## NBER WORKING PAPER SERIES

# ADVANCED TECHNOLOGIES ADOPTION AND USE BY U.S. FIRMS: EVIDENCE FROM THE ANNUAL BUSINESS SURVEY

Nikolas Zolas Zachary Kroff Erik Brynjolfsson Kristina McElheran David N. Beede Cathy Buffington Nathan Goldschlag Lucia Foster Emin Dinlersoz

Working Paper 28290 http://www.nber.org/papers/w28290

# NATIONAL BUREAU OF ECONOMIC RESEARCH 1050 Massachusetts Avenue Cambridge, MA 02138 December 2020

Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this data product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. (DRB Approval Number CBDRB-FY20-095, CBDRB-FY20-331 and CBDRB-FY21-041). We thank Scott Ohlmacher, John Eltinge, John Jankowski, Rob Seamans, John Haltiwanger, Frank Li, Susan Helper and Pascual Restrepo for excellent comments and feedback, as well as participants in the 2020 AEAStat session. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Nikolas Zolas, Zachary Kroff, Erik Brynjolfsson, Kristina McElheran, David N. Beede, Cathy Buffington, Nathan Goldschlag, Lucia Foster, and Emin Dinlersoz. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Advanced Technologies Adoption and Use by U.S. Firms: Evidence from the Annual Business Survey Nikolas Zolas, Zachary Kroff, Erik Brynjolfsson, Kristina McElheran, David N. Beede, Cathy Buffington, Nathan Goldschlag, Lucia Foster, and Emin Dinlersoz NBER Working Paper No. 28290 December 2020 JEL No. M15,O3,O47,O51

# ABSTRACT

We introduce a new survey module intended to complement and expand research on the causes and consequences of advanced technology adoption. The 2018 Annual Business Survey (ABS), conducted by the Census Bureau in partnership with the National Center for Science and Engineering Statistics (NCSES), provides comprehensive and timely information on the diffusion among U.S. firms of advanced technologies including artificial intelligence (AI), cloud computing, robotics, and the digitization of business information. The 2018 ABS is a large, nationally representative sample of over 850,000 firms covering all private, nonfarm sectors of the economy. We describe the motivation for and development of the technology module in the ABS, as well as provide a first look at technology adoption and use patterns across firms and sectors. We find that digitization is quite widespread, as is some use of cloud computing. In contrast, advanced technology adoption is rare and generally skewed towards larger and older firms. Adoption patterns are consistent with a hierarchy of increasing technological sophistication, in which most firms that adopt AI or other advanced business technologies also use the other, more widely diffused technologies. Finally, while few firms are at the technology frontier, they tend to be large so technology exposure of the average worker is significantly higher. This new data will be available to qualified researchers on approved projects in the Federal Statistical Research Data Center network.

Nikolas Zolas Center for Economic Studies U.S. Census Bureau 4600 Silver Hill Road Washington, DC 20233 nikolas.j.zolas@census.gov

Zachary Kroff Center for Economic Studies U.S. Census Bureau zachary.kroff@census.gov

Erik Brynjolfsson Stanford Digital Economy Laboratory 366 Galvez Street, Room 238 Stanford, CA 94305 and NBER erik.brynjolfsson@gmail.com

Kristina McElheran University of Toronto Scarborough 105 St. George Street Toronto, ON M5S 3E6 k.mcelheran@utoronto.ca David N. Beede Center for Enterprise Dissemination U.S. Census Bureau david.n.beede@census.gov

Cathy Buffington Special Projects and Research Economic Directorate U.S. Census Bureau Catherine.D.Buffington@census.gov

Nathan Goldschlag Center for Economic Studies U.S. Census Bureau Nathan.Goldschlag@census.gov

Lucia Foster Center for Economic Studies U.S. Census Bureau lucia.s.foster@census.gov

Emin Dinlersoz Center for Economic Studies U.S. Census Bureau emin.m.dinlersoz@census.gov

### 1 Introduction

Advances in technology are important drivers of productivity growth and living standards (see e.g. Solow, 1957; Stiroh, 2001). Thus, measuring advanced technology adoption is critical for understanding the current state of the U.S. economy and for planning for the future. However, owing to a lack of comprehensive data on firms' adoption and use of such technologies, we are "flying blind into what has been called the Fourth Industrial Revolution" (Brynjolfsson and Mitchell 2017b).<sup>2</sup> This data gap hinders evidence-based decision making at all levels of government and society. Datasets that provide detailed information on the diffusion of new technologies are rare, and those that are available often suffer from coarse aggregation (to industry classification levels), have response and sampling biases, fail to capture the non-manufacturing economy, and/or miss key emerging technologies. Consequently, we have limited knowledge about decisions to adopt these technologies or how adoption relates to firm characteristics such as industry, firm size, and firm age.

We describe the U.S. Census Bureau's recent efforts to fill this data gap by collecting information on adoption and use of several advanced technologies from a large, nationally representative sample of firms covering the private nonfarm sectors of the economy.<sup>3</sup> Our contributions are threefold. First, we provide a detailed first look at the adoption rates for several rising technologies that may shape the future, including artificial intelligence (AI), robotics and digitization. Second, our discussion concerning the challenges and opportunities faced in this collection provide context in analyzing the results and will be informative for other surveys that seek to measure rapidly evolving technologies. Third, we provide guidance for future researchers who seek to use this data in their own research through the Federal Statistical Research Data Center (FSRDC) network.<sup>4</sup>

Despite increasingly widespread discussion in the press of machine learning, robotics, automated vehicles, natural language processing, machine vision, voice recognition and other advanced technologies, we find that their adoption rates are relatively low. Furthermore,

<sup>&</sup>lt;sup>2</sup> The National Academy of Sciences (NAS) panel report in 2017 titled "Information Technology and the U.S. Workforce: Where Are We and Where Do We Go from Here?" makes a similar point about the paucity of data in this area and calls for a comprehensive and holistic approach to filling this data gap.

<sup>&</sup>lt;sup>3</sup> The ABS covers all nonfarm employer businesses filing the 941, 944, or 1120 tax forms (see <u>https://www.census.gov/programs-surveys/abs/about.html</u> for more details about the ABS).

<sup>&</sup>lt;sup>4</sup> Qualified researchers on approved projects can use these data and other Census Bureau micro data through the FSRDC network. See https://www.census.gov/fsrdc.

adoption is quite skewed, with heaviest concentration among a small subset of older and larger firms. We also find that technology adoption exhibits a hierarchical pattern, with the mostsophisticated technologies being present most often only when more-basic applications are as well. For instance, digitization of business information is very widely adopted. Adoption of cloud services displays an intermediate level of adoption, with a large share of firms electing to host at least one or more IT functions in the cloud. Intensive use of the cloud is slightly lower but still quite prevalent. Cloud users overwhelmingly also display a higher level of digitization. Likewise, firms reporting adoption of advanced business technologies also have predominantly adopted both digitization and cloud services, suggesting interdependencies between technology applications and potentially a cumulative progression of adoption.

While the adoption of key technologies is closely associated with a number of readily observed firm characteristics, it is worth noting that even after controlling for many of these characteristics, there is substantial heterogeneity in adoption that cannot be fully explained by available data. This suggests that our understanding of how and why firms adopt new technologies is still rather imprecise, providing further motivation for the collection and use of future survey data.

An extensive literature amassed over the last decade argues that technology adoption and use by firms significantly impacts labor markets and the economy overall.<sup>5</sup> Despite this growing literature and an accelerating pace of technological change, measurement of technology use at the firm-level has lagged considerably. This is particularly concerning in light of mounting evidence for increasing differences among firms in productivity, growth, and contributions to economic dynamism in recent years.<sup>6</sup> Scarcity of firm-level data has been cited as a central bottleneck in developing a better understanding of these technologies' impacts on workers, firms, and market dynamics (e.g., Seamans and Raj 2018).

In the absence of detailed, firm-level data, researchers have relied primarily on aggregated data or small-scale surveys focused on specific types of firms and technologies. For

<sup>&</sup>lt;sup>5</sup> Acemoglu and Autor (2011), Brynjolfsson et al. (2019), Brynjolfsson and Mitchell (2017a), Acemoglu and Restrepo (2019), and Agrawal, Gans, and Goldfarb (2019) all provide excellent reviews of current and future research issues pertaining to the diffusion of various advanced technologies.

<sup>&</sup>lt;sup>6</sup> While high cross-sectional heterogeneity in firm performance is long-established (e.g., Syverson 2004 & 2011; Hopenhayn 2014), recent studies point to increasing firm heterogeneity along a number of economically important dimensions (Andrews et al. 2015; Van Reenen 2018; Song et al. 2019; Decker et al. 2020; Autor et al. 2020; Bennett 2020a), some of which has been linked empirically to IT use (Barth et al. 2020; Lashkari et al. 2020).

instance, recent research has utilized nationwide or industry-level measures of robot diffusion (Acemoglu and Restrepo 2020, Graetz and Michaels 2017), information technology (Bessen 2002), or industry-level total factor productivity and patents (Autor and Salomons 2018) as proxies for automation. Yet, the relationship between labor and capital that is at the heart of recent contributions to the literature varies both across and within highly heterogeneous firms.<sup>7</sup> Thus, observing variation among firms in adoption and use of technology is critical for understanding the underlying mechanisms at work. Only with a higher-resolution lens will it be possible to accurately characterize broader technology adoption effects at higher levels of aggregation that reflect equilibrium outcomes of heterogeneous product, labor, and capital markets, not to mention differentiated organizational and production contexts within those markets.

We introduce the technology module from the Annual Business Survey (ABS) and present the first results from this module. This survey represents a partnership between the Census Bureau and the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF) and consolidates three surveys: the Survey of Business Owners (SBO), the Annual Survey of Entrepreneurs (ASE), and the Business R&D and Innovation Survey for Microbusinesses (BRDI-M) and adds a new and expanded collection on nationally representative business innovation statistics.<sup>8</sup> The ABS captures many important characteristics of the business context in which technology adoption and use take place. In addition to the technology questions that are the focus of this "first look" paper, it asks businesses about company information (such as the type of ownership), owner characteristics, innovation, research and development (R&D), intellectual property, and finance and other business characteristics. The current plan is to conduct the ABS for five years, with rotating modules about particular topics (such as technology or globalization) allowing for longitudinal analysis. The sample in the initial year of ABS was larger than is planned for subsequent years,

<sup>&</sup>lt;sup>7</sup> Worker earnings, for instance, vary significantly by both firm (Song et al. 2019) and establishment (Barth et al. 2016), with firm-level variation in information technology investment explaining important between- and withinfirm differences in wages and other labor market outcomes in recent years (Barth et al. 2020). Dixon et al. (2020) document striking firm-level heterogeneity in labor market outcomes associated with investment in robots. <sup>8</sup> The SBO and ASE were conducted by the Census Bureau, while the BRDI-M was conducted by the Census Bureau under a partnership with NCSES.

with approximately 850,000 employer businesses.<sup>9</sup> In subsequent years, the ABS will be mailed to a nationally representative sample of approximately 300,000 employer companies across all private nonfarm sectors in the United States.

The ABS was mailed in June 2018 with data collection taking place through the end of the calendar year (the primary reference period is calendar year 2017). The sample size and timeliness make it one of the largest and most up-to-date data set available on advanced technology adoption in the world. Response to the ABS is required by law, reducing selection bias, though certainly not eliminating it. In contrast, many privately funded surveys typically used in empirical work on technology adoption suffer from low response rates and significant selection bias, limiting the generalizability of many findings. Finally, the ABS also includes many small and young firms, which are often underrepresented in surveys of R&D and technology use. This is important because these firms are central to economic growth and dynamism (e.g., Haltiwanger et al. 2013 & 2016; Decker et al. 2014 & 2017), yet we have evidence that they respond differently to new advances in information technology in general (Hitt and Tambe 2016) and to new technologies such as cloud computing, in particular (Jin and McElheran 2017).

The economy-wide, large sample of firms allows for the tabulation and publication of novel data on nascent and growing technology use by U.S. businesses that smaller samples and/or imputation might not support. Also, the large sample size has the potential to identify at a highly disaggregated level (e.g., 6-digit NAICS) the industries and locations in which these nascent technologies are being adopted at a higher or lower rate.<sup>10</sup> Thus, the large sample size may help inform sampling strategies for future data collections and the development of survey instrument skip patterns for industries and firms for which these technologies are less relevant, thereby reducing respondent burden.

<sup>&</sup>lt;sup>9</sup> The reason for the large sample in the first year of the ABS is that it coincided with the quinquennial Economic Census and is intended to provide data that had previously been provided by the SBO, which was conducted in Economic Census years.

<sup>&</sup>lt;sup>10</sup> One aspect of the data that we have yet to properly analyze is the geography of technology adoption. A recent paper by Bloom et al. (2020) uses text analysis of earnings calls, newspaper announcements, patents and job postings to document the spatial rise of 20 new and important technologies. Our analysis gets complicated by the fact that the survey is enterprise-level and not establishment-level, so that it becomes difficult to pinpoint adoption location by firms with multiple establishments. However, it may be possible combine the enterprise-level findings with some more recent establishment-level surveys, like the Annual Survey of Manufactures (ASM) or Economic Censuses (EC) to specify adoption of certain technologies, like robotics, by location.

This first technology module contains three detailed questions: the availability of information in digital format (digitization), expenditure on cloud computing services, and use of several advanced business technologies. Taking these in turn, the first question explores firms' reliance on digital information, which is widely regarded as a necessary input to more-advanced uses of digital technologies (Brynjolfsson and McAfee 2014, Brynjolfsson and McElheran 2019). Brynjolfsson and McAfee (2014) argue that in order for firms to adopt artificial intelligence technologies, the necessary ingredients are massive amounts of digital information ("big data") and sufficient computing power. In broad terms, digital information is defined as "the representation of information in bits."<sup>11</sup> In addition to its importance as a prerequisite for AI adoption, digital representation of information is a key ingredient for more effective management of several business functions, such as electronic commerce, supply chain management, customer relations and marketing, and human resources. Recent research based on the U.S. Census Bureau's Management and Organizational Practices Survey (MOPS) has shown that decision making that relies on digital information ("data-driven decision making") has been rapidly diffusing in the U.S. and has important implications for firm performance (Brynjolfsson and McElheran 2016 and 2019).

The second question explores the extent to which firms rely on cloud computing, which has shifted the cost structure and use of IT by a broad range of firms (Armbrust et al. 2010, Brynjolfsson et al. 2010) and is widely viewed as a key enabler of digital transformation (Forrester 2017, Sunyaev 2020). Also, a second necessary ingredient for adopting AI technologies is sufficient computing power to handle and exploit massive quantities of digital information. Until the mid-2000's, much of this computing power has been beyond the reach of all but the largest, most technologically advanced firms. The advent of cloud computing services made scalable computing resources available on-demand, fundamentally changing the economics of IT use from an ownership model with high up-front fixed costs to an outsourced model with highly elastic variable costs (Armbrust et al. 2010).

Cloud services, therefore, may substitute for firms' fixed investments in their own physical data centers and owned IT resources, including software. Yet, while cloud services provide firms with the flexibility to scale up and down the volume of IT services they purchase, the solutions offered are less tailored to an individual firm's needs (Staten 2008, Schneier 2015).

<sup>&</sup>lt;sup>11</sup> Goldfarb and Tucker (2019) provide a useful review of the economics of digital technologies.

To the extent that cloud services enable outsourcing of major IT functions, smaller and younger firms have become able to access computing capabilities that are too costly to implement inhouse. Consistent with this mechanism, Jin and McElheran (2017) find that outsourced IT since the rise of cloud computing is significantly correlated with improved survival and productivity of young establishments.

Finally, the third question asks directly about the use of advanced "business technologies," including those typically categorized as "AI." These technologies include automated guided vehicles, machine learning, machine vision, natural language processing, and voice recognition software. Respondents are presented with a list that covers robotics (i.e., "automatically controlled, reprogrammable, and multipurpose machines"), various cognitive technologies (i.e., applications that help machines to "perceive, analyze, determine response and act appropriately in [their] environment", a standard definition of AI), radio frequency identification, touchscreens/kiosks for customer interface, automated storage and retrieval systems, and automated guided vehicles. This question is aimed at pinning down where the frontier of technology use actually lies and understanding interrelationships among different applications, while informing future surveys' sampling methodology and content.

Many of these frontier business technologies have been the focus of recent research. For instance, Acemoglu and Restrepo (2020), Graetz and Michaels (2018), Dinlersoz and Wolf (2018) and Dixon et al. (2020) find that the diffusion of robots has had important labor market and productivity consequences across regions and nations. Similarly, using data on the introduction of a machine translation system in a digital international trade platform, Brynjolfsson, Hui, and Liu (2019) find significant economic effects arising from automation and AI technologies. Their analysis indicates that machine translation has so far had economically significant impact on trade volume on the platform by reducing language-related frictions.

Using the data collected on AI, robotics, and digitization technologies, this paper provides, as a first look, some key statistics on their diffusion across firms and sectors. The diffusion rates are presented using tabulation weights constructed from the Longitudinal Business Database (LBD) to give estimates for the entire population of U.S. firms. In addition, estimates using both tabulation and employment weights are provided to offer a picture of the fraction of workers employed by firms using the technologies. This is critical for understanding the "future of work," given the increasing concentration of both employment and advanced

technology adoption in fewer, larger firms (e.g., Autor et al. 2020). The analysis documents the prevalence of the technologies across firm size and age categories, as well as the co-presence patterns for the technologies at the firm level. It also identifies which technologies are in the early stages of diffusion as indicated by the rates of testing versus the rates of actual use of technologies by firms. The firm-level connection between innovation and advanced technology presence is also explored to better understand how technology adoption may enhance the firm's ability to innovate, either indirectly, by allowing the firm to reallocate resources toward innovation, or directly by building upon existing technology.

Since the new data introduced here may provide a valuable resource to the research community, we also discuss possible interpretations of our results as they relate to a number of open questions about firm technology use and adoption. We highlight just a few of these potential areas of inquiry including: (a) dynamics and diffusion, (b) technology hierarchy and value chains, (c) technology as the "great equalizer" removing barriers for small and young firms, (d) technology complementarities, (e) technology and innovation, and (f) technology and the worker, specifically, the macroeconomic and distributional effects of new technology adoption, including how new technologies may substitute for or complement workers in various occupations. We do not attempt to draw definitive conclusions regarding these increasingly salient topics, nor do we address causal links between firm characteristics and technology adoption or between technology adoption and firm performance. Instead, our focus is to highlight preliminary findings and how ABS data might be used to advance the research frontier in these areas.

The remainder of the paper is organized as follows. Section 2 provides context for the ABS by reviewing related surveys and measures. Section 3 discusses the technology module and high-level findings. We then turn to providing detailed results from each of the three questions in Sections 4-6. In Section 7 we discuss our findings in light of open questions concerning technology and key economic outcomes. We provide concluding remarks in Section 8.

## 2 Review of Other Business Technology Surveys

We start by reviewing other business technology surveys. We restrict our attention to business surveys, nevertheless acknowledging that household surveys can provide important

complementary information.<sup>12</sup> We start by discussing Census Bureau surveys and then discuss other closely related private data collections that emphasize technology adoption.

#### A. Surveys Administered by the Census Bureau

The Census Bureau has collected data on advanced technologies over the past three decades through various surveys (see Table 1). Three relatively short-lived surveys focused on technology adoption and use are important to mention here, if only to discuss what was learned and how this module fits within that framework. The Survey of Manufacturing Technology (SMT) was an extensive survey on the adoption and use of advanced technologies in the manufacturing sector. The SMT collected information from establishments within selected manufacturing subsectors<sup>13</sup> about current and planned use of 17 technologies across 5 categories: design and engineering (e.g., computer-aided design); fabrication/machining and assembly (e.g., numerically controlled/computer numerically controlled machines); automated materials handling (e.g., automated guided vehicle systems); automated sensor-based inspection or testing; and communications and control (e.g., local area networks). While providing rich details on technology adoption, use, benefits, and costs at the plant level, the SMT was collected only for three years (1988, 1991, 1993) before it was discontinued. It remains, however, one of the most comprehensive surveys on advanced technologies, and in particular, on automation in the manufacturing sector (see Dunne 1991; Doms et al. 1995; Dunne and Schmitz 1995; Jensen and McGuckin 1996; McGuckin, et al. 1998; Dinlersoz and Wolf 2018).

