hype-cycle-2017-artificial-intelligence-peak-hype

technologies. New evidence of the right kind is essential. Next, examples of challenges, successes, and lessons learned from studying the internet foreshadow core concerns for empirical studies of AI. Recent lessons from studying practices related to the use of “big data” in firms and cloud computing follow. These examples, while distinct from machine learning and AI, have key similarities that can inform empirical research on AI, moving forward. Throughout, I highlight key “take-aways” from these recent studies that should inform future data collection as well as the use of that data by researchers and policy makers. Finally, I provide an early glimpse into a new data collection effort undertaken by the U.S. Census Bureau that promises to greatly advance our understanding of the diffusion and impacts of AI in the coming year or two.

Two caveats are worth mentioning up front. The first is that this article focuses almost exclusively on *firm* use of AI and the economic, organizational, and process implications of this technological change. How individual consumers encounter and respond to AI is an important and interesting question. But it is subject to distinct considerations from those that govern firm behavior and how we might measure it. Consumer-level dynamics will also in large part be influenced by how firms adopt and deploy these technologies higher up in the value chain. Key economic implications such as changes to the nature of work or boundary of the firm require a focus on firms.

This focus is also influenced by the second caveat, which is that this article leans heavily on my own research and that of my co-authors, advisors and co-located colleagues. This is not to imply a lack of other meaningful contributions concerning the economics of technology and technological change. However, many of the key issues of interest for measurement and data collection receive short shrift in the published versions of most papers. Thus, I have less visibility to the key tradeoffs and data quality issues other endeavors have faced. Also, to properly give credit where credit is due would culminate in a review article of unruly proportions spanning a rapidly growing field – a worthy undertaking, but beyond the scope of this paper.

II. Why AI Requires New Measurement

The past decade has seen swift advances in digital information technologies that are interacting to produce exponential rates of improvement in speed and capacity. These technologies are diffusing rapidly among firms (Tambe 2014; Brynjolfsson and McElheran 2016). Moreover, they are being applied in ways that were unimaginable outside of science fiction just a few years ago. Digital assistants that shop for you on demand, self-driving cars, and deployment of police offers in pursuit of anticipated crimes have stepped off of movie screens and into our daily lives.

Less cinematically yet more economically relevant, these technologies possess tremendous flexibility in both their cost structure and application. So much, so, in fact, that new measurement – and new approaches to measurement – are required. In addition, we know more about important contingencies in the economic impacts of new digital technologies. In essence, having studied firms using the data equivalent of microscopes, it is difficult to return to coarser levels of observation. Yet the cost and difficulty of the type of data collection proposed here will, for many audiences, require some justification.

Prior Intuitions Need Not Apply

First, while we have a number of frameworks from the study of technological change that are useful for grappling with the rise of AI (e.g., Agrawal, Gans, Goldfarb 2018a & b), we cannot always apply the things we learned about prior technologies to the current context. For instance, the cannon of IT productivity research has consistently found that large incumbents tend to be the primary beneficiaries of new IT advances (e.g., Tambe and Hitt 2012). However, recent work points to stumbling blocks for leading incumbents with internet-based processes (McElheran 2015) and distinct advantages for young, small establishments when it comes to cloud computing (Jin and McElheran 2018, Retana et al., forthcoming). Reasonable people differ about whether AI will reinforce the scale advantage of large firms or lead to more "democratization" of IT. Both forces are likely in play, as nimble entrants leverage the lower fixed costs of cloud computing and established incumbents mine vast proprietary training data sets. Ultimately, the net effect in the economy – or perhaps even in a given firm – will depend on how the AI is actually implemented, in practice.

The General Purpose Technologies Framework: Implementation Matters

Implementation and variance in specific applications of technology are core themes in the large and salient literature on general purpose technologies, or GPTs. These transformational technologies are typically identified as being *pervasive* in that they are applicable across a broad range of uses, *malleable* in that they are subject to significant adaptation and improvement by the firms that use them, *innovation-promoting* by lowering the costs of product and process innovation in many areas of the economy (e.g. David 1990; Bresnahan and Trajtenberg 1995). Examples of prior GPTs include the steam engine, electrification, the combustion engine, and many waves of information technology (Milgrom and Roberts 1990, 1995; Bresnahan and Greenstein 1996; Rosenberg and Trajtenberg 2004; Jovanovic and Rousseau 2005; Forman, Goldfarb, and Greenstein 2012; Cardona, Kretschmer, and Strobel 2013). Many argue that AI falls into this category and leading scholars have leveraged insights from these past waves of innovation to help anticipate what is coming next (e.g., Mokyr 2017; Trajtenberg 2018).

An important insight from studying GPTs in the past is that the true value and economic implications of these technologies – for good or ill – only manifest after significant upheaval and re-adjustment. In the past, firms often had to relocate, change their business process, adjust complementary investments, and hire workers who themselves had to re-tool their skill sets before the full impact of these technologies became manifest in the economy (Rosenberg and Trajtenberg 2004; Jovanovic and Rousseau 2005). Some of these changes can happen within relatively short time frames, but some are subject to considerable inertia and uncertainty, both from within and outside the firm (Reinganum 1989; Bresnahan and Greenstein 1996; Henderson 1993; Christensen 1997; Afuah 2000, 2004; McElheran 2015; Gans 2016, *inter alia*). Often, the "dominant designs" (Abernathy and Utterback 1978) that a GPT will ultimately take are unclear while this adjustment process takes place, making data collection challenging simply because definitions are vague and in flux. Also, as prices for the technology fall, standard measures such as the magnitude of investment may cease to be informative.

Finally, these adjustments may even change the very nature of the central unit of observation – the firm (Hitt 1999; Forman and McElheran 2018), further complicating the task of observing and evaluating these developments in the economy.

