

**The Power of Prediction:
Predictive Analytics, Workplace Complements, and Heterogeneous Firm
Performance¹**

Erik Brynjolfsson
Stanford University and
NBER
erikb@stanford.edu

Wang Jin
MIT Sloan School of
Management
jwangjin@mit.edu

Kristina McElheran
University of Toronto
k.mcelheran@utoronto.ca

Abstract

Anecdotes abound concerning the promise of predictive analytics tools and techniques in an age of increasingly ubiquitous digital information. However, large-scale microdata on this fast-emerging practice and its nuances has been lacking. To address this gap, we worked with the U.S. Census Bureau to collect detailed responses from over 30,000 establishments in the U.S. manufacturing sector on predictive analytics and key workplace complements. This effort yielded the first statistics on the widespread adoption of predictive analytics in a large, representative panel of establishments. Focusing on plant-level productivity over time, we find a significant increase in firm performance among adopters – but almost entirely among plants that have also pursued specific tangible and intangible investments. For workplaces with high IT investment, educated employees, a high-efficiency production strategy, and certain management practices, multi-factor productivity is more than two to four percent higher with the use of predictive analytics. Instrumental variables estimates and the timing of performance gains with respect to adoption suggest a causal relationship. While the use of predictive analytics is diffusing rapidly across a wide range of firms, there are understudied strategic and organizational contingencies that determine how productive it is, in practice.

Keywords: predictive analytics, productivity, complementarity, strategic commitment, management practices, information technology, digitization

¹ Disclaimer: Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. All errors are our own.

“What information consumes is rather obvious: it consumes the attention of its recipients...The real design problem is not to provide more information to people, but [to design] intelligent information-filtering systems.”

- Herbert A. Simon 1971 *Designing Organizations for an Information-Rich World*

1. Introduction

According to Ransbotham et al. (2016), “the hype around data and analytics has reached a fever pitch.” The increased volume, variety, and timeliness of digital information produced by sources ranging from social media to the Internet of Things (IoT) have dramatically shifted what is measurable within firms and markets (McAfee and Brynjolfsson 2012). At the same time, exponential growth in computing power and the rise of new computational methods (in particular, new machine learning algorithms) have fueled excitement about tools and practices to extract more value from this data. In a 2019 IDC report, the worldwide forecasted revenues for big data and business analytics will reach \$274.3 billion by the year of 2022 with a five-year compound annual growth rate of 13.2%.² Research in strategic management is increasingly orienting towards understanding the impacts of digitization for firm strategy and firm performance (e.g., Greenstein et al. 2013, Adner et al. 2019, Bennett 2020).

Ironically, however, rising big-data enthusiasm is not grounded in big-sample evidence. Critical gaps persist in our understanding of the prevalence, business impacts, and limitations of these tools and techniques among heterogeneous firms.

We worked to address these gaps by collaborating with the US Census Bureau to gather the first large scale, representative data on the adoption of predictive analytics and potential tangible and intangible complements such as IT investment, worker education, production strategy, and structured management practices. We link this new survey to a detailed panel of administrative data to estimate productivity at the establishment level. Leveraging these novel measures, instrumental variables estimates based on plant-level variance in government-mandated data collection, and key details of timing for distinct waves of adopters, we provide new insights into the prevalence and variance in predictive analytics use, the causal relationships at work, and the

² For more details, please see IDC website: https://www.idc.com/getdoc.jsp?containerId=IDC_P33195.

workplace complements that both constrain and augment the benefits achievable with these rapidly-emerging tools and techniques.

In our representative sample of over 30,000 U.S. manufacturing establishments, we find that adoption of predictive analytics is both widespread and significantly associated with at least two to four percent higher multi-factor productivity. This shift in how firms extract value from data translates into an average \$830,000 difference in sales between adopters and non-adopters, accounting for production inputs and a wide range of other factors. Critically, these gains are only present among establishments that have also pursued high levels of IT capital, employ educated workers at high rates, have committed to a high-efficiency production strategy, or rely on certain management practices centered on tracking key performance indicators.

Predictive analytics, unlike other methods for making sense of data inputs,³ is a set of techniques ranging from data mining to statistical modeling (including machine learning and “artificial intelligence”) to analyze historical and current data in order to make predictions about future events.⁴ Long-standing theories in economics and management have modeled how better information can support and improve human decision-making, creating value at both the individual and organizational levels (Raiffa 1968; Blackwell 1953; Gorry and Scott 1971; March 1994). By extracting core information from large and complex data, predictive analytics shrinks the dimensions of relevant information, reduces the cognitive costs of decision-making, and speeds the execution of higher-quality decisions, thereby boosting returns from digital information.

Significant research attention has been paid in recent years to “Data-Driven-Decision Making” (DDD) practices among managers and the rise of predictive analytics as a related but distinct phenomenon. The adoption of DDD and data analytics have been linked to higher Tobin’s q and profits (Brynjolfsson et al. 2011; Saunders and Tambe 2015), better productivity (Müller et al. 2018; Brynjolfsson and McElheran 2019), and innovation (Wu et al. 2019). However, the causal link between predictive analytics and organizational performance in large-scale, representative studies remains lacking. Building on prior literature, we address this gap by exploring the

³ For instance, a recent study by Berman and Israeli study the impacts of descriptive analytics (2020). All of our results separately control for management practices that are “data-driven” according to other questions on the survey (Brynjolfsson and McElheran 2019), so we are specifically isolating this particular set of techniques.

⁴ Please see Wikipedia and citations within at: https://en.wikipedia.org/wiki/Predictive_analytics.

performance implications of predictive analytics in a sample covering more than 50% of the U.S. manufacturing economy. We address open questions about causality by employing both instrumental variables estimation and timing tests to rule out reverse causality. Our findings provide evidence that predictive analytics is causally related to higher productivity, on average. However, important boundary conditions apply related to firm heterogeneity and strategic commitments that have received little or no attention to date.

In particular, our study reveals how these non-trivial returns on the use of predictive analytics depend almost entirely on the presence of other tangible and intangible firm investments. Anecdotal evidence is mounting that many firms struggle to realize benefits from their investments in data analytics (Schrage 2014). Lack of complementary skills and unaligned corporate strategy are reported to be key challenges (Ransbotham et al. 2015). Earlier academic work on IT productivity emphasizes significant heterogeneity across industries and firms in the benefits of generic IT investment (e.g., Stroh 2002, Brynjolfsson and Hitt 1995). A robust theoretical literature argues that this heterogeneity may arise from investments in complementary assets and managerial practices (Kandel and Lazear 1992; Milgrom and Roberts 1990, 1995; Holmstrom and Milgrom 1994; Brynjolfsson and Milgrom 2013; Brynjolfsson et al. 2018). A number of empirical studies have confirmed this intuition with respect to general-purpose IT and computer use (Black and Lynch 2001; Caroli and Van Reenen 2001; Bresnahan et al. 2002; Aral and Weill 2007; Bloom et al. 2012; Aral et al. 2012) as well as for data-centered management practices (Tambe 2014; Brynjolfsson and McElheran 2019; Dubey et al. 2019). Yet tools and techniques for leveraging digital information are evolving rapidly, detailed data on their use in practice is difficult to obtain for large samples of firms, and many open questions remain. In particular, large-scale evidence on what specific complements should be considered – much less how much they matter – is almost entirely lacking. Also missing is a rigorous empirical approach for disentangling causality and potential confounds across a large and economically important sample of firms.⁵ Our study makes progress on these dimensions and sheds light on heretofore understudied organizational complements such as elements of the firms strategic positioning day-to-day management practices.

⁵ A few studies in the management information systems literature provide conceptual frameworks for understanding how predictive analytics can create synergies in organizations and improve performance (Côte-Real et al. 2014; Grover et al. 2018).

This study builds on the frameworks for organizational complementarity developed in Athey and Stern (1998) and Brynjolfsson and Milgrom (2013) and contributes to a small but growing empirical literature on complementarities among technology, workplace characteristics, and managerial practices (e.g., Black and Lynch 2001; Aral et al. 2012; Tambe et al. 2012; Bloom et al. 2012; Brynjolfsson and McElheran 2019). Insights from these and related works guided our development of new survey questions specifically designed with this research question in mind.

Our findings contribute to a growing strategic management literature emphasizing not only the average impacts of digitization and digital strategies, but also the heterogeneity in those effects related to varying organizational and market-based contingencies and constraints. This literature dates at least as far back as Tushman and Anderson's (1986) exploration of how technology interacts with a given firm's competence to more-recent explorations of internal and external influences on digital strategy choices and their implications (e.g., Mithas et al. 2013, McElheran and Jin 2020).

This effort is part of a small but non-trivial set of collaborations between academics and administrative agencies to combine the coverage and consistency of administrative data collection with insights and questions from the frontier of academic research (see Buffington et al. 2017 and Zolas et al. 2020). Our partnership with Census yielded detailed establishment-level information on management and organizational practices, worker education, the availability and use of data, extensive and intensive margins of predictive analytics use, specific use cases, and related questions on the organizational design of data collection – including government influence on this process. Linking this new survey with longstanding administrative data on production inputs and outputs, we achieve unusually-detailed insights for such a large sample (over 30,000 establishments).

