Do IT Capabilities still drive Productivity and Innovation in the Digital Age?

Erik Brynjolfsson[1], Wang Jin[2], Sebastian Steffen[3]

Abstract

In light of the increased prevalence of new information technologies, such as cloud computing and machine learning, traditional IT measures based on physical IT capital have become unreliable, while IT complementary human capital has become the new bottleneck. New IT technologies have thus, somewhat paradoxically, made the measurement of IT capabilities and their impact on firm productivity significantly harder than they already were. We create novel IT measurements based on industry-, and firm-level demands for IT skills and occupations from 2010 until 2022. Strong correlations with "official" productivity measures at the industry level validate our approach and suggest their usefulness at the firm level, where no official and reliable measures currently exist. We demonstrate that our measures are robustly associated with higher productivity at both the industry and firm levels, based on a battery of estimation techniques from the productivity literature. Our preferred firm-level estimation implies that a one percent increase in IT skills is associated with a 0.009 percent increase in total sales, which translates to an average gain of $540,000. Our measures are also positively associated with firm innovation, as measured by the total number of patents, citations, and real value of patents, suggesting that IT human capital drives productivity growth through innovation. Our methodology to define these human capital IT measures is general and simple enough to allow for future and backward-compatible extensions.

Keywords: IT capabilities, IT capital, Human capital, Measurement, IT metrics, Productivity, Innovation

---

[1] Erik Brynjolfsson, Stanford Digital Economy Lab, 366 Galvez Street, Room 238 Stanford, CA 94305 and NBER. Email:erik.brynjolfsson@gmail.com. Erik Brynjolfsson thanks the Stanford Digital Economy Lab (https://digitaleconomy.stanford.edu/) for financial support for this research. The views expressed herein are those of the authors.
[2] Wang Jin, Stanford Digital Economy Lab, Email: jwangjin@stanford.edu.
[3] Sebastian Steffen, Boston College and Stanford Digital Economy Lab, Email: sebastian.steffen@bc.edu.

1. Introduction

In 1978 Kravis, Heston, and Summers published the very first paper of the now 10th version of the Penn World Tables to more reliably measure GDP per capita (Kravis, Kravis, Heston, & Summers, 1978). They aimed “to fill, in an approximate way, a gap in the world statistical system arising from the absence of comparative data on real GDP per capita.” We now have the opportunity to offer the Information System (IS) community similar measurements for IT capabilities, measured in various manners, of industries and firms in the digital age.

IT is among the most important and productive technologies of our era and has fundamentally transformed how firms operate. A large body of academic work has connected the adoption of such technologies to improved firm productivity (Bharadwaj, 2000; Tambe & Hitt, 2012), innovation (Brynjolfsson & Saunders, 2009), customer reach, stock market returns, and supply chain management, among many other aspects. The application of IT is so ubiquitous across industry sectors that it has been argued to be a general-purpose technology (GPT) (Bresnahan & Trajtenberg, 1995). These technologies include machine learning (Brynjolfsson, Mitchell, & Rock, 2018), data-driven decision-making (Brynjolfsson & McElheran, 2019), and cloud computing (Jin & McElheran, 2017).

However, measuring the ’technological stack’ of firms is very challenging — never mind causally identifying the impact of IT on firm productivity. Besides measurement challenges, one major obstacle arises from the necessity of complementary physical and human capital for IT (Brynjolfsson & Milgrom, 2012), along with the imperative for organizational changes and managerial practices (Bloom et al., 2019) to enable productive use. Given the need for these organizational adjustments, the productivity effect of IT often requires a substantial amount of time to materialize (Brynjolfsson et al., 2020).

In the past, counting physical IT equipment, such as PCs or central processor computing capacity, served as a reasonable proxy for firm-level IT data (Bresnahan, Brynjolfsson, & Hitt, 2002; Brynjolfsson, Hitt, & Yang, 2002). With the rise of cloud storage and computing as well as machine learning, and the Everything-as-a-service paradigm, these types of measures break down and may even indicate a lack of IT capability. Renting computing and storage capacity has become increasingly cheap and easy, but measuring it reliably is impossible. Even if one could measure firms' usage of such virtual hardware, it is much more important to understand *how* — instead of if — firms leverage it. Currently, existing measures are unable to capture such detail reliably.

Traditional IT capital measures are unable to capture firms' technological progress. For example, IT capital investment shares by industry and others[4] have been virtually unchanged over the last decade and are heavily skewed towards the IT sector, as other industries outsource their IT infrastructure to AWS, Google Cloud, and Azure, among others. Bob Solow's productivity paradox, namely that "you can see the computer age everywhere but in the productivity statistics", has only worsened with modern technology innovation (Solow, Bob, 1987). In other words, traditional IT capital data predates the age of the cloud and cannot capture the meaningful IT advances that have happened across most industries (see Figure 1) and within firms. New IT technologies have thus made the measurement of industry and firm-level IT capabilities and productivity significantly harder than they already were.[5]

An additional problem is due to accounting challenges for zero-price goods. IT is commonly used to provide free services, for example as platforms for communication, advertising,

[4] These include various types of IT capital expenditures and stocks based on (but are not limited to) the Annual Capital Expenditure Survey (ACES) by the US Census and the National Income and Product Accounts (NIPA) by the Bureau of Economic Analysis (BEA) under the Department of Commerce.