Early Data Matters

Waiting until these adjustments settle out is a questionable approach, however. Certain firms invest in response to the "hype cycle"[7] and will respond to new technological opportunities quickly. They are best served by realistic expectations about the rate and direction of change: not just about what can technically be accomplish in a lab, but what real organizations can achieve by applying the new tech in certain ways. Other firms prefer to wait until technology is better developed and complementary resources and process flows are better understood.[8] Either way,

[7] https://www.gartner.com/technology/research/methodologies/hype-cycle.jsp
[8] For instance, Bai, Jin, McElheran, and Williams (2018) find that later adopters of Enterprise Resource Planning technologies did better than earlier adopters.

with distorted incentives due to hype on the one hand, and reluctance to plan in the face of uncertainty on the other, optimal investment decisions are unlikely. Policy makers, likewise, are flying blind without better facts to lean on, yet they face pressure to respond to these seemingly inexorable forces of change. Thus, waiting until the precise direction and rate of change become clear can have real economic and social costs.

Fine-Grained Data is Essential

To fill the current fact void, many researchers, policymakers, and managers turn to aggregate statistics to get a sense of the magnitude of the change underway. Headway has been made in robotics (Raj and Seamans 2018; Acemoglu and Restrepo 2018). However, even as we are developing a sense that these technologies may be contributing to growth and productivity at the aggregate level, we still have very little understanding about the mechanisms at work *within* firms. Without micro data, for instance, it is impossible to pin down the extent to which AI might substitute for or even complement labor – or under what conditions. The key complementary investments that make AI productive – or that hinder its application – cannot be identified in aggregate statistics. Perhaps even more important, a theme that arises from the current frontier of research into technological change is the importance of data that varies not just at the firm level, but *within* firms, as well. Looking within the firm, at the business unit – or even specific processes or tasks – yields critical insights when it comes to studying new technologies, because the applications and key complements or substitutes also vary at this level.[9]

[9] See, for example Brynjolfsson, Mitchell, and Rock (2018).

Measure Complements

Recent work on the complementary investments and design choices that enable cutting edge technologies point to critical dependencies with difficult-to-observe features of firms such as the organization of decision rights, management practices, human capital, and culture (Brynjolfsson, Hitt, and Yang 2002; Aral, Brynjolfsson, and Wu 2012; Tambe, Hitt, and Brynjolfsson 2012; McElheran 2014; Saunders and Brynjolfsson 2016; Brynjolfsson and McElheran 2017). The challenge, therefore, is that new data on AI, alone, will be insufficient to yield more than cursory insights. These important complementarities and correlates of intangible capital are traditionally very difficult to observe directly in large data sets. In sum, the data lacuna is deeper and wider than commonly understood – and the task of filling it much greater.

Beyond Robotics

Raj and Seamans (2018) argue convincingly for the value of measurement of robotics. However, while there are important links between robotics and AI (e.g., the term "automation" is used for both across contexts), there are important differences as well. In particular, a key difference between robotics and un-embodied AI such as machine learning is the sectors in which they will matter. This has implications for where to target scare resources for data collection and where to look for early evidence of the economic impacts. For instance, in the U.S. Census Bureau's program of data collection, the manufacturing surveys have historically been the quickest to respond to changes in technology– but that is not where automation of white collar work via algorithms and machine learning is likely to happen first – or with the biggest impact. Robotics may tell us a lot about manufacturing, but will be incomplete in informing the impacts of AI in services, which makes up the bulk of many modern economies. This is why new data collection efforts must target multiple sectors, and is a chief advantage of the Census

Bureau's Annual Business Survey, which is discussed in more detail in section VII. Before arriving there, the next few sections outline insights we have derived from non-AI technologies that have important implications for the measurement of AI, moving forward.

III. What we learned from studying The Commercial Internet

The example of the commercial internet is instructive for thinking about data collection and measurement issues in the early years of new technology diffusion. The network technology we know as the Internet became available to firms in 1994 and diffused so quickly and widely that, by 1995, most researchers consider availability to be nearly ubiquitous. Use of the internet varied among firms in important ways. For instance, the use and benefits were concentrated in urban areas, despite widespread expectations that it would lead to a "death of distance" (Forman, Goldfarb, and Greenstein 2005b, 2012).

Studying this period sheds light on two key modalities for collecting data on emerging technologies and their relative strengths and weaknesses. The primary windows into this phenomenon in the U.S. have been: 1) proprietary data sets collected for non-statistical purposes and, 2) administrative data collected by the U.S. government. The different methodologies possess distinct strengths and weaknesses.

Method 1: Proprietary Survey Data

Much of what we know about the early days of the internet relies heavily on proprietary data. A key resource in this stream of research has been the Harte Hanks or Computer Intelligence Infocorp (variously referred to as HH or CI), currently owned by Aberdeen. This data set is designed to support market research by vendors of a wide range of information and communication technologies (ICTs). For years, it has provided extremely rich information on a wide range of information technologies, and it possesses a number of key virtues. First, it collects establishment-level data, so it is possible to observe within-firm variation and get accurate geographic breakdowns for widely-dispersed multi-divisional firms. It is relatively representative of the economy as a whole (Forman, Goldfarb, and Greenstein 2005b), and is quick to pick up on new technology trends, sometimes with very detailed measures. It has been used extensively in research on IT diffusion and productivity (e.g.,Greenstein 1993; Bresnahan, Brynjolfsson, and Hitt 2002; Bai et al. 2018) including the internet (Forman, Goldfarb, and Greenstein 2002, 2003, 2005a & b, 2008, 2012).

Some key drawbacks of the survey, however, are less-emphasized in the literature. These flaws all derive from the core fact – which applies to most proprietary data sets – that the data are not collected with research in mind. Documentation can be poor-to-non-existent, with a lack of visibility to important elements of the data-generating process such as precise wording of the survey questions, methodologies for following up on non-response, coding of missing observations, and sampling frames. In the Harte Hanks data, in particular, linkages between establishments within firms have historically been poorly maintained (requiring extensive investments in data cleaning), and linkages within firms across years are not always reliable.

The applications focused on in this data set change over time, so that the only measure of IT use that spans long periods is personal computers. This makes tracing the lifecycle of certain applications and the shift between technologies challenging.