Thus, a key contribution of this study is to provide the first-ever descriptive statistics into the prevalence and contexts in which these practices have diffused in the U.S. Through systematic analysis, our study further reveals how core dimensions of process design and workplace practices may determine the returns to this increasingly widespread set of tools. In addition to contributing to the frontier of academic research in this area, our findings provide evidence-based guidance for managers looking to achieve better returns on investment in data and predictive analytics and help them stay competitive in an increasingly digital age.

2. Conceptual Motivation

The Performance Effect of Predictive Analytics

Given the high ratio of interest-to-evidence surrounding predictive analytics, it is more important than ever to understand what impact predictive analytics has on business performance and how firms can achieve a higher return on their investments in “big data” (Grover et al. 2018). Existing theory in economics and management predicts performance gains from increasingly sophisticated tools for extracting useful insights from large sets of structured and unstructured data. All else equal, better information is argued to support and improve human decision-making, creating value at both the individual and organizational levels (Raiffa 1968; Blackwell 1953; Gorry and Scott 1971 March and Simon 1994). By extracting essential insights from large and complex data, predictive analytics shrinks the dimensions of relevant information, reduces the cognitive costs of decision-making, and speeds execution of higher-quality decisions, thereby boosting returns from digital information. Following prior literature focusing on the business value of data and analytics (Brynjolfsson et al. 2011; Tambe 2014; Saunder and Tambe 2015; Müller et al. 2018; Brynjolfsson and McElheran 2019), we argue that the adoption of predictive analytics is likely to have a positive impact on business performance and can lead to a significant gain in efficiency and productivity, all else equal.

Predictive Analytics and Potential Complements

In practice, important factors are typically not equal across firms. A prevalent theme from the frontier of economics and strategic management research is the widespread and increasing heterogeneity in firm performance and workplace conditions.⁶ Prior theory and empirical studies highlight how complementarities among IT investments and organizational characteristics can lead to differences among firms that may persist and even grow over time (Milgrom and Roberts 1990, 1995; Black and Lynch 2001; Caroli and Van Reenen 2001; Bresnahan et al. 2002; Aral and Weill 2007; Bloom et al. 2012; Aral et al. 2012; Brynjolfsson and Milgrom 2013). While laggard firms

⁶ 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). This phenomenon is not restricted to the United States (e.g., Berlingieri et al. 2017).

may catch up along certain dimensions (Brynjolfsson and McElheran 2019), trends towards increasing dominance of “superstar firms” (Autor et al. 2020), as well as rising concentration and inequality in workplace conditions and employee earnings have been linked to technology investment at the industry and firm level (Bessen 2017; Bennet 2020b; Lashkari et al. 2020; Barth et al. 2020). In addition, attention is increasingly focused on difficult-to-measure intangible features of firms and markets that may amplify these dynamics (Saunders and Brynjolfsson 2016; Haskel and Westlake 2018). Building on these insights and techniques, we explore potential tangible and intangible complements to predictive analytics that might be expected to drive economically important heterogeneity in returns to these tools and techniques.

IT Capital Stock

The successful implementation of predictive analytics is widely expected to depend on the firms’ IT infrastructure. The collection, storage, and communication of input data for predictive modeling all require robust tangible investments in sensors, transmission equipment (e.g. routers), data storage hardware, and computing resources for analyzing data. For instance, Bosch’s manufacturing factories use predictive analytics tools for the maintenance of equipment such as robots, heat exchangers, and spindles using data generated from sensors.⁷ General Electric, another manufacturing giant, augments its jet engines with sensors to monitor temperature, fuel consumption, and many other daily operation tasks while generating 500 gigabytes of data per engine for each flight.⁸ Unsurprisingly, it is also investing heavily in predictive analytics to utilize such volume of data (Winnig 2016). Furthermore, building, training, and implementing these analytic tools require corresponding data processing hardware and software. IT-savvy firms that are more prepared for the industrial Internet of Things (IoT) – a particularly important dimension of the “big data” trend for our manufacturing context – may have fully-depreciated investments in sensors and other IT infrastructure to collect and analyze data, giving them a tangible advantage when it comes to analytics.

Intangible complements may also be associated with these easier-to-measure IT capital

⁷ See reports at <https://blog.bosch-si.com/industry40/industry-4-0-predictive-maintenance-use-cases-in-detail/> and <https://amfg.ai/2019/03/28/industry-4-0-7-real-world-examples-of-digital-manufacturing-in-action/> for more details.

⁸ <https://venturebeat.com/2015/06/18/here-comes-the-industrial-internet-and-enormous-amounts-of-data/>

investments over time. Firms with long-standing experience in developing and deploying IT resources and related practices are expected to enjoy disproportionate returns due to accumulated IT capabilities, appropriate organizational design, human capital accumulation, and other dimensions of “learning by doing” (e.g., Bharadwaj 2000; Saunders and Brynjoflsson 2016). In addition, firms with greater accumulation of IT capital may enjoy a greater capacity to adopt and effectively deploy new techniques and related business processes (e.g., Han et al. 2011). We hypothesize that firms with higher measured IT capital stock will have lower costs and greater opportunities to leverage predictive analytics to boost productivity in their workplaces.

Educated Workers

While IT capital stocks may proxy for a range of critical complements to predictive analytics, it is useful to disentangle this cluster of interrelated activities as finely as possible. In particular, it is widely accepted that more-skilled and better-educated workers are key drivers of growth in both manufacturing productivity (Black and Lynch 2001; Moretti 2004; Waldman 2016) and returns to IT (Brynjoflsson and Hitt 2000; Bresnahan et al. 2002). With the increasing digitalized organization and growing prevalence of business applications for data and predictive analytics, demand for workers with complementary analytic skills is also growing. Firms increasingly demand workers that know how to deploy “smart” technologies in production settings (Helper et al. 2019) as well as those who can help translate digital information into business insights. While competition for these workers may drive up wages to a point where they do not manifest excess returns in the production function, we expect that their presence will boost the estimated returns from adopting predictive analytics due to these complementarities.

Production Strategy

Production strategy has received limited attention – even in the strategic management literature – and has not, to our knowledge, been linked to IT or digitization. Yet the relevance of this concern is conceptually very straightforward and arguably first-order in our manufacturing setting.

At its most fundamental level, predictive analytics uses historical and current data to predict future outcomes. Thus, the effectiveness of this forward-looking approach depends on: 1) the richness of past data inputs to characterize meaningful states of the world, and 2) sufficient stationarity of the environment so that past performance may be a useful indicator of future

conditions. Not all production processes will score high on these criteria. In general, workplaces with greater automation and reduced variance will provide richer data inputs and better alignment with the value proposition of predictive analytics compared to other, more dynamic, production environments.

Yet dynamism versus stationarity of a production environment is rooted in a host of economic and strategic considerations that are very difficult to adjust, potentially limiting or augmenting the value of predictive analytics in particular settings in persistent ways. For organizational, strategic, cognitive, and technological reasons, the orientation of a business unit's production process towards efficiency versus flexibility tends towards one extreme or the other and is fundamentally difficult to reverse (e.g., Ghemawat and Ricart Costa 1993). It is correlated with a range of specific management practices (McElheran et al. 2020), and typically aligned with key dimensions of the firm's market context and positioning (Hayes and Wheelwright 1979, Safizadeh et al. 1996). Thus, investments in more-automated and more-efficient production processes versus more innovative and dynamic activities is indicative of an organizational environment that is subject to lower variance. This holds for features of both the external market context (e.g., market demand and supply chain processes), as well as for internal operations (equipment utilization, product quality, cycle times, etc).

Following the influential work of Hayes and Wheelwright (1979), we conceptualize a high-efficiency production process in the manufacturing sector as one where a plant operates primarily on a continuous-flow production basis.⁹ Compared to other processes (e.g. job shops for custom work or prototyping for product development), continuous-flow processes tend to be more capital-intensive¹⁰ continuously-monitored, and managed so as to minimize variance. This promotes stationarity and the value of predictions based on past information.

The richness of information available to be analyzed also tends to be higher. Higher-volume, lower-mix production tends to involve more tightly-coupled systems where downtime and low

⁹ The comparison is between the establishments with continuous flow production process and the job shops and establishments with a batch process. See Hayes and Wheelwright (1979) for a more complete discussion of the process design features captured by the measures we use here.

¹⁰ Based on the industry level productivity database from the National Bureau of Economic Research (NBER), these industries have the highest average capital stock within the manufacturing sector. Database and statistics are available upon request.

capital utilization are more costly or even dangerous, promoting systematically greater use of instrumentation and sensors. The data generated from these sensors are increasingly digitalized, transmitted, and analyzed at a high frequency, providing rich inputs to predictive analytics practices when they are adopted. Thus, we anticipate lower costs and higher returns to predictive analytics in workplaces with a high-efficiency production strategy.