[5] From an accounting perspective, cloud services allow firms to access modern IT infrastructure with a highly flexible variable cost (e.g., operating cost) and through pay-as-you-go service modes rather than through IT capital investments (e.g., Byrne, Corrado, & Sichel, 2018; Jin, 2022).

or purchasing. A price of zero implies no (direct) profit and therefore leads to these services not showing up in GDP and productivity measurements.[6]

Yet another reason why IT productivity increases may not have shown up in official statistics is that there may be significant lags in adoption or unlocking the technology's potential. Previous GPTs, such as electricity, required a decades-long learning and adjustment period before manufacturers were able to take significant advantage of them (Brynjolfsson & McAfee, 2014). Could it be the same for recent information technologies? Given that there are significant IT productivity differences across countries, industries, and firms, learning and adjusting seem to play an important role (Bloom, Sadun, & Van Reenen, 2012).

In many ways, recent information technologies are different from previous innovations (Agrawal, Gans, & Goldfarb, 2019). These technologies have achieved super-human performance on a wide range of (albeit relatively narrow and domain-specific) tasks and have the potential to directly or indirectly displace a large number of workers. Theories of routine-biased technological change (Goos, Manning, & Salomons, 2014) suggest that routine tasks will be most easily and quickly replaced, though the large cost savings associated with displacing more expensive and complex tasks and workers implies higher returns for investing in IT that can displace 'high-skill' work. Notably, large language models and generative AI have the potential to impact even cognitive and creative tasks, such as writing and painting and suggest a theory of cognitive-biased technological change.

The interconnectedness of today's digital world also induces significant externality and network effects. IT may be adopted to steal market share from competitors instead of to increase profits. While improvement for selfish reasons and competition is in the vein of Adam

[6] See (Brynjolfsson, Collis, Diewert, Eggers, & Fox, 2019) for an in-depth discussion and a proposed alternative, namely GDP-B.

Smith’s ’invisible hand’, IT may have significant negative externalities that turn market competition into a zero-sum game instead of a win-win. For example, many IT innovations are aimed at grabbing consumers' attention and time and lead to distraction and reduced worker productivity. Certain types of IT may therefore be relatively less productive compared to past innovations, such as electrification, and infrastructure developments.

Part of the difficulty of measuring IT stocks and productivity is due to the inherent nature of these technologies: they are rapidly changing and innovating, at a pace virtually unmatched by earlier technologies, and much of their value and impact derives from complementary, often intangible investments in skills, business processes, and organizational capital. In the age of AI and low-cost, and easily accessible cloud-based services — also referred to as XaaS to encompass software (SaaS), platforms (PaaS), and infrastructure (IaaS) as a service, among others — the physical number of computers that a firm or an industry owns are no longer a meaningful measure of IT sophistication (Brynjolfsson et al., 2002). Even in 1987, Robert Solow summarized that ’you can see the computer age everywhere except in the productivity statistics. The rise of the productivity-enhancing effects of intangible IT has been documented through indirect measurement exercises of firm productivity residuals (Bartel, Ichniowski, & Shaw, 2007; Brynjolfsson et al., 2002) as well as market-based estimates of Tobin’s q.

We argue that firms’ demand for IT human capital – IT skills and labor – is (i) a major component of IT sophistication, (ii) a critical complement to IT capital, and (iii) more easily measurable than IT capital and is therefore a better way to estimate firms’ IT capabilities. Specifically, we measure industries’ and firms’ demand for IT skills and labor through their US online job postings between 2010 and 2020.

2. Data & Methodology

We create three different types of definitions of IT metrics based on (i) occupational demands of job postings, (ii) skill demands of job postings, and (iii) combining occupation-level measures based on prior literature with occupational demands of job postings.

The first set of metrics is directly calculated from the occupational demands of U.S. online job postings. This data, which we acquire from Lightcast, consists of roughly 330 million job postings and contains information on each job posting's occupational code (based on the Standard Occupational Classification (SOC) system), industry code (based on the North American Industry Classification System (NAICS)), posting date, skill demands, and firm name, among many others. We then aggregate the job postings' occupational demands to the firm (industry) level by taking a weighted average, with the weights being defined by the firm-level (industry-level) shares of job postings with a corresponding occupational code in a given year.

The second set of metrics is similarly derived from Lightcast but leverages the skill demands instead of the occupation associated with each job posting. We derived this set of metrics based on over 16,000 skills in Lightcast's legacy skill taxonomy.[7] These are mapped into roughly 800 skill clusters, which themselves are mapped into 27 skill cluster families. For example, the skill Python falls into the skill cluster 'Coding and Scripting Languages', which itself falls into the skill cluster family 'Information Technologies'. The 'Information Technologies' skill cluster family defined by the Lightcast taxonomy naturally extends to a definition for identifying information technology skills within job postings, and thus firms and industries, but we also define skill-based measures that capture more narrow IT capabilities, such as cloud, machine learning (ML/AI), natural language processing (NLP), and cybersecurity. Some additional care is required

[7] Lightcast has expanded its taxonomy to cover over 32,000 skills since then.

to count and deal with skills, as some firms appear to engage in strategic signaling and demand significantly more skills for some job postings than can reasonably be supplied by job applicants.[8]

The third, and final, set of metrics is derived by combining the Lightcast data with occupation-level metrics defined by prior literature, such as the Suitability for Machine Learning (SML) (Brynjolfsson et al., 2018), among others. Using the number of job postings for each SOC code demanded by each firm and industry, we are then able to aggregate these occupation-level metrics to firm-, and industry-level metrics using a weighted average.