The Computer Network Use Supplement (CNUS), which asked manufacturing plants about their e-commerce activities and e-business processes, was a supplement to the 1999 Annual Survey of Manufactures (ASM). The CNUS provided the most detailed insights to date

<sup>&</sup>lt;sup>12</sup> The Census Bureau through the American Community Survey (ACS) and periodic supplements to the Current Population Survey (CPS, jointly sponsored by the Bureau of Labor Statistics and the National Telecommunications and Information Administration) collects data on household technology (i.e., computer and Internet) adoption and use. The ACS and CPS also collect data on workers in technology-intensive occupations (e.g., computer programmers) and industries (e.g., computer systems design and related services). The Bureau of Labor Statistics collects data on workers by occupation, including technology-intensive occupations, in its Occupational Employment Survey. The Pew Research Center also collects household data on computer and internet use.
<sup>13</sup> Sampled establishments were from one of the following subsectors: fabricated metal products (SIC 34); industrial machinery and equipment (SIC 35); electronic and other electric equipment (SIC 36); transportation equipment (SIC 37); instruments and related products (SIC 38). These major industry groups accounted for about 43% of all employees and value added in this sector as reported in the 1987 Census of Manufactures.

on early applications of the commercial internet (see Atrostic and Nguyen 2005, 2006, 2007; Fichman and Melville 2014; McElheran 2015; and Fort 2017), including unprecedented insight into the supply chain relationships mediated by the new technology (Forman and McElheran 2019). However, it was also restricted to the manufacturing sector and was only conducted once. E-shipments data continues to be collected in the ASM (Angle and Forman 2018), but the original, more detailed survey was not repeated after its initial year.

Lastly, the Information and Communication Technology Survey (ICTS) was a supplement to the Annual Capital Expenditure Survey (ACES) from 2003 to 2013.<sup>14,15</sup> The ICTS asked for information regarding purchases and expenses for four types of ICT equipment and software: computer and peripheral equipment; ICT equipment excluding computer and peripheral equipment; electro-medical and electrotherapeutic apparatus; and computer software. In contrast to the SMT and CNUS, ICTS covered non-manufacturing as well as manufacturing firms. Of note, it separately measured firm-specific ("own-account") software investment by firms, a particular type of investment that has been linked to innovation (Bessen and Righi 2020) and impacts on workers (Barth et al. 2020a).

In addition to the Census surveys geared specifically toward measuring technology, there are many other Census collections with selected questions about technology. Most of these questions ask about firms' or establishments' software use or e-commerce activities. For example, the following contain at least one question about either software use or e-commerce activities: Annual Retail Trade Survey (ARTS), Annual Survey of Entrepreneurs (ASE), Annual Wholesale Trade Survey (AWTS), Business R&D and Innovation Survey (BRDIS), Census of Construction Industries (CCN), Census of Manufacturing (CMF), Service Annual Survey (SAS), and Survey of Business Owners (SBO). The Annual Survey of Manufactures (ASM) and the CMF provide some of the most detailed breakdowns of establishment-level investments in technology, separating out capitalized IT investment in computers and data processes equipment beginning in 2002 and splitting operating expenses dedicated to IT into expensed computer hardware and equipment, purchased software, and data and other purchased computer services (Jin and McElheran 2017).

<sup>&</sup>lt;sup>14</sup> The ACES is a nationally representative (covering manufacturing and non-manufacturing industries) annual survey of around 50,000 firms and, from 2002 onward, includes capitalized IT expenditure.

<sup>&</sup>lt;sup>15</sup> The ICTS for 2012 was suspended for budgetary reasons, was briefly reinstated for the year 2013, and then discontinued.

Other surveys ask about complementary subjects. The 2014 ASE module on R&D and Innovation asks about process innovations (including automation). The Management and Organizational Practices Survey (MOPS) asks how establishments use data to support decision making.<sup>16</sup> The 2016 Business R&D and Innovation Survey for Microbusinesses (BRDI-M)<sup>17</sup> (conducted in partnership with the National Center for Science and Engineering Statistics (NCSES) which was responsible for its predecessor Microbusiness Innovation Science and Technology (MIST) Survey) asks whether processes were improved by increasing automation. The BRDIS (newly renamed the Business Enterprise Research & Development Survey (BERD)) has targeted technology questions for a number of years, with special emphasis on R&D expenditures related to specific technologies, including biotechnology and nanotechnology (2008-2019) and artificial intelligence (2018-2019) (see Foster et al. 2020). Finally, some information on robotics and automation-related imports by firms can be obtained from the Longitudinal Firm Trade Transactions Database (LFTTD) which contains administrative data on trade transactions by U.S. firms.

As the above descriptions make clear, earlier Census Bureau data collections on technology were generally not repeated over time and did not necessarily focus on emerging technologies that might have the largest impact on business operations.<sup>18</sup> These data collections also did not directly ask about the consequences of technology adoption and use for worker outcomes at the business level (something that the ABS is scheduled to do in the 2019 version), though certain insights are possible by linking complementary data sets (e.g., Barth et al. 2020a). Furthermore, most technology data collection efforts have focused on manufacturing and not on the diffusion of emerging technologies in other sectors. For example, data was not systematically collected on the diffusion of radio-frequency identification (RFID) and barcodes in retail and wholesale trade and other services that now collectively account for a larger share of GDP than manufacturing and have implications for substitutions of different types of labor (in particular, customer labor) in the production function (Basker et al. 2017).

<sup>&</sup>lt;sup>16</sup> MOPS is a relatively newer ASM supplement that has been collected for years 2010 and 2015. See Brynjolfsson and McElheran (2016 & 2019). New questions on artificial intelligence are planned for 2021.

<sup>&</sup>lt;sup>17</sup> The BRDI-M is a follow up to the NCSES/Westat collected MIST survey conducted in 2014 that looks at the innovation activities for firms with fewer than 10 employees

<sup>&</sup>lt;sup>18</sup> Consensus is lacking on the technologies to focus on in a limited survey form and it is very difficult to predict in advance which emerging technologies will have the largest impact on business operations.

The Census Bureau has recently begun multiple efforts to further advance its measurement of technology. In addition to the module on the 2018 ABS, beginning in reference year 2018, the ASM began collecting expenditures on industrial robotics (Buffington, Miranda and Seamans 2019); in the same year the Annual Capital Expenditures Survey (ACES) began collecting expenditures on service and industrial robotics. The 2021 MOPS is currently being developed and will likely include some questions about the use of AI in decision making. The ASM and MOPS are establishment-based surveys; while the ACES and ABS are firm-based surveys. The content on the respective surveys attempt to take this difference in scope into account (for example, the MOPS asks about the locus of decision making) and we keep this difference in mind when interpreting results (for example, each firm is assigned to a sector but may encompass many disparate activities).

Lastly, the 2019 ABS includes a technology module with a focus on workers. It asks firms about the effects of technology adoption and use on their workers' numbers, types, and skills. These questions offer a unique opportunity to document firms' own assessment of how technology impacts their workforce. While workforce-related questions only have qualitative response categories, firms' responses can be compared with the quantitative responses provided by the same firms in other survey and administrative data. This last point highlights an important strength of Census data: it is possible to link data from multiple surveys using standardized record identifiers. Finally, in order to address the shortcoming of a lack of a longitudinal component, which is especially important for understanding the diffusion of technology, the ABS technology modules are scheduled to be repeated over three-year cycles.<sup>19</sup> While the sample of respondents will differ over this time period, Census expects there to be considerable overlap, thereby providing a glimpse for how technology adoption changes over time for a select group of firms.

#### B. Other Surveys

We highlight a few alternative, privately-run surveys in this section (see Table 2). Helper (1995) uses a survey of 499 automotive suppliers regarding their use of computer numeric

<sup>&</sup>lt;sup>19</sup> That is, the first technology module from 2018 will be repeated in 2021 and the second technology module from 2019 will be repeated in 2022.

control (CNC) machines and the applicability of this technology in their typical production<sup>20</sup> to show that arm's-length supplier/customer relationships inhibit the adoption of CNC technology. Aral et al. (2012) use survey data collected from customers of a major enterprise resource planning system to look at complementarities between IT, HR analytics and performance pay. The global marketing services company Harte Hanks administered an international survey to make up the Computer Intelligence Technology Database (CITDB) from 1996 until 2015. This survey samples more than 150,000 establishments of firms across 20 European countries and the United States. The CITDB contains information on IT adoption in areas such as PCs/laptops, servers, network infrastructure, IT employees, software and hardware, and (more recently) some cloud computing.<sup>21</sup>

Some recent surveys focus on various automation technologies. The Georgia Tech Survey of Advanced Technology and Robotics in U.S. Manufacturing was conducted in 2018 by Nancey Green Leigh and others of the Georgia Institute of Technology (Green Leigh et al 2020). This survey "was conducted to better understand U.S. manufacturers' use of robotics and automation technology and to generate real knowledge about their impacts on employment and manufacturing competitiveness."<sup>22</sup> Green Leigh surveyed 428 U.S. manufacturing firms regarding their use of rapid prototyping, additive manufacturing, computer-aided design and manufacturing (CAD/CAM), machine vision and real-time monitoring, advanced materials, CNC machines, and robots.

Finally, Helper, Seamans, Reichensperger, and Bessen collected responses to the National Survey of Auto Suppliers for 2018. The National Survey of Auto Suppliers includes a plant, human resources, and sales survey form administered to firms in "any tier of the supply chain for new cars or light trucks." This survey asks auto suppliers about their use of various automation technologies and how automation has impacted their employment, robots' effects on certain performance outcomes, and their gathering and analysis of operations data.<sup>23</sup>

<sup>&</sup>lt;sup>20</sup> Helper excluded 213 of these respondents because the technology was either unknown to them or reportedly not applicable to their business. Details about this survey can be found in Helper (1995).

<sup>&</sup>lt;sup>21</sup> For more information about the CITDB, see Bloom et al. (2014), McElheran (2014), Bloom et al. (2016), and Haug et al. (2016). Other notable works using CITDB include Bresnahan et al. (2002), Forman et al. (2002), Brynjolfsson and Hitt (2003), Forman et al. (2009), Mahr (2010), Bloom et al. (2012), Forman et al. (2012), and Kretschmer et al. (2012).

<sup>&</sup>lt;sup>22</sup> Details about this survey can be found at <u>https://planning.gatech.edu/gatech\_survey\_mfg\_tech/</u>.

<sup>&</sup>lt;sup>23</sup> For more information on this survey, visit <u>http://sites.bu.edu/tpri/auto-survey/</u>.

Researchers have also relied on data collections by trade associations. For instance, country- and industry-level data on robot installations are published by the International Federation of Robotics. UN Comtrade provides data on robot imports, and country-level numbers of robotics patents filed are available from the U.S. Patent and Trademark Office (USPTO). The European Commission (EC) also released an enterprise-level survey that found that 42% of enterprises adopt at least one AI technology (European Commission 2020). In addition, some consulting firms have also collected data on technology adoption and use. For example, Deloitte collected information from about 1,100 US-based companies representing 10 industries in 2018 and published adoption rates for advanced technologies in *State of AI in the Enterprise*<sup>24</sup>, while both McKinsey (McKinsey 2018) and PwC (PwC 2019) surveyed a number of executives on their expected adoption of AI, with 20-30% of executives planning to adopt some form of AI.

A growing trend relies on machine learning and related techniques to measure the spread and impact of new digital technologies by analyzing large data sets of job postings, employment profiles and similar data on digital platforms. For instance, digitized information on workers and job postings have been used to track the supply and demand of IT labor as a proxy for IT diffusion in the economy (e.g., Tambe and Hitt 2012, 2020; Webb 2019; Bana et al, 2020; Tambe et al. 2020; Atalay et al. 2020). Bennett (2020b) uses machine learning algorithms to match the Port Import/Export Reporting Services (PIERS) data on sales of automation equipment to the National Establishment Time Series (NETS) data, which is a panel-data version of Dun and Bradstreet publications. In general, these large data sets improve on the power and coverage of readily-available data sets such as COMPUSTAT, yet they still lack the representativeness and accurate representation of dynamic patterns (particularly for young and small establishments) available in Census data (Crane and Decker 2020).

Data collection efforts to measure the diffusion of new technologies exist outside of the U.S. and provide useful complementary views on this global phenomenon. For example, the European Commission (EC) released an enterprise-level survey that found that 42% of enterprises adopt at least one AI technology (European Commission 2020). Statistics Canada

<sup>&</sup>lt;sup>24</sup> For more information see the report at <u>https://www2.deloitte.com/content/dam/insights/us/articles/4780\_State-of-</u><u>AI-in-the-enterprise/DI\_State-of-AI-in-the-enterprise-2nd-ed.pdf</u>.

collects firm-level data on a range of automation technologies, including robots, as well. (Dixon et al. 2019).

#### **3** Technology Module and Overall Results

With this context serving as background, we now turn to how we developed the technology module for the ABS. As the ABS was being developed, a team of researchers within and outside of Census<sup>25</sup> had the opportunity to propose a limited set of questions as a module. Three main criteria were considered in the development process: appropriateness, consistency, and optimality. Generally, content must be appropriate with regard to the Census Bureau's mission and its role in the larger Federal Statistical System, consistent with the instrument on which it would appear (in terms of goals of the instrument and its format), and optimal in terms of weighing the benefits (i.e., filling existing gaps in our knowledge) against the costs (i.e., respondent burden) of the additional data collection.

Three questions were developed for inclusion on the 2018 ABS. As is standard practice with the Census Bureau collections, these questions were subject to cognitive testing. The decision to add the technology module to the ABS came relatively late in the survey cycle, which meant weighing the benefit of quickly responding to an emerging data need –and the opportunity that the large sample presented –against the drawbacks of only conducting one round of cognitive testing.<sup>26</sup> One set of interviews took place in August 2017 during the second round of ABS cognitive testing, and a second set took place during the same month as part of a debriefing of high tech companies who participated in the Business R&D and Innovation Survey (BRDIS). A detailed description of the results and recommendations from cognitive testing are in the Appendix.

The final resulting questions are shown in Text Box 1. The first question asks about the intensity of digitization of six critical business activities (personnel, financial, customer feedback, marketing, supply chain, and production) and allows for a write-in for "other"

<sup>&</sup>lt;sup>25</sup> Catherine Buffington, Lucia Foster, and Scott Ohlmacher from Census, along with Erik Brynjolfsson and Kristina McElheran.

<sup>&</sup>lt;sup>26</sup> Census typically requires two rounds of cognitive testing for new survey content. The first round is exploratory and is used to identify problems with the content including cognitive difficulty or excessive burden as well as whether the respondent has the information needed to answer the question. The second round is confirmatory and is used to confirm that changes made to the content based on findings and recommendations from the first round do in fact correct the problems uncovered in testing.

activities. The second question asks about the intensity of cloud services purchases for eight business functions and allows for a write-in for "other" functions.<sup>27</sup> These two questions have check boxes for four percentage ranges (None, Up to 50%, More than 50%, and All) along with an option for "Don't know" and one indicating that a particular type of information is not collected by the business, or that the particular IT function is not used. This final response option is useful for distinguishing firms that are not at risk of adopting a technology from those that might be potential adopters at some point in the future, but which have not adopted as of the survey year.<sup>28</sup> The third question concerns the testing and intensity of use of nine advanced business technologies (including augmented reality, machine learning, and robotics) in the production of goods or services. For each of the nine technologies, there are checkboxes for "Testing," four intensity percentage ranges (No use, Less than 5%, 5%-25%, and More than 25%), and "Don't know" options.

The ABS was collected from June through December 2018. The response rate for the portion of the survey used in the paper was 68.7%, slightly lower than the usual response rate for Census Bureau surveys (see Table 3). However, as shown in Table 4, when weighted by LBD-derived tabulation weights, the size and age distributions of responders align closely with the national population of firms.<sup>29</sup> The firms included in the sample for the 2018 ABS had a mean employment of about 89 (or 26 by LBD tabulation weight), and a mean age of 16 years (the same as with LBD tabulation weights).<sup>30</sup> About 67% of the firms sampled had fewer than 10 employees (75% by LBD tabulation weight), and 3% had 250 or more (1% by LBD tabulation weight). The oldest firms (21+ years) represented about 33% of the sample (31% by LBD tabulation weight). These distributions line up well with the national size and age distributions for firms. The LBD's 19 two-digit NAICS sectors were aggregated to form 13 sectors for the purposes of the subsequent analysis. The largest shares of firms in the sample fall into Professional Services,

<sup>&</sup>lt;sup>27</sup> Cloud services are "[information technology [IT] services provided by a third party that [a] business accesses ondemand via the internet."

<sup>&</sup>lt;sup>28</sup> When we consider overall adoption rates of a technology type (e.g. cloud computing), we only consider firms that are "at risk" of adopting, meaning that as long as they respond that they are "at risk" for at least one of the technologies listed within a category, then we include them in our denominator.

<sup>&</sup>lt;sup>29</sup> The firm size, age and industry composition of the non-responses also closely aligns with the size, age and industry distribution of responders.

<sup>&</sup>lt;sup>30</sup> Firm size, age, and industry are derived from the Longitudinal Business Database (LBD).

Retail Trade, and Other (Arts, Food, Other Services) sectors. In the remainder of this paper, we rely primarily on the LBD-derived tabulation weights when describing response rates as well as extensive and intensive technology adoption rates.<sup>31</sup>

#### A. Item Non-Response

One challenge in analyzing the results of the survey is the item non-response rate and the share of "Don't Know" responses to the technology module questions. The item response rate to the technology module does not differ dramatically from other parts of the survey (nearly 95%). The cumulative responses are displayed in Figure 1. For item non-response, approximately 45,000 firms skipped the technology module, while answering other parts of the survey. For these firms, the responses to each technology question are labeled as "Missing". For the item responses, between 40,000 and 105,000 respond with "Don't Know", depending on the technology and technology function (for instance, Digital Information had fewer firms responding "Don't Know" than Cloud Computing or Business Technologies). Firms who skipped the technology module and firms who responded with "Don't Know" are skewed toward larger and slightly older firms when compared to the overall distribution of survey respondents.<sup>32</sup> This means that while the "Don't Know" responses may have relatively little effect on either observed or imputed firm digitization rates, they may have a significant effect on one of our key findings which estimates the association between adoption and firm size. Compared to digitization, cloud services had a larger number of firms responding "Don't Know" to all question categories (i.e., all categories of IT functions) and their size distribution is even more skewed toward larger firms. Business technologies had the most firms respond "Don't Know" to all question categories (roughly three times as many as for digitization). However, the size distribution of the firms that responded "Don't Know" to the business technology questions is very similar to that of the overall sample of survey respondents. While the overall levels of business technology adoption may —of the three technology questions—be most affected by the

<sup>&</sup>lt;sup>31</sup> Note that for this reason, tables and statistics in this paper may not line up with published Census Bureau tabulations of ABS data (which rely on the survey weights used for sampling).

<sup>&</sup>lt;sup>32</sup> The reason why this particular set of questions receives higher than usual "Don't Know" responses may be due to the persons filling out the survey (usually financial analysts), who are unlikely to have reliable measures of intensive (or extensive) use for the various technologies, as this is not typically a line-item found in financial statements. Note also that Census was unable to perform a follow-up to the item non-response for the technology module in the 2018 ABS. However, in the following year of the ABS (2019 ABS), we find that the number of "Don't Know" responses declined.

rate of "Don't Know" responses, we expect the association between size variables of business technology use to be the least affected by "Don't Know" responses.