KPI monitoring

The old adage that “what gets measured gets managed” holds in the digital age as well as ever. Manufacturing firms have long been developing and monitoring key performance indicators (KPIs) in a variety of operational practices such as Lean manufacturing and Total Quality Management. Firms track capacity utilization rate, manufacturing cycle time, and production downtime, among many others. These KPIs help firms quantify and monitor the performance of operations and thus are highly correlated with the breadth and intensity of data collection and analysis at the establishment (Brynjolfsson and McElheran 2019). Schrage and Kiron (2018) stress the importance of KPI use in the digital era with the rise of big data and computational and algorithmic innovation. Firms with well-established KPI practices will tend to have richer inputs to feed their predictive models. In other words, “what gets measured gets into predictive modeling.” Given the advancement in reinforcement learning and neural networks, predictive analytics have the potential to help make better decisions in rather complex business processes with its increasing capability and precision (Khurana et al. 2018; Khan et al 2018).

Yet, more-sophisticated predictions will only provide performance gains to the extent that they are acted on in practice. Again, we anticipate that prior use of managerial routines around KPI monitoring will support greater organizational capacity for leveraging outputs from predictive models, leading to important production dependences between this class of managerial practices and new predictive analytics techniques.

3. Data and Key Measures

The lack of large-scale data on business adoption of predictive analytics has been a significant barrier to empirical research in this area. We made progress by collaborating with the U.S. Census Bureau to design and field a purpose-designed survey that was sent to a large and representative sample of establishments in the U.S. manufacturing sector. Leveraging the extensive reach,

expertise, and administrative infrastructure of the Census Bureau, we developed, tested, and deployed a set of questions that were added to the 2015 Management and Organizational Practice Survey (MOPS).¹¹ Our sample includes roughly 30,000 establishments that provided data on predictive analytics adoption and frequency of use in 2015, along with recall data for 2010. Responses to the survey are required by law, yielding a response rate of 70.9%.¹² Therefore, our study is significantly less affected by the respondent bias and lack of representativeness that typically plague other custom-designed surveys.¹³

Predictive Analytics Use

Adding questions to Census surveys requires that they be applicable across a wide range of industry and firm settings and that they pass rigorous cognitive testing designed and implemented by dedicated Census employees and reviewed by both stakeholders across a range of positions and backgrounds. At the end of multiple rounds of development and cognitive testing, the focal question on the MOPS survey asks, “How frequently does this establishment typically rely on predictive analytics (statistical models that provide forecasts in areas such as demand, production, or human resources.” They may mark all that apply among *Never*, *Yearly*, *Monthly*, *Weekly*, and *Daily*, with separate columns for 2010 (recall) and 2015.

We construct a measure of the extensive margin of predictive analytics use by identifying establishments that adopt with any frequency and compare them to those that report “*Never*.” To capture the intensive margin, we first assign numeric values from 0 to 4 to each frequency category, with higher values representing higher frequencies of predictive analytics use. In cases where plants report multiple frequencies, we default to the highest one. We also explore using a normalized score based on taking the average of multiple responses for a given establishment (see Bloom et al. 2019).¹⁴

To explore different use cases for predictive analytics, we further combine information from

¹¹ This survey is conducted on both 2010 and 2015 as a supplement of the Annual Survey of Manufactures (ASM). See Bloom et al. (2019) and Buffington et al. (2017) for more details on the MOPS survey.

¹² For more details, please see the website for the MOPS: [https://www.census.gov/programs-surveys/mops/technical-documentation/methodology.html#:~:text=The%202015%20MOPS%20had%20a,\(URR\)%20of%2070.9%25](https://www.census.gov/programs-surveys/mops/technical-documentation/methodology.html#:~:text=The%202015%20MOPS%20had%20a,(URR)%20of%2070.9%25).

¹³ See Zolas et al. (2020) for a more-comprehensive discussion of the issues and related survey work.

¹⁴ The normalization is helpful for easy interpretation of coefficients in the regression analysis and taking an average of the scores helps to account for the multiple responses given by a given establishment. Nevertheless, we find the results from both measures are consistent.

another MOPS question regarding where data analytics was applied at the plant. This question focuses on three main business functions: supply chain management (SCM), demand forecasting (DF), and product design (PD). We construct the corresponding use cases for predictive analytics conditional on the plants reporting having adopted predictive analytics and using data analytics in the corresponding function.¹⁵

Figures 1 and 2 present the adoption of predictive analytics by states and industries (3-digit NAICS). Both figures reveal a surprising fact: predictive analytics was widely diffused among manufacturing plants across almost all states and industries as early as 2010, with average adoption rates well over 70%. Further exploration using a balanced sample of roughly 18,000 establishments (those with data for both years)¹⁶ indicates a rather small change in the adoption of predictive analytics between 2010 and 2015. That is, the average rate of growth in use of predictive analytics observed in our sample is under 1.5% per year.¹⁷ This slow rate of change is due in no small part to the high baseline adoption in 2010. If we think about predictive analytics as a “management technology” following Bloom et al. (2016), its diffusion likely reached a relatively level of maturity – “latter majority stage” in the technology adoption S curve (see Rogers 2010) - and thus has slowed its growth among the remaining population of firms. Remaining non-adopters will tend to be firms with low net benefits of adoption, due either to high costs or low anticipated returns.

In addition to being informative about the timing and penetration of predictive analytics adoption, this empirical fact has important implications for our research approach. In particular, standard models for addressing time-invariant organizational heterogeneity through the use of firm or establishment fixed effects are of limited use in our empirical setting. Recall that panel data methods only identify coefficients off of establishments that transition from not using predictive analytics to adopting in our sample. Thus, they would be applicable to a very small and highly-selected group of plants whose defining features is being in this “laggard” group of adopters (those that adopted between 2010 and 2015). Frontier users of these practices would be differenced out

¹⁵ Both questions are presented in Appendix Table A1.

¹⁶ The smaller number of observations in 2010 as well as in the balanced panel is due to the sample rotation of the ASM. The ASM sample frame changes in years ended with “4” and “9”, which means that the establishments in our 2010 and balanced samples have to be in both 2009-2013 and 2014-2018 sample frames.

¹⁷ The adoption rate of predictive analytics increased from 73% in 2010 to 80% in 2015, implying a 1.4% growth rate per year.

of any panel data estimation, likely biasing the average productivity attributed to these practices.

Moreover, the 5-year gap in our two-year panel generates considerable measurement imprecision concerning the exact year of adoption. Therefore, we focus our empirical approach primarily on pooled OLS models for our two discrete years of data and include both industry-year fixed-effects as well as a rich set of controls (many of which capture fixed or quasi-fixed organizational characteristics). For completeness, we report the less-informative results from plant fixed-effects models in the online appendix.

[Insert Figures 1 and 2 here]

Performance Data and Workplace Complements

Using plant-level identifiers maintained by the Census Bureau, we merge the MOPS with the Annual Survey of Manufactures (ASM), the Census of Manufactures (CMF), and the Longitudinal Business Database (LBD) to bring in information on detailed inputs (including accumulated and depreciated capital stocks, as well as input costs for labor, materials, and energy), outputs (total value of shipments and value-added), age, and whether the establishment belongs to a multi-establishment firm. We restrict the sample to observations with complete information on sales, labor, materials, energy, and the total number of employees for technical and disclosure-avoidance reasons.

To explore potential organizational complements, we leverage several other measures in the ASM and the MOPS. To explore the relationship to tangible IT investments, we calculate IT capital stocks using capital expenditure on computer and peripheral data processing equipment from the ASM and CMF panel dating back to 2002, using a standard perpetual inventory approach and an industry-level deflator for hardware from the Bureau of Economic Analysis (BEA). We impute values for years in which values are missing¹⁸ and depreciate at the rate of 35% per year following Bloom et al. (2014), Brynjolfsson and McElheran (2016), and Jin and McElheran (2019).¹⁹ We

¹⁸ We impute the missing values using the average of the IT investment from the closest before and after years that have non-missing values for the establishment-years where the capital expenditure on computer and data processing equipment is missing. For instance, if IT investment in 2008 is missing, we impute it using the average IT investment for the establishment in 2007 and 2009 or using the 2007 and 2010 values if 2009 is missing. Similar logic is applied to missing values from other years. A similar method has been used in Bloom et al. (2014), see appendix for more detail.

¹⁹ Also note that the Non-IT capital is calculated using a similar approach but with total non-IT capital expenditure (total capital expenditure less IT capital investment) instead. The industry level deflator for the non-IT capital stock is obtained from the ASM Total Factor Productivity databases.

test the complementarity between IT and predictive analytics by utilizing the IT capital stock variable. The advantage of this measure is that it accounts for the overall stock. If it takes time for firms to adjust and utilize novel IT investment, we will be able to capture the lagged effect. A reasonable concern here is that we are unable to capture the effect of capitalized software (e.g. ERP investment) which might also play a significant role in facilitating the implementation of predictive analytics or otherwise boost productivity (Barth et al. 2020). To address this concern, we conduct several robustness tests in both the baseline performance analysis and the complementarity tests. First, we control separately for software and IT services expenditures from the ASM, in addition to IT capital stock. Alternatively, we use a measure summing up all IT investments (hardware, software, and services) instead of the IT capital stock variable. In both cases, we find that our results stay robust and consistent.