In recent years, the firm that purchased Harte Hanks, Aberdeen, has dramatically reduced their direct data collection and turned to a proprietary algorithm for predicting technology use at firms, limiting the data's future usefulness for economic research. Finally, these data sets can be expensive to acquire, with pricing models that incentivize researches to omit key co-variates or restrict the number of observations to relatively small samples.

Another drawback that applies particularly to technologies diffusing among young firms is that it lacks data that can be used to study performance outcomes. Matching to public data such as Compustat is typically used, which limits samples to large, public firms.

The large-firm bias was less of a problem for older, high-fixed cost technologies, as a certain scale was typically required to derive economic benefits. Over time, however, the shift to client-server architectures, internet-based networks, and now on-demand computing services via cloud computing has moved many new technologies to a variable-cost, pay-as-you-go model with dramatic implications for smaller organizations (Jin and McElheran 2018). Data sets that skimp on this increasingly dynamic segment of the economy may prove distorted insights into the spread and impact of AI.

Method 2: Administrative Data

Administrative data corrects many of these shortcomings by being far more transparent and systematic in its data collection, with far more resources (such as statutory authority to compel survey response and dedicated personnel to follow up on non-response) to deploy. We have learned many important facts about the digital economy from official statistics collected by the U.S. Census Bureau that could not have been discovered using proprietary data.

Administrative data collection on the use of the internet in the U.S. began when the Census Bureau conducted the Computer Network Use Supplement (CNUS) to the 1999 Annual Survey of Manufactures. This rich survey not only provided novel aggregate statistics on e-commerce in the U.S. (U.S. Department of Commerce "E-Stats"), but revealed important variation in how firms actually applied internet-based technologies (Atrostic and Nguyen 2006).

This survey had the additional critical advantage of being straightforward to link to other administrative data on firm accounts. This made possible detailed establishment-level analysis of productivity among users of internet applications in the dot-com era (Atrostic and Nguyen 2005) and beyond (Angle and Forman 2018).

However, there was a considerable lag between the introduction of the internet and official data collection. We do not have any systematic visibility to internet use in the critical early years. Also, this rich survey was only done in one year and only in manufacturing, and can be difficult to access due to administrative hurdles. Few researchers outside the Census Bureau have used this data to study the impacts of the internet.

One exception is Forman and McElheran (2018), which examines the extent to which firms shifted transactions from inside the firm to market-based exchange in the wake of the commercial internet. This study leaned heavily on the rare visibility to *how the technology was put into use* by adopting firms, making it possible to disentangle internal versus external coordination over the internet and to separate older networks (such as Electronic Data Interchange) from the newer technology. These types of details concerning the processes to which new technologies apply are essential – yet they remain quite rare and are often "one-off" snapshots rather than ongoing data collection efforts.

Key Learnings from Research on the Internet

The Internet Era revealed many important insights into how firms respond to and take advantage (or not) of new technological advances. Those insights can best be gleaned by going to those papers or review articles such as Forman, Goldfarb, and Greenstein (forthcoming). More to the focus of this article, the process of achieving those insights yielded important lessons for what is important and not from a data-collection standpoint:

1. **Panel micro data is important:** At the aggregate level, it may be possible to get a sense of early trends. However, capturing key elements of the phenomenon relies on firm or establishment micro data, and, in particular, panel data that keeps a reasonably large and

stable sampling frame. Proprietary data sets are often collected for short-term "market intelligence," rely on small samples that over-emphasize large firms, and do not invest in maintaining panel linkages or stable sampling frames. Administrative data is no panacea, unfortunately. Census annual surveys rotate their sample frame every five years. Also, useful surveys may not be repeated year-over-year due to the need to limit the response burden imposed by government data collection.

2. **Within-firm variation is an issue.** While a lot of our theorizing and evidence concern firms as the unit of analysis, there is important within-firm heterogeneity in the use and impacts of technology. Forman et al. (2008) find that not all units of a firm have the complementary human resources that they need to deploy internet technology, but that they can leverage internal resources elsewhere in the firm. McElheran (2014) shows that decision rights over IT purchasing can vary significantly among units of multi-establishment firms, which has important implications for where and how new technology is applied (and how well-suited it is to its local operating environment).

3. **Applications Matter.** The ways in which GPTs are applied can vary at the process level. In addition to the examples discussed above, a useful distinction in prior work was between "basic" and "advanced" internet use, which required combining a number of distinct measures of internet applications (e.g., Forman et al. 2012).

IV. Big Data, Data-Driven Decision Making and Analytics

While it seems like ancient history already, the hot thing in technology before AI was "big data" and predictive analytics. Large detailed data sets and analytical techniques based on regression analysis and data visualization promised a new ability to measure things that we have never before been able to measure. New efficiencies and innovation were expected to ensue (McAfee and Brynjolfsson 2014). Firms began concerted efforts to focus decision making more on data and evidence (Brynjolfsson, Hitt and Kim 2011). Broad-based facts concerning this phenomenon are just beginning to emerge.

Measurement was – and continues to be – a central concern. Advances in understanding the diffusion and economic implications of big data and analytics have come primarily from the deployment of new surveys specifically targeted at the phenomenon and new data collection techniques that themselves lean on analytics and big data.

Method 3: Administrative Survey Data goes Big(ish)

As mentioned above, data collection performed by administrative agencies differs in important ways from private data collection. Typically, a key advantage is size and representativeness. Usually, however, there is a tradeoff in terms of richness or the extent to which is connected to core research questions. This has very recently begun to change.