The volume of "Missing" responses is another sample characteristic requiring consideration. "Missing" responses are distinct from the "Don't Know" responses in that firms leaving the technology questions blank generally do so consistently across each of the digitization, cloud services, and business technology questions. These firms are much more likely to be large. The share of firms with 250 or more employees skipping this module was more than two times the corresponding share in the overall sample. For this reason, "Missing" responses may have a significant effect on our estimates of how size predicts adoption of digitization, cloud services, and business technologies.<sup>33</sup>

As we highlight below, our preliminary analysis reveals that both size and age appear to be highly associated with various forms of technology adoption. Thus, excluding the "Don't Know" or "Missing" responses from the sample may bias our findings and paint an inaccurate picture of the types of firms that adopt frontier technologies. At the same time, we want to avoid imposing strong assumptions on subsamples for which we have relatively little information. This is particularly important in the context of firm technology adoption, which is known to be skewed (e.g., Tambe and Hitt 2012), highly idiosyncratic due to organization-specific complementary investments (e.g., Brynjolfsson and Saunders 2016; Brynjolfsson and McElheran 2019; Brynjolfsson, Jin, and McElheran 2020) and is also implied in well-known discussions of why productivity and performance among firms varies so widely (e.g., Syverson 2011, Song et al. 2019).

## B. Item Non-Response Imputation

To address the issue of item non-response, we take two approaches. The first is to report findings using only the raw data of certainty responses (i.e., excluding all "Missing" and "Don't Know" responses). The second is to report findings after imputing values for "Don't Know"

<sup>&</sup>lt;sup>33</sup> The public use tables reporting technology adoption levels (found here:

<sup>&</sup>lt;u>https://www.census.gov/data/tables/2018/econ/abs/2018-abs-digital-technology-module.html</u>) do not impute for item non-response, nor do they correct for ABS sample weights as we do in this paper. If using the public-use tables to report aggregate adoption *rates*, those rates may be considered a lower bound.

responses and for "Missing" items across the usage categories based on identifiable firm characteristics.<sup>34</sup>

Note first that we leverage information from responses indicating that a particular information type or IT function is not applicable to or used by the firm. We consider these firms to not be at risk of adopting digitization or cloud services in these settings (e.g., a firm without supply chain information cannot adopt digitization of supply chain data). Thus, when reporting adoption rates based on observed data, we exclude them from the denominator.<sup>35</sup> In our imputations, we include this as one of the possible categories and assign firms probabilistically to this not-at-risk category. This preserves a consistent ratio of adopters to potential adopters across imputed and non-imputed statistics.

It is important to note two things about these ratios. The first is that they are designed to shed light on the diffusion of technologies within the set of firms *that might potentially adopt them.* This is distinct from calculating the ratio of adopters to all firms in the economy, which is informative about the presence of a technology to date but might understate meaningful technology use by including firms that are not candidates for adoption in the denominator.<sup>36</sup> This is a particular concern for the ABS, which has unusually comprehensive coverage of the entire population of U.S. firms – as opposed to a smaller survey that may only go to firms in a certain industry or of a certain size, which automatically (often implicitly) constrains the risk set. We report unconditional adoption rates in Appendix A.3.

For imputed statistics, the imputed response for each firm is determined by performing an ordered probit estimation using each of the usage categories as the dependent variable and firm size (broken into 12 size bins, similar to the size bins in the legacy<sup>37</sup> Business Dynamic Statistics (BDS)) interacted with the firm's primary industry (2-digit NAICS based on the largest share of

<sup>&</sup>lt;sup>34</sup> For the digitization and cloud services questions, the usage categories consist of "None", "Less than 50%", "More than 50%" and "All". For the business technologies questions, the usage categories consist of "None", "Testing", "Up to 5%", "Between 5% and 25%" and "More than 25%".

<sup>&</sup>lt;sup>35</sup> Note, however, that in Appendix A.3 we do also report adoption rates unconditional on being at risk (i.e., including not-at-risk firms in the denominator).

<sup>&</sup>lt;sup>36</sup> This is similar to calculating the unemployment rate, which only considers the number of participants in the labor force and not the total adult population.

<sup>&</sup>lt;sup>37</sup> In 2018, the BDS underwent a redesign which updated the size bins used in reporting. For more information, see <u>https://www.census.gov/programs-surveys/bds/news-updates/2018-bds-release.html</u>.

the firm's payroll in the LBD), average payroll per employee, age (broken out into 8 size categories), and state indicators, as the explanatory variables. For cloud services, we also use the responses for digitization as an explanatory variable; for business technologies, we further use the responses for both digitization and cloud services.<sup>38</sup> The ordered probit provides a set of predicted probabilities for each of the informative usage categories for all three technology questions. These corresponding probabilities are then combined with the LBD tabulation weights and tabulated. Note that values are kept as-is for firms that responded to one of the usage categories as either "None" or some usage. It is only firms that left the item blank or responded with "Don't Know" for which probabilities are tabulated in the analysis below.<sup>39</sup> Details of the imputation process, along with additional background information and post-imputation checks can be found in the Appendix section A.2.

Lastly, we describe our process for creating LBD tabulation weights using the 2017 LBD. The survey weights and thus subsequent tabulation weights assigned to firms in the ABS result from a complex survey frame and sampling design used to produce nationally representative statistics by business owner characteristics including race, sex, ethnicity, and veteran status.<sup>40</sup> The resulting sample and subsequent set of responding firms do not produce nationally representative weighted distributions of certain firm characteristics when compared to the LBD, such as size, age or industry distribution. We compare the distributions of weighted firm counts (see Table 4) and employment across NAICS sectors to their respective distributions in the 2017

<sup>&</sup>lt;sup>38</sup> We first group all of the response options into the usage categories, counting the "Not Collected" option from the Digitization question as "None" and the "Do not use IT function" from the Cloud Services question into "None". As we demonstrate later in the paper, the technology categories appear hierarchical, implying that firms will be more (less) likely to purchase (not purchase) cloud services if they have (do not have) information in a digital format and so forth.

<sup>&</sup>lt;sup>39</sup> In subsequent validation exercises, we calculated the firm-level precision (share of imputed responses that are correct) and recall (share of actual responses that we imputed correctly) for the full sample by randomly splitting the data into a 70/30 split. For 70% of the data, we estimate our imputation equation ("training set") and generate probabilities on the remaining the 30% ("test set"), providing us with a set of validation statistics in each of the response categories across all technologies and sub-technologies at the firm-level. Our validation across all of the response categories found somewhat low precision and recall rates for each of the different "use" categories in digitization and cloud, but performs better when the use categories are limited to "use" and "not use", indicating that the firm controls used in the imputation have a hard time distinguishing between intensity. The validation performs better for the "business technologies", where the majority of responses fall into the "no use" category. When tabulating the imputation probabilities in our subsequent analysis and tables, we rely primarily on the extensive margin (use/no use) measures.

<sup>&</sup>lt;sup>40</sup> For more information, see <u>https://www.census.gov/programs-surveys/abs/technical-</u> <u>documentation/methodology.html</u>.

LBD.<sup>41</sup> We construct tabulation weights based on the universe of firms in the 2017 LBD to adjust for these deviations from national estimates. Our weights are calculated by stratifying the firms in the 2017 LBD and our final sample of firms in the ABS on firm size, age, and industry. These strata are defined by the 19 two-digit NAICS sectors and the 12 firm size and 12 firm age groups used in the BDS. All firms in a given stratum are assigned the same weight, which is calculated by dividing the number of firms in that stratum from the 2017 LBD by the number of firms in that stratum from the ABS sample.<sup>42</sup> With these new tabulation weights, we match much more closely the firm and employment distributions in the 2017 LBD.

We now turn to detailed analysis of the three technology questions in the next three sections of the paper. Given the abundance of results, we have organized each of the sections in the same manner. Each section has the same set of subsections: summary of responses (subsection A), adoption and use rates (subsection B), industry breakdown (subsection C), relation to firm size and age (subsection D), and summary remarks.

## 4. Digital Share of Information by Business Activity

The first question on the 2018 ABS technology module queried firms on the type of information stored digitally.

#### A. Summary of Responses

Figure 1 contains the frequencies of responses for the digital share of information by business activity or function. Digitization is widely adopted across the majority of firms, with most firms electing to store their personnel and financial information digitally. Other types of information, such as supply chain and production information, were less likely to be stored digitally mainly due to firms not collecting those types of information in any format. It is worth emphasizing that, had we not restricted attention to whether the information existed at the firm in any format, we would dramatically underestimate the move from "analog" to digital representations of information by U.S. firms. An open – but distinct – question is whether firms

<sup>&</sup>lt;sup>41</sup> The LBD is the underlying data used in creating the Business Dynamics Statistics (BDS), which reports economywide statistics on firms and establishments.

<sup>&</sup>lt;sup>42</sup> To account for unusually large weights caused by too few ABS firms in a stratum, we also winsorize the LBD tabulation weights at the 99.99th percentile.

should collect additional types of information and/or whether the availability of digital technologies expands or shrinks the scope of their activities.<sup>43</sup> We take the boundary of firm activities as given for our core estimates.

For firms that did respond in the affirmative for each type of information, the most frequent responses are "more than 50%" or "all", followed by "up to 50%." There are a large number of firms indicating full digital information use for personnel and financial business activities. These categories represent the highest overall use of digital information, followed by customer feedback and marketing. Table 5 lists the three most common information types digitized by sector. In all sectors, financial and personnel information are the most likely to be digitized.

## B. Adoption and Use Rates

Table 6 contains the adoption and intensive use shares for digital information by business function. A firm is considered an adopter if its response indicates at least some use of digital information in a business function. According to our LBD tabulation weights, 90.2%<sup>44</sup> (75.6% non-imputed) of firms that collect at least one type of information stored at least one type of information in digital format. Consistent with the frequencies in Table 5, the highest rates of adoption are in financial, personnel, and marketing activities. The relatively high rates of adoption in these categories are not surprising, as most firms rely on basic financial and personnel functions, regardless of sector. On the other hand, the lowest rates of adoption are observed in production and supply chain activities, in part driven by the fact that these activities are more concentrated in the manufacturing sector, which only makes up 4% of national firm counts. Also shown in Table 6, the incidence of intensive usage (more than 50% or all) parallels the basic adoption rates by function. The most intensity. In contrast, supply chain function has the lowest incidence of intensive use at about 32.6%.

<sup>&</sup>lt;sup>43</sup> While conventional wisdom suggests that digital technologies lower the costs of collecting information, thereby expanding the scope of firm information collection and processing, work on early waves of technology adoption has suggested a countervailing mechanism whereby the firm boundary itself contracts (Brynjolfsson et al. 1994; Forman and McElheran 2019).

<sup>&</sup>lt;sup>44</sup> Across business functions – not in table

Table 7 highlights the five business function pairs that have the highest co-presence (correlation) with adoption of digital information. The correlations indicate the extent of potential complementarities between the information types. The highest correlation (0.70) occurs in the financial and personnel pair, which is consistent with the fact they have the top two adoption rates in Table 6. The next pair is marketing and customer feedback, which have a correlation of 0.69. The high correlation for this pair is sensible if firms that use digital information in their marketing activities also tend to have digitized customer feedback platforms. The relatively high correlation for supply chain and production is also intuitive. Production and supply chain activities complement one another and the reliance on digital information in production pairs well with digitization of information in the supply chain.

## C. Industry Breakdown

Figure 2 is a butterfly chart of adoption and use rates for digital information by sector. The right panel of the chart represents, by sector, the adoption rates of digital information across all surveyed information types. The adoption rate for "at-risk-of-adoption" firms is highest (96.5%) in the Information sector, followed by Professional Services and Education. The sectors with the lowest adoption rates (85%-88%) are Transportation & Warehousing, Retail Trade, and Other (Arts, Food, Other). The segments within each bar in the chart capture adoption rates by the number of information types in digital format. The leftmost segment in the right panel indicates the share of firms that have digitized at least three types of information, the next segment adds firms with exactly two information types digitized, and the final segment represents with only one type of information digitized. In all sectors, a large share of adopters report having three or more types of information digitized.

The left panel of Figure 2 represents intense use of digitization (defined as "50% or more" or "all") by sector. Once again, Information leads with more than 91% of "at-risk" firms indicating intensive use. In general, the ranking of sectors by adoption and intensity of use rates parallel each other. Similar to the extensive margin measures, most firms report digitizing at least two types of information, regardless of sector. In fact, the fraction of firms digitizing only one type of information intensively is relatively small in each sector. Overall, digitization appears to be highly prevalent across sectors.

#### D. Relation to Firm Size and Age

Table 8 shows digitalization rates by firm size and age. Specifically, the table reports the estimated coefficients<sup>45</sup> from a linear probability model (LPM) where the dependent variable is an indicator of whether a firm uses at least one type of digital information, and the independent variables are pairwise interactions of four size categories and four age categories (16 size-age cells in total). <sup>46</sup> The estimated coefficients for size-age cells indicate that for a given size level, the presence of digital information slightly increases with firm age, with the exception of the smallest size category, where the use indicator actually declines with age. This observation is consistent with research that old, small firms are a relatively low-growth and less-advanced segment of the firm population, and thus may be less reliant on digital information in business functions. This finding is supported in Jin and McElheran (2017) and is consistent with prior work on low- versus high-growth entrepreneurship (Guzman and Stern 2020). In general, the adoption rate is monotonic in size, with the largest firms having consistently higher rates of adoption relative to smaller firms.

The general pattern with age is more nuanced. For smaller firms (1-9 employees), age is negatively associated with adoption rates, with a smaller proportion of old firms adapting their information digitally than young firms. However, as the size increases, this pattern changes, with age positively associated with increased adoption of digitization.

To summarize, the vast majority of firms (>90%) that collect at least one form of information store at least one form of information digitally. Both financial and personnel information are the most likely sources of information to be digitized, with both information types intensively digitized. Manufacturing, Information and Professional Services are among the highest adopters of digitization, with size being a primary correlate of adoption. The next section looks at cloud service purchases.

<sup>&</sup>lt;sup>45</sup> All coefficients are statistically significant at a 0.1% level of significance.

<sup>&</sup>lt;sup>46</sup>The results from a Probit specification yield essentially identical estimates for the digital information technology, and for most of the other technologies discussed below. Hence, we uniformly use an LPM for all specifications that follow later in the paper. For further information on limitations and potential biases introduced from estimating a probability model using ordinary least squares (OLS) see Horrace and Oaxaca (2006).

#### **5. Cloud Service Purchases by IT Function**

This section describes the adoption patterns for cloud service purchases across size, age and sector. We see similar but lower rates of adoption across the categories of cloud services, with firms electing to host multiple IT functions in the cloud rather than concentrating cloud purchases within a single IT function.

### A. Summary of Responses

Using our LBD-derived weights, the majority of firms (78.4%; 54.1% non-imputed) that perform at least one IT function also purchase at least one cloud service. These purchases vary across several functions, with little concentration in a specific function, as found in Figure 3. A large number of firms indicated that they did not know whether they had any cloud purchases. For many IT functions, among those reporting cloud use, the frequency of responses was typically highest for the "up to 50%" category followed by "more than 50%" and "all". Exceptions to this are in Security or Firewall and Billing and Account functions and to some extent in Servers. For these functions, more adopters of cloud reported in the "all" category compared to "more than 50%." The Data Analysis function has the lowest number of firms reporting some cloud purchase, whereas Billing and Account Management shas the highest number of firms, closely followed by Security or Firewall and Collaboration and Synchronization functions.

## B. Adoption and Use Rates

Table 9 contains the extensive margin of adoption and intensive use shares in terms of expenditure for purchased cloud services by business function. A firm is considered an adopter if its response indicates at least some expenditure on cloud services in a business function. First, when all functions are considered, both the adoption rates of cloud services and their intensive-use rates tend to be generally lower than the adoption of digital information reported in Table 6. The highest adoption and intensive-use rates are observed for Billing and for Security, whereas the lowest rates are in Customer Relations and in Data Analysis. The rates in Table 9 indicate that purchased cloud services are more prevalent in relatively standard IT functions and adoption and intensity are lower in more-specialized functions. Overall, about a quarter of the firms have indicated adoption of purchased cloud services in All IT functions.

One notable pattern for cloud services purchased is the lack of a single mass concentration for a particular service. This suggests that firms are utilizing the cloud for a variety of business processes and perhaps speaks to the flexibility and/or diversity of cloud services available by the time of the survey. The ranking of the intensive-use rates (defined as "50% or more" and "all") is similar to the ranking of the adoption rates with a few exceptions. Intensive usage rates in Data Storage and Servers services are higher than Synchronization and All IT cases. Data Analysis is lowest ranked in terms of intensive use as with its ranking in the adoption rates.

Table 10 shows the top five business function pairs that have the highest co-presence (correlation) in terms of adoption of purchased cloud services. As in Table 7, the correlations indicate the extent to which the two functions in a pair tend to rely or not on cloud service purchases. The highest correlation (0.73) occurs in the Servers and Security pair which may be driven by the fact that Security, which has the highest rate of adoption for cloud services, could go hand-in-hand with a need to protect servers. The relatively high correlations between All IT functions and Security, Servers, and Storage may not be too surprising, as these functions are often bundled together by cloud-hosting providers. The third-highest correlation is for the pair Servers and Storage (0.67), which is consistent with firms unsurprisingly needing storage services to harness digital information created through online transactions on those servers.

## C. Industry Breakdown

Like Figure 2, Figure 4 provides the butterfly chart for adoption and intensive-use rates across sectors in the case of purchased cloud services. Like digitization, the highest adoption and intensive-use rates are in Information, followed closely by Professional Services and Education. The lowest rates are in Agriculture, Mining, Utilities, Retail Trade, and Transportation and Warehousing, in addition to the Other category. These ranking are consistent with a prior that IT-intensive sectors such as Information, Professional Services, Education and Health Care may have more reliance on cloud services. Figure 4 also reveals that cloud services purchases have much lower diffusion rates compared to those for digital information in any given sector. For instance, in the Information sector, the digital information diffusion rate based on at least some use in a business activity is about 97%, compared to the diffusion rate of purchased cloud services, which is about 89%.

As in the case of digital information use, for all sectors there is a large fraction of firms relying on cloud services for three or more IT functions, indicating that conditional on using some cloud services firms tend to use those services for many IT functions, regardless of sector. Table 11 shows that Billing and Security are the most common IT functions for most sectors, with certain sectors predominantly relying on the cloud to perform collaborative or synchronized tasks. Roughly one-third the sectors list "All IT functions" as the third most common use in the cloud. This seems to be suggestive that digitization and cloud usage are most efficient when used and shared across multiple platforms. This may be due to a variety of reasons such as shared infrastructure and personnel, price breaks from vendors offering bundled services, or complementarities in functionality.

### D. Relation to Firm Size and Age

The relationship between adoption of purchased cloud services to firm size and age is shown in Table 12. As in Table 8, the cells contain the magnitudes of the estimated coefficients<sup>47</sup> from a linear probability model, where the dependent variable is an indicator for a firm purchasing cloud services for at least one IT function and the independent variables are the 16 size-age bins. While the relationship between cloud services expenditure and both size and age are generally lower than for digital information across all cells, the patterns are broadly similar. For the smallest size category (1-9 employees), the adoption rate declines with age, from 0.74 to 0.58, moving from the youngest (0-5 years old) to the oldest (21+ years old) firms. For the middle size categories (10-49 or 50-249 employees), the variation across age categories is much more limited. For the largest size category (250+ employees), there is a non-monotonic pattern with the adoption rate first increasing with age and then declining. The smallest-oldest category has the lowest adoption rate, consistent with the pattern for digital information. The highest rates occur for firms in the higher size categories, again consistent with the digitization patterns. The relatively higher rates of adoption for middle- and large-size categories may indicate that these firms are most likely to outsource IT services to a cloud computing provider, while smaller firms may either perform IT functions internally, not have a need for cloud computing and storage

<sup>&</sup>lt;sup>47</sup> All coefficients are statistically significant at a 0.1% level of significance.

services, or face higher frictions to adoption (e.g., such as a dedicated IT staff for migrating IT functionality into the cloud).<sup>48</sup>

To summarize, the adoption rates for business IT functions in the cloud is significantly lower than the adoption rates of storing information digitally. However, this technology is fairly widespread across various applications, as nearly a third of each different type of IT function is being performed in the cloud and being used intensively. We see a monotonic pattern of adoption by size similar to digitization, where the largest firms are the most likely to adopt some form of cloud computing services. The relationship between cloud use and firm age is more nuanced and varies with firm size while being non-monotonic. Typically, the lowest adoption rates are among the oldest firms. The next section looks at business technologies and their patterns of adoption.