Next, we use information from the background characteristics in the MOPS regarding the percentages of managers and non-managers with a bachelor's degree. Combined with the total number of employees (from the ASM) and the number of managers (from the MOPS), we calculate the weighted average of the percentage of employees (both managers and non-managers) with a bachelor's degree following Bloom et al. (2019). This measure is similar to the prior literature that uses education as a proxy for human capital in labor studies (Card 1999; Bresnahan et al. 2002; Arcidiacono et al. 2010).

Another measure we exploit in the MOPS is an indicator of how the production process is organized. We focus on establishments with high-flow production processes (including both cellular and continuous-flow manufacturing).²⁰ These processes are characterized by low product mix and high volume per product, which also tend to be more capital-intensive, with higher levels of automation (Safizadeh et al. 1996). This is in contrast to establishments operating as job shops, batch manufacturing facilities, or R&D focused establishments that are all more likely to have a “jumbled flow” process design, supporting flexible, high-mix but generally low-volume operations. The latter is typically central to product design and prototyping (Hayes and Wheelwright 1979) and appear less amenable to very structured approaches to management (McElheran et al. 2020).²¹

²⁰ See Kiran (2019) for detailed description of cellular manufacturing.

²¹ We assigned the establishments characterized by batch production along with job shops. These establishments lie in-between the “job shops” and continuous flow production in terms of their capital intensity, level of machine

We test empirically whether establishments with high-flow process designs are better positioned to seize larger benefits from the adoption of predictive analytics. Note that the production process is usually quasi-fixed since the transition cost is high. This is largely aligned with the empirical evidence from our data, where we observed almost no change for this variable. That is, the percentage of plants transitions into the continuous-flow production process is less than 0.7% per year between 2010 and 2015. Arguably, this is a feature for testing complementarity since it allows us to test the effect of predictive analytics on the performance given the pre-existing cross establishment differences (Aral et al. 2012).

Lastly, we construct an indicator for establishments that report monitoring 10 or more KPIs. Based on the MOPS survey, establishments are categorized by the number of KPIs monitored, ranging from 0 KPI, 1-2 KPI, 3-9 KPI, and 10 or more KPI. We take the top category as a proxy for establishments with high intensity of KPI monitoring to test our hypothesis on whether the practice of KPI monitoring is a complement for predictive analytics.

Table 1 presents the summary statistics for our baseline sample.²² Overall, we have a pooled sample of 51,000 observations from both 2010 and 2015 with a complete set of information for our empirical analysis (about 20,000 for 2010 and 31,000 for 2015). Table 2 reports the summary statistics for all key variables in this baseline sample. The average adoption rate for predictive analytics is well above 70%, with the majority of establishments reporting annually and/or monthly use (i.e. frequency value is roughly 1.12). The average annual total value of shipments and total employment in log terms for establishments in our sample are around 10.37 and 4.56, respectively, which can be translated into about \$32,000,000 and an average size of 96 employees.²³ The mean age of these establishments is more than 24 years old. These establishments have on average about \$175,000 of IT capital stock and slightly over 15% of workers with a bachelor's degree. Finally, about 35% of the establishments reported having the continuous-flow production process, and

automation, as well as product standardization. Adding them in the control group is likely to add measurement error, biasing our estimates downwards. In so doing, we might be estimating a low bound of the effect on instead of suffering an omitted variable bias (if we do not control it in the base group). We also control for R&D facilities separately.

²² Table A1 presents the key question from the MOPS regarding predictive analytics in the survey. Also, please see the survey questionnaire for the 2015 MOPS at the U.S. Census website:https://www2.census.gov/programs-surveys/mops/technical-documentation/questionnaires/ma-10002_15_final_3-2-16.pdf

²³ To cover a larger percentage of the manufacturing economy and all industries, the ASM selects a set of establishments in the certainty sample frame. For instance, establishments with employment for the establishment is greater than or equal to 1,000 in the 2012 Census or the establishment that classified within an industry with less than

around 44% of them reported tracking 10 or more KPIs. With this discussion in mind, we now turn to the empirical method.

4. Empirical Methods

Adoption and Correlation Test

After providing descriptive statistics on adoption, our empirical analysis explores correlations between predictive analytics use and potential workplace complements including IT capital stock, an indicator for plants having a more continuous-flow production process, percentage of educated workers, and top KPI monitoring. If complementarities exist, we should observe higher adoption of predictive analytics for establishments that also report the presence of these characteristics (Brynjolfsson and Milgrom 2013). We test both linear probability and probit models, and explore a wide range of establishment and firm characteristics to address potential omitted variable bias. We also control for geographic differences and industry-year fixed-effects to account for any transitory industry-specific shocks.

Performance Analysis

To estimate the performance impact of predictive analytics, we take a conventional approach to modeling the establishment's production function (Brynjolfsson and Hitt 2000; Brynjolfsson and Hitt 2003; Bloom et al. 2012; Tambe and Hitt 2012). As our baseline specification, we consider a log-transformed Cobb-Douglas production function in equation (1):

$$\text{Log}(Y_{ijt}) = \beta_0 + \beta_{pa} \log(PA_{ijt}) + \beta_k \log(K_{ijt}) + \beta_l \log(L_{ijt}) + \beta_m \log(M_{ijt}) + \mu X_{ijt} + w_{ijt} + \varepsilon_{ijt} \quad (1)$$

Y_{ijt} is the total value of shipment of establishment i in industry j at time t , K denotes the establishment's capital stock at the beginning of the period, PA is an indicator (or a frequency measure) for the establishment's adoption of predictive analytics, L is labor input, M is the establishment's consumption of material and energy inputs, and X is a vector of additional factors such as IT capital stock at the beginning of the period, age of the establishment, percentage of educated workers, and indicator for those that belong to multi-establishment firms. Both w_{ijt} - the "technical productivity" and ε_{ijt} - the "shock to productivity" are unobservable econometrically (but w_{ijt} might be observable by establishments). The estimate we are primarily interested in here is β_{pa} , where we can interpret as the average effect of predictive analytics on establishments' productivity, all else equal.

Identification via Government Mandates

One of the primary concerns in the approach described above is that the adoption of predictive analytics can be endogenously determined, preventing a causal interpretation of β_{pa} .²⁴ To address this issue, we explore IV estimation using an indicator that equals one when the establishment is required to collect data due to government regulation or agencies, and zero otherwise.²⁵

The underlying logic for this approach is based on the “Porter hypothesis” (Porter 1991; Porter and Van der Linde 1995; see also Ambec et al. 2013 for a useful review) arguing that well-designed government regulations can stimulate firms to innovate and adopt new technology and practices. Of relevance in our setting, data collection is often mandated by federal and local governments to demonstrate compliance with a wide array of environmental and safety regulations. Abundant anecdotes support the prevalence of this phenomenon. For instance, as early as 1970, the Occupational Safety & Health Administration (OSHA) required about 1.5 million employers in the United States to keep records of their employees’ work-related injuries and illnesses under the Occupational Safety and Health Act of 1970.²⁶ The United States Congress established the Toxic Release Inventory (TRI) program under the Emergency Planning and Community Right-to-Know Act (EPCRA) to request firms to collect and report their emission level on hundreds of chemicals since 1986.²⁷

The unexpected consequences of implementing such regulations for firms that are not already data-savvy can be striking. A well-known case of how a firm transformed its data capabilities, managerial practices, and ultimately, productivity by focusing on workplace safety is the story of Alcoa Corporation under the leadership of Paul O’Neill in the late 1980s and 1990s. While the catalyst in this specific case was a top-management dictate to monitor and dramatically reduce

²⁴ More productive establishments might choose to adopt predictive analytics and create a spurious relationship and upward bias. Tambe and Hitt (2012) both provide a useful discussion and suggest that such concerns may warrant less attention they they typically receive in the context of IT productivity studies.

²⁵ The information comes from a question that was added to the MOPS in 2015 specifically to improve our understanding of establishments’ data collection activity, which is a key input to Data-Driven-Decision (DDD) making studied in Brynjolfsson and McElheran (2016 and 2019). Labro et al. (2019) also leverages our addition to the survey in a similar IV method to explore how predictive analytics is related to other MOPS measures.

²⁶ For more details on OSHA Recordkeeping rule, please see the OSHA website <https://www.osha.gov/recordkeeping2014/records.html>

²⁷ For more details on TRI program, please see the EPA website at <https://www.epa.gov/toxics-release-inventory-tri-program/what-toxics-release-inventory>

worker safety incidents, the mechanism we have in mind is quite similar. In the case of Alcoa, an unexpected mandate to prioritize safety concerns over productivity resulted in an organization where data became more abundant, new performance metrics were devised and analyzed with increased frequency, and managerial incentives became tied to new performance metrics. The end results were not only improved worker safety but also soaring productivity as a result of improved process monitoring and management (Clark and Margolis 1991; Duhigg 2012).