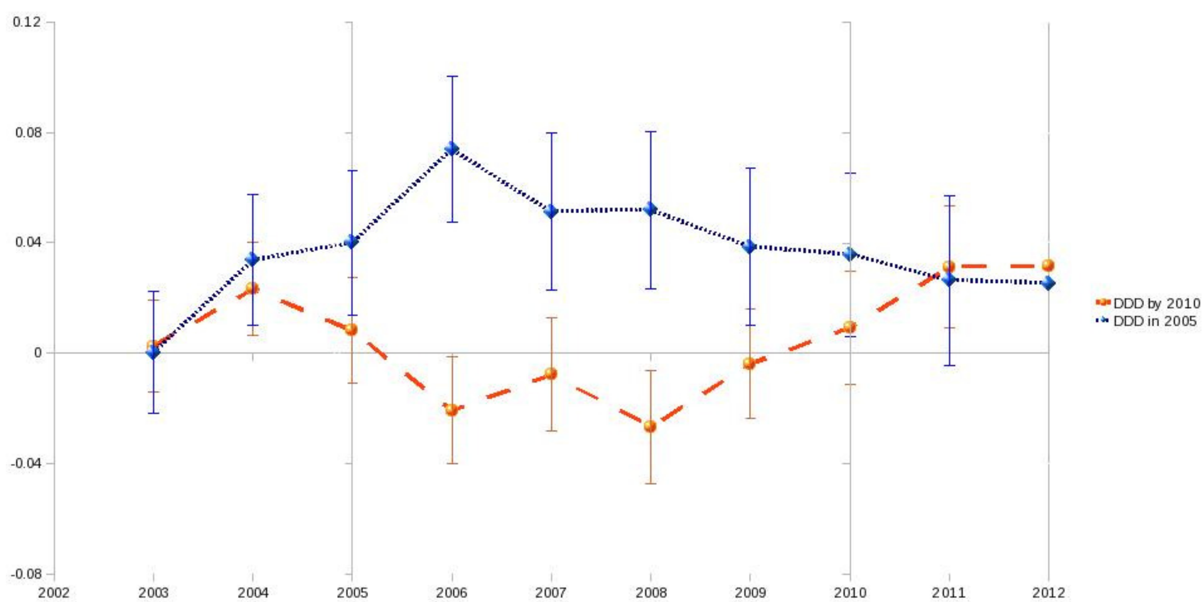
In 2010, the Census Bureau teamed up with economists and private funding sources to collect novel direct data on a range of management practices (Bloom et al. 2016 & 2018). The new Managerial and Organizational Practices Survey (MOPS) went to a sample of over 50,000 establishments in the manufacturing sector whose response was required by law. This is significant, as much of the survey data used in economic and management studies is subject to sampling and response bias that is difficult to correct for – and often overlooked or excused in the name of novelty.

The first wave of the survey asked about data-related management practices, which gave insights into how firms were making use of their data resources (Brynjolfsson and McElheran 2016 & 2017). The survey was repeated in 2015, and new questions were added about predictive analytics. Because this survey was conducted as a supplement to the Annual Survey of Manufactures, it was straightforward to link these new questions to existing panel data on firm operations. This made it possible to estimate plant-level productivity and track it over time.

Using this data set, which combined the virtues of detailed surveys of emerging technology with large and representative administrative data, we learned that the use of data in decision-making was surprisingly widespread as early as 2005. For instance, just looking at individual questions on the availability of data for decision making or use of data in decision making, roughly 45% of plants reported that they were in the highest categories by 2005. In order to identify the frontier of these practices, we combined these qualitative questions with other questions about the number of metrics tracked and whether targets were used to evaluate these metrics. The combination of questions identified a cluster of practice that only applied to 11% of plants in 2005. Note that this is very far away from a clean “pre-treatment” year – understanding what took place before 2005 and when these practices first started to diffuse is forever beyond the reach of this data.

These practices diffused very rapidly in the manufacturing sector (Brynjolfsson & McElheran 2016), and were strongly associated with productivity gains at the plant level (Bryjolfsson & McElheran 2017). The gains enjoyed by early adopters, however, were quickly achieved by establishments adopting later on in the 2005-2010 window (Figure 1).

Figure 1: Returns to Early and Late Adoption of Data-Driven Decision Making (DDD)



(Brynjolfsson and McElheran, 2017)

Had this survey take place for the first time much later, we might have missed the average effects, entirely.

Critically, we also were able to discover important complements driving variation within these average effects and over time. These productivity benefits depended in many firms on having a robust IT infrastructure prior to adopting the data-centric management practices.

The MOPS survey was repeated in 2015, and we added considerably to the section on data-driven decision making. In particular, we added questions on the use of predictive analytics within firms. These results are currently undergoing internal disclosure avoidance and content review by Census. In particular, we will soon be able to report on the evolution of data-related management practices towards frontier techniques relying on predictive analytics - both in terms of adoption and the relationship to firm performance.

Things we learned from the MOPS that apply to measuring AI:

A few insights from these recent data collection efforts are worth reporting and adding to the list of "take-aways" that have implications for the measurement of AI:

4. **Recall questions can help (with caveats).** Asking respondents in 2011 to not only report on activity from the prior year but also from 2005 allowed to us to "go back in time" to catch an earlier point on the diffusion curve. While we were successful in identifying the critical adoption and productivity gains in the 2005 – 2010 time period, many would consider this a "near-miss" in terms of data collection. A little bit later or omitting the recall questions (which increase respondent burden and come at a significant opportunity cost in terms of other data items) would have caused us to miss these insights and would have hampered our ability to use important econometric approaches such as exploring plant fixed effects (which, unsurprisingly, turn out to matter – see Brynjolfsson and McElheran 2017).

5. **Multiple measures are often critical.** Both our work and other research based on the MOPS survey emphasizes the use of multiple measures of the constructs of interest. Measurement error is a significant concern when only one respondent per organization can be queried. Using an index of questions or principal components extracted from the survey is a powerful way to reduce measurement error. However, this can make data collection much more expensive and requires foresight during the survey development stage. It also requires a strategic commitment on the part of researchers who often have

to give up collecting weaker measures of many additional interesting dimensions of firms in order to gain fewer yet more reliable measures.[10]

6. **Extensive margins are far easier to measure than intensive margins – but possibly confusing.** While it is relatively straightforward to ask about the presence of a technology (this is what Aberdeen/Harte Hanks has historically done), there are some pitfalls to be aware of. The first is experimentation and "tinkering" – firms that are experimenting with an early-stage technology (particularly AI) are not doing the same things as a firm that has deployed the tech throughout their business. Also, adoption is not necessarily an absorbing state. Early evidence from the second wave of the MOPS shows that certain plants "de-adopted" frontier data-driven decision making between waves of the survey in ways that are associated with giving up productivity gains (Brynjolfsson, Ohlmacher, Jin, McElheran, and Yang 2018).