Before explaining these three sets of metrics, we give an overview of the underlying job posting data.

2.1 Job Postings from Lightcast

We obtain detailed, annotated job postings data from Lightcast - a high-quality data source with comprehensive coverage of job posting portals beginning in 2010 and with increasing popularity in recent research (e.g., Acemoglu, Autor, Hazell, & Restrepo, 2020; Hershbein & Kahn, 2018).[9] Using these data we can create simple, yet insightful (and computationally-expensive) aggregations. We classify each job posting according to different categorical definitions and then, for each firm and industry, calculate its share of all job postings that fall within each category.

2.2 Metrics Derived from Job Postings' Occupational Codes

Using the SOC taxonomy, we derive several occupation-based IT metrics. Currently, we aim to capture (i) all computer occupations (SOC codes beginning with 15), (ii) Cybersecurity occupations, (iii) Machine-Learning-related occupations, (iv) Cloud-related occupations, and (v)

---

[8] Some job postings demand up to 400 skills according to Lightcast. This highlights a potential weakness of these data: they do not contain information on how important each skill will be for the actual job. Employers may also list substitute skills along with the skills they actually require - for example, they may demand R even though their codebase relies on Python.

[9] Lightcast provides partial coverage for 2007, but does not provide any coverage of job postings in 2008 or 2009. We therefore use Lightcast data starting from 2010.

NLP-related occupations. For each definition we then identify the shares of an industry's (firm's) job postings that do and do not fall into it:

$$\mathrm{it}(i,t)=\frac{\sum_{p_j}1[\mathrm{soc}(p_j)\in\{\text{IT Def}\},\mathrm{ind}(p_j)=i,\mathrm{year}(p_j)=t]}{\sum_{p_j}1[\mathrm{ind}(p_j)=i,\mathrm{year}(p_j)=t]}$$

where p_j is the j^{th} job posting, time, $t\in\mathcal{T}\equiv\{2010,\ldots,2020\}$ and industry $\mathrm{ind}\ \mathrm{i}\in\mathcal{I}\equiv\{11,21,22,23,31{-}33,42,44{-}45,48{-}49,51,52,53,54,55,56,61,62,71,72,81,92\}$, as defined by NAICS.[10] For example, our definition of general IT skill capabilities is based on the 'Information Technology' skill cluster family, while our definitions for cloud and cybersecurity are based on more narrowly defined skill clusters within our data.[11] The definition for firms is analogous, except that calculations are done across firms instead of NAICS industries.

2.3 Metrics Derived from Job Postings' Skill Demands

The Lightcast skills data is annotated via Lightcast's industry-leading skill parser, which identifies skills in the job posting's text and maps them into a detailed skills taxonomy. The skill taxonomy covers over 16,000 skills, which are grouped into over 600 skill clusters, which themselves are further grouped into 28 skill cluster families. The taxonomy was initially assembled from online resumes and is continuously updated through client feedback, research, and forums. K-Means clustering along with additional qualitative checks were employed to create meaningful skill clusters. Whenever new skills are added to the taxonomy, the labels are refit to the entire history of job posting data, which minimizes potential biases that may arise through Lightcast's time-varying ability to capture new skills.

[10] We omit the lists of SOC codes corresponding to each of our definitions here but these are available in the working paper and upon request.
[11] Several of our IT capability measures can be defined both through demand for skills, as well as occupations. See a use case of the cybersecurity IT capability measure in (Bana, Brynjolfsson, Jin, Steffen, & Wang, 2021).

Notably, the Lightcast taxonomy is significantly more detailed than other skill taxonomies, and thus allows micro-level insights into which specific skills employers demand. The O*Net skill taxonomy, which contains 2 levels, with 35 skills grouped into 6 skill groups, is unable to provide such detail.[12]

The methodology to derive industry-level IT measures based on job posting-level skill demands is similar to that of the previous section but requires an additional aggregation step. Specifically, we first create occupation-level skill demands as in Brynjolfsson and Steffen (2021). This aggregation accounts for the large heterogeneity in the number of skills that each job posting demands. This way, job postings that demand a very large number of skills do not outweigh job postings that demand fewer skills, which reduces potential biases due to job postings that may cast too wide of a net. Thus, for each job posting, p_j, we start with the binary Lightcast skill vectors over the entire set of skills, $S = S_1, \ldots., S_k$, which we call $S(p_j) = (1[S_1], \ldots, 1[S_K])$, and normalize them to derive the weighted skill demands of each job posting:

$$s(p_j) = \left(\frac{S_1(p_j)}{\sum_k S_k(p_j)}, \ldots, \frac{S_K(p_j)}{\sum_k S_k(p_j)} \right)$$

To put it in words, for our preferred skill-based measure, we first derive a weighted skill demand vector for each job posting. Before we can aggregate these job-posting-level skill demands to the industry-year (as well as firm-year) level, just like before, we define the following skill-based IT metrics: Information Technology Skills, Cloud Skills, Cybersecurity Skills, and AI Skills.