Consistent with this qualitative example and growing empirical evidence for the “Porter Hypothesis” in a range of settings, we expect that regulatory pressure could have a similar unintended consequence when it comes to data collection and use. When required to collect new data and devise new reports by statute, firms are more likely to build infrastructure and managerial systems for data collection, storage, and analysis.²⁸ Workers and managers are trained in new systems and techniques for capturing, analyzing, and communicating insights based on objective data. Increased volumes of data on production processes, human resources, and the like – regardless of the motivation for collecting it—may constitute useful inputs for the predictive analytics process (Raghupathi and Raghupathi 2014). This capability development should lower the costs of adopting predictive analytics, all else equal.

Moreover, regulation mandated by the U.S. government has historically been unrelated to productivity-increasing initiatives. This is likely the case for data collection as well in the manufacturing sector: for instance, the objective of the OSHA’s recordkeeping rule is to improve worker safety and health benefits while the TRI program was established for public awareness of the release of toxic chemicals and emergency planning. Objections to these types of regulation have been non-trivial and typically rooted in arguments that they divert resources from other productivity-enhancing managerial activities and investments historically (Gollop and Roberts 1983; Gray 1987). While some empirical evidence suggests that well-designed regulations tend not to significantly reduce manufacturing firm’s competitiveness and overall performance (Jaffe et al. 1995; Lanoie et al. 2011), we expect the direct effect to work against a positive relationship between government mandates and productivity, with productivity enhancements arriving only out

²⁸ For instance, the EPA requires firms in wide range of industries (e.g. pulp and paper, petroleum and Coal product, and Chemical manufacturing) to install the continuous emission monitoring systems (CEMS) to collect and analyze emission data.

of the indirect effect of increased data capabilities and resources on other aspects of the production process.

Following this logic, government-mandated data collection should satisfy both the relevance condition and exclusion restrictions for a valid instrumental variable. A key concern, however, is that such an instrument will need to vary at a highly granular level in order to identify the mechanism of interest with sufficient power. We therefore designed a question for the MOPS that captures government-mandated data collection at the plant level. Prior validation of this new measure demonstrates considerable plant-level variability, even controlling for a host of observable characteristics such as industry and size (McElheran et al. 2020). It has been used as an instrument in related studies of data-driven decision making and predictive analytics, as well (Brynjolfsson and McElheran 2019).²⁹

Timing and Causality

We further explore the causal relationship between predictive analytics and productivity via the timing of adoption and productivity changes, with the goal of ruling reverse causality. Leveraging the annual data on inputs and output from the ASM and the CMF, we construct a panel from 2010 to 2016 for establishments in the 2015 MOPS sample. Within this panel, we leverage the recall question on the MOPS to identify establishments that adopted predictive analytics by 2010 (“early” adopters) and those that adopted between 2010 and 2015 (“late” adopters). Leaning on evidence that many (if not most) of the organizational practice measures in the MOPS are quasi-fixed over this time period, we can extend our productivity estimation approach to track performance differences among the early adopters, late adopters, and “laggards” (non-adopters by 2015). This approach allows us to rule out that productivity precedes adoption by verifying a few patterns we expect to see in the data if this relationship is, indeed causal. First, whether early adopters have a performance premium compared to both late adopters and laggards. Second, whether and when late adopters become more productive compared to both early adopters and

²⁹ As noted in Brynjolfsson and McElheran (2019), establishments may collect data at their own will and collect data to comply with government regulation at the same time, so firms that already are data and analytics- oriented may not be as responsive to our instrument.

laggards. Later adopters should outperform the never-adopting laggards.³⁰

Performance Tests of Complementarity

After addressing the questions of causality in the baseline performance model, we proceed to a formal performance test to explore complementarities. Following the empirical strategy in Athey and Stern (1998), Novak and Stern (2009), and Brynjolfsson and Milgrom (2012), our empirical specifications follow equation (2):

$$\begin{aligned} \text{Log}(Y_{ijt}) = & \beta_0 + \beta_{pa} \log(PA_{ijt}) + \beta_c C_{ijt} + \beta_{interaction} \log(PA_{ijt}) \times C_{ijt} + \beta_k \log(K_{ijt}) + \\ & \beta_l \log(L_{ijt}) + \beta_m \log(M_{ijt}) + \mu X_{ijt} + w_{ijt} + \varepsilon_{ijt} \end{aligned} \quad (2)$$

All variables in equation (2) are identical to those in equation (1) except C_{ijt} , which denotes respectively the indicators for high IT capital stock, continuous-flow production process, high percentage of educated workers, and top KPI monitoring. We interact predictive analytics use with these potential complements to identify the heterogeneous return with and without the presence of such a complement. If complementarity exists between predictive analytics and any of these potential complements, we expect $\beta_{interaction}$ to be positive and significantly different from zero.

Also, we focus our analysis using primarily the industry-year fixed-effect models (in contrast to the plant fixed-effect models). While this approach does not strip out time-invariant features of a given plant's production process that might be correlated with both productivity and the use of predictive analytics, it has other important advantages for our research question here. That is, it allows us to disentangle the effects of organizational practices for performance that might likewise be fixed or change very little over the period we observe. It also accommodates the reality in our context that diffusion was already quite widespread by the time of our study. A fixed-effects estimator would only yield insights for the subsample of later adopters, providing a fundamentally incomplete view into this fast-moving phenomenon.

5. Results

Correlation Tests for Workplace Complements

³⁰ While system GMM and other semi-structural estimation methods (see Arellano and Bond 1991; Blundell and Bond 2000; Levinsohn and Petrin 2003; Akerberg et al. 2015) have performed well in recent studies of IT productivity (e.g., Tambe and Hitt 2012; Nagle 2019), the structure of our core data data source (two-year panel with recall values) limits our ability to implement these methods, which generally require longer lag periods.

We start our empirical exploration by exploring conditional correlations between the adoption of predictive analytics and potential workplace complements. Figure 3 presents relationships by quintile of IT capital stock and percentage of workers with a bachelor's degree (our proxy for skilled labor), whether or not the establishment has a continuous-flow production process, and categories for the number of KPIs monitored, controlling for size, age, and industry fixed effects (at the 6-digit NAICS level).³¹ The base group for all panels is the bottom category (e.g. the base group for panel a is the bottom quintile of IT capital stock).

[Insert Figure 3 here]

The IT capital stock relationship is sharpest at the top of the distribution (Panel a). Establishments with top-quintile IT capital are the most likely to use predictive analytics. The coefficient for this IT-intensive group is the only one that is positive and significantly different from zero (at the five percent level). These establishments invested heavily in IT over time (this is a stock, not a flow measure of investment) and are equipped with infrastructure to provide better or lower-cost data to be analyzed, as well as computational capabilities. This lowers the cost and increases the returns to predictive analytics, promoting higher returns to both. Lower quintiles of IT capital are significantly less associated with predictive analytics, with the second quintile associated with the lowest likelihood (significant at the one percent level), but a generally noisy relationship for the bottom 80% of the IT capital distribution.³²

In Panel b, we show that both cellular manufacturing and continuous-flow production have the highest average adoption rates for predictive analytics while job shops have the lowest rate of adoption (base group). Finally, the likelihood of adopting predictive analytics is monotonically increasing with the percentage of educated workers at the plant (Panel c), though confidence intervals for many of the coefficients overlap. A similar relationship applies for the practices related to monitoring operations using key performance indicators (Panel d); these coefficients are significant at the one percent level.

³¹ Quintiles were chosen to explore possible non-linearities and comply with the Census Bureau disclosure avoidance guidelines.

³² We explored potential non-linearities or alternative explanations at the bottom of the distribution, as well, probing whether these establishments rely primarily on other approaches such as pre-packaged software or outsourced IT services to support predictive analytics with less owned IT capital stock. However, we found no evidence for this in our sample.

While these conditional correlations suggest potential complementarity among predictive analytics and these workplace characteristics, we delve into adoption patterns more deeply using a richer multivariate model in Table 3. The dependent variable in column 1 is a binary indicator of the use of predictive analytics. Using a linear probability model that includes all of the potential workplace complements as well as a rich set of additional controls including size, age, and industry fixed effects, the results are largely consistent with the patterns identified in Figure 3. This correlation test is a crucial step in understanding whether complementarities are at work in our setting (Brynjolfsson and Milgrom 2013), and all four potential complements pass. We test for complementarity in the production process, below.

Columns 2 to 4 report results from specifications similar to column 1 but employ the use cases of predictive analytics as our dependent variables.³³ Organizational contingencies become apparent when we split by use case: predictive analytics in supply chain management and demand forecasting are more similar in the sense that all determinants in columns 2 and 3 have similar signs and magnitudes. The story is somewhat different for product design (see results in column 4): establishments with R&D focus and establishments that are also firm headquarters are more likely to adopt predictive analytics for product design, but not for other applications. In contrast, efficiency and production focused establishments (with continuous flow production processes) are more likely to adopt predictive analytics in supply chain management and demand forecasting, but not in product design. These results suggest that firms might be strategically adopting predictive analytics in specific use cases where they fit the most. This heterogeneity in adoption also leads us to further study the returns by use cases in the performance context.