## 6. Advanced Business Technologies

In this section we analyze firm responses to the business technologies question. Due to their wide technological scope, we link the responses here with the previous technology adoption questions and perform a deeper set of analyses assessing the range of response categories.

#### A. Summary of Responses

The frequency of responses in Figure 5 indicates that very few firms use the business technologies included in the module, and many answered, "Don't know". Based on our LBD tabulation weights, only 10.3% (8.5% non-imputed) of firms adopt at least one of the listed advanced business technologies. Given the advanced and specialized nature of at least some of the technologies, it is not surprising that only a relatively small number of firms indicate any type of use. The highest use frequencies are observed in automated storage, touchscreens and machine learning. However, our analysis of the responses to the automated storage question indicated that firms most likely interpreted automated storage as mainly data storage and not the physical storage and retrieval systems the question was intended to measure.<sup>49</sup> As a result, we do

<sup>&</sup>lt;sup>48</sup> In the case of the latter, we may be able to capture this by looking at adoption patterns in future ABS modules that focus on cloud computing.

<sup>&</sup>lt;sup>49</sup> We performed multiple sets of analyses identifying the sectors and industries most likely to use automated storage and retrieval systems and looked at technological similarities that were correlated with automated storage and retrieval systems. We found the publishing sector is the largest adopter and most likely to adopt automated storage. We also found a significantly higher correlation between automated storage and "data storage" in cloud computing than would be predicted. As a result, we concluded that subsequent analyses using automated storage as an outcome variable are likely to be invalid.

not consider the responses for this technology suitable for addressing our research questions, and they are dropped from the analysis.

#### B. Adoption and Use Rates

Table 13 provides the use and testing rates for each business technology. Perhaps not surprisingly, the highest use and testing rates are observed for the Touchscreen/Kiosks technology. The adoption rate is 6.1% for this technology and the testing rate is quite small (0.9%). Machine learning comes second in use and testing rates, but the rates are quite low at 2.9% and 0.7%, respectively. Voice Recognition and Machine Vision, which are can be considered examples of Machine Learning applications, have the next two highest use and testing rates.

While robots are usually singled out as a key technology in studies of automation, the overall diffusion of robotics use and testing is very low across firms in the U.S. The use rate is only 1.3% and the testing rate is 0.3%. These levels correspond relatively closely with patterns found in the robotics expenditure question in the 2018 ASM. Robots are primarily concentrated in large, manufacturing firms. The distribution of robots among firms is highly skewed, and the skewness in favor of larger firms can have a disproportionate effect on the economy that is otherwise not obvious from the relatively low overall diffusion rate of robots. The least-used technologies are RFID (1.1%), Augmented Reality (0.8%), and Automated Vehicles (0.8%). Looking at the pairwise adoption of these technologies in Table 14, we find that use of Machine Learning and Machine Vision are most coincident. We find that use of Automated Guided Vehicles is closely associated with use of Augmented Reality, RFID, and Machine Vision.

Next, we turn to testing-versus-use rates across different technologies to assess which technologies are in earlier phase of diffusion, that is, where testing is high relative to use. In Figure 6, the vertical axis represents the ratio of the fraction of firms testing to the fraction of firms using. The technologies are represented by the circles. The size of each circle corresponds to the use rate for that technology with larger circles representing higher rates of use. Technologies are ordered in the figure by usage rate, low to high. As shown in panel a, the technology with the highest testing-to-use ratio is Augmented Reality, where nearly half as many firms as those using the technology report testing it. The next highest ratios are observed in RFID and Natural Language Processing and the lowest ratios are in technologies that are

relatively more diffused (and hence, used), such as Touchscreens, Machine Learning and Machine Vision. For Touchscreens, for instance, only about 15 firms report testing the technology for every 100 that use it. It is notable that most testing-to-use ratios are below 0.3, indicating that there are fewer than 30 firms testing the technology for every 100 using it.

The remaining panels of Figure 6 plot the testing-to-use ratio for technologies by firm size, age, and manufacturing status. Panel b displays ratios by firm size, where small firms are defined as those with 1-9 employees and large firms are those with at least 250 employees. The blue circles capture usage among large firms and the orange circles represent usage among small firms. The sizes of the circles are smaller for small firms for each technology, consistent with the earlier finding that larger firms tend to use the business technologies at a higher rate, in general. Interestingly, testing-to-use ratios are higher for small firms for most business technologies (and particularly for Robotics and RFID technologies). The ratios are similar for large and small firms for Automated Vehicles and Machine Learning.

Panel c in Figure 6 shows the ratios by firm age. Young firms are defined as those that are 0-5 years old and old firms are the ones that are 21 years or older. Strikingly, testing-to-use ratios are almost uniformly higher for young firms compared to the older firms, in some cases substantially so (e.g., Augmented Reality, Automated Vehicles, and Robotics). The only technology where the two ratios are similar is Touchscreens. It is interesting to note that circle sizes are similar across the two age groups, reflecting the earlier finding that firm age is less of a predictor for technology use than is firm size. Overall, the patterns in Figure 6, panel c indicate that within the population of young firms there is a high rate of testing compared to use, whereas older and larger firms tend to either experiment less with these technologies or the diffusion of these technologies among the set of older and larger firms is relatively greater by the time of our survey.

Finally, panel d in Figure 6 presents the testing-to-use ratios for manufacturing versus non-manufacturing industries. Despite popular conceptions that Services are at the frontier of new technology adoption, more of the technologies covered in the ABS tend to be adopted within manufacturing, as indicated by the much larger circles for manufacturers, with the exception of Voice Recognition, Touchscreens, and Natural Language Processing. For the technologies most closely associated with automation, such as Machine Learning, Robotics and

Machine Vision, manufacturing firms show significantly higher usage compared to nonmanufacturing firms. This seems to indicate that these specific technologies may have started to find their place within the production process for manufacturing firms, while non-manufacturing firms are still experimenting with how these technologies can be implemented for their businesses. The largest differences in testing-to-use ratios across the two sector-based groups are in Voice Recognition, where testing is relatively more prevalent in manufacturing, and Natural Language Processing, where testing is relatively more intense in non-manufacturing.

#### C. Industry Breakdown

The butterfly chart in Figure 7 provides sectoral diffusion rates for all business technologies considered together. Manufacturing leads with about 15% of firms indicating use of at least one business technology, followed by Health Care (14%), Information (12%), Education (11%) and Professional Services (10%). The lowest diffusion rates for the technologies are in Construction, Agriculture, Mining and Utilities, Management and Administrative, and Finance, Insurance and Real Estate sectors. The extensive margin of use in these sectors hover around 5%. Note that, conditional on at least some advanced technology adoption, most firms across all sectors report using only one technology, in stark contrast to the use of digital information and cloud services. It may be that the benefits of digital information and cloud services across multiple business functions are much higher than for the use of these frontier technologies, either because they are more "general or flexible" in their application or because there are economies of scope in their adoption. In addition, the costs of adoption for digitization and cloud services may also be lower.

The testing rates on the left panel of the figure reveal an interesting pattern. While Manufacturing leads sectors in rates of adoption, the testing rate in Manufacturing is not the highest. In fact, Information has the highest testing rate at about 8.1%, followed by Professional Services. These two sectors also have the highest testing-to-use ratio (0.48 and 0.36, respectively). The lowest ratios are in Health Care, Retail Trade and the Other (Arts, Food, and others).

Looking at the most common types of business technologies adopted by sector in Table 15, we find that there is substantial variation. The trade sectors (Retail, Wholesale and others) primarily adopts Touchscreens followed by Machine Learning. Manufacturing is most likely to

adopt Machine Learning followed by Touchscreens and Robotics. RFID technology is most commonly used in the Retail, Wholesale, and Transportation and Warehousing sectors, consistent with these industries tracking physical goods through supply chains.

The three industries (4-digit NAICS) with the highest adoption rate for a given technology are shown in Table 16. Not surprisingly, robotics use is highest in three manufacturing industries, with the highest rate in Motor Vehicle Parts Manufacturing (17%).<sup>50</sup> RFID use is most common in Warehousing and Storage (6%).<sup>51</sup>

#### D. Relation to Firm Size and Age

How does the use of business technologies vary by firm size and age? Table 17 provides the estimated coefficients from a Linear Probability Model where the left-hand-side variable is whether a firm uses at least one of the business technologies, and the right-hand-side variables are 16 size-age categories, as in Tables 8 and 12. A clear pattern emerges. First, the smallest firms have the lowest use rates and the use rates tend to increase with size, even when controlling for firm age. Second, for small firms (less than 50 employees) use rates tend to decline with age (though not monotonically), with oldest small firms having the lowest adoption rates in general. For larger firms (50+ employees), use rates exhibit the opposite pattern: as firm age increases, so does the use rate (especially for 50-249 employee firms). The highest use rates are in the largest and oldest firms. Another notable feature of the table is that for each age category the use rate increases with size. Overall, these patterns suggest that size is an important predictor of business technology use and the connection between age and the use of these technologies depends on size.

## 7. Discussion of Results in the Context of Open Research Questions

The technology module of the 2018 ABS reveals several interesting patterns of technology adoption across firm size, age, and sector. Most notably, we find that adoption for the latest advanced technologies appears to be quite low, overall, with adoption mainly being led by the

<sup>&</sup>lt;sup>50</sup> This finding is also consistent with the preliminary findings from the Robotics question in the 2018 ASM on which manufacturing industries are the largest adopters.

<sup>&</sup>lt;sup>51</sup> One interesting, but potentially concerning finding is that Machine Learning is most prevalent in Metalworking Machinery Manufacturing (12.3%) and Machine Shops (11.6%), which may be the result of cognitive error around the term "machine" or reflect the growing trend of embedding algorithms in production technology. Software Publishers, less surprisingly, are in the top-3 industries using machine learning. Further follow-up is needed.

largest and oldest firms. This contrasts with popular discussion but it is consistent with much prior work on IT adoption, which documents advantages for incumbent firms (albeit sometimes with a lag) due to mechanisms such as economies of scale or complementary organizational capital (Bertschek and Kaiser 2004, Tambe and Hitt 2012, Saunders and Brynjolfsson 2016). We also capture technologies at different states of diffusion, with digitization and cloud computing already taking on relatively large and significant roles in business and, within adopting firms, across business functions.

The technology module, in its current state, can be used to address several open empirical questions related to technology adoption across firms.<sup>52</sup> These include first-order questions, such as how pervasive AI use is among US firms, to more deeper questions relating to dynamics and diffusion, sequence of technology adoption, and the organizational capabilities and/or infrastructure required to adopt the most-advanced technologies. This data is also informative about complementarities among different technologies (Milgrom and Roberts 1990 & 1995, Bertschek and Kaiser 2004, Brynjolfsson and Milgrom 2013), which have already been mentioned in describing some of the notable pairwise correlations. In addition to these questions, the module linked to the 2018 ABS innovation module sheds light on technology's role in stimulating innovation. Finally, and most relevant, the technology module provides a glimpse into the macro-economic and distributional effects of technology use. Each of these questions are discussed below.

## A. AI Adoption

One of the primary goals of the ABS technology module is to provide the first comprehensive look into the adoption rates of Artificial Intelligence (AI) by US firms. As AI rapidly advances its capabilities and becomes more integrated into the workplace, there is an ongoing debate as to whether these technologies will lead to further prosperity and enhance our productivity or whether they will lead to mass joblessness and wage stagnation (Brynjolfsson and McAfee, 2014; Aghion et al. 2017). Central to this debate is obtaining reliable measures on the

<sup>&</sup>lt;sup>52</sup> This is in contrast to some of the open macroeconomic questions related to technology adoption described in Comin & Mestieri (2014). These questions also include how to measure technology adoption and what are some of the key drivers of adoption, but with emphasis on the macro and across-country effects of adoption.

pervasiveness of this technology and identifying which firms are already using this technology so that we can begin to gauge its impact.

In a 2017 report by the National Academies of Science, Engineering, and Medicine (NASEM) entitled, "Information Technology and the U.S. Workforce: Where Are We and Where Do We Go From Here?", an expert panel acknowledges the potential of AI to transform the economy, but notes that accurate forecasts of these effects are limited due to the absence of reliable data on the use of AI in the economy. The ABS can potentially rectify this gap in the data by providing a first glimpse as to the spread of AI across a nationally representative set of firms and identifying a set of firms that are leading the race in its adoption.

From the set of business technologies, we have identified five potential technologies that incorporate elements of AI. These include: Automated Guided Vehicles, Machine Learning, Machine Vision, Natural Language Processing and Voice Recognition. Across all AI-related technologies, the aggregate adoption rate for all firms in the economy is 6.6% (4.5% non-imputed), meaning that approximately 1 in 16 firms in the US are utilizing some form of AI in the workplace. This adoption rate is significantly lower than the adoption rate highlighted in the AI survey by the European Commission and other private surveys by McKinsey, Deloitte, and PwC. However, it is important to consider the sampling methods of those surveys. Neither of the other surveys claim to be nationally representative and tend to focus on larger, publicly traded companies. In contrast, ABS sample includes many small firms where AI adoption is very low.

This is important because AI adoption rate varies greatly by firm size. Figure 8 charts the adoption rate of AI across 12 different size categories used in the BDS. Adoption rates (defined as usage or testing) monotonically increase from 5.3% for the group of firms with the smallest number of employees to 62.5% for firms with 10,000+ employees.

In other words, scale appears to be a primary correlate of AI usage, likely due to both the large quantities of data and computing power required to fully realize the most popular types of AI currently available. This may potentially have far-reaching implications on topics such as inequality, competition and the rise of "superstar" firms (Autor et al 2020; Lashkari et al 2019), especially if AI is shown to have widespread productivity benefits (Tambe et al, 2020). If only a select group of firms are able to fully realize the benefits of AI, we can expect further divergence

for the "frontier" and most productive set of firms. Alternatively, early adoption of AI by some firms may generate rapid rates of growth.

#### B. Technological Hierarchies

The three technology categories listed in the technology module seem to require different levels of technological sophistication for adoption, with digitization being the first step, and culminating in one of the advanced business technologies. This "hierarchy" in technological sophistication is apparent in Figure 9, which plots a Sankey diagram of firm counts that adopt each of the different technology categories. From the diagram, we can see that the vast majority of firms who utilize the cloud for their IT services also digitize their information. Similarly, we see that for the vast majority of firms that adopt at least one advanced business technology, they almost always purchase cloud services.

This clear evidence of hierarchy points to some of the challenges that firms may face, as well as barriers to technology adoption. In the exercises that relate technology use to firm size and age, we found that the largest and oldest firms are by far the most likely to adopt at least one business technology, implying that scale effects may be an important determinant of technology adoption. And while it has been speculated that cloud services can "open" up unlimited computing power to smaller firms, the uptake of this technology and other advanced, nascent technologies is still very low, suggesting that the rates of return to investing in these technologies for smaller firms is not high enough to justify the costs. This may be due to the return being dependent on the scale of the data being used. Small firms are likely to have only small amounts of digital information and so have little reason to invest in high-powered computing or apply the latest machine learning algorithms to their data. On the other hand, the returns to access to high-powered computing and the latest advances in AI become much more apparent as digitization scales.

In panel b. of Figure 9, we plot a similar Sankey diagram for firms who test or use "Machine Learning", one of the core technologies associated with artificial intelligence. In this example, the majority of firms who adopt Machine Learning have multiple cloud computing purchases (usually 3 or more IT functions hosted on the cloud) and multiple pieces of their data digitized. Related to this idea of technological hierarchies, are technological complementarities, which we discuss in the next subsection.
#### C. Technological Complementarities

We started to explore the notion of technological complementarities when we listed the top five pairwise technological correlations within each category of technology. The idea behind technological complementarities is that adopting one kind of technology is likely to lead to adopting another type, or that certain technology types require adoption of multiple technologies and complementary investments in order to fully benefit (Brynjolffson et al. (2018)).

To further explore this idea, we look at the highest cross-category pairwise correlations with each of the advanced business technologies in Table 18. We find that for many of the core elements of artificial intelligence, namely Machine Learning, Machine Vision, and Natural Language Processing, the associated technology categories associated with these include digitizing production information and performing their data analysis on the cloud. These correlations suggest that certain technologies may need to be adopted in tandem to fully reap the benefits of the technology.

#### D. Technology and Innovation

Adopting technologies is often associated with improvements in productivity and efficiency. However, it is not clear what mechanism within the firm causes these improvements, as adoption may also be associated with higher labor costs (for higher skilled workers), training and learning, significant capital investments and perhaps changing some of the underlying fundamentals for the firm. These factors may also force the firm to innovate so that it can effectively adopt the new technology.

As an exercise, we look at how technology adoption is associated with measures of both product and process innovations within the innovation module of the ABS.<sup>53</sup> We estimated a linear probability model looking at whether a firm responded positively to producing a product or process innovation in the last three years based on their technology adoption, controlling for their size, age and industry. We group each of the firms into eight different technology categories: None (no use across all technology categories), Digitization Only, Cloud Only,

<sup>&</sup>lt;sup>53</sup> Product innovation is defined as the business having introduced or significantly improved a new good or service, while process innovation is defined as the business having introduced or significantly improved their method of manufacturing, logistics, delivery or distribution methods or support activities.

Business Technology Only, Digitization and Cloud, Digitization and Business Technology, Cloud and Business Technology, All three technologies.

We plot the coefficients for these technology groupings in Figure 10. We see that as the technological sophistication increases, the magnitudes of the coefficients for both product and process innovations increases, indicating a positive association between technology adoption and innovation. We further decompose these regressions to each of the subcategories in panel b, looking at the adoption of each technology type on any (product or process) innovation. We see a surprising amount of heterogeneity across each technology type, with certain technologies having stronger and positive associations with innovation than others.

#### E. Macro/Distributional Outcomes of Technology Use

While a survey can tell us firm exposure to certain technologies, it does not tell the whole story for the worker and the technologies to which workers are exposed. In this section, we weigh some of the key statistics by employment and demonstrate how some of the key advanced business technologies, despite having relatively low adoption rates at the firm-level, have significantly higher worker exposure rates.

Weighted by the LBD tabulation weights, the imputed adoption rates of digitization, cloud services and at least one advanced business technology are 90.2%, 78.4% and 10.3% respectively. However, in Table 19, if we assume that each worker within the firm has at least some exposure to the firm's technology, then the "exposure" rates for digitization, cloud services<sup>54</sup> and at least one business technology change to 98.9%, 94.3% and 57.2%. The employment-weighted shares of adoption for all technology types differ quite dramatically from the weighted results, with the employment-weighted adoption rates for advanced business technologies being nearly five times higher. Therefore, while the firm-level adoption rates for advanced business technologies. Second, more than 4 of 10 workers are in firms that have adopted at least one form of digital information and purchased cloud services. These findings raise some important questions regarding the macroeconomic/distributional impacts of these technologies, especially if we believe that these technologies will substitute for labor.

<sup>&</sup>lt;sup>54</sup> Conditional on collecting and storing at least one type of information and utilizing at least one IT function

Furthermore, the discrepancies between firm versus employment-weighted adoption rates hint at the reallocation that may be at play in shifting inputs and resources from non-adopting to adopting firms.

Taking a closer look at each of the business technologies in Table 19, we find that the business technologies with the highest discrepancies between the firm-weighted adoption rates and employment-weighted adoption rates are Robotics and RFID, each of which has an eight times higher employment-weighted adoption rate. One of the limitations of the survey is that it is a firm-level survey, while adoption of certain technologies may take place only at the establishment level, suggesting that the employment-weighted exposure measures listed here are an upper bound.