Again, there are many other possible, subjective definitions. Equipped with these different definitions of IT metrics, just like before, we aggregate from the job-posting to the corresponding industry-year (or firm-year) level, for each (i, t) combination, by summing the relevant skills:

[12] While the distinction between skill, task, ability, knowledge, tool, work context, or experience is not entirely clear, we believe that the Lightcast taxonomy gets closest to the layman understanding of skill and certainly closer than education or wage, which were commonly used in the past.

$$it(i,t) = \frac{\sum_{p_j} \sum_{k \in \text{IT Deg}} s_k(p_j) 1[\text{ind}(p_j) = i, \text{year}(p_j) = t]}{\sum_{p_j} \sum_k s_k(p_j) 1[\text{ind}(p_j) = i, \text{year}(p_j) = t]}$$

2.4 Metrics derived from Scholarly Work

Our final set of IT metrics leverages measures defined by prior scholarly work. Specifically, we take the Work-from-Home (WFH) scores from Bai, Brynjolfsson, Jin, Steffen, & Wan (2021), which are based on Dingel and Neiman (2020); the Suitability for Machine Learning (SML) scores from Brynjolfsson et al. (2018); the abstract, routine, and manual task scores from Autor & Dorn (2013) , the Artificial Intelligence (AI), Software, and Robot technology scores from Webb (2020), and the AI scores from Felten et al. (2023). All of these scores are defined at the occupational, 6-digit SOC level, which allows us to merge them with our industry-year (and firm-year) panel of occupational counts. To aggregate, we then take a weighted average of the occupation-level scores, where the weights are defined by the occupational counts.

3. Results

3.1. Industry-Level Analysis

We start our analysis by examining the correlations of our Lightcast IT metrics with the official IT capital measures from the Bureau of Labor Statistics (BLS) and the Bureau of Economic Analysis (BEA). Overall, we find that while the correlation is positive and high at around 0.5 across most measures, there are notable differences, which highlight the advantage of our new measure.

Specifically, traditional IT measures suggest that the healthcare sector is the second-most IT capable, behind the IT sector, but ahead of the professional services sector and manufacturing. While there is no ground truth for the IT capabilities of industries, these numbers do not seem to represent anecdotal evidence on hospitals struggling with ransomware or other examples of bad tech) and highlight the importance of considering human capital instead of just physical and software IT capital. Firms in the IT, manufacturing, and wholesale sectors invest in both physical capital, such as data centers, and automation machinery, as well as software and human capital. The professional services sector is known to be one of the leading investors in software and IT human capital but tends to outsource physical IT capital, such as through cloud services and data warehousing, leading to a significant downward bias in traditional measures when measuring IT investment.

Importantly, this type of mismeasurement also significantly distorts the estimates of returns and productivity of IT. We first demonstrate in Figure 2 that our general IT skill measure, which is unavailable in traditional data, is positively and significantly correlated with total factor productivity (TFP). Further, our results indicate that this correlation increased over the last decade, while the correlation between BLS physical IT capital and TFP decreased. This suggests an increasing importance of IT human capital and highlights the contribution of our IT metrics. Overall, our measures are validated at the industry level, but we also created analogous measures of IT capabilities at the firm level, where no official measures exist.

3.2. Firm-Level Analysis

We start our firm-level analysis by matching our firm-level IT metrics from the Lightcast data with the Compustat data, which gives us access to public firms' operational data including annual revenue, total employment, cost of goods sold, and total assets.

Since Compustat and Lightcast do not share a common firm identifier, we use a combination of name and address fuzzy matching to construct a bridge between Compustat and Lightcast data. We also follow Campello, Gao, & Xu (2019) and match the employers to the subsidiaries of Compustat firms using information extracted from historical Orbis data provided by Bureau van Dijk (BvD). We manually check the links identified to ensure the accuracy of our matching.

We estimate the impact of IT human capital on firms' productivity following conventional methods in the IT productivity literature (Brynjolfsson & Hitt, 2003; Tambe & Hitt, 2012). Essentially, we estimate a revenue-based Cobb-Douglas production function in the baseline models. Controlling for key inputs including total capital stock (i.e., total assets), total number of employees, cost of goods sold, and industry and year fixed effects, the estimated coefficient in Column 1 is positive and significant with a magnitude of 0.030. This indicates that a one percent increase in the total number of IT skills is associated with a 0.03 percent increase in total sales. This can be translated into 1.76 million dollars.