Lastly, anticipating our exploration into the causal impact of predictive analytics on establishments' performance using instrumental variables, we find that plants whose data collection is mandated by the government are significantly more likely to adopt predictive analytics regardless of use case. This promising first-stage correlation will be useful for interpreting the results of the IV estimation, below.

³³ We omitted coefficients of some control variables in our reported tables to save space but detailed tables are available upon request. The correlations between various use cases of predictive analytics and potential complements are also explored and presented in figures in the appendix.

The Effect of Predictive Analytics on Establishment Performance

Table 4 explores the performance implications of predictive analytics. Following prior work (e.g., Brynjolfsson and Hitt 2003; Bloom et al. 2012), we estimate revenue-based total factor productivity (TFP), using the log sales as the dependent variable and controlling for production inputs such as non-IT capital stocks and the cost of labor, materials, and energy. We similarly include controls for plant-level characteristics that have been correlated with productivity in prior work such as multi-unit status, headquarter status, age, and data-driven decision making (Evans 1987; Majumdar 1997; Collis et al. 2007; Brynjolfsson and McElheran 2019).

All columns further include industry-year fixed-effects (at 6-digit NAICS level) to control for potential differences across industry and year (e.g., prices). It is important to understand how our approach differs in this regard from prior studies limited to less-granular controls such as 3-digit NAICS (e.g., NAICS 311 is the Food Manufacturing and 322 is Paper Manufacturing). These fine-grained industry definitions allow for comparisons within very-narrowly defined markets where products and technologies are often much more similar. For instance, NAICS 311351 is Chocolate and Confectionery Manufacturing from Cacao Beans versus 311352, which is Chocolate and Confectionary Manufacturing from Purchased Chocolate. NAICS 322212 is Folding Paperboard Box Manufacturing while 32213 is Setup Paperboard Box Manufacturing.

Results from column 1 indicate that the adoption of predictive analytics is associated with roughly 2.87 percent (significant at the 1% level) higher productivity, all else equal. Compared to the distribution of sales in our sample, this can be translated into an \$830,000 difference in the total value of shipments between adopters and non-adopters, holding other factors constant. Column 2 adds in the potential complements.³⁴ This shrinks the coefficient for predictive analytics by almost half, consistent with an important role for organizational complements in the productivity of predictive analytics. That said, the coefficient for the extensive margin of predictive analytics use remains positive and significant at the 1% level. And, although the magnitude is much smaller, the economic impact remains large: the coefficient in column 2 suggests that the adoption of predictive analytics is associated with 1.45 percent higher productivity, commensurate

³⁴ We also include a structured management practice index based on MOPS data but exclude data-related management practices including KPI tracking and target setting (Brynjolfsson and McElheran 2019) and control for them separately in the performance tests of complementarity. See Bloom et al. (2019) for a detailed description of the index.

with approximately \$460,000 higher sales on average.³⁵

In column 3, we explore the intensive margin of predictive analytics use. The coefficient for the use frequency of predictive analytics is positive and significant at the one-percent level.³⁶ Based on the mean and standard error for the frequency measure of predictive analytics (1.12 and 1.06 respectively, see Table 1), this indicates that moving from yearly frequency to monthly, for instance, is associated with 0.89 percent higher productivity, equivalent to about \$284,000 higher sales.

Up to now, we have explored the pooled OLS regressions without considering possible endogeneity in plant adoption of predictive analytics. This could bias our estimate of the returns to predictive analytics upwards and lead to a spurious positive correlation between the two. To address this potential concern, we use government-mandated data collection as an instrument in an IV estimation in column 4.³⁷

Consistent with the results in the adoption models, the result from the first stage of our two-stage least square (2SLS) estimation shows that government-mandated data collection is indeed highly correlated with predictive analytics. In the second stage, the effect of predictive analytics on establishments' productivity is still large, positive, and statistically significant. The larger magnitude of the coefficient in this estimation is consistent with either a downward bias in our OLS model, for instance, due to measurement error, which has been found to present in the MOPS data (Bloom et al. 2019). Not mutually exclusive, this pattern is also consistent with strong local treatment effects, whereby plants most-receptive to the influence of the government mandate also experience the greatest productivity shift. This could arise, for instance, if a large mass of less-data-savvy and less-productive plants had to implement big shifts in data collection required by regulators that then spilled over into relatively larger productivity gains compared to plants that maintained high levels of both data sophistication and productivity, mandate or no (see, e.g., Porter

³⁵ The mean total value of shipment in our sample is around \$32 M.

³⁶ Recalled that we constructed this measure by first assigning numeric values from 0 to 4 to all adoption frequencies (i.e. *Never*, *Yearly*, *Monthly*, *Weekly*, and *Daily*), with higher values representing higher frequencies. In cases where plants report multiple frequencies, we default to the highest one.

³⁷ Similar to Brynjolfsson and McElheran (2019), we use the frequency of predictive analytics for the IV estimations to avoid potential complications with non-linear first stage estimation for the indicator of adoption of predictive analytics.

1991 or Ambec et al. 2013).³⁸

Another oft-cited concern in the IT productivity literature is reverse causality. If establishments with better performance are more likely to adopt predictive analytics, then this would confound our interpretation of the patterns in our data. To address this issue, we test the timing of the performance-enhancing effect through the adoption of predictive analytics. If the identified effect runs from the adoption of predictive analytics to better performance, we should observe the adoption before observing improved performance.

A practical hurdle to this approach in our data is that we do not observe the precise year when adoption takes place. We can only observe shifts along the extensive margin sometime in the 2010-2015 window. We exploit the slow-moving nature of many organizational features to extrapolate outside of our core sample window, constructing a panel from 2010-2016 with annual information on inputs and output using the ASM and the CMF databases. We fix the organizational complements and create time-invariant indicators for early and late adopters and then compare productivity for these two groups against plants that had not adopted by 2015.

Figure 4 plots the coefficients for the early and late adopter indicators in the performance model between 2010 and 2016. The patterns of the two plot lines are quite consistent with our hypothesized causal relationship. Early adopters perform significantly better from the start of our panel and retain their advantage vis-a-vis non-adopters through 2016. The performance for plants identified as late adopters is not significantly different from that of non-adopters through 2013. We know that these plants all adopted in the 2010-2015 window, but not precisely when. Consistent with steadily increasing diffusion over time (a well-established pattern in the technology diffusion literature (Hall 2003 and 2004)), performance in this group begins to rise and is significantly different from non-adopters on average by 2014 and remains statistically indistinguishable from the effects for early adopters throughout 2016.³⁹ It is noteworthy that the late adopters start to close the performance gap with early adopters, due in part to declining

³⁸ Another possibility could be that the IV estimation suffers an omitted variable bias if the controls in our IV specification (e.g. costs of labor, capital, material, energy, and IT capital) do not account for the cost of mandatory data collection. To address this concern, we leverage the rich information in the ASM and conduct a robustness test adding other controls such as cost of contract workers and cost of professional services. The results are robust and consistent to those in column (4) and available up on request.

³⁹ The adoption rate for later adopters should reach 100% by 2015 assuming no withdrawn after the adoption.

performance advantages for the early adopters towards the end of the sample. One potential explanation is that competitive convergence in these practices for deriving value from data erodes excess returns over time, but that technical productivity persists at higher levels compared to plants that never employ these techniques. Unpacking these details is beyond the scope of our study. Nevertheless, the overall patterns identified in this figure provide supporting evidence against reverse causality.⁴⁰

[Insert Figure 4 here]

Before moving on to the performance tests of complementarity, we explore the robustness of these patterns to alternative measures for both the dependent and independent variables. Results are presented in Table 5. In column 1, we find that the significant positive effect of predictive analytics is robust when using the value-added as our output measure. Columns 2 and 3 report results from specifications that employ an alternative measure for the use frequency of predictive analytics (as discussed in the data section) using both the pooled OLS and the IV regressions. These results suggest that the identified performance-enhancing effect of predictive analytics is neither isolated to one particular output measure (i.e. sales) nor does it depend on a specific measure of predictive analytics.⁴¹

Not reported, we also explored the inclusion of plant-year fixed-effects. This is identical to a first-difference model since we only have two periods in the baseline sample. This yields a positive but noisy coefficient for predictive analytics, possibly due to reasons including high adoption rates by 2010, insufficient variations and lack of precision on adoption timing for late adopters, and lastly, the suppression on the estimated return of predictive analytics absorbing most time-invariant organizational complements.⁴²

⁴⁰ Regression results for figure 2 is available upon request.

⁴¹ We also examine the effects of predictive analytics on labor productivity and use a translog production function (instead of Cobb-Douglas form). In both cases, the findings are qualitatively and quantitatively similar to the results in our baseline model (pending disclosure review).