7. **Distinguishing the application from the underlying technology can be challenging**. In developing the measure on predictive analytics, we did not forecast the convergence of "analytics" and "AI" in how the words are commonly used. To the extent that firms have been using old techniques (such as multivariate regression) to build models that predict what firms can expect in the future, they are doing predictive analytics. This is far more prevalent – but also very different in terms of the techniques, inputs, and insights – compared to predictive analytics based on machine learning. To the extent that our question captures the practice without distinguishing the techniques or inputs (notably, the size or quality of data required) we are probably missing important shifts in what "analytics" means for firms. Also, the rate of substitution between technologies will not be homogeneous. Mara Lederman succinctly identifies the problem in her note from the first NBER Economics of AI volume: "Finally, as ML and other artificial intelligence technologies diffuse across organizations, they are likely to diffuse at different rates. This means that, at least in some datasets, we are likely to observe a mix of ML -based and traditional decision-making which creates another potentially important source of unobserved heterogeneity." (Lederman 2018).

Method 5: Mine Publicly Available Data

Another key disadvantage of survey methods, be they researcher-driven or conducted with the assistance of government agencies, is the long lead-time involved. Designing a good survey that has been sufficiently piloted, subjected to cognitive testing, sent into the field, followed up to reduce non-response, and coded and cleaned prior to analysis can take months, if not years. For a

[10] Being aware of this tradeoff is probably useful for referees of these papers who often ask for richer controls or complements without understanding the tradeoffs involved.

brand-new technology that diffuses quickly, that may take too long, and key dimensions may get missed.

Another successful approach to tracking technological change in the economy is to use frontier analytical techniques to extract signals from public or quasi-public data. Tambe and Saunders (2014) used analytics methods to find analytics-based practices by using topic modelling techniques on publically available 10K filings. They linked firm reporting of key words such as "enterprise data" and "data warehouse" to performance indicators in Compustat to find a significant performance benefit for firms that report data-related activities. In another example, Tambe (2014) and Tambe and Hitt (2012) analyzed text from job postings (not necessarily public, but not exactly private, hence "quasi-public") to track changes in demand for certain IT-related skills. By measuring the complements to certain technological inputs, they estimate the use of those technologies and the labor market and firm performance implications. To the extent that particular programming approaches (neural nets, reinforcement learning, deep learning, etc.) can be specified in job postings, this might be a particularly useful approach to studying AI. At the very least, it might be a way to observe important complementary inputs that are rarely collected in other data sets or are difficult to access and could profitably be combined with direct measures of AI use. [11]

V. IT Outsourcing and Cloud Computing

Another relevant advance in measuring the diffusion and the economic implications of new technologies comes out of research into cloud computing. Understanding the diffusion and applications of cloud computing are essential for gauging the impact of AI in the economy due to

[11] Mention the LEHD and related work here? I have a project scoped out to do this, but I am not sure how much of my research agenda I should give away!

the important infrastructure it currently provides – particularly for young firms -- for advances in machine learning and access to training data for AI. Insights into this phenomenon, however, have been tremendously hampered by lack of data.

Cloud computing first started to diffuse in 2006 with the introduction of Amazon Web Services. At the same time, for unrelated reasons, the U.S. Census Bureau asked for the most fine-grained breakdown of IT expenditure to date, in the Annual Survey of Manufactures. Capitalized IT investments were separated out from other IT expenditure as early as 2002, but in 2006, Census asked firms to additionally separate out expenditures on hardware and equipment, software, and purchased IT services. This distinction starts to matter in 2009 onward, when cloud computing took off (Jin and McElheran 2018).

Catching the early wave of this new technology happened by accident, and was another “near-miss.” By 2008, we see a significant shift in the share of IT expenditure away from owned IT capital and into purchased services like cloud computing. We also observe that this shift was accentuated for young plants, which receive better coverage in the Census data than in other data sets.