In Columns 2 and 3, we add firm and year fixed effects to further control for firm-level time-invariant and slow-moving unobservable variables, such as organizational structure and/or management practices as well as other inputs including software investments, leverage, cash holdings, and R&D investments. The estimated coefficients for Log Total IT Skills are smaller in magnitude but stay positive and statistically significant at the 1% level. Further robustness tests using alternative IT measures (i.e., Log Total IT Job Postings), Translog production function, and estimating a value-added specification provide similar results. It is important to point out that although controlling for firm fixed effects addresses issues with time-invariant unobservable variables, it does not control for time-varying identification threats. Meanwhile, the firm fixed effects model may bias our estimates downward by stripping out slow-moving organizational complements that are important for IT to realize their productivity effect (e.g., Brynjolfsson & Hitt, 1995) as well as by magnifying measurement errors in the panel study.

One particular concern in the IT productivity literature centers on unobserved time-varying productivity shocks that may simultaneously boost output and IT investment. This kind of shock would lead to an upward bias on the estimated impact of IT on productivity. We address this concern using a series of widely adopted econometric models (Ackerberg, Caves, & Frazer, 2015; Arellano & Bond, 1991; Blundell & Bond, 2000; Levinsohn & Petrin, 2003) and report results in Table 2. These models have been demonstrated to perform well in the prior IT productivity literature (e.g., Tambe and Hitt 2012).

Column 1 in Table 2 uses the semiparametric method developed by Olley & Pakes (1996), which uses capital investment (both structure and equipment) as a proxy for unobservable shocks that could lead to a spurious correlation between IT investment and productivity. Column 2 instruments for unobserved shocks with the cost of materials (Levinsohn and Petrin 2003). Ackerberg et al. (2015) discuss the limitations of the aforementioned approaches and propose an alternative. We present their estimator in column 3. Column 4 uses a System GMM estimator, relying on 2-period lagged differences for all variables to instrument for current-period investment levels (Arellano and Bond 1991, Blundell and Blond 2000), which addresses potential endogeneity concerns of IT human capital in the productivity estimation. The coefficients on Log Total IT Skills remain positive and significant at the 1% level. In addition, they are also consistently higher compared to the result from the Pooled OLS model in Column 1 Table 1. This pattern across models with varying and distinct identifying assumptions is consistent with a downward bias on the baseline estimates, further suggesting that our findings are likely causal.

While so far we built a causal link between firms' IT skill acquisitions and increased productivity, the mechanism through which this increase occurs is unclear. One channel we explore is the impact of IT skill acquisition on innovation (e.g., Brynjolfsson & Saunders, 2009;

Ravichandran, Han, & Mithas, 2017). We match our analytics sample with firms' patent and citation data from the USPTO based on data provided by Kogan, Papanikolaou, Seru, & Stoffman (2017). In doing so, we gain access to information on the number of patents applied for, the number of patents issued, the real value of innovation, and forward citations. We estimate the association between IT skills and firms' innovation outcomes using a model similar to the one used in Column 3 Table 1 and present the results in Figure 3. Although the magnitudes of the coefficients vary, they are all positive and significant at the 1% level, suggesting that firms' IT skills are positively associated with innovation outcomes.

4. Conclusion

In light of the increased prevalence of new information technologies, such as cloud computing and machine learning, traditional IT metrics based on physical IT capital have become less reliable. The measurement of industries' and firms' IT capabilities and productivity has become increasingly challenging. IT-complementary skills and human capital have become operation-critical and firms' demand for them is an important component of firms' IT capabilities. We revisit the IT productivity debate and overcome these data issues by leveraging large and granular data from online job postings to derive novel industry-, and firm-level measurements of IT capabilities, validate them against official numbers, where they exist, and derive novel IT productivity estimates from TFP regressions. We further provide evidence that our measure of IT skills is positively associated with firms' innovation outcomes including patent filed and issued, and innovation values.

We believe that making these metrics publicly available offers significant value to the IS/IT community and can help to shed light on the IT productivity paradox, firm productivity, and related questions. Our methodology to define these measures is general and simple enough to allow for future, and backward-compatible, extensions and we plan to build and release future versions in correspondence with the academic community.

References

Acemoglu, D., Autor, D., Hazell, J., & Restrepo, P. (2020). AI and Jobs: Evidence from Online Vacancies. *National Bureau of Economic Research Working Paper Series, No. 28257*. https://doi.org/10.3386/w28257

Ackerberg, D. A., Caves, K., & Frazer, G. (2015). Identification Properties of Recent Production Function Estimators. *Econometrica : Journal of the Econometric Society.*, *83*, 2411–2451.

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction. *Journal of Economic Perspectives*, *33*, 31–50.

Autor, D. H., & Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US Labor Market. *American Economic Review*, *103*, 1553–1597.

Bai, J. (Jianqiu), Brynjolfsson, E., Jin, W., Steffen, S., & Wan, C. (2021). Digital Resilience: How Work-From-Home Feasibility Affects Firm Performance. *NBER Working Paper*.

Bana, S., Brynjolfsson, E., Jin, W., Steffen, S., & Wang, X. (2021). Cybersecurity Hiring in Response to Data Breaches. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.3806060

Bartel, A., Ichniowski, C., & Shaw, K. (2007). How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills. *Quarterly Journal of Economics*, *122*, 1721–1758.

Bharadwaj, A. S. (2000). A Resource-Based Perspective on Information Technology Capability and Firm Performance: An Empirical Investigation. *MIS Quarterly*, *24*, 169–196. Business Source Complete.

Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R., Patnaik, M., Saporta-Eksten, I., & Reenen, J. V. (2019). What Drives Differences in Management Practices? *American Economic Review*, *109*, 1648–1683.

Bloom, N., Sadun, R., & Van Reenen, J. (2012). Americans do IT better: US multinationals and the productivity miracle. *American Economic Review*, *102*, 167–201.

Blundell, R., & Bond, S. (2000). GMM Estimation with persistent panel data: An application to production functions. *Econometric Reviews.*, *19*, 321–340.

Bresnahan, T. F., Brynjolfsson, E., & Hitt, L. M. (2002). Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *The Quarterly Journal of Economics*, *117*, 339–376.

Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies "Engines of growth"? *Journal of Econometrics*, *65*, 83–108.

Brynjolfsson, E., Collis, A., Diewert, W. E., Eggers, F., & Fox, K. J. (2019). GDP-B: Accounting for the Value of New and Free Goods in the Digital Economy. *National Bureau of Economic Research Working Paper Series*, *No. 25695*. https://doi.org/10.3386/w25695

Brynjolfsson, E., & Hitt, L. (1995). Information technology as a factor of production: The role of differences among firms. In *Information technology as a factor of production: The role of differences among firms*. Cambridge, Mass: Industrial Performance Center, Massachusetts Institute of Technology.

Brynjolfsson, E., & Hitt, L. M. (2003). Computing Productivity: Firm-Level Evidence. *The Review of Economics and Statistics*, *85*, 793–808.

Brynjolfsson, E., Hitt, L. M. (Lorin M., & Yang, S. (2002). Intangible Assets: Computers and Organizational Capital. *Brookings Papers on Economic Activity*, *2002*, 137–198.

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

Brynjolfsson, E., & McElheran, K. (2019, July 6). *Data in Action: Data-Driven Decision Making and Predictive Analytics in U.S. Manufacturing* [SSRN Scholarly Paper]. Rochester, NY. https://doi.org/10.2139/ssrn.3422397

Brynjolfsson, E., & Milgrom, P. (2012). Complementarity in Organizations. In R. Gibbons & J. Roberts (Eds.), *The Handbook of Organizational Economics* (pp. 11–55). Princeton University Press.

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What Can Machines Learn and What Does It Mean for Occupations and the Economy? *AEA Papers and Proceedings*, *108*, 43–47.

Brynjolfsson, E., & Saunders, A. (2009). *Wired for innovation: How information technology is reshaping*

the economy /. Cambridge, Massachusetts : MIT Press,.
Byrne, D., Corrado, C., & Sichel, D. E. (2018). *The Rise of Cloud Computing: Minding Your P's, Q's and K's* (Working Paper No. 25188). National Bureau of Economic Research.
Campello, M., Gao, J., & Xu, Q. (2019). Personal Income Taxes and Labor Downskilling: Evidence from 27 Million Job Postings. *Kelley School of Business Research Paper*.
Felten, E. W., Manav, R., & Seamans, R. (2023). Occupational Heterogeneity in Exposure to Generative AI. *SSRN Electronic Journal*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4414065
Goos, M., Manning, A., & Salomons, A. (2014). Explaining Job Polarization: Routine-Biased Technological Change and Offshoring. *American Economic Review*, *104*, 2509–2526.
Griliches, Z., & Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics.*, *31*, 93–118.
Hershbein, B., & Kahn, L. B. (2018). Do Recessions Accelerate Routine-Biased Technological Change? Evidence from Vacancy Postings. *American Economic Review*, *108*, 1737–1772.
Jin, W. (2022, July 25). *Cloud Adoption and Firm Performance: Evidence from Labor Demand* [SSRN Scholarly Paper]. Rochester, NY. https://doi.org/10.2139/ssrn.4082436
Jin, W., & McElheran, K. (2017). Economies Before Scale: Survival and Performance of Young Plants in the Age of Cloud Computing. *SSRN Electronic Journal*, 1–74.
Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological Innovation, Resource Allocation, and Growth. *Quarterly Journal of Economics*, *132*, 665–712.
Kravis, I. B., Kravis, I. B., Heston, A. W., & Summers, R. (1978). Real GDP Per Capita for More Than One Hundred Countries. *The Economic Journal : The Journal of the British Economic Association.*, *88*, 215–242.
Levinsohn, J., & Petrin, A. (2003). Estimating Production Functions Using Inputs to Control for Unobservables. *The Review of Economic Studies.*, *70*, 317–341.
Olley, G. S., & Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica : Journal of the Econometric Society.*, *64*, 1263–1297.
Ravichandran, T., Han, S., & Mithas, S. (2017). Mitigating Diminishing Returns to R&D: The Role of Information Technology in Innovation. *Information Systems Research*, *28*, 812–827. Business Source Complete.
Solow, Bob. (1987, July 12). We'd better watch out. *New York Times Book Review*, 36.
Tambe, P., & Hitt, L. M. (2012). The Productivity of Information Technology Investments: New Evidence from IT Labor Data. *Information Systems Research*, *23*, 599–617.
Webb, M. (2020). The Impact of Artificial Intelligence on the Labor Market. In *Working Paper*.