#### F. Technology as Equalizer

One reason some economists believe that technology adoption is important is that technology may be viewed as a "great equalizer." In this view, young and small firms, that could be more nimble, may be able to quickly scale up using the cost savings and efficiency improvements from adopting the latest technologies as compared to old, large incumbents. One aspect that this survey reveals, however, is that the latest technology adoption is mostly being done by the larger and older firms, potentially leading to increased separation between the typical firm and "superstar" firms.

This may actually be driven by the technology itself, as much of the latest technology relies on scale effects to be useful. For instance, leading manifestations of machine learning and artificial intelligence benefit from larger amounts of data and smaller firms may be unable to provide the necessary data where adopting these technologies proves efficient. This could create a positive feedback loop creating "superstar" firms as large firms reap efficiency gains from adopting and refining the latest technologies, which in turn, makes them larger and making the technology even more productive.

#### 8. Conclusion

We have provided an introduction to the technology module in the 2018 ABS and placed it in the larger context of related work at the Census Bureau to collect comprehensive data on technology adoption and use by U.S. firms in order to provide a more accurate picture of the

38

state of advanced technology use in the U.S. economy. Because of the large pool of respondents (about 850,000 firms) in the 2018 ABS, the module represents a unique opportunity to offer insights on technology adoption and use across all sectors of the economy and across a variety of key firm characteristics.

Using this new collection, we provide a first look at the diffusion of digital information use, cloud computing purchases, and several new and emerging business technologies. A few key observations emerge. While the use of digital information in business functions and cloud computing purchases for many IT functions are highly prevalent across firms, the diffusion of the business technologies is very limited. However, employment-weighted diffusion rates indicate that the presence of many of these technologies in large firms exposes many workers in the economy to these technologies. Further insights into how these new technologies impact the skill composition and demand for these workers will be unveiled in the 2019 ABS. In addition, there are important differences in the diffusion and intensive use rates across sectors. The analysis of the connection between the prevalence of different technologies and firm life-cycle indicators (firm size and age) reveals that technology adoption and use is not always monotonically related to these indicators.

In general, the business technologies explored in the module's third question are more prevalent in larger and older firms. This skewness in technology prevalence suggests that these technologies may have a disproportionate economic impact despite their generally low adoption rates. If the concentration of economic activity in larger and older firms in the U.S. economy continues to increase as it has in the recent past, then the effects of technology adoption by these firms will have growing influence on key economic aggregates, such as employment and productivity. It is our hope that this paper serves as an impetus for further research using this new data set to help answer these important questions.

Looking towards the future, we will continue to validate the responses from the survey by incorporating and comparing output from existing Census data on technology use such as the 2018 ASM, 2018 BRDIS (and 2019 BERD) and 2018 ACES. We also plan to utilize administrative data, such as patents linked to Census data (see Graham et al. 2018) to help validate responses and look outside towards external researchers utilizing the Federal Statistical Research Data Centers (FSRDC) network to contribute their ideas on validating and improving

39

the data. Finally, looking even further in the future, the same technology module is expected to be a part of the 2021 Annual Business Survey, providing a panel dimension for the set of firms queried on both modules.

#### References

- Acemoglu, D, and Autor D. 2011. "Skills, tasks and technologies: Implications for employment and earnings." *Handbook of labor economics*. Vol. 4. Elsevier. 1043-1171.
- Acemoglu, D. and Restrepo, P., 2019. "Automation and new tasks: how technology displaces and reinstates labor." *Journal of Economic Perspectives*, 33(2), pp.3-30.
- Acemoglu, D. and Restrepo, P., 2020. "Robots and jobs: Evidence from US labor markets." *Journal of Political Economy*, 128(6), pp.2188-2244.
- Aghion, P., Jones, B.F. and Jones, C.I., 2017. "Artificial intelligence and economic growth". National Bureau of Economic Research Working Paper No. w23928.
- Agrawal, A, Gans J, and Goldfarb A. 2019. "The Economics of Artificial Intelligence" NBER Conference Report.
- Andrews, D, Criscuolo, C and Gal, P. 2015. "Frontier Firms, Technology Diffusion and Public Policy." OECD Future of Productivity Background Paper.
- Angle, P and Forman C. 2018. "Does it level the playing field for small establishments? Evidence from manufacturing. Evidence from Manufacturing". Georgia Tech Scheller College of Business Research Paper. 18-37):
- Aral, S., Brynjolfsson, E. and Wu, L., 2012. "Three-way complementarities: Performance pay, human resource analytics, and information technology." *Management Science*, 58(5), pp.913-931.
- Armbrust, M, Fox A, Griffith R, Joseph AD, Katz R, Konwinski A, Lee G, Patterson D, Rabkin A and Stoica I. 2010. "A view of cloud computing." *Communications of the ACM*. 53(4): 50-58.
- Atalay, E, Phongthiengtham P, Sotelo S and Tannenbaum D. 2020. "The evolution of work in the united states." *American Economic Journal: Applied Economics*. 12(2): 1-34.
- Atrostic, BK and Nguyen SV. 2005. "IT and productivity in U.S. Manufacturing: Do computer networks matter?" *Economic Inquiry*. 43(3): 493-506.
- Atrostic, BK and Nguyen SV. 2006. "How businesses use information technology: Insights for measuring technology and productivity." U.S. Census Bureau Center for Economic Studies Working Paper CES 06-15.
- Atrostic, BK and Nguyen SV. 2007. "Computer input, computer networks, and productivity." Hard-to-measure goods and services: Essays in honor of Zvi Griliches. University of Chicago Press.
- Autor, D and Salomons A. 2018. "Is Automation Labor-Displacing? Productivity Growth, Employment, and the Labor Share." Brookings Papers on Economic Activity, (Spring): 1-87.
- Autor, D, Dorn D, Katz LF, Patterson C and Van Reenen J. 2020. "The fall of the labor share and the rise of superstar firms." *The Quarterly Journal of Economics*. 135(2): 645-709.
- Bana, S., Brynjolfsson, E., Rock. D., and Steffen, S. 2020. "Job2vec: Learning a Representation of Jobs", Available at SSRN.

- Barth, E., Bryson, A., Davis, J.C. and Freeman, R., 2016. "It's where you work: Increases in the dispersion of earnings across establishments and individuals in the United States." *Journal of Labor Economics*, 34(S2), pp.S67-S97.
- Barth, E., Davis, J.C., Freeman, R.B. and McElheran, K., 2020a. "Twisting the Demand Curve: Digitalization and the Older Workforce." NBER Working Paper w28094.
- Barth, E., Roed, M., Schøne, P. and Umblijs, J., 2020b. "How Robots Change Within-Firm Wage Inequality."
- Basker, E, Foster L and Klimek S. 2017. "Customer-employee substitution: Evidence from gasoline stations." *Journal of Economics & Management Strategy*. 26(4): 876-896.
- Bennett, VM. 2020a. "Changes in persistence of performance over time." *Strategic Management Journal*. 41(10): 1745-1769.
- Bennet, VM. 2020b. "Automation and market dominance". *Mimeo*, Fuqua School of Management, Duke University.
- Bertschek, I and Kaiser U. 2004. "Productivity Effects of Organizational Change: Microeconometric Evidence." *Management Science*. 50(3): 394-404.
- Bessen, J. 2002. "Technology adoption costs and productivity growth: the transition to information technology." *Review of Economic dynamics* 5.2: 443-469.
- Bessen, J. E. and C. Righi. 2020. "Information Technology and Firm Employment." Boston Univ. School of Law, Law and Economics Research Paper No. 19-6. Available at SSRN: https://ssrn.com/abstract=3371016 or http://dx.doi.org/10.2139/ssrn.3371016.
- Bloom, N., Hassan, T.A., Kalyani, A., Lerner, J. and Tahoun, A., 2020. "The Geography of New Technologies." Institute for New Economic Thinking Working Paper Series, (126).
- Bloom, N, Sadun R and Van Reenen J. 2012. "Americans do IT better: US multinationals and the productivity miracle." *American Economic Review*, 102(1), pp.167-201.
- Bloom, N, Garicano L, Sadun R, and Van Reenen J. 2014. "The distinct effects of information technology and communication technology on firm organization." *Management Science* 60, no. 12: 2859-2885.
- Bloom, N and Draca M and Van Reenen J. 2016. "Trade induced technical change? The impact of Chinese imports on innovation, IT and productivity." *The Review of Economic Studies*, 83(1), pp.87-117.
- Bresnahan, T.F., Brynjolfsson, E. and Hitt, L.M., 2002. "Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence." *The Quarterly Journal of Economics*, 117(1), pp.339-376.
- Brynjolfsson E, Hofmann P and Jordan J. 2010. "Cloud computing and electricity: Beyond the utility model." *Communications of the ACM*. 53(5): 32-34.
- Brynjolfsson, E, Jin W and McElheran K. 2020. "The power of prediction: Predictive analytics, organizational complements, and firm performance." Working paper available at SSRN.
- Brynjolfsson, E, Rock D, and Syverson C. 2017. "Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics." National Bureau of Economic Research, No. w24001.

- Brynjolfsson, E., Hui, X. and Liu, M., 2019. "Does machine translation affect international trade? Evidence from a large digital platform." *Management Science*, 65(12), pp.5449-5460.
- Brynjolfsson, E, Malone TW, Gurbaxani V and Kambil A. 1994. "Does information technology lead to smaller firms?" *Management Science*. 40(12): 17.
- Brynjolfsson, E, and McAfee A. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies.* WW Norton & Company, 2014.
- Brynjolfsson, E and Mitchell T. 2017a. "What Can Machine Learning Do? Workforce Implications." *Science* 358(6370): 1530-1534.
- Brynjolfsson, E and Mitchell T. 2017b. "Track How Technology Is Changing Work." *Nature*, 544(7650): 290-291.
- Brynjolfsson, E and McElheran K. 2016 "The Rapid Adoption of Data-Driven Decision-Making." *American Economic Review Papers and Proceedings*, 106(5): 133-139.
- Brynjolfsson E and McElheran K. 2019. "Data in Action: Data-Driven Decision Making and Predictive Analytics in U.S. Manufacturing." Working paper, Rotman School of Management, Toronto, Canada.
- Brynjolfsson, E and Milgrom P. 2013. "Complementarity in organizations." *The handbook of organizational economics*, pp.11-55.
- Brynjolfsson, E., Rock, D. and Syverson, C., 2018. "The productivity J-curve: How intangibles complement general purpose technologies".National Bureau of Economic Research Working Paper No. w25148.
- Brynjolfsson, E. and Saunders, A., 2009. "Wired for innovation: How information technology is reshaping the economy". Mit Press.
- Buffington, C, Miranda J and Seamans R. 2018. "Development of Survey Questions on Robotics Expenditures and Use in U.S. Manufacturing Establishments." U.S. Census Bureau Center for Economic Studies Working Paper, CES 18-44.
- Comin, D. and Mestieri, M., 2014. "Technology diffusion: Measurement, causes, and consequences." In *Handbook of Economic Growth* (Vol. 2, pp. 565-622). Elsevier.
- Crane, L.D, and Decker, RA. 2020. "Research with private sector microdata: The case of NETS/D&B." Conference paper, NBER Summer Institute. https://conference.nber.org/conf\_papers/f142811.pdf.
- Decker, R, Haltiwanger J, Jarmin R and Miranda J. 2014. "The role of entrepreneurship in us job creation and economic dynamism." *Journal of Economic Perspectives*. 28(3): 3-24.
- Decker, RA, Haltiwanger J, Jarmin RS and Miranda J. 2017. "Declining dynamism, allocative efficiency, and the productivity slowdown." *American Economic Review Papers and Proceedings*. 107(5): 322-26.
- Decker, RA, Haltiwanger J, Jarmin RS and Miranda J. 2020. "Changing business dynamism and productivity: Shocks versus responsiveness." *American Economic Review*. 110(12): 3952-90.

- Dinlersoz, EM and Wolf Z. 2018. "Automation, labor share, and productivity: Plant-level evidence from U.S. Manufacturing." U.S. Census Bureau Center for Economic Studies Working Paper, CES 18-39.
- Dixon, J., Hong, B. and Wu, L., forthcoming. "The Robot Revolution: Managerial and Employment Consequences for Firms." *Management Science*.
- Dunne, T. 1991. "Technology usage in u.S. Manufacturing industries: New evidence from the survey of manufacturing technology." U.S. Census Bureau Center for Economic Studies Working Paper, CES 91-7.
- Dunne, T and Schmitz JA, Jr. 1995. "Wages, employment structure and employer size-wage premia: Their relationship to advanced-technology usage at us manufacturing establishments." *Economica*. 62(245): 89-107.
- Doms, M, Dunne T and Roberts MJ. 1995. "The role of technology use in the survival and growth of manufacturing plants." *International Journal of Industrial Organization*. 13(4): 523-542.
- Fichman, RG and Melville NP. 2014. "How posture-profile misalignment in it innovation diminishes returns: Conceptual development and empirical demonstration." *Journal of Management Information Systems*. 31(1): 203-240.
- European Commission. 2020. "European enterprise survey on the use of technologies based on artificial intelligence". https://ec.europa.eu/digital-single-market/en/news/artificial-intelligence-first-quantitativ.e-study-its-kind-finds-uptake-businesses-across-europe
- Forman, C, Goldfarb A and Greenstein S. 2002. "Digital dispersion: An industrial and geographic census of commercial internet use". National Bureau of Economic Research No w9287.
- Forman, C, Goldfarb A and Greenstein S. 2009. "The Internet and Local Wages: Convergence or Divergence?" National Bureau of Economic Research No. w14750.
- Forman, C, Goldfarb A and Greenstein S. 2012. "The Internet and local wages: A puzzle." *American Economic Review* 102.1: 556-75.
- Forman, C and McElheran K. 2019. "Firm organization in the digital age: It use and vertical transactions in U.S. Manufacturing." University of Toronto Working Paper, https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3396116.
- Forrester. 2017. "Unlock the value of cloud: How to expand your hybrid cloud with consistency, high performance, and security everywhere." Report, Forrester Consulting, Cambridge, MA.
- Foster, L., Grim, C. and Zolas, N., 2020. "A portrait of US firms that invest in R&D." *Economics of Innovation and New Technology*, 29(1), pp.89-111.
- Fort, TC. 2017. "Technology and production fragmentation: Domestic versus foreign sourcing." *The Review of Economic Studies*. 84(2): 650-687.
- Goldfarb, A and Tucker C. 2019. "Digital economics." *Journal of Economic Literature*, 57(1), pp.3-43.

- Graetz, G and Michaels G. 2018. "Robots at work." *The Review of Economics and Statistics*, 100(5):753-768.
- Graham, S.J., Grim, C., Islam, T., Marco, A.C. and Miranda, J., 2018. "Business dynamics of innovating firms: Linking US patents with administrative data on workers and firms." *Journal of Economics & Management Strategy*, 27(3), pp.372-402.
- Green Leigh, N., Kraft, B. and Lee, H., 2020. "Robots, skill demand and manufacturing in US regional labour markets." *Cambridge Journal of Regions, Economy and Society*, 13(1), pp.77-97.
- Guzman, J. and Stern, S., 2020. "The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 32 US States, 1988–2014." *American Economic Journal: Economic Policy*, 12(4), pp.212-43.
- Haltiwanger, J, Jarmin RS, Kulick RB and Miranda J. 2016. "High growth young firms: Contribution to job, output and productivity growth." US Census Bureau Center for Economic Studies Paper No. CES-WP-16-49.
- Haltiwanger, J, Jarmin RS and Miranda J. 2013. "Who creates jobs? Small versus large versus young." *Review of Economics and Statistics*. 95(2): 347-361.
- Haug, K.C., Kretschmer, T. and Strobel, T., 2016. "Cloud adaptiveness within industry sectors-Measurement and observations." *Telecommunications policy*, 40(4), pp.291-306.
- Helper, S. 1995. "Supplier relations and adoption of new technology: results of survey research in the US auto industry." National Bureau of Economic Research No. w5278.
- Hopenhayn, H.A., 2014. "Firms, misallocation, and aggregate productivity: A review." *Annual Review of Economics.*, 6(1), pp.735-770.
- Horrace, W.C. and Oaxaca, R.L., 2006. "Results on the bias and inconsistency of ordinary least squares for the linear probability model." *Economics Letters*, 90(3), pp.321-327.
- Hitt, LM and Tambe P. 2016. "Health care information technology, work organization, and nursing home performance." *ILR Review*, 69(4), pp.834-859.
- Jensen, JB and McGuckin RH. 1996. "Firm performance and evolution: Empirical regularities in the U.S. Microdata." U.S. Census Bureau, Center for Economic Studies Working Paper CES 96-10.
- Jin, W and McElheran K. 2017. "Economies before Scale: Learning, Survival, and Performance of Young Plants in the Age of Cloud Computing", Rotman School of Management Working Paper No. 3112901.
- Kretschmer, T, Miravete EJ, and Pernías JC. 2012. "Competitive pressure and the adoption of complementary innovations." *American Economic Review* 102.4: 1540-70.
- Lashkari, D., Bauer, A. and Boussard, J., 2019. "Information technology and returns to scale." Available at SSRN https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3458604.
- Mahr, F. 2010. Aligning Information Technology, Organization, and Strategy: Effects on Firm Performance. Springer Science & Business Media, 2010.
- McElheran, K. 2014. "Delegation in multi-establishment firms: Evidence from it purchasing." *Journal of Economics & Management Strategy*, 23(2), pp.225-258.

- McGuckin, RH, Streitwieser ML and Doms M. 1998. "The effect of technology use on productivity growth." *Economics of Innovation and New Technology*. 7(1): 1-26.
- McKinsey. 2018 "AI Adoption Advances, but foundational barriers remain." https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advancesbut-foundational-barriers-remain#
- Milgrom, P and Roberts J. 1995. "Complementarities and fit strategy, structure, and organizational change in manufacturing." *Journal of Accounting and Economics*, *19*(2-3), pp.179-208.
- Milgrom, P and Roberts J. 1990. "The economics of modern manufacturing: Technology, strategy, and organization." *The American Economic Review*, pp.511-528.
- National Academies of Sciences, Engineering, and Medicine, 2017. "Information technology and the US Workforce: Where are we and where do we go from here?" National Academies Press.
- PwC. 2019. "2019 AI Predictions." https://www.pwc.com/us/en/services/consulting/library/artificial-intelligence-predictions-2019
- Schneier B. 2015. "Should companies do most of their computing in the cloud? (Part 1), blog post." <u>https://www.schneier.com/blog/archives/2015/06/should\_companie.html</u>, accessed May 1, 2017.
- Seamans, R, and Raj M. 2018. "AI, labor, productivity and the need for firm-level data." National Bureau of Economic Research No. w24239.
- Solow, R. M. 1957. "Technical change and the aggregate production function." *The Review of Economics and Statistics*, 312-320.
- Song, J, Price DJ, Guvenen F, Bloom N and Von Wachter T. 2019. "Firming up inequality." *The Quarterly Journal of Economics*. 134(1): 1-50.
- Staten J. 2008. "Is cloud computing ready for the enterprise?" Forrester Research Report. <u>https://www.forrester.com/report/Is+Cloud+Computing+Ready+For+The+Enterprise/-/E-RES44229#</u>.
- Stiroh, K. J. 2001. "What drives productivity growth?" Economic Policy Review, 7(1).
- Sunyaev A. 2020. "Cloud Computing. In: Internet Computing." Springer, Cham. https://doi.org/10.1007/978-3-030-34957-8\_7
- Syverson, C., 2004. "Product substitutability and productivity dispersion." *Review of Economics* and Statistics, 86(2), pp.534-550.
- Syverson, C. 2011. "What determines productivity?" *Journal of Economic Literature*. 49(2): 326-365.
- Tambe, P., Hitt, L.M. and Brynjolfsson, E., 2012. "The extroverted firm: How external information practices affect innovation and productivity." *Management Science*, 58(5), pp.843-859.