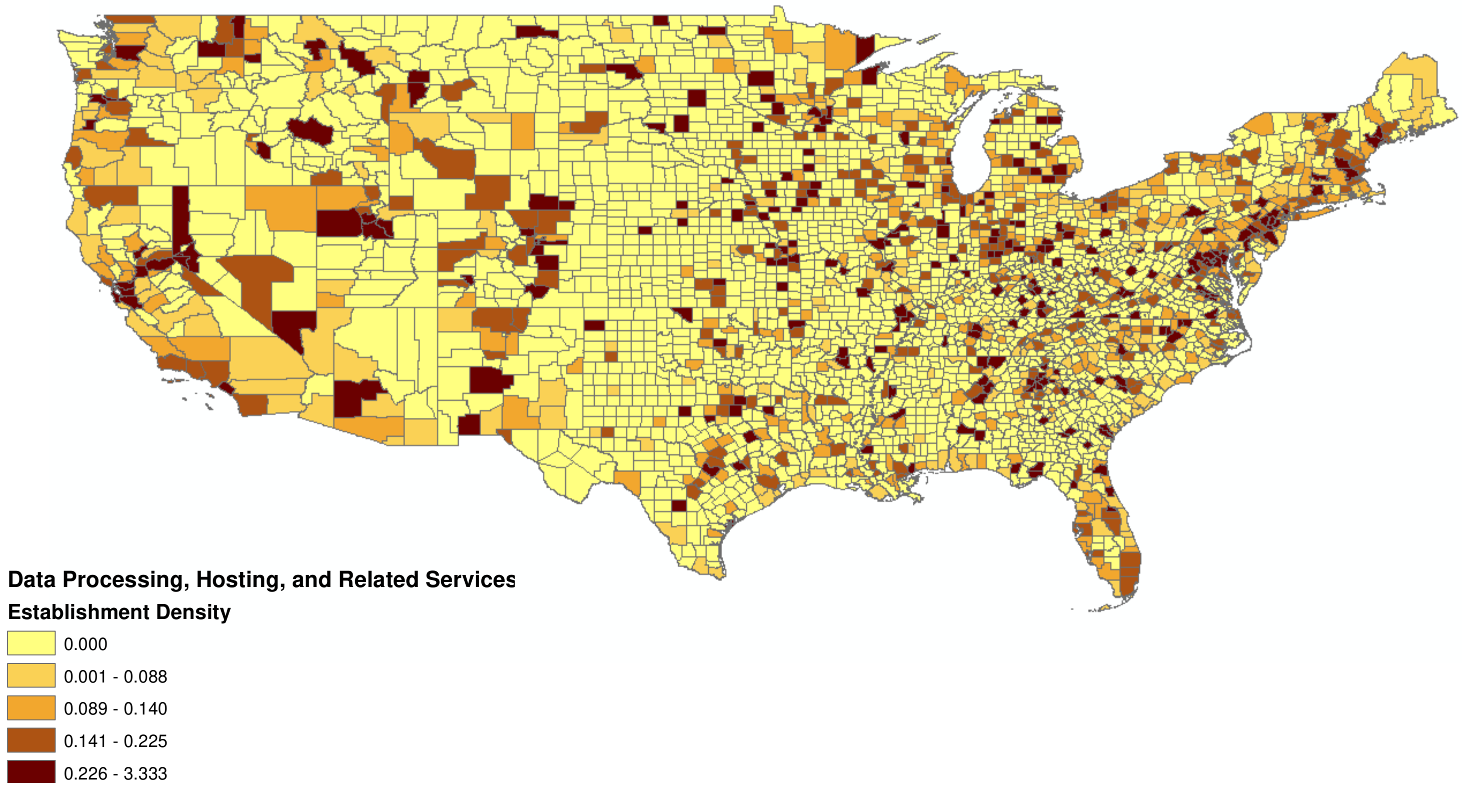
Table 1. IT Expenditure Breakdown by Type for Young and Older Plants, 2006 – 2014

	(1)	(2)	(3)	(4)
Dependent Variables	**% Expenditure on IT Capital Flows**	**% Expenditure on IT Services**	**% Expenditure on Software**	**% Expenditure on Equipment**
Young	0.009** (0.004)	-0.004 (0.003)	-0.015*** (0.003)	0.010*** (0.003)
Late	-0.054*** (0.002)	0.021*** (0.002)	0.036*** (0.001)	-0.003* (0.002)
Young x Late	-0.006 (0.004)	0.011** (0.003)	0.002 (0.003)	-0.007* (0.004)
Industry Fixed Effects	Y	Y	Y	Y
N	~239,700	~239,700	~239,700	~239,700
R-Squared	0.152	0.086	0.097	0.131

Note: Table 4 in Jin and McElheran (2018). Results are from OLS regressions controlling for industry (6-digit NAICS) fixed effects. The dependent variables are the percentage of each type of IT spending with respect to the total expenditure reported on IT. All columns include an unreported indicator for whether the plant reported zero IT expenditure (note that these are reported, not imputed zeroes). Late is the indicator for the sample years 2008 through 2014. There is no statistical difference between the Great Recession years (2008-2009) and the post-Recession period (2010-2014), so they are combined for ease of exposition. Statistical significance is denoted as follows: * 10%, ** 5%, *** 1%.

We find a strong geographic component to the diffusion of cloud computing. Using county-level prevalence of cloud services providers (see Figure 2), we found a strong correlation between nearby suppliers and use of these services. We are currently investigating the extent to which this may be due to local spillovers versus shared complements such as skilled IT labor. Some may simply be due to geographic features associated with both lower cooling costs and flatter terrain for broadband infrastructure (Kolko 2012). To the extent that these factors may be exogenous shifters of cloud computing use, they could prove to be useful instruments for the adoption of the technological infrastructure and other complements such as AI.

Figure 2. Data Processing, Hosting, and Related Services



(Jin and McElheran, 2018)

Key Learnings from Studying Cloud Computing:

8. **Age matters.** The economic benefits of the cloud show up, at least in the early years, almost entirely among young plants – both young stand-alone firms and young plants of older firms (Jin and McElheran 2018). Without visibility to the behavior of these typically smaller and non-public establishments, an important new trend in technology-driven productivity would have been impossible to observe. To the extent that AI relied on cloud infrastructure and key advances are coming from start-ups that can leverage the cloud to enter at unprecedented rates (Ewens et al. forthcoming), this distinction is likely to be quite important for measuring AI.

Method 7: All of the Above

Finally, we may also be able to make the most headway by combining these different methods. It should be possible to match up data on the same phenomenon but collected in different ways to reduce measurement error and provide complementary views on the technology, organizational and market context in which AI is diffusing. Aside from the more well-known links to Compustat, prior work has also combined Harte Hanks data with Census data to link IT adoption with organizational design and financial data (McElheran 2014). Job-posting data has been merged with publicly available reviews of firms to link IT use and IT-related practices (Tambe et al. 2018). This approach, while more costly, also has the most potential to limit the drawbacks of any one method, alone.

VII. New Measures of AI: The Annual Business Survey

A very recent data collection effort offers a new hope for making significant headway on the direct measurement of AI. Beginning in June of 2018, the U.S. Census Bureau fielded the new Annual Business Survey (ABS).[12] This survey, which replaces a number of prior surveys such as the Annual Survey of Entrepreneurs, targets 850,000 firms of all sizes and industry types. Among many questions, it contains a module on new technologies that asks "*In 2017, to what extent did this business use the following technologies in producing goods or services? Select one for each row*." The columns range from "*testing, but not using in production or service*" to "*in use for more than 25% of production or service*." The rows include a range of advanced technologies, specifically including various types of automation and machine learning. In addition to the critical new questions on the use of business technologies, the survey goes far to capturing potentially important complements to the use of these new technologies such as the amount and type of digital information, the goals and organization design of innovation, strategic positioning, and what types of customers it serves, . The availability of standard identifiers should enable linkages across Census data sets to provide additional insights, such as levels of

[12] See https://www.census.gov/newsroom/press-releases/2018/annual-business-survey.html for additional details.

investment and other operating expenses, as well. Data collection closes in late 2018, with official statistics available sometime in late 2019. Additional waves of the survey are planned.