Figures and Tables

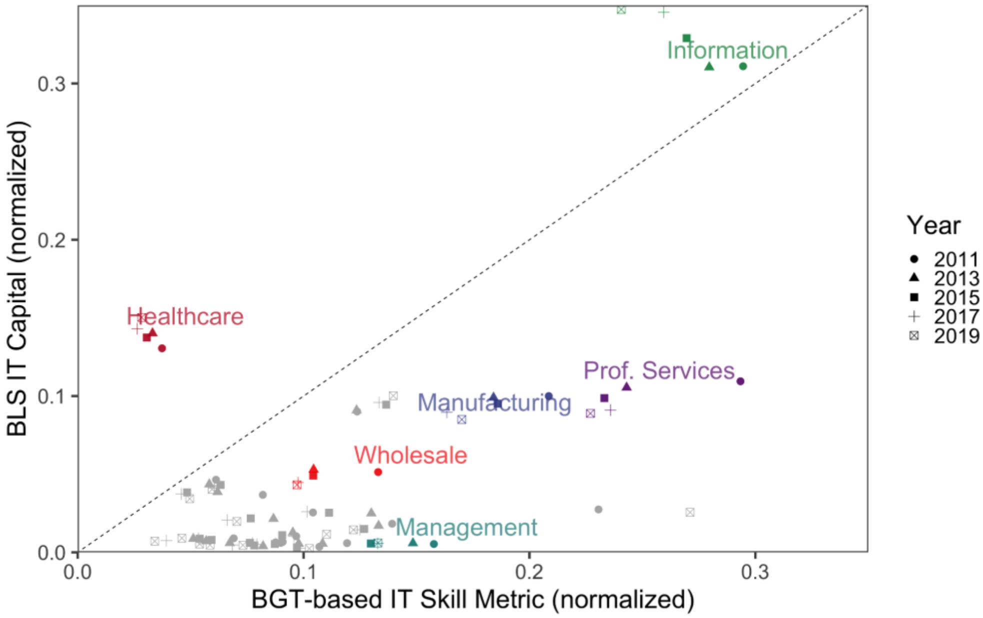

Figure 1: IT Skill Metric and IT Capital by Sectors

Notes: Traditional IT capital investment shares (y-axis) have remained relatively constant over time and do not show the increasing IT capabilities across industries. It overrepresents the IT capabilities of the healthcare and information industries compared to our IT skill demand measure (x-axis).

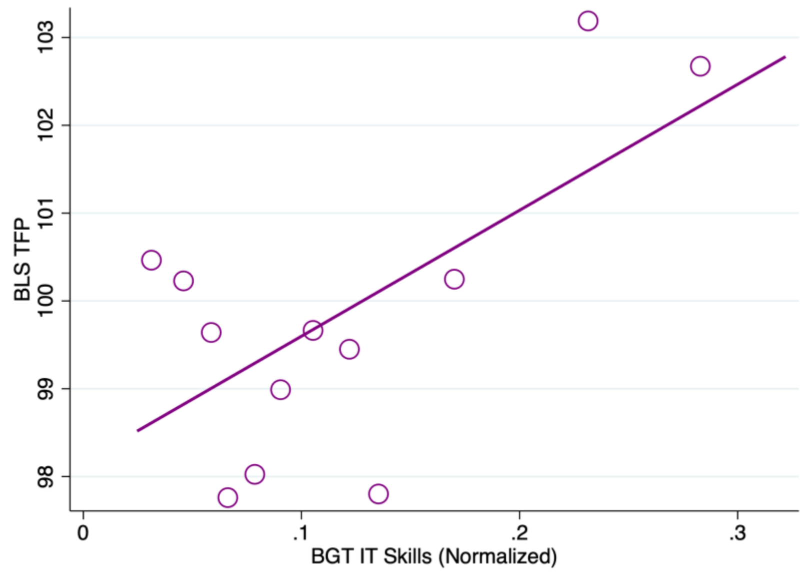

Figure 2 Correlation between Lightcast IT Metrics and BLS Total Factor Productivity

Notes: Binscatter plot using STATA 16. The Y-axis is the BLS annual total factor productivity for major industries (see https://www.bls.gov/productivity/tables/ for more information). The X-axis is the normalized IT skills measure calculated based on the Lightcast (formerly BGT) data.