- Tambe, P and Hitt LM. 2012. "The productivity of information technology investments: New evidence from it labor data." *Information Systems Research*. 23(3-part-1): 599-617.
- Tambe, P., Hitt, L., Rock, D. and Brynjolfsson, E. 2020. "Digital Capital and Superstar Firms", NBER Working Paper.
- Van Reenen, J. 2018. "Increasing differences between firms: Market power and the macro economy." CEP Discussion Paper Working Paper, 1576.
- Webb, M. 2019. "The impact of artificial intelligence on the labor market." Available at SSRN 3482150.

# Text Box 1: 2018 Annual Business Survey Technology Questions

#### DIGITAL SHARE OF BUSINESS ACTIVITY

In 2017, how much of each type of information was kept in digital format at this business? *Select one for each row.* 

		None	Up to 50%	More than 50%	All	Don't know	This type of information not collected by this business
A. B. C. D. E. F.	Personnel Financial Customer Feedback Marketing Supply Chain Production Other: (specify)						

#### CLOUD SERVICE PURCHASES

Considering the amount spent on each of these IT functions, how much was spent on cloud services? (Cloud services are services provided by a third party that this business accesses ondemand via the internet.) *Select one for each row.* 

		None	Up to 50%	More than 50%	All	Don't Know	Don't use this IT
							function
Α.	All IT functions						
Β.	Security or firewall						
C.	Servers						
D.	Data storage and						
	management (Examples:						

E.	Amazon Web Services, IBM Bluemix, Microsoft Azure) Collaboration and file synchronization (Examples: Dropbox, OneDrive, Google			
F. G.	Drive) Data Analysis Billing and account			
Н.	management Customer relationship management			
I.	Other: (specify)			

#### **BUSINESS TECHNOLOGIES**

In 2017, to what extent did this business use the following technologies in producing goods or services? *Select one for each row.* 

		No use	Testing, but not using in production or service	In use for less than 5% of production or service	In use for between 5% – 25% of production	In use for more than 25% of production or service	Don't know
A.	Augmented reality						
B.	Automated guided vehicles (AGV) or AGV systems						
C.	Automated storage and retrieval systems						
D.	Machine learning						
E.	Machine vision software						
F.	Natural language						
G.	Radio-frequency identification (RFID) inventory system						
H.	Robotics						
I.	Touchscreens/kiosks for customer interface (Examples: self-checkout, self-						

Augmented reality	Technology that provides a view of a real- world environment with computer- generated overlays.
Automated guided vehicles (AGV) or AGV systems	A computer-controlled transport vehicle that operates without a human driver. AGVs navigate facilities through the use of software and sensors.
Automated storage and retrieval systems	Technology that locates, retrieves, and replaces items from predetermined storage locations.
Machine learning	Computer algorithms that use data to improve their predictive performance without being reprogrammed.
Machine vision	Technology used to provide image-based automatic inspection, recognition or analysis.
Natural language processing	Technology that allows a computer to process human speech or text.
Radio-frequency identification (RFID) system	A system of tags and readers used for identification and tracking. Tags store information and transmit them using radio waves. Readers maybe be mobile or fixed in place.
Robotics	Reprogrammable machines capable of automatically carrying out a complex set of actions.
Touchscreens/kiosks for customer interface (Examples: self-checkout, self- check-in, touchscreen ordering)	A computer with a touchscreen that allows a customer to receive information or perform tasks related to the business such as registering for a service or purchasing items.
Voice recognition software	Software that converts speech to text or executes simple commands based on a limited vocabulary or executes more complex commands when combined with natural language processing.

Text Box 2: 2018 Annual Business Survey Technology Definitions

Survey	Years	Topics
Annual Business Survey (ABS)*	2017-2018	Software, Data Processing, Digitization, Cloud Services, Automation, AI, Robotics
Annual Capital Expenditures Survey	2018	Robotics
(ACES)	2002-2018	Software
Appuel Detail Trade Survey (ADTS)	1999-2018	E-Commerce
Annual Retail Trade Survey (ARTS)	2002, 2017	Software, Data Processing
Appual Survey of Entropropours (ASE)	2014-2016	E-Commerce, Tech on Profits
Annual Survey of Entrepreneurs (ASE)	2014	Software, Automation
	2018	Robotics
Annual Survey of Manufactures (ASM)	2000-2018	Data Processing
Annual Wholesale Trade Survey (AWTS)	1999-2018	E-Commerce
Business Research & Development and Innovation Survey (BRDIS)*	2008-2016	Software
Business R&D and Innovation Survey - Microbusiness (BRDI-M)*	2014, 2016	Software, Automation
Census of Construction Industries	2012	Software, Data Processing
(CCN)	2002	E-Commerce
Census of Manufacturing (CMF)	2002, 2007, 2012	E-Commerce, Data Processing
Census of Retail Trade (CRT)	2017	Self-service technologies
Computer Network Use Supplement (CNUS)	1999	E-Commerce
Information and Communication Technology Survey (ICTS)	2003-2011, 2013	ICT, Software
Management and Organizational Practices Survey (MOPS)	2010, 2015	Data-Driven Decisions
Service Annual Survey (SAS)	2005-2016	E-Commerce
Survey of Business Owners (SBO)	2007, 2012	E-Commerce
Survey of Manufacturing Technology (SMT)	1988, 1991, 1993	Software, Automation, Robotics

# Table 1: Technology Information in Surveys Conducted by the U.S. Census Bureau\*

\* Note: The ABS, BRDIS, and BRDI-M surveys are conducted by the U.S. Census Bureau for the National Center for Science and Engineering Statistics (NCSES), which provides funding as well as expertise on the surveys' subject matter, content, and sampling strategy.

Source (Authors)	Years	Observation Count/Type	Topics
Helper, Seamans, Reichensperger, and Bessen	2018	Ongoing/Establishment	Automation, Robotics, Data Processing
Nancy Green Leigh	2017	428/Establishment	Automation, Robotics
Harte Hanks	1996-2015	160,000/Establishment	ICT, Software, Cloud Services
Susan Helper	1989	286/Firm	Automation
Deloitte - State of AI in the Enterprise	2018	1,100/Firm	AI
Narrative Science in Partnership with National Business Research Institute	2017	197/Firm	AI
Accenture – Technology Vision	2019	6,672/Firm	AI
McKinsey Digital Manufacturing Global Expert Survey	2018	<700/Firm	Digital Manufacturing
McKinsey – Global Lighthouse Network	2018-2019	44/Site	Business processes, management for manufacturing establishments that have scaled "4 <sup>th</sup> industrial revolution" solutions

 Table 2: Relevant Other Surveys on Technology Use

#### Table 3: Response Rates and Sample Construction\*

	# of Responses
Initial Mailout (June 2018)	850,000
Response <sup>55</sup>	583,000
Linked to 2017 Longitudinal	573,000
Business Database (LBD) <sup>56</sup>	

<sup>&</sup>lt;sup>55</sup> Survey response is determined by whether the respondent answered question 1 of the survey, which asks whether the business has ceased operations. Note that there are several instances and sections of the survey where responses are "Missing" or left blank. These are classified as "item non-response". If the firm ceased operations at the time of the survey and are matched to the LBD, we retain the records for that firm and that firm is still included in our main sample.

<sup>&</sup>lt;sup>56</sup> The match rate between the ABS and 2017 LBD is not perfect, due to a variety of unforeseen reasons, including that the ABS firm transitioned to a non-employer firm (no payroll), firm exits, merger and acquisition activity and more.

#### Table 4: Summary Statistics and Distributions of ABS Respondents\*

#### a. Firm Level Statistics

Mean	ABS Sample (Raw)	ABS-LBD Sample (Weighted)	National (2017 BDS)
Employment	89.32	26.28	24.24
Age	16.33	15.61	n.a.

#### **b.** Firm Distributions (in %)

By Size	ABS Sample (Raw)	ABS-LBD Sample (Weighted)	National (2016 BDS)**
1 to 9	67	75	76
10 to 49	21	20	20
50 to 249	8	4	4
250+	3	1	1

	ABS Sample	ABS-LBD Sample	National
By Age (in %)	(Raw)	(Weighted)	(2017 BDS)
0 to 5	25	27	33
6 to 10	16	17	16
11 to 20	25	25	23
21+	33	31	28

#### c. Sectoral Distribution (in %)

	ABS Sample	ABS-LBD Sample	National
Sector	(Raw)	(Weighted)	(2017 BDS)
Agriculture, Mining, Utilities	2	1	0
Construction	10	11	11
Education	1	2	2
Finance, Insurance, Real Estate	10	9	9
Health Care	9	11	11
Information	2	1	1
Management & Administrative	5	6	6
Manufacturing	8	4	4
Other (Arts, Food, Other Services)	14	23	23
Professional Services	17	13	13
Retail Trade	13	11	11
Transportation & Warehousing	4	3	3
Wholesale Trade	5	5	5

\*Note: Tables tabulated from linked 2018 ABS data with the 2017 Longitudinal Business Database (LBD). The 2017 size, age and industry figures from the LBD are the figures listed in the tables. Firms that did not respond to any of the 2018 ABS survey are excluded. Industry tabulations for multi-unit firms are generated from the largest payroll industry within the firm (if there is a tie, then the largest employer is used).

\*\*The firm size categories reported in Table 4b were created based on size categories used in a previous (legacy) BDS. With the release of the 2017 and 2018 BDS came updated size categories which make it impossible to compare the publicly available size categories to the size categories used here. For this reason, we compare to the legacy 2016 BDS instead of the new 2017 BDS.

	Business Function		
Sector	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>
Agriculture, Mining, Utilities	Financial	Personnel	Production
Construction	Financial	Personnel	Marketing
Manufacturing	Financial	Personnel	Production
Wholesale Trade	Financial	Personnel	Marketing
Retail Trade	Financial	Personnel	Marketing
Transportation & Warehousing	Financial	Personnel	Marketing
Information	Financial	Personnel	Marketing
Finance, Insurance, Real Estate	Financial	Personnel	Marketing
Professional Services	Financial	Personnel	Marketing
Management & Administrative	Financial	Personnel	Marketing
Education	Financial	Personnel	Marketing
Health Care	Financial	Personnel	Marketing
Other (Arts, Food, Other)	Financial	Personnel	Marketing

#### Table 5: Top Use Digitized Information by Business Function by Sector

Notes: "Use" is defined as having responded with "Up to 50%", "More than 50%" or "All" for the information category listed on "Digital Share of Business". Shares are computed using the LBD tabulation weights of firm counts, divided by the total number of firms (including those that left the responses as "Don't Know" or missing). Imputed responses for "Missing" and "Don't Know" categories are used in the numerator.

<b>Business Function</b>	% Use	% Intensive Use
Financial	83.8 (87.5)	70.8 (73.7)
Personnel	72.0 (76.2)	53.2 (56.4)
Marketing	61.2 (67.8)	44.1 (49.4)
Feedback	53.4 (59.9)	38.5 (43.6)
Production	46.4 (54.5)	36.4 (42.3)
Supply Chain	44.6 (53.8)	32.6 (39.0)
Other	14.2 (19.4)	11.7 (15.9)

#### Table 6: Use of Digital Information by Business Function – Conditional

Notes: "Use" is defined as having responded with "Up to 50%", "More than 50%" or "All" for the information category listed on "Digital Share of Business". "Intensive Use" is defined as having responded with "More than 50%" or "All". Shares are computed using the LBD tabulation weights of firm counts, divided by the total number of firms, conditional on collecting and storing the business function (including those that left the responses as "Don't Know" or "Missing" whose responses were imputed). Listed shares are imputed shares, with raw weighted values in parentheses.

Function 1	Function 2	Correlation
Financial	Personnel	0.695 (0.737)
Marketing	Feedback	0.678 (0.678)
Supply Chain	Production	0.597 (0.598)
Supply Chain	Marketing	0.516 (0.521)
Supply Chain	Feedback	0.480 (0.484)

 Table 7: Top Pairwise Correlated Digitized Information Use by Business Functions

Notes: "Use" is defined as having responded with "Up to 50%", "More than 50%" or "All" for the information category listed on "Digital Share of Business". Correlation is defined as cross-category responses (Use/No Use) at the firm-level. Imputed values are listed, while raw weighted values are in parentheses.

		Firm Size		
Firm Age	1 to 9	10 to 49	50 to 249	250+
0 to 5	0.897 (0.871)	0.963 (0.942)	0.980 (0.960)	0.974 (0.924)
6 to 10	0.866 (0.835)	0.956 (0.939)	0.973 (0.954)	0.987 (0.966)
11 to 20	0.907 (0.880)	0.957 (0.935)	0.971 (0.957)	0.987 (0.984)
21+	0.797 (0.760)	0.951 (0.931)	0.980 (0.965)	0.980 (0.964)

Table 8: Size-Age Coefficients for Digital Share of Business Activity

Notes: Size-Age coefficients generated from linear probability model (LPM) where the outcome variable is "Use/No Use" for at least one type of information that is digitized. Respondents who answered with "Do not collect this information" for *all* of the information types are excluded. Independent variables are the 16 size-age categories assigned to each firm and the LPM is weighted by the LBD tabulation weights. All coefficients are significant to the 0.1%. All firms are included (including "Don't Know" and missing) with the dependent variable being imputed for firms whose responses are "Missing" or "Don't Know" (with raw weighted values in parentheses).

Cloud Service Purchased	% Use	% Intensive Use
Billing	50.8 (56.6)	33.8 (37.5)
Security	49.7 (58.1)	31.2 (36.3)
Synchronization	46.5 (53.6)	23.6 (27.2)
All IT	45.8 (54.0)	25.4 (29.9)
Data Storage	44.1 (51.7)	26.1 (30.5)
Servers	41.6 (48.5)	26.0 (30.1)
Customer Relations	37.6 (44.0)	22.4 (26.1)
Data Analysis	30.8 (37.6)	18.0 (22.0)
Other	8.7 (11.8)	5.9 (8.0)

#### **Table 9: Cloud Service Purchases by IT Function - Conditional**

Notes: "Use" is defined as having responded with "Up to 50%", "More than 50%" or "All" for the category listed on "Cloud Service Purchases". "Intensive Use" is defined as having responded with "More than 50%" or "All". Shares are computed using the LBD tabulation weights of firm counts, divided by the total number of firms conditional on the firm performing the IT function (including those that left the responses as "Don't Know" or missing, whose values were imputed). Listed shares are imputed shares, with raw weighted values in parentheses.

Table 10, 100 I all white Colleanons allong Cloud Del vice I di chases
--

Technology 1	Technology 2	Correlation
Servers	Security	0.727 (0.731)
Security	All IT	0.721 (0.716)
Servers	Data Storage	0.666 (0.669)
Servers	All IT	0.661 (0.659)
Data Storage	All IT	0.661 (0.664)

Notes: "Use" is defined as having responded with "Up to 50%", "More than 50%" or "All" for the category listed on "Cloud Service Purchases". Correlation is defined as cross-category responses (Use/No Use) at the firm level. Imputed values are listed, while raw weighted values are in parentheses.

	Cloud Services		
Sector	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>
Agriculture, Mining,			
Utilities	Billing	Security	Synchronization
Construction	Billing	Security	Synchronization
Manufacturing	Security	Billing	Synchronization
Wholesale Trade	Security	Billing	Synchronization
Retail Trade	Billing	Security	Synchronization
Transportation &			
Warehousing	Billing	Security	Synchronization
Information	Synchronization	All IT	Billing
Finance, Insurance, Real			
Estate	Security	Billing	Synchronization
Professional Services	Synchronization	Security	All IT
Management &			
Administrative	Billing	Security	Synchronization
Education	Synchronization	Billing	Security
Health Care	Billing	Security	All IT
Other (Arts, Food, Other)	Billing	Security	All IT

#### Table 11: Top Use Categories for Cloud Services by Sector

Notes: "Use" is defined as having responded with "Up to 50%", "More than 50%" or "All" for the category listed on "Cloud Service Purchases". Shares are computed using the LBD tabulation weights of firm counts, divided by the total number of firms (including those that left the responses as "Don't Know" or missing). Imputed responses for "Missing" and "Don't Know" categories are used in the numerator.

	Firm Size			
Firm Age	1 to 9	10 to 49	50 to 249	250+
0 to 5	0.744 (0.698)	0.867 (0.825)	0.891 (0.849)	0.878 (0.822)
6 to 10	0.689 (0.640)	0.842 (0.800)	0.874 (0.828)	0.912 (0.886)
11 to 20	0.754 (0.706)	0.829 (0.785)	0.852 (0.811)	0.886 (0.854)
21+	0.579 (0.525)	0.796 (0.752)	0.855 (0.822)	0.852 (0.814)

**Table 12: Size-Age Coefficients for Cloud Service Purchases** 

Notes: Size-Age coefficients generated from linear probability model (LPM) where the outcome variable is "Use/No Use" for at least one type of cloud service purchase. Respondents who answered with "Do not use IT function" for *all* of the cloud purchase types are excluded. Independent variables are the 16 size-age categories assigned to each firm and the LPM is weighted by the LBD derived tabulation weights. All coefficients are significant to the 0.1%. All firms are included (including "Don't Know" and missing) with the dependent variable being imputed for firms whose responses are "Missing" or "Don't Know" (with raw weighted values in parentheses).

<b>Business Technology</b>	% Use	% Testing
Touchscreens	6.1 (6.0)	0.9 (0.9)
Machine Learning	2.9 (2.9)	0.7 (0.7)
Voice Recognition	2.5 (2.6)	0.7 (0.7)
Machine Vision	1.8 (1.8)	0.4 (0.4)
Robotics	1.3 (1.4)	0.3 (0.3)
Natural Language	1.3 (1.3)	0.4 (0.4)
RFID	1.1 (1.1)	0.3 (0.3)
Augmented Reality	0.8 (0.8)	0.4 (0.4)
Automated Vehicles	0.8 (0.8)	0.2 (0.2)

#### Table 13: Business Technology Use Rates by Type

Notes: "Use" is defined as having responded with "In use for less than 5% of production or service", "In use for between 5% - 25% of production or service" or "In use for more than 25% of production or service" for the category listed on "Business Technologies" (excluding "Automated Storage and Retrieval Systems"). "Testing" is defined as having responded with "Testing but not using in production or service". Shares are computed using the LBD tabulation weights of firm counts, divided by the total number of firms (including those that left the responses as "Don't Know" or missing). Listed shares are imputed shares, with raw weighted values in parentheses.

Table 14: Top P	Pairwise Co	orrelations among	<b>Business</b> '	Technologies
-----------------	-------------	-------------------	-------------------	--------------

Technology 1	Technology 2	Correlation
Machine Learning	Machine Vision	0.522 (0.516)
Automated Vehicles	Augmented Reality	0.491 (0.489)
Machine Vision	Natural Language	0.398 (0.395)
RFID	Automated Vehicles	0.389 (0.386)
Machine Vision	Automated Vehicles	0.383 (0.380)

Notes: Correlations are for whether a firm lists "Use" for a technology category. "Use" is defined as having responded with "In use for less than 5% of production or service", "In use for between 5% - 25% of production or service" or "In use for more than 25% of production or service" for the category listed on "Business Technologies" (excluding "Automated Storage and Retrieval Systems"). Imputed values are listed, while raw weighted values are in parentheses.

		<b>Business Technology</b>	
Sector	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>
			Automated
Agriculture, Mining, Utilities	Touchscreens	Machine Learning	Vehicles
Construction	Touchscreens	Machine Learning	Voice Recognition
Manufacturing	Machine Learning	Robotics	Touchscreens
Wholesale Trade	Touchscreens	Machine Learning	RFID
Retail Trade	Touchscreens	Machine Learning	RFID
Transportation & Warehousing	Touchscreens	Machine Learning	RFID
Information	Touchscreens	Machine Learning	Voice Recognition
Finance, Insurance, Real Estate	Touchscreens	Voice Recognition	Machine Learning
Professional Services	Touchscreens	Voice Recognition	Machine Learning
Management & Administrative	Touchscreens	Machine Learning	Voice Recognition
Education	Touchscreens	Machine Learning	Voice Recognition
Health Care	Touchscreens	Voice Recognition	Machine Learning
Other (Arts, Food, Other)	Touchscreens	Machine Learning	Machine Vision

Notes: "Use" is defined as having responded with "In use for less than 5% of production or service", "In use for between 5% - 25% of production or service" or "In use for more than 25% of production or service" for the category listed on "Business Technologies" (excluding "Automated Storage and Retrieval Systems"). Shares are computed using the LBD tabulation weights of firm counts, divided by the total number of firms (including those that left the responses as "Don't Know" or missing). In this scenario, "Use" for a business technology includes "Testing". Imputed responses for "Missing" and "Don't Know" categories are used in the numerator.