VIII. Conclusion

The economics profession typically lags with detection and tracking of new technologies. The attention now turning to AI in the popular media, policymaking, and academic realms will help ameliorate this. However, we have to be cognizant –and transparent – concerning the limitations of all of the different measurement approaches described in this article. Throughout this article, I provide justifications for the costly and uncertain investments that will be needed to fill the current data gap and considerations that should be kept in mind while doing so. We have made considerable progress in recent years, but much remains to be done and significant obstacles lie in the path.

In particular, our shaky measurement foundation is not well-addressed by current publishing incentives and norms in the economics profession. Data collection methodologies, cleaning and processing choices, and survey validity are given relatively short shrift in seminars. Space constraints limit attention to this important dimension in journal articles, and referees focus far more on identification than measurement. Publications resting on unstable data foundations are publishable if they address "novel" questions – which poisons the well for later, more careful efforts. The creation and cultivation of large novel data sets is time-consuming – and difficult to achieve within timeframes used for promotion at many institutions.

As a result, our current measurement of how firms use technology at every level is poor. This article touches on a subset of the measurement considerations that apply, and yet the number of things requiring careful consideration is large. Moreover, the tradeoffs are binding –

given limitations of budgets and respondent attention, advances on one front typically entail retreating on another one.

Acting on the insights summarized in this article will require efforts on multiple fronts. The first is to devote far more time and attention in our research discussions to measurement. Improving the training of graduate students (and ourselves) on not only new techniques such as machine learning, but also on older disciplines such as survey methodology would improve our tools and educate the next generation of researchers and gatekeepers. Collaborations will likely be necessary to harness the funding and painstaking effort required to acquire and curate data of sufficient scope and richness. We also need to discover opportunities to complement efforts underway (e.g., in the area of robotics) and create public goods that can attract new scholars to the field and accelerate insights into this important phenomenon.

REFERENCES

Acemoglu, Daron, and Pascual Restrepo. 2017. 'Robots and jobs: Evidence from US labor markets'.

Acemoglu, Daron, and Pascual Restrepo. 2018. 'The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment', *American Economic Review Papers and Proceedings* 108: 1488-542.

Afuah, Allan. 2000. "How Much Do Your *Co-Opetitors'* Capabilities Matter in the Face of Technological Change?", *Strategic Management Journal*, 21: 387-404.

———. 2004. "Does a focal firm's technology entry timing depend on the impact of the technology on co-opetitors?", *Research Policy*, 33: 1231-46.

Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2018a. *The Economics of Artificial Intelligence: An Agenda* (University of Chicago Press: Chicago).

———. 2018b. "Exploring the Impact of Artificial Intellgience: Prediction Versus Judgment." In.

Angle, P., C. Forman. 2018. Does IT Level the Playing Field for Small Establishments? Evidence from Manufacturing. Working Paper, Georgia Tech.

Aral, Sinan, Erik Brynjolfsson, and Lynn Wu. 2012. 'Three-Way Complementarities: Performance Pay, Human Resource Analytics, and Information Technology', *Management Science*, 58: 913-31.

Atrostic, B.K., and Sang V. Nguyen. 2005. "IT and Productivity in U.S. Manufacturing: Do Computer Networks Matter?", *Economic Inquiry*, 43: 493-506.

———. 2006. "How Businesses Use Information Technology: Insights for Measuring Technology and Productivity." In *Center for Economic Studies Working Paper CES 06-15*. U.S. Bureau of the Census.

Bloom, Nicholas, and John Van Reenen. 2011. "Human Resource Management and Productivity." *Handbook of Labor Economics* (Elsevier) volume 4b.

Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt. 2002. "Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence", *Quarterly Journal of Economics*, 117: 339-76.

Bresnahan, Timothy F., and Shane Greenstein. 1996. "Technical Progress and Co-Invention in Computing and in the Uses of Computers", *Brookings Papers on Economic Activity. Microeconomics*, 1996: 1-83.

Bresnahan, Timothy F., and Manuel Trajtenberg. 1995. "General Purpose Technologies – 'Engines of Growth?'', *Journal of Econometrics*, 65: 83-108.

Brynjolfsson, Erik, Lorin M. Hitt, and Shinkyu Yang. 2002. 'Intangible assets: Computers and Organizational Capital', *Brookings Papers on Economic Activity*: 137-98.

Brynjolfsson, Erik, Tom Mitchell, and Daniel Rock. 2018. "What Can Machines Learn, and What Does It Mean for Occupations and the Economy?" In *AEA Papers and Proceedings*, 43-47.

Brynjolfsson, Erik, Lorin M. Hitt, and Heekyung Hellen Kim. 2011. "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?" MIT working paper.

Brynjolfsson, Erik, and Andrew McAfee. 2014. *The Second Machine Age* (W.W. Norton Company: New York).

Brynjolfsson, Erik, and Kristina McElheran. 2016. “The Rapid Rise of Data-Driven Decision Making”, *American Economic Review Papers and Proceedings*, 106: 133-39.

Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2018. “Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics,” in *The Economics of Artificial Intelligence: An Agenda* (University of Chicago Press: Chicago).

Brynjolfsson, Erik, Scott Ohlmacher, Wang Jin, Kristina McElheran, Mu-Jeung Yang. 2018. “Why Don’t Firms Sutain Productive Management Practices? The Case of Data-Driven Decision Making.” University of Toronto working paper.

Cardona, M., T. Kretschmer, and T. Strobel. 2013. “ICT and productivity: conclusions from the empirical literature', *Information Economics and Policy*, 25: 109-25.

Christensen, Clayton M. 1997. *The Innovator's Dilemma* (Harper Collins: New York).

David, Paul A. 1990. 'The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox', *The American Economic Review*, 80: 355-61.