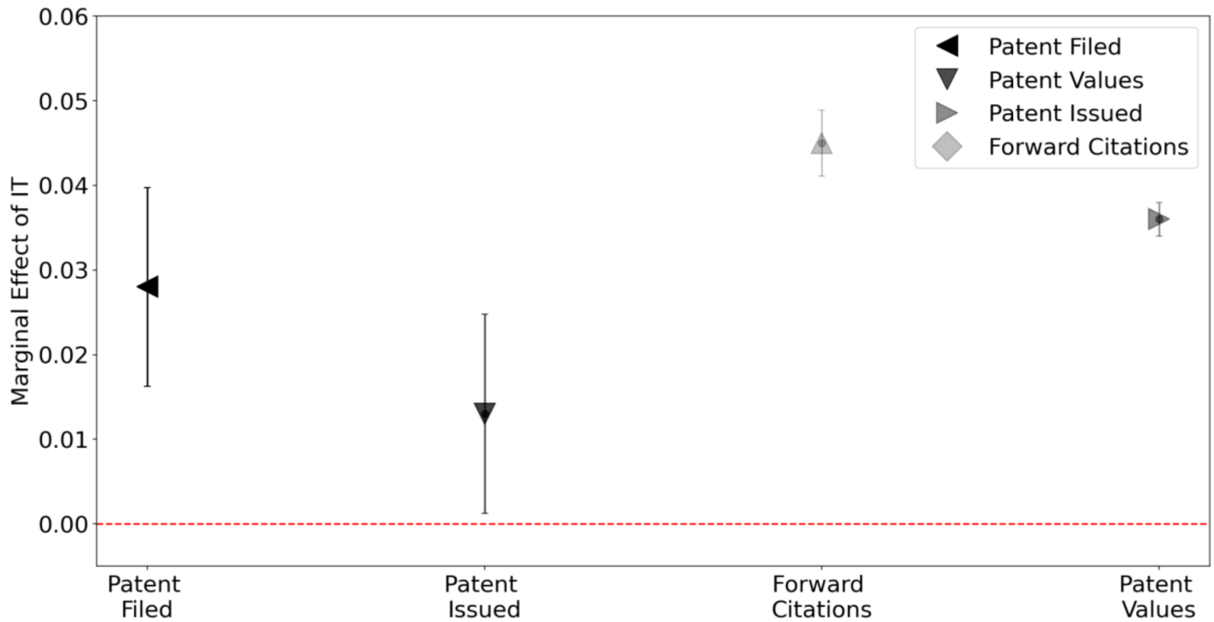

Figure 3 Marginal Effects of IT on Innovation Outcomes

Notes: Reported results are based on model specifications similar to column 3 Table 1. The control variables include all key inputs as well as logged software expenditure and R&D expenditure. The error bars indicate 95% confidence intervals.

Table 1 Productivity Effect of IT Human Capital (Baseline)

Models	**(1) Industry FX**	**(2) Firm FX**	**(3) Added Controls**	**(4) Alternative IT**	**(5) Translog**	**(6) Value Added**
LHS Variable			**Log Sales**			**Log Value Added**
Log Total IT Skills	0.030***	0.010***	0.009***		0.008***	0.014***
	(6.94)	(4.30)	(3.55)		(3.16)	(3.66)
Log Total IT Job Postings				0.009***		
				(3.15)		
Log Total Assets	0.153***	0.117***	0.093***	0.094***	0.087***	0.147***
	(12.39)	(7.43)	(6.36)	(6.37)	(6.53)	(6.41)
Log Total Employment	0.177***	0.397***	0.439***	0.440***	0.550***	0.742***
	(8.06)	(10.45)	(10.29)	(10.28)	(7.92)	(17.66)
Other Controls			Y	Y	Y	Y
Year FX	Y	Y	Y	Y	Y	Y
Industry FX	Y	N	N	N	N	N
Firm FX	N	Y	Y	Y	Y	Y
Number of Observations	24,646	24,451	18,985	18,985	19,433	18,487
adjR2	0.944	0.987	0.989	0.989	0.972	0.989

Notes: Log Total IT skills is the logged sum of the IT skills among all job postings. Log Total IT job postings is the logged total number of job postings in all IT-related occupations. Total assets is measured as the total Property, Plant, and Equipment. Columns 3-6 also control for lagged one-period cash holdings, dividend payout, and R&D expenditure over total assets. t statistics are reported in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 2 Productivity Effect of IT Human Capital (Identification)

Models	**(1) Olley-Pake**	**(2) Levin-Petrin**	**(3) ACF**	**(4) System GMM**
LHS Variable		**Log Sales**		
Log Total IT Skills	0.069***	0.060***	0.067***	0.009***
	(0.018)	(0.016)	(0.002)	(0.002)
Log Total Assets	0.111***	0.103***	0.120***	0.051***
	(0.010)	(0.012)	(0.012)	(0.013)
Log Total Employment	0.228***	0.217***	0.222***	0.370***
	(0.007)	(0.012)	(0.005)	(0.036)
Other Controls	Y	Y	Y	Y
Year FX	Y	Y	Y	Y
Firm FX	Y	Y	Y	Y
N	19,216	19,216	15,527	13,779

Notes: Column 1 uses the semiparametric method developed by Olley and Pakes (1996), which uses capital investment (both structure and equipment) as a proxy for unobservable shocks that could lead to spurious correlation between IT and productivity. Column 2 follows the approach in Levinsohn and Petrin (2003), using expenditure on intermediate inputs (cost of goods sold) as a proxy for unobservable productivity shocks. Column 3 employs the method developed by Ackerberg, Caves, and Frazer (2006) to further account for collinearity problems when estimating productivity using the Levinsohn-Petrin techniques. Column 4 employs the system GMM estimator following Arellano and Bond (1991) and Blundell and Bond (2000). This specification passes both over-identification and autocorrelation tests. * p<0.10, ** p<0.05, *** p<0.01.