Augmer	ited Reality	Mean (All Industries)	0.010	(0.007)
5121	Motion Picture and Video Industries	· · · · · · · · · · · · · · · · · · ·	0.042	(0.035)
5112	Software Publishers		0.032	(0.025)
5414	Specialized Design Services		0.028	(0.023)
Automa	ted Guided Vehicles	Mean (All Industries)	0.009	(0.007)
1151	Support Activities for Crop Production		0.066	(0.059)
4245	Farm Product Raw Material Merchant Wholesalers		0.041	(0.039)
2379	Highway, Street and Bridge Construction			(0.028)
Machin	e Learning	Mean (All Industries)	0.034	(0.025)
3335	Metalworking Machinery Manufacturing		0.127	(0.108)
3327	Machine Shops; Turned Products; Screw, Nut and Bo	olt Manufacturing	0.119	(0.105)
5112	Software Publishers	-	0.104	(0.083)
Machin	e Vision	Mean (All Industries)	0.021	(0.016)
3344	Semiconductor and Other Electronic Component Ma	nufacturing	0.142	(0.123)
3335	Metalworking Machinery Manufacturing		0.115	(0.098)
3363	Motor Vehicle Parts Manufacturing		0.107	(0.084)
Natural Language ProcessingMean (All Industries)			0.015	(0.011)
5112	Software Publishers		0.065	(0.049)
5191	Other Information Services		0.058	(0.046)
5182	Data Processing, Hosting, and Related Services		0.052	(0.042)
RFID		Mean (All Industries)	0.016	(0.011)
4931	Warehousing and Storage		0.064	(0.056)
4248	48 Beer, Wine, and Distilled Alcoholic Beverage Merchant Wholesalers		0.061	(0.048)
3363	Motor Vehicle Parts Manufacturing		0.061	(0.045)
Robotic	S	Mean (All Industries)	0.019	(0.014)
3363	Motor Vehicle Parts Manufacturing		0.189	(0.158)
3261	Plastics Product Manufacturing		0.176	(0.151)
3335	Metalworking Machinery Manufacturing		0.145	(0.128)
Touchso	ereens	Mean (All Industries)	0.066	(0.049)
6231	Nursing Care Facilities (Skilled Nursing Facilities)		0.212	(0.155)
3121	Beverage Manufacturing		0.172	(0.147)
7139	Other Amusement and Recreation Industries		0.159	(0.125)
Voice R	Voice RecognitionMean (All Industries)		0.028	(0.021)
6215	Medical and Diagnostic Laboratories		0.183	(0.154)
6211	Offices of Physicians		0.144	(0.122)
6214	Outpatient Care Centers		0.087	(0.072)

#### Table 16: Top 3 Industry Use Rates for Each Business Technology

Notes: "Use" is defined as having responded with "In use for less than 5% of production or service", "In use for between 5% - 25% of production or service" or "In use for more than 25% of production or service " for the category listed on "Business Technologies" (excluding "Automated Storage and Retrieval Systems"). Shares are computed using the LBD tabulation weights of firm counts, divided by the total number of firms (including those that left the responses as "Don't Know" or missing). Raw weighted values are in parentheses. Means generated from cross-industry means. In this scenario, "Use" for a business technology includes "Testing". Imputed responses for "Missing" and "Don't Know" categories are used in the numerator.

	Firm Size			
Firm Age	1 to 9	10 to 49	50 to 249	250+
0 to 5	0.098 (0.094)	0.232 (0.210)	0.267 (0.224)	0.347 (0.262)
6 to 10	0.076 (0.073)	0.216 (0.203)	0.258 (0.235)	0.334 (0.281)
11 to 20	0.118 (0.113)	0.194 (0.179)	0.244 (0.224)	0.330 (0.271)
21+	0.062 (0.059)	0.182 (0.171)	0.329 (0.312)	0.379 (0.356)

Table 17: Size-Age Coefficients for Business Technologies

Notes: Size-Age coefficients generated from linear probability model (LPM) where the outcome variable is "Use/No Use" for at least one type of business technology. Independent variables are the 16 size-age categories assigned to each firm and the LPM is weighted by the LBD tabulation weights. All coefficients are significant to the 0.1%. All firms are included (including "Don't Know" and missing) with the dependent variable being imputed for firms whose responses are "Missing" or "Don't Know" (with raw weighted values in parentheses).

Business Technology	Most Correlated Digital Information	Correlation	Most Correlated Cloud Service	Correlation
Dusiness Teemology				
Augmented Reality	Other Information	0.086 (0.083)	Other IT Functions	0.108 (0.105)
Automated Vehicles	Other Information	0.081 (0.079)	Other IT Functions	0.104 (0.102)
Machine Learning	Production	0.127 (0.129)	Data Analysis	0.148 (0.147)
Machine Vision	Production	0.114 (0.114)	Data Analysis	0.124 (0.122)
Natural Language	Other Information	0.084 (0.083)	Data Analysis	0.116 (0.114)
RFID	Supply Chain	0.101 (0.099)	Data Analysis	0.103 (0.101)
Robotics	Production	0.108 (0.107)	Data Analysis	0.086 (0.085)
Touchscreens	Feedback	0.165 (0.171)	Data Analysis	0.168 (0.17)
Voice Recognition	Feedback	0.092 (0.098)	All IT Functions	0.122 (0.124)

# **Table 18: Technological Complementarities with Business Technologies**

Notes: Correlations are for whether a firm lists "Use" for a technology category with "Use" in another technology category (Digitization or Cloud Computing). In this scenario, "Use" for a business technology includes "Testing". Imputed values are listed, while raw weighted values are in parentheses.

Business Technology	% Use (Tab- Weighted)	% Use (Employment- Weighted)	Difference Ratio
Touchscreens	6.1 (4.6)	25.4 (13.6)	4.2 (2.9)
Machine Learning	2.9 (2.2)	8.9 (5.2)	3.1 (2.4)
Voice Recognition	2.5 (2)	7.5 (5.9)	3 (3)
Machine Vision	1.8 (1.4)	5.6 (3.1)	3.2 (2.3)
Robotics	1.3 (1.0)	10.4 (6.4)	7.8 (6.1)
Natural Language	1.3 (1.0)	4.3 (3.5)	3.3 (3.5)
RFID	1.1 (0.9)	9.6 (4.9)	8.4 (5.6)
Augmented Reality	0.8 (0.6)	2 (1.4)	2.4 (2.2)
Automated Vehicles	0.8 (0.6)	2.2 (1.6)	2.7 (2.5)

 Table 19: Firm-Weighted versus Employment-Weighted Adoption Rates for Business

 Technologies

Notes: "Use" is defined as having responded with "In use for less than 5% of production or service", "In use for between 5% - 25% of production or service" or "In use for more than 25% of production or service" for the category listed on "Business Technologies" (excluding "Automated Storage and Retrieval Systems"). "Testing" is defined as having responded with "Testing but not using in production or service". Shares are computed using the LBD tabulation weights of firm counts, divided by the total number of firms (including those that left the responses as "Don't Know" or missing). Employment weights are combined with the LBD tabulation weights and the difference ratio is computed by dividing the Employment Weighted Use rate by the LBD Tabulation-Weighted Use rate. Imputed values are listed, while raw weighted values are in parentheses.





# Digital Share of Business Activity

\*Note: Tabulations based on unweighted and unimputed responses.

# Figure 2: Extensive and Intensive Margin Measures of Digitized Information by Sector -Conditional



Firms with Over Half of Information in Digital Format

Firms with Information in Digital Format

Notes: "Use" is defined as having responded with "Up to 50%", "More than 50%" or "All" for the information category listed on "Digital Share of Business". "Intensive Use" is defined as having responded with "More than 50%" or "All". Shares are computed using the LBD tabulation weights of firm counts, divided by the total number of firms that did not respond "Do Not Collect This Information" for *all* information types (including those that left the responses as "Don't Know" or missing). Sectors are defined by 13 aggregated LBD 2-digit NAICS codes and assigned for multi-unit firms based on the industry with the largest payroll share within the firm. Original responses classified as "Missing" or "Don't Know" are imputed.



# Figure 3: Firm Responses to Purchases of Cloud Service

Note: Tabulations based on unweighted and unimputed responses.

# Figure 4: Extensive and Intensive Margin Measures of Use Rates for Cloud Service Purchases by Sector - Conditional



Firms Intensively Using Cloud Services for IT Functions

Firms Using Cloud Services for IT Functions

Notes: "Use" is defined as having responded with "Up to 50%", "More than 50%" or "All" for the category listed on "Cloud Service Purchases". "Intensive Use" is defined as having responded with "More than 50%" or "All". Shares are computed using the LBD tabulation weights of firm counts, divided by the total number of firms that did not respond "Did not use this IT function" for *all* cloud purchase types (including those that left the responses as "Don't Know" or missing). Sectors are defined by combined 2-digit NAICS and assigned for multi-unit firms by largest payroll industry by firm. Original responses classified as "Missing" or "Don't Know" are imputed.



# Figure 5: Firm Responses to Business Technologies

Note: Tabulations based on unweighted and unimputed responses.





a. Testing-to-Use Ratios for all Business Technologies (All Firms)



b. Testing-to-Use Ratios for all Business Technologies (By Size)





c. Testing-to-Use Ratios for all Business Technologies (By Age)

d. Testing-to-Use Ratios for all Business Technologies (By Manufacturing Status)

Notes: "Use" is defined as having responded with "In use for less than 5% of production or service", "In use for between 5% - 25% of production or service" or "In use for more than 25% of production or service" for the categories listed on "Business Technologies" (excluding "Automated Storage and Retrieval Systems"). "Testing" is defined as having responded with "Testing but not using in production or service". Bubble size is determined by number of firms who respond to "Use" for the listed technology. Categories are sorted by use rates for large firms in panel b., old firms in panel c. and manufacturing firms in panel d. All ratios here are calculated using imputed response for "Missing" responses.

# Figure 7: Extensive and Intensive Margin Measures of Use and Testing Rates for Business Technologies by Sector



Notes: "Use" is defined as having responded with "In use for less than 5% of production or service", "In use for between 5% - 25% of production or service" or "In use for more than 25% of production or service " for the category listed on "Business Technologies" (excluding "Automated Storage and Retrieval Systems"). "Testing" is defined as having responded with "Testing but not using in production or service". Sectors are defined by combined 2-digit NAICS and assigned for multi-unit firms by largest payroll industry by firm. Original responses classified as "Missing" are imputed.



# Figure 8: AI Usage across size Categories (Weighted)

Notes: The figures visually represent the share of firms (weighted by survey-LBD weights) that indicate use or test at least 1 of the following business technologies: Automated Guided Vehicles, Machine Learning, Machine Vision, Natural Language Processing and Voice Recognition. Values are based on the imputed probabilities for respondents who answered "Missing" or "Don't Know" to 1 or more of the aforementioned business technologies.
## Figure 9: Technological Hierarchies<sup>57</sup>



a. Sankey Diagram for Business Technologies (Any)



#### b. Sankey Diagram for Machine Learning

<sup>&</sup>lt;sup>57</sup> The Sankey diagrams visually represents firm counts as they progress from no technology adoption to business technology and machine learning technology adoption. The size of the grey area is representative of the number of firm counts progressing to the next stage. Note that the calculations are made using imputed responses for "Missing".

#### **Figure 10: Technology Adoption and Innovation**



#### a. Technology Adoption and Innovation Outcomes



#### b. Technology Adoption and Innovation Outcomes (Any)

Notes: The figures visually represent the coefficients from an LPM where the dependent variable is a 1/0 for whether the firm conducted a Product or Process Innovation (defined as having responded "Yes" to one of the categories listed under Product or Process Innovation in the ABS). Panel b. considers the subset of firms who use a technology category (i.e. Digitization or Cloud Computing) and looks at how the individual subcategories is associated with any type of innovation (either product or process). The independent variables include the technology use categories, along with firm size and age controls, and industry fixed effects. All coefficient plots are calculated using imputed independent variables for "Missing".

## Appendix: A1. Cognitive Testing First Week

The second round of cognitive testing for the ABS, which was the first set of interviews to include the three technology questions, occurred in Nashville in mid-August. Nine companies participated in the testing. Each was given three questions relating to its use of digital information, cloud services, and business automation technologies. Two versions of Question 1 were tested to determine both the most preferred wording (i.e., "digital" vs. "digitized"), as well as the best method to measure the use of digitized data. Version A used the term "digitized" and asked what *shares* of different types of data were digitized; and version B used the term "digital" and asked how the business *changed* its availability of different types of digital data. Question 2 asked about the extent to which the establishment purchased cloud services for various IT functions. Question 3 asked about how intensively the business used various technologies (mainly relating to automation; e.g., automated guided vehicles (AGVs), machine learning, robotics, etc.). The column choices for each of the three questions included both qualitative (i.e., "slight," "moderate," "extensive," etc.) and quantitative (i.e., percentage range) measures of technology use. Appendix A contains the original versions of the questions tested.

Results from testing Question 1 showed that respondents generally preferred version A (i.e., they preferred being asked about *how much* of their data was digitized, as opposed to being asked about how its availability changed), but they also typically preferred the word "digital" over "digitized." Consequently, the version A measure of digitized data use was recommended, replacing the term "digitized" with "digital," and including a definition of digital data. Including examples of personnel data to clarify its meaning was also recommended. For Question 2, respondents overall understood the concept of cloud services and were able to answer confidently whether they used them. In some cases, however, respondents either were not sure in which category to include a service or did not know whether the cloud service was free or purchased. Recommendations for this question included adding a clearer definition of cloud services and including examples of name-brand products in applicable categories. In answering Question 3, many of the technological terms were unfamiliar to some respondents, however they were generally able to answer confidently and accurately regarding the technologies they *did not* 

74

use. For technologies that *were* utilized by participant firms, respondents occasionally misunderstood the question's intended scope (e.g., employees using touchscreen technology in settings other than labor substitution). Here, the sole recommendation was to include examples that would clarify the intended scope of the question.

#### Second Week

The second set of interviews for the technology questions took place in Boston in late-August. Six companies included in the BRDIS were given revised versions of the three technology questions (see Text Box A2). Question 1 now had a single version, asking about the share of different types of "information" (term substituted for "data") that was "digital" (term substituted for "digitized"). The columns no longer included qualitative measures, but rather solely percentage ranges. The percentage ranges themselves were also adjusted, and a "Not Applicable" column was added. Similarly, the revision to Question 2 removed all qualitative descriptors and asked simply for percentage ranges of cloud services shares (with respect to spending). For clarity, the revision also included product examples for the "Data storage and management" and "Collaboration and file synchronization" categories. In revising Question 3 the phrase "for customer interface," along with some examples, was added for clarification to the "Touchscreens/kiosks" category. Also, unlike the previous two questions, the revision of Question 3 removed the quantitative descriptors (i.e., percentage ranges) of automation technology use, in addition to adding a column with "Not currently using, but planning future use."

The responses to the revised version of Question 1 seemed generally more confident. With the revised wording, the overall understanding of the question was good. Almost all respondents easily answered, "More than 50%" for every category, and many stated confidently that simply "Everything is digital." The revised version of Question 2 was better understood than the original, however all but one individual responded that they would need to consult their firm's IT group to accurately answer the question. The revised version of Question 3 was relatively difficult for respondents to answer. The fact that there were only qualitative descriptions to measure overall use of different technologies made it unclear how to answer the questions if, for example, a technology is used extensively in one part of the company but little or not at all in another.

75

## **Text Box A1: First Version of the Technology Questions**

	Data Not	Slightly Digitized	Moderately	Heavily	Entirely
	Digitized	(less than 5% of	Digitized (6-50%	Digitized (50-	Digitized
		this kind of data)	of this kind of data)	99% of this kind	
				of data)	
Personnel					
Financial					
Customer					
feedback					
Marketing					
Supply chain					
Production					
Other (write-in)					

**1A**) What share of each of the following types of data is digitized at this business?

**1B**) During the three years from 2015 to 2017, to what extent did this business change the availability of digital data of each of the following types?

	Digital Data Not Used	Decreased Availability	No Change in Availability	Slight Increase in Availability	Moderate Increase in Availability	Extensive Increase in Availability
Personnel						
Financial						
Customer						
feedback						
Marketing						
Supply						
chain						
Production						
Other (write-in)						

**2)** In 2017, to what extent did this business purchase **cloud services** for the following information technology (IT) functions?

	Not	Did not	Slight	Moderate	Intensive
	Used	purchase	Use	Use	Use
		as cloud	(less than	(6-50% of	(More than
		service	5% of	spending for	50% of
			spending	this	spending
			for this	function)	for this
			function)		function)
Security or firewall					
Servers					
Data storage and management					
Collaboration and file					
synchronization					
Data analysis					
Billing and account management					
Customer relationship management					
Other (write-in)					

3) In 2017, to what extent did this business use the following technologies?

	Not Used	Slight	Moderate Use	Intensive Use
		Use (Piloting	(In use for	(In use for
		or using in	between 5-	more than
		less than 5%	25% of	25% of
		of production	production or	production or
		or service)	service)	service)
Augmented reality				
Automated guided vehicles (AGV)				
or AGV system				
Automated storage and retrieval				
systems				
Cloud-based servers, storage and				
data management				
Machine learning				
Machine vision software				
Natural language processing				
Radio-frequency identification				
(RFID) inventory system				
Robotics				
Touchscreens/kiosks				
Voice recognition software				

#### **Text Box A2: Revised Version of the Technology Questions**

Question 1

In 2017, what share of each of the following types of information was **digital** at this business?

#### Mark one box for each row

	Data not	Less than	More than 25%,	More than	Not
	digital	25%	but less than 50%	50%	Applicable
Personnel					
Financial					
Customer feedback					
Marketing					
Supply chain					
Production					
Other (write-in)					

#### Question 2

In 2017, what share of spending on the following information technology (IT) functions at this business was used to purchase **cloud services**?

Cloud services are services provided by a third party that this business accesses on-demand via the internet.

#### Mark one box for each row

	Did not	Less than	More than	More	Not
	purchase as	25%	25%, but less	than 50%	applicable
	cloud service		than 50%		••
Security or firewall					
Servers					
Data storage and management					
(Examples: Amazon Web					
Services, IBM Bluemix, Microsoft					
Azure)					
Collaboration and file					
synchronization (Examples:					
Dropbox, OneDrive, Google					
Drive)					
Data analysis					
Billing and account management					
Customer relationship					
management					
Other (write-in)					

# Question 3

In 2017, to what extent did this business use the following technologies?

### Mark one box for each row

	Not currently	Some	Limited use	Extensive use	Not
	using, but	experimental	in production	in production	applicable
	planning future	use	or services	or services	•••
	use				
Augmented reality					
Automated guided vehicles					
(AGV) or AGV systems					
Automated storage and					
retrieval systems					
Cloud-based servers, storage					
and data management					
Machine learning					
Machine vision software					
Natural language processing					
Radio-frequency					
identification (RFID)					
inventory system					
Robotics					
Touchscreens/kiosks for					
customer interface					
(Examples: self-checkout,					
touchscreen ordering)					
Voice recognition software					

# Table A3: Summary of Recommended Changes based on Cognitive Testing

Section	Question	Recommendation	Accepted?	Notes
Business Structure	For the person(s) owning the largest percentage(s) in this business in 2017, please list the percentage owned by each person and his or her name.	Put the column designated for a name before the column for percent ownership.	Yes	
	In 2017, did two or more members of one family own the majority of this business?	Consider including "step" relationships to this definition.	No	
	Did spouses/unmarried partners jointly own this business; Was this business operated equally by both spouses/unmarried partners?	Consider specifying what time frame is relevant to the questions.	Yes	
Owner Characteristics	What was the highest degree or level of school Owner X completed prior to establishing, purchasing, or acquiring this business?	For individuals who have received an associate degree, it may be helpful to explicitly state that associate degrees are to be excluded.	Yes	
	Prior to establishing, purchasing, or acquiring this business, what was the field of the highest degree completed for Owner X?	Consider changing or adding to "field of the highest degree completed" to include "major" or some other simplifying term.	No	
	How important to Owner X are each of the following reasons for owning this business?	Consider additional response categories to include: "carrying on the family business," and "Helping and/or becoming more involved in my community."	Yes	Both suggested categories were included.