Forman, Chris, Avi Goldfarb, and Shane Greenstein. 2002. “Digital Dispersion: An Industrial and Geographic Census of Commerical Internet Use,” in Steven Wildman and Laurie Cranor (eds.), *Rethinking Rights and Regulations, Institutional Responses to New Communication Technologies* (MIT Press: Cambridge).

———. 2003. 'Which Industries Use the Internet?' in Michael Baye (ed.), *Organizing the New Industrial Economy: Advances in Applied Microeconomics* (Elsevier: Bristol, UK).

———. 2005a. 'Geographic Location and the Diffusion of Internet Technology', *Electronic Commerce Research and Applications*, 4: 1-10.

———. 2005b. 'How Did Location Affect the Adoption of the Commercial Internet? Global Village vs. Urban Leadership', *Journal of Urban Economics*.

———. 2008. 'Understanding the Inputs into Innovation: Do Cities Substitute for Internal Firm Resources?', *Journal of Economics and Management Strategy*, 17: 295-316.

———. 2012. 'The Internet and Local Wages: A Puzzle', *American Economic Review*, 102: 556-75.

Forman, Chris, Avi Goldfarb, and Shane Greenstein (Forthcoming) “How Geography Shapes—and is Shaped by—The Internet” *The New Oxford Handbook of Economic Geography.*

Forman, Chris, and Kristina McElheran. 2018. "The Digital Reorganization of Firm Boundaries: IT Use and Vertical Integration in U.S. Manufacturing." Working Paper, University of Toronto.

Forman, Chris, and Nicolas van Zeebroeck. 2012. 'From Wires to Partners: How the Internet Has Fostered R&D Collaborations Within Firms', *Management Science*, 58: 1549-68.

Fort, Teresa C. 2017. 'Technology and Production Fragmentation: Domestic versus Foreign Sourcing', *The review of economic studies*, 84: 650-87.

Gans, Joshua. 2016. *The Disruption Dilemma* (MIT Press).

Greenstein, Shane M. 1993. 'Did Installed Base Give an Incumbent Any (Measurable) Advantages in Federal Computer Procurement?', *RAND Journal of Economics*, 24: 21.

Henderson, Rebecca. 1993. 'Underinvestment and incompetence as responses to radical innovation: evidence from the photolithographic alignment equipment industry', *RAND Journal of Economics*, 24.

Hitt, Lorin M. 1999. 'Information Technology and Firm Boundaries: Evidence from Panel Data', *Information Systems Research*, 10: 134-49.

Jin, Wang, and Kristina McElheran. 2018. "Economies before Scale: Learning, Survival and Performance of Young Plants in the Age of Cloud Computing." SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3112901

Jovanovic, Boyan, and Peter L. Rousseau. 2005. 'Chapter 18 - General Purpose Technologies.' in Philippe Aghion and Steven N. Durlauf (eds.), *Handbook of Economic Growth* (Elsevier).

Kolko, Jed. 2012. 'Broadband and local growth', *Journal of Urban Economics*, 71: 100-13.

McElheran, Kristina. 2014. 'Delegation in Multi-Establishment Firms: Evidence from I.T. Purchasing', *Journal of Economics & Management Strategy*, 23: 225-57.

———. 2015. 'Do market leaders lead in business process innovation? The case (s) of e-business adoption', *Management Science*, 61: 1197-216.

Mesenbourg, Thomas L. 2001. "Measuring the Digital Economy." In. U.S. Census Bureau Report.

Milgrom, Paul, and John Roberts. 1990. 'The Economics of Modern Manufacturing: Technology, Strategy, and Organization', *American Economic Review*, 80: 511-28.

———. 1995. 'Complementarities and fit strategy, structure, and organizational change in manufacturing', *Journal of Accounting and Economics*, 19: 179-208.

Mokyr, Joel. 2017. "The Past and the Future of Innovation: Some Lessons from Economic History." In.

Raj, Manav, and Robert Seameans. 2018. 'AI, Labor, Productivity, and the Need for Firm-Level Data.' in Ajay Agrawal, Joshua Gans and Avi Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda* (University of Chicago Press: Chicago).

Reinganum, Jennifer F. 1989. 'The Timing of Innovation: Research, Development, and Diffusion.' in Richard Schmalensee and Robert D. Willig (eds.), *Handbook of Industrial Organization* (North-Holland: New York).

Retana, German F., Chris Forman, Sridhar Narasimhan, Marius Florian Niculescu, and D. J. Wu. forthcoming. 'Technology Support and Post-Adoption IT Service Use: Evidence from the Cloud', *MIS Quarterly*.

Rosenberg, Nathan, and Manuel Trajtenberg. 2004. 'A General-Purpose Technology at Work: The Corliss Steam Engine in the Late-Nineteenth-Century United States', *The Journal of Economic History*, 64: 61-99.

Saunders, Adam, and Erik Brynjolfsson. 2016. 'Valuing IT-Related Intangible Assets', *MIS Quarterly*, 40: 83-110

Saunders, Adam, and Prasana Tambe. 2013. "A Measure of Firms' Information Practices Based on Textual Analysis of 10-K Filings." In.: Sauder School of Business, University of British Columbia.

Tambe, Prasanna, Lorin M. Hitt, and Erik Brynjolfsson. 2012. 'The Extroverted Firm: How External Information Practices Affect Innovation and Productivity', *Management Science*, 58: 843-59.

Tambe, Prasanna. 2014. 'Big data investment, skills, and firm value', *Management Science*, 60: 1452-69.

Tambe, Prasanna, and Lorin M Hitt. 2012. 'The productivity of information technology investments: New evidence from IT labor data', *Information Systems Research*, 23: 599-617.

Trajtenberg, Manuel. 2018. 'AI as the Next GPT: A Political Economy Perspective.' in, *The Economics of Artificial Intelligence: An Agenda* (University of Chicago Press: Chicago).

U.S. Department of Commerce. 1999. "E-Stats." *https://www.census.gov/programs-surveys/e-stats.html*