⁴² To take advantage of the annual input and output data (in between 2010 and 2015), we employ a sample similar to the panel for the timing test from 2010 to 2015. Then we re-run the establishment and year fixed-effects model assuming an exact adoption year for all late-adopters (e.g. assuming all of them adopted in 2011, or 2012, and so forth) since the precise year of adoption is unobservable (by us). By providing additional information on annual inputs and output we achieve more precision in the establishment and year fixed-effects models. The magnitude and the precision of the identified effect for the adoption of predictive analytics increase over time as the percentage of the adoption increases, consistent to those in the timing test. Results are presented in the appendix figure A10.

Variation in Effects by Use Case

Anecdotal evidence and case studies have shown large return heterogeneity across the applications of predictive analytics such as supply chain management, demand forecasting, among many others (Schlegel 2015; LaRiviere et al. 2016). However, large scale empirical evidence is severely lacking. Filling in this gap, we explore the differential effects of predictive analytics accounting for its use case. We estimate the performance regression using the measures for predictive analytics in supply chain management, demand forecasting, and product design respectively. These results are presented in Table 6. The coefficient for the adoption of predictive analytics in supply chain management is the largest and the most precisely estimated one among all three. The coefficient for predictive analytics in demand forecasting is smaller, but stays positive and significant (at 10% level), while the one for product design is close to zero and noisy. These results indicate that most of the identified effect of predictive analytics on performance likely comes from its uses in supply chain management and demand forecasting for the manufacturing plants.

We further conduct a robustness test to explore the differential returns of predictive analytics across the use cases conditional on whether the establishment belongs to a multi-unit firm. If the majority of benefits for predictive analytics are derived by its capability to lower uncertainty and improve efficiency for decision-making primarily in operation, logistics, and coordination, the establishments from multi-unit firms would benefit more as they are likely to have higher coordination costs in the supply chain for both internal and external markets. In contrast, it is unlikely that predictive analytics in product design would have the same effect. Empirically, we interact the use cases of predictive analytics with an indicator for establishments belonging to the multi-unit firms in our baseline model. The results are presented in Table 7. Consistent with our expectation, predictive analytics in supply chain management and demand forecasting are significantly more productive in the establishments from multi-unit firms while the adoption of predictive analytics in product design provides no such effect. These results are consistent with anecdotal evidence reporting wide adoptions and large benefits of predictive analytics in these areas (Waller and Fawcett 2013; Schlegel 2015; LaRiviere et al. 2016) and provide explanations for findings in Labro et al. (2019) where the adoption of predictive analytics is correlated with lower inventory and a more focused product mix (fewer products), and thus strengthen the causal

interpretation of the identified effect.⁴³

Formal Performance Test for Complementarity

Up to now, we have focused on the average effect of predictive analytics on productivity and presented a collage of evidence to argue that the relationship is causal. However, this does not explain the heterogeneous returns discussed in the anecdotal data or address the widely-discussed hurdles to a successful deployment of this frontier practice. Organizational complements feature significantly in the prior IT productivity literature (Black and Lynch 2001; Bresnahan et al. 2002; Tambe 2014; Brynjolfsson and McElheran 2019). The richness of our new survey and the linked administrative data allows us to make significant progress on this front.

Following Brynjolfsson and Milgrom (2013), we explore the extent to which the factors that help explain the adoption of predictive analytics are associated with disproportional returns on the practice, as complementarity theory would predict.

The results are presented in Table 8. Overall, all four potential complements pass the performance tests as all interaction terms are positive and significant. Joint effects combining the direct effect of predictive analytics with the interaction terms are also reported in the table. The returns to predictive analytics are significantly higher in the presence of complementary workplace features such as high IT capital stock (column 1), a high percentage of employees with a Bachelor's degree (column 2), a production strategy focused on high-efficiency continuous-flow production (column 3) and the top category for the number of KPIs tracked at the plant (column 4). Notably, not only is the magnitude of these joint effects consistently larger than those of the effects for predictive analytics without complements, the differences are also statistically significant.

To present these results more intuitively, Figure 5 provides a visual representation of the average effects with and without key complements. The y-axis indicates the magnitude of the coefficients and x-axis labels indicate the categories for each complement (e.g. predictive analytics with high IT capital stock and without high IT capital stock).⁴⁴ Confidence intervals (at 95%) are

⁴³ Overall, we believe that the variation in the effects of predictive analytics by use case is interesting and also has important implications for organizational practices. However, we recognize that more details on implementation are yet needed to disentangle the underlying mechanisms and are beyond the reach of the data in this particular study. We document these results and leave further work for future research.

⁴⁴ To highlight the difference, we further use the size of the plotted dots to represent the values of the marginal effects (large sizes represents higher values of marginal effect).

plotted to indicate statistical significance. The striking pattern that emerges from this figure is that the marginal effects of predictive analytics are never statistically different from zero, *unless* they are combined with these other tangible and intangible workplace investments. These results not only provide evidence in support of complementarities, but they provide clear boundary conditions on the phenomenon and practical guidance for managers of organizations considering these practices.

[Insert Figure 5 here]

6. Conclusion

With the explosion of digital information and substantial growth in business expenditure on data and analytics, it is more crucial than ever to understand the magnitude and mechanism of effects of these practices on business performance. Although compelling anecdotal and small-sample evidence exists that predictive analytics is associated with improved performance in a number of settings, stories of frustration and unrealized potential also abound. Large sample microdata, controls for a growing list of potential confounds, and support for causal interpretations has heretofore been lacking. Detailed understanding of complementarities with workplace investments in tangible and intangible infrastructure and management practices has been even more difficult to come by, even in smaller studies, due to data limitations.

To address these gaps, we worked with the U.S. Census Bureau to field a purpose-designed survey, resulting in large-scale on both data-related practices and key organizational characteristics in a sample designed to be representative of the U.S. manufacturing sector. Contrary to popular belief, this sector has historically been a leading adopter of frontier technologies, and continues to be so (Zolas et al. 2020), and is one of the longest-standing contexts for economic research. Thus, our inferences are likely applicable to a large distribution of firms with relatively well-understood economic dynamics. This is critical at a time of widespread enthusiasm for data and digital technologies, a high ratio of hype to evidence, concerns over what digitalization implies for the “future of work,” and no sign of decreasing technological advancement.

We find that plants reporting use of predictive analytics show over two to four percent higher productivity on average, which is worth over \$800,000 in increased sales for the average plant in our sample. This effect is statistically significant, economically large, and robust to different ways

of measuring both the dependent and explanatory variables.

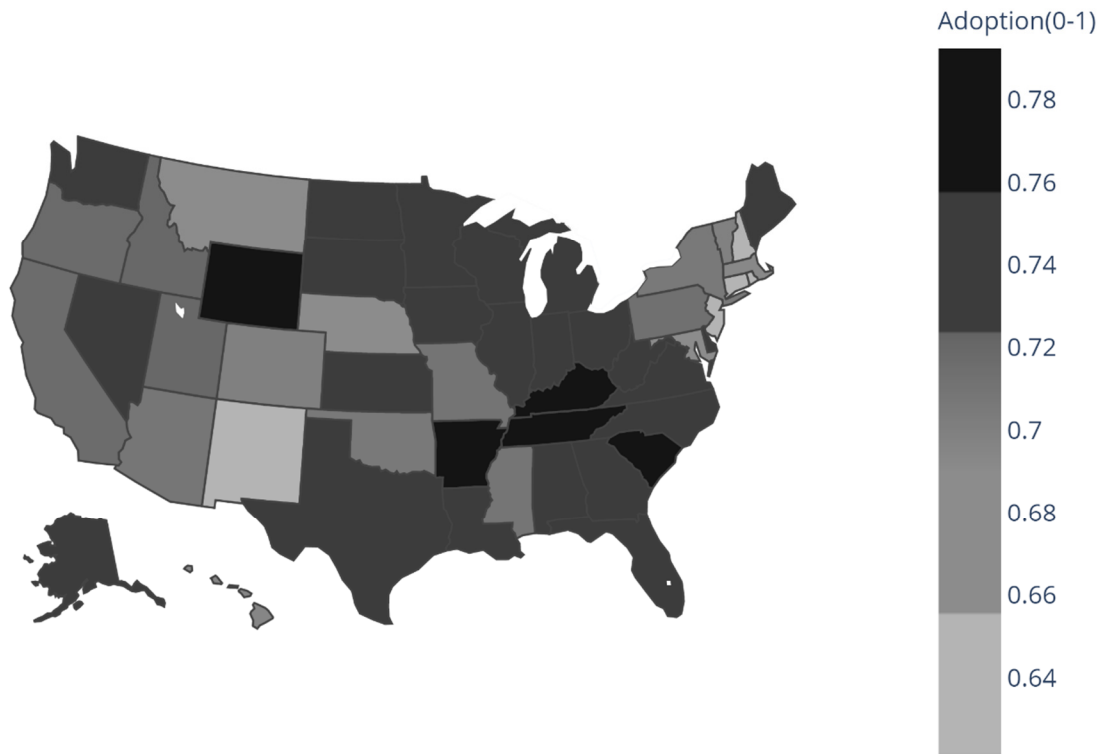
Strict causality in productivity studies of IT and related workplace practices is notoriously difficult to come by. We make some progress by collecting plant-level information on government mandates for data collection as a plausibly exogenous driver of data-related practices aimed at extracting value from digital information-- including predictive analytics.