	Does Owner X have any of the following difficulties?	Add response category for "None of the Above" to differentiate between having no disability and item nonresponse. Additionally, specifically mention "even when using a hearing aid" for the hearing difficulty response category.	N/A	This question was not included in the final version.
Other Business Characteristics	In 2017, which of the following types of workers were used by this business? Select all that apply.	Consider adding category for workers who receive a commission, seasonal employees, workers who are on call/demand, etc.	No	
	In 2017, did this business use any of the following to promote or conduct business?	Since participants selected the option for "company website" even if they didn't use it to promote or conduct business, perhaps the words "promote" and "conduct" should be in bold print or underlined to emphasize what this question is asking.	N/A	This question was not included in the final version.
Business Financing	For 2017, what was the total amount of money that the owner(s) personally put into the business? Your best estimate is fine.	Due to the fact that there was some confusion regarding whether to report the amount that each owner put in to the company, or to sum those amounts, it may be beneficial to bold and/or italicize the word "total" in the question.	No	
	In 2017, did this business attempt to establish any new funding relationships (for example, loans, investments, or gifts) with any of the following sources?	Specify whether new financing with a bank the owner already has a relationship with would still qualify as an affirmative response. Consider adding a definition for the term "crowdfunding."	N/A	This question was not included in the final version.

For 2017, what was the total amount of money this business received from angel investors, venture capitalists, or other businesses in return for a share of ownership in this business?	Consider adding a definition for the term "angel investor."	Yes	
For 2017, what was the total amount of money this business borrowed from a bank or other financial institutions, including business loans, a business credit card carrying a balance, or a business line of credit? Include all draws on a business line of credit, even if paid off during the year.	This question may benefit by having a list for respondents to reference of specifically what to include and exclude in their response.	No	
For 2017, what was the total amount of money this business received from family, friends, and employees?	Instead of "total amount of money" consider using "capital or investment funds" from family etc.	Yes	The term "investment funds" was chosen.
At any time during 2017, did this business need additional financing?	Perhaps consider moving this question sooner if including the previous question or specify not to include what was already covered by the previous question if that is the intent of this question.	No	

Business	For 2017, which of the following	Prior to asking this question, consider	No	
Performance	negatively impacted the profitability of	asking respondents first a yes/no question		
	this business?	regarding whether they experienced any		
		negative impacts on their profit.		
		Regarding the subjectivity of the question,		
		it may help to include a qualifier stating	Yes	
		something akin to "please only include		
		factors that impacted profit."		

## Appendix A.2 – Imputation of Item Non-Response and "Don't Know"

This section describes the steps and intuition for our imputation of item non-response and "Don't Know" responses to Technology questions Q1, Q2 and Q3. For further details on the methodology, please contact the authors.

## A.2.1 Measuring Item Non-Response and "Don't Know" Across Technology Questions

A primary contribution for the paper is to develop a nationally representative set of technology adoption and use measures based on the survey results, which in public use tabulations report aggregate response counts for each technology question (see public use tabulations at: <a href="https://www.census.gov/data/tables/2018/econ/abs/2018-abs-digital-technology-module.html">https://www.census.gov/data/tables/2018/econ/abs/2018-abs-digital-technology-module.html</a> ). We extend the public use tabulations in two distinct ways: a.) we utilize our LBD-derived tabulation weights that corrects for the survey non-response and more closely matches the Business Dynamic Statistics, b.) We impute technology responses for firms with item non-response or for firms that answered, "Don't Know". We have demonstrated that the number of total respondents, combined with our LBD-derived tabulation weights gives us a reasonable sample that is representative of the economy.

Surveys suffer from non-response and item non-response. Survey non-response are those cases in which firms do not respond to the request. Survey non-response is typically handled by an adjustment of the survey weight resulting in new tabulation weights for those firms that did respond. Item non-response occurs when a survey respondent answers some, but not all, survey items. We address survey non-response by adjusting to national LBD totals as documented in Section 3. We address item non-response for the three technology questions here. We adopt a broader definition of item non-response and include those responses that include "Don't know" as well. Table A4 provides the rounded number of item non-response ("Missing") and "Don't Know" responses across each of the technology questions.

		Unweighted (Mean)	Weighted
	Missing	45,500	375,000
Digital Information	Don't Know	56,000	502,000
	Total	101,500	877,000
Cloud Purchases	Missing	46,000	378,000
	Don't Know	91,000	794,000
	Total	137,000	1,072,000
	Missing	46,000	378,000
Business Technologies	Don't Know	90,000	838,000
	Total	136,000	1,116,000

Table A4: Approximate Weighted and Unweighted Firm Counts of Item Non-Response and "Don't Know"

**Note:** The figures are authors' calculations. They consist of approximate means across all of the sub-categories for each technology question (see Figures 1, 3 and 5 for more exact figures across each of the sub-categories). "Missing" response options have very little variation across all of the sub-categories (range from 45-48,000 unweighted firm counts across all technologies), while "Don't Know" response options have more variation (between 40,000 and 105,000 unweighted firm counts across all technologies).

Across the three technology questions, "Missing" and "Don't Know" make up between 17-23% of our full set of responses. Excluding the firms from our analysis is problematic for multiple reasons: a.) it is not accepted Census Bureau procedure to adjust weights for item non-response, b.) the distribution of firms with "Missing" and/or "Don't Know" responses differ from the set of responders.

Excluding the firms from our analysis is problematic, as the distribution of firms with "Missing" and "Don't Know" responses differ from the remaining set of responses. In our analysis of the size, industry and age distribution of firms who respond with "Missing" or Don't Know, we find significant differences in the size distribution (but not age or sectoral distribution) relative to the full set of responses: firms who respond with "Missing" or "Don't Know" are significantly more likely to be large (250+ employees). These differences get more skewed at the top of the distribution (10,000+ employees). Finally, assigning each item non-response an arbitrary value (e.g. "no use" or "use") would bias our high-level estimates.

## A.2.2 Approach for addressing Item Non-Response and Don't Know

Our approach for managing the "Missing" and "Don't Know" responses relies on utilizing the existing responses for all of the technological categories, combined with a number of observed firm characteristics that we can gather from linking the ABS to the LBD. These include firm measures of size interacted with 2-digit NAICS indicator variables, age, multi-unit status, mean payroll per employee, responses to existing technology questions and geographic controls, taken as the location of the largest employment establishment for multi-unit firms.

We can then estimate an ordered probit<sup>58</sup> across the different use categories that specifically tell us whether or not a firm used a technology. For Q1 and Q2, these categorical options include:

- None
- Less than 50%
- More than 50%
- All
- Information not collected/IT function not used

While for Q3, the categorical response options include:

- None
- Testing
- Less than 5%
- 5 to 25%
- More than 25%

See Tables A5 - A7 (listed at the end of the section) to review the adoption patterns for these response options when we remove the "Missing" and "Don't Know" answers.

<sup>&</sup>lt;sup>58</sup> We also ran an ordered logit and obtained similar results

For each of the sub-technology categories, we estimate the LHS use category with our set of approximately 450,000 non-missing responses. We can then predict the probability for each of the LHS use categories for the set of "Missing" and "Don't Know" responses. More formally, our estimating equation is:

$$Y_{ij} = \sum_{k=1}^{12} SIZE_{ik} \times IND_i + \sum_{l=1}^{8} AGE_{il} + MU_i + \ln PAY_i + STATE_i + \varepsilon_{ij}$$
(1)

Where  $Y_{ij}$  is each of the response options for technology j,  $\sum_{k=1}^{12} SIZE_k \times IND_m$  are 12 employment size categories (based on the 2016 Business Dynamic Statistics size categories) interacted with a 2-digit NAICS industry indicator variable,  $AGE_l$  are 8 age indicator categories,  $PAY_i$  is the mean payroll per employee of firm *i*, and a state-level indicator variable for the location of the largest employment establishment within the firm. For the cloud computing question, we included an indicator variable for the number of digital information types the firm has. For business technologies, we included a similar indicator variable for both cloud computing and digital information.

We run this estimation across each of the technologies embedded within each technology question and obtain a set of predicted probabilities for each of the "Missing" and "Don't Know" respondents across the extensive margin response options. We leave the responses for the firms that answered the question in one of the 4 (or 5) allowable response options as-is. We also estimate this equation separately for the "Missing" and Don't Know, as there are slight differences in the size and age distribution of the firms that did not answer the question (Missing) and those that responded with Don't Know.<sup>59</sup>

### A.2.4 Alternative Approaches

In addition to estimating the separate response categories for each of the sub-technologies, we also ran several robustness checks on our estimation by experimenting with other estimation types (e.g., ordered logit), removing different coefficients and separating the interactive terms, and pooling our estimation for the "Missing" and Don't Know. The results are highly consistent, with some minor variation in the probabilities.

The imputed probabilities are tabulated across the use and no-use categories for each subtechnology to create the tables and figures. For the size/age regressions (Tables 8, 12 and 17), which looked at the extensive margin for use of an entire technology group, a separate probit imputation (1/0) was performed for the set of firms with at least one "Missing" or "Don't Know" response. This probit contains the exact same set of covariates as the ordered probit estimation. Part of the reason for doing this is that by summing the imputed probability of use, we may be overstating the likelihood of usage for firms that skipped the entire section (or responded with

<sup>&</sup>lt;sup>59</sup> The "Don't Know" responders tend to be slightly smaller than the "Missing" and skewed more towards smaller firms, but larger on average than the full sample. "Missing" firms also tended to be slightly younger relative to "Don't Know" and the full sample.

Don't Know) for mechanical reasons. For instance, there are 9 sub-categories for cloud usage, and suppose a firm who skipped each cloud question has a relatively low probability of usage for each cloud technology (say 10-15% probability). By summing the probability across the 9 categories, the firm is almost guaranteed to be counted as having using at least 1 cloud technology, when the "true" outcome is likely to be much smaller.

### A.2.3 Post-Imputation Checks (Precision and Recall)

The results from the imputation provide us with a set of probabilities to each of the usage categories for firms that skipped a technology question or marked it with "Don't Know". When we tabulate the distribution of the imputed results and compare them with Tables A5-A7, we find that the shares for each of the usage categories by technology question closely matches the shares for the non-imputed, with slightly higher levels of "All" usage for Digital Information and Cloud Computing. This is not all too surprising as the size distribution of firms that underwent the imputation tended to be larger than the non-imputed sample.

As a secondary check, we performed a Precision/Recall analysis on the imputed responses to gauge the accuracy of our estimation. We performed this by doing the following: we started with our set of non-imputed firms and randomly sampled them into two groups (Test Set (30%) and Training Set (70%)). We then implemented the exact same estimation for the training set and predict the probability of use for the test sample. We are then able to calculate the Precision (share of imputed values that match the response), Recall (share of responses that were imputed correctly) and Accuracy (share of both).

The results from our Precision and Recall checks are mixed. For digital information and cloud computing, our precision and recall levels for each of the response 5 response options was somewhat lower than we would have liked, despite our accuracy being pretty high. This is likely due to having too many response options (3 separate usage responses, 1 non-use and 1 not applicable). Admittedly, the imputation model has a difficult time distinguishing firms by their intensive margin, thus contributing to the relatively low precision and recall rates we find for digital information and cloud computing. When we limit the responses to use/no use, our precision and recall rates uniformly increase to higher levels that provide greater confidence in our imputation methodology.

For the business technologies, our precision and recall perform significantly better, but this is mostly due to the very high levels of non-usage found across each of the categories. When we exclude the "none" from our analysis, our precision and recall drop dramatically.

At first, our precision and recall rates may seem problematic, but we believe they represent an interesting finding, when we remember our sampling frame and initial research question of who adopts advanced technologies. The precision and recall rates were computed for our full sample, which largely consists of relatively small firms, whose adoption rates are very heterogeneous. Our data and explanatory variables are simply too crude to provide a reliable predictor for the precise types of firms that adopt certain technologies and those that do not. While we can claim that size is a reliable predictor of adoption, even amongst large firms, we see heterogeneous patterns of adoption depending on the technology type. In other words, there are simply too

many unknown factors that cannot be measured by traditional metrics (such as firm size, age and industry) that appear to drive technology adoption.

	Mean	Feedback	Financial	Marketing	Personnel	Production	Supply	Other
							Chain	
None	21.6%	25.7%	10.6%	20.9%	18.8%	23.2%	23.2%	28.6%
Less than 50%	9.1%	10.2%	11.5%	11.6%	15.6%	6.2%	7.1%	1.3%
More than 50%	14.6%	12.9%	27.4%	16.5%	22.5%	10.6%	10.0%	2.7%
All	15.3%	13.9%	34.2%	14.2%	21.8%	11.4%	8.8%	2.9%
Information not								
Collected	39.4%	37.4%	16.3%	36.7%	21.4%	48.7%	50.9%	64.4%

Table A5: Pre-Imputation Usage Rates for non-Missing and non-Don't Know - Digital Information (Weighted)

Table A6: Pre-Imputation Usage Rates for non-Missing and non-Don't Know - Cloud Purchases (Weighted)

	Mean	All IT	Data	Billing	Customer	Security	Servers	Storage	Synch.	Other
			Analysis	_	Service	-		_	-	
None	30.7%	27.6%	34.2%	29.3%	32.1%	28.0%	32.6%	29.7%	27.8%	35.1%
Less than 50%	11.3%	14.4%	8.1%	13.0%	10.5%	13.6%	11.3%	12.6%	16.9%	1.6%
More than										
50%	8.5%	10.5%	6.2%	12.3%	8.4%	9.9%	8.6%	9.6%	9.4%	1.6%
All	8.2%	7.3%	5.2%	13.6%	7.0%	12.9%	10.0%	8.6%	7.9%	1.6%
IT Function										
not Used	41.2%	40.2%	46.4%	31.7%	42.0%	35.7%	37.6%	39.5%	38.0%	60.1%

Table A7: Pre-Imputation Usage Rates for non-Missing and non-Don't Know - Business Technologies (Weighted)

	Mean	Aug.	AGV	Mach.	Mach.	NLP	RFID	Robotics	Touch	Voice
		Reality		Learn	Vision					Recog.
None	97.4%	98.7%	99.0%	96.4%	97.9%	98.3%	98.5%	98.3%	93.0%	96.8%
Testing	0.5%	0.4%	0.2%	0.7%	0.4%	0.4%	0.3%	0.3%	0.9%	0.7%
Less than 5%	0.7%	0.4%	0.3%	1.0%	0.5%	0.5%	0.4%	0.4%	1.7%	1.2%
Between 5-										
25%	0.6%	0.2%	0.2%	0.9%	0.5%	0.3%	0.3%	0.4%	1.5%	0.7%
More than 25%	0.8%	0.3%	0.3%	1.0%	0.7%	0.5%	0.4%	0.5%	2.8%	0.7%

## **Appendix A.3 – Unconditional Adoption Rates by Technology Type for Digital Information and Cloud Computing**

One of the challenges in computing shares of technology adoption is identifying the "correct" denominator. For Tables 6 and 9, the shares of adoption for each of the digital information types and cloud computing purchases were calculated conditional on the firm *not* responding in the affirmative to "This type of information not collected by this business" (in the case of digital information) or "Don't use this IT function" (in the case of cloud purchases), meaning that the firm was "at risk" to adopt a technology, but selected not to. While Figures 2 and 4 compute shares based on the number of firms that did *not* respond in the affirmative for each of the above for *all* of the subcategories. We do this to minimize selection and to get a more accurate sense of, "for the firms that would potentially use this technology, what percentage of them actually are?"

We realize that many may want the adoption rate to measure the overall pervasiveness of a certain technology in the economy, regardless of demand, which we provide here. We start by providing the unconditional adoption rates (all firms are in the denominator) for Tables 6 and 9 and re-do our butterfly chart figures (Figures 2 and 4) with the unconditional adoption rates.

Business Function	% Use	% Intensive Use
Financial	72.2 (73.3)	61.0 (61.7)
Personnel	59.1 (59.9)	43.6 (44.3)
Marketing	42.3 (42.7)	30.5 (31.1)
Feedback	36.9 (31.7)	26.6 (23.1)
Production	27.6 (28.1)	21.6 (21.8)
Supply Chain	25.7 (26.2)	18.7 (19.0)
Other	6.8 (6.9)	5.6 (5.7)

 Table A8: Use of Digital Information by Business Function – Unconditional

Notes: "Use" is defined as having responded with "Up to 50%", "More than 50%" or "All" for the information category listed on "Digital Share of Business". "Intensive Use" is defined as having responded with "More than 50%" or "All". Shares are computed using the LBD tabulation weights of firm counts, divided by the total number of firms. Listed shares are imputed shares, with raw weighted values in parentheses.

Cloud Service Purchased	% Use	% Intensive Use
Billing	38.0 (38.8)	25.3 (25.7)
Security	36.3 (37.0)	22.8 (23.1)
Synchronization	33.0 (33.6)	16.8 (17.0)
All IT	31.7 (32.2)	17.6 (17.8)
Data Storage	30.5 (31.1)	18.1 (18.3)
Servers	29.5 (30.1)	18.5 (18.7)
Customer Relations	25.3 (25.7)	15.0 (15.3)
Data Analysis	19.5 (19.8)	11.4 (11.6)
Other	4.7 (4.8)	3.2 (3.2)

#### **Table A9: Cloud Service Purchases by IT Function - Unconditional**

Notes: "Use" is defined as having responded with "Up to 50%", "More than 50%" or "All" for the category listed on "Cloud Service Purchases". "Intensive Use" is defined as having responded with "More than 50%" or "All". Shares are computed using the LBD tabulation weights of firm counts, divided by the total number of firms. Listed shares are imputed shares, with raw weighted values in parentheses.

For each of the technologies within digital information and cloud computing services, the unconditional adoption rates are on average between 10-15% lower than the conditional counterparts. These differences are magnified for industry-specific items, such as "Supply Chain" or "Production" data, which highlights the importance of demand in terms of adopting certain technologies. For the butterfly charts, we see similar 10-15% drops in the aggregate adoption rates by industry.

### Figure A1: Extensive and Intensive Margin Measures of Digitized Information by Sector -Unconditional



Notes: "Use" is defined as having responded with "Up to 50%", "More than 50%" or "All" for the information category listed on "Digital Share of Business". "Intensive Use" is defined as having responded with "More than 50%" or "All". Shares are computed using the LBD tabulation weights of firm counts, divided by the total number of firms. Sectors are defined by combined 2-digit NAICS and assigned for multi-unit firms by largest payroll industry by firm. Original responses classified as "Missing" or "Don't Know" are imputed.

92

#### Figure A2: Extensive and Intensive Margin Measures of Use Rates for Cloud Service **Purchases by Sector - Unconditional**



Notes: "Use" is defined as having responded with "Up to 50%", "More than 50%" or "All" for the category listed on "Cloud Service Purchases". "Intensive Use" is defined as having responded with "More than 50%" or "All". Shares are computed using the LBD tabulation weights of firm counts, divided by the total number of firms. Sectors are defined by combined 2-digit NAICS and assigned for multi-unit firms by largest payroll industry by firm. Original responses classified as "Missing" or "Don't Know" are imputed.

#### Firms Using Cloud Services for IT Functions