A risk in studies such as ours is reverse causality: do predictive analytics *cause* higher performance, or are higher performing firms simply more likely to adopt predictive analytics for other reasons. We address this issue not only through the quasi-experiments created via our instrument variable analyses, but also by testing the performance gap among early, later adopters, and non-adopters by 2015 in a panel setting. Our findings show that early adopters have a performance premium since 2010 while later adopters catch up during the 2010-2015 period as more and more of them converted. The timing of these effects works against the possibility of reverse causality. Therefore, we contribute to the literature by being among the first to provide evidence that the effect of predictive analytics on performance is likely causal in a large and representative sample. Moreover, we document the heterogeneous returns among use cases of predictive analytics within a manufacturing setting. That is, the positive effect of predictive analytics on establishments' productivity comes primarily from its uses in supply chain management and demand forecasting. This is largely consistent with the anecdotal evidence and further supports our causal interpretation for the productivity-enhancing effect of predictive analytics.

In addition, prior studies find that workers with complementary skills are essential in realizing the performance-enhancing effect of data analytics (Tambe 2014; Wang et al. 2019). However, few have systematically examined organizational complements for predictive analytics. We contribute to the literature by exploring the heterogeneous returns of predictive analytics and emphasizing how they vary by both tangible and intangible complementary investments. We identify that IT capital stock, educated workers, continuous-flow production process, and practice of monitoring KPI are complements for predictive analytics. Establishments are more likely to adopt predictive analytics with the presence of these complements and enjoy a significant productivity premium post-adoption. Our results strongly support our hypotheses and suggest that predictive analytics are significantly more productive when properly aligned with complementary

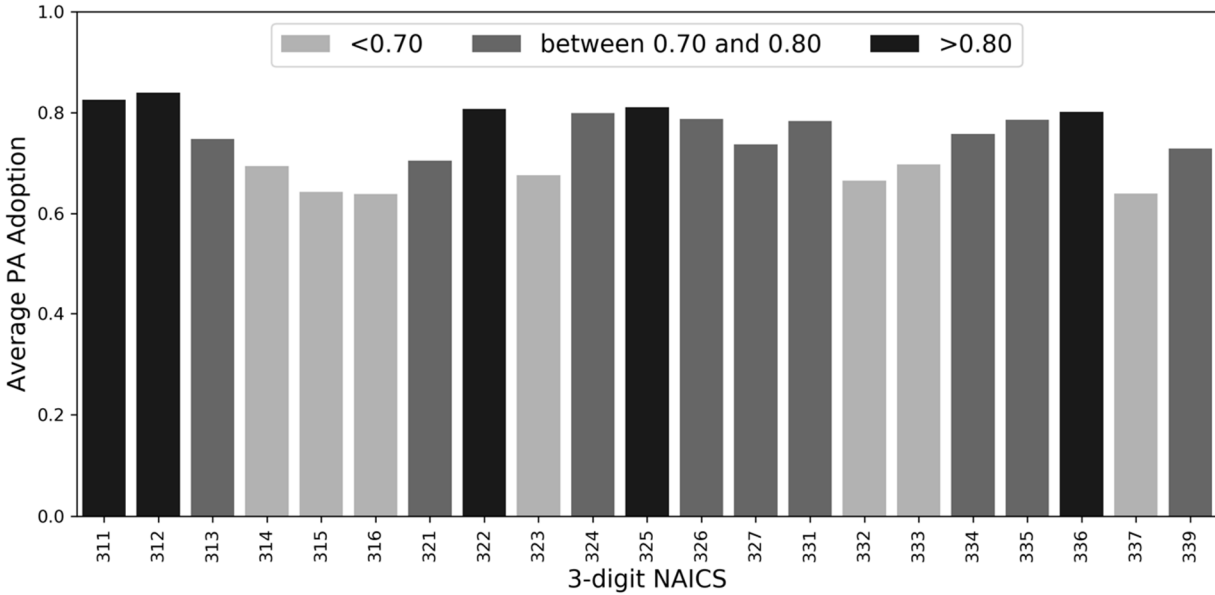
organizational settings. These findings build on prior work exploring complementarities between organizational characteristics and IT (e.g., Bresnahan et.al. 2002, Brynjolfsson and Milgrom 2009) and provide a foundation for practical insights into the mechanisms by which predictive analytics can better provide business value.

**Figure 1. Adoption of Predictive Analytics by State
(US Manufacturing in 2010)**



Notes: Reported statistics in the legend are the average adoption of predictive analytics across the states in the United States based on the baseline sample in 2010. This sample consists of establishments in 2015 MOPS samples (with 2010 recall) that can be merged with the Annual Survey of Manufactures (ASM), Census of Manufactures (CMF), and the Longitudinal Business Database (LBD), excluding administrative records, non-tabbed observations, and plants with negative value-added. Please see the data section for more details about the sample selection criterion. Darker color indicates a higher average adoption among the establishments within a particular state. The adoption patterns by states are similar to this figure if either balanced or baseline 2015 data are used.

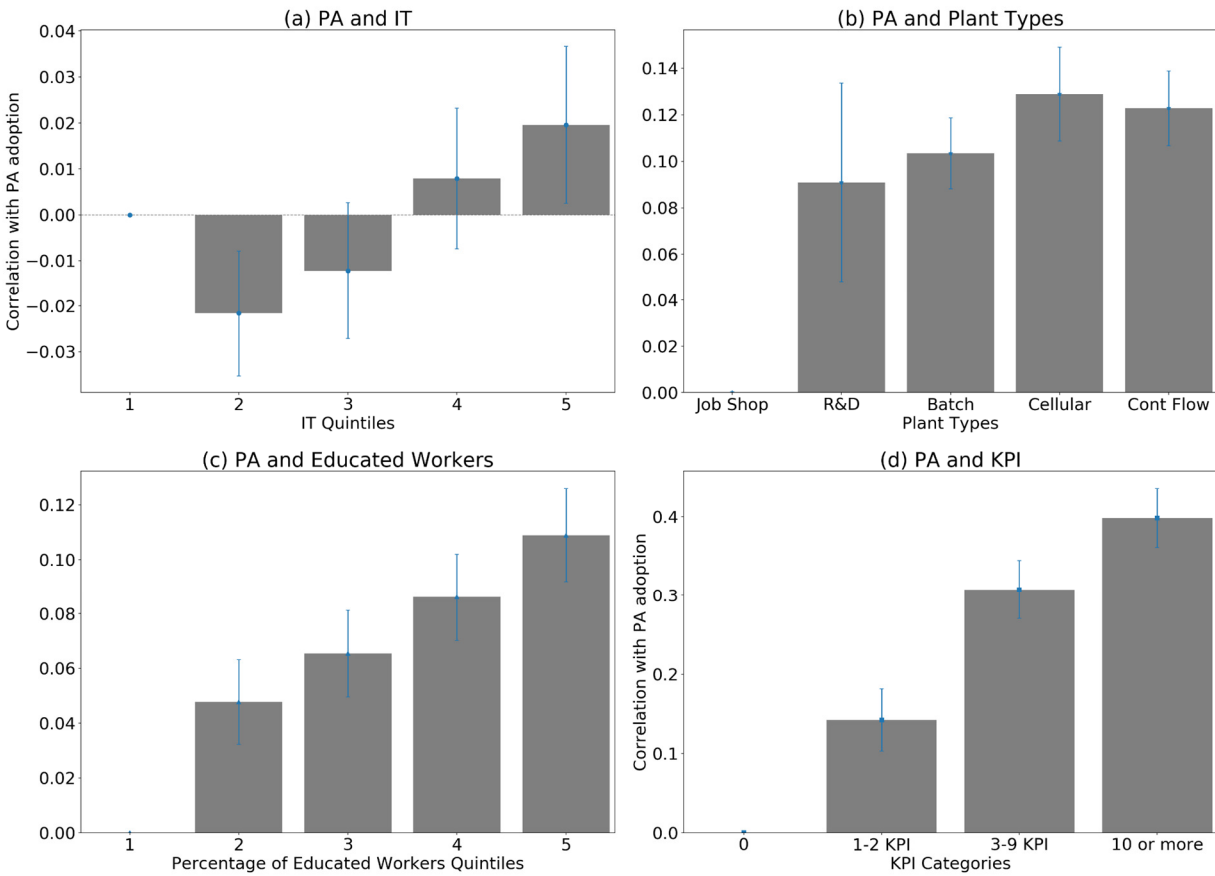
**Figure 2. Adoption of Predictive Analytics by Industry
(US Manufacturing in 2010)**



Notes: Reported statistics are based on the baseline sample for the adoption of predictive analytics in 2010. The average adoption rate is shown on the Y-axis. The 3-digit NAICS codes are shown in the X-axis and the corresponding industry definitions are listed in the table below. Darker color indicates higher average adoption for a particular industry. The rankings across industries for PA adoption are similar to the figure above if either balanced or baseline 2015 data are used. Detailed statistics are provided in the online Appendix table A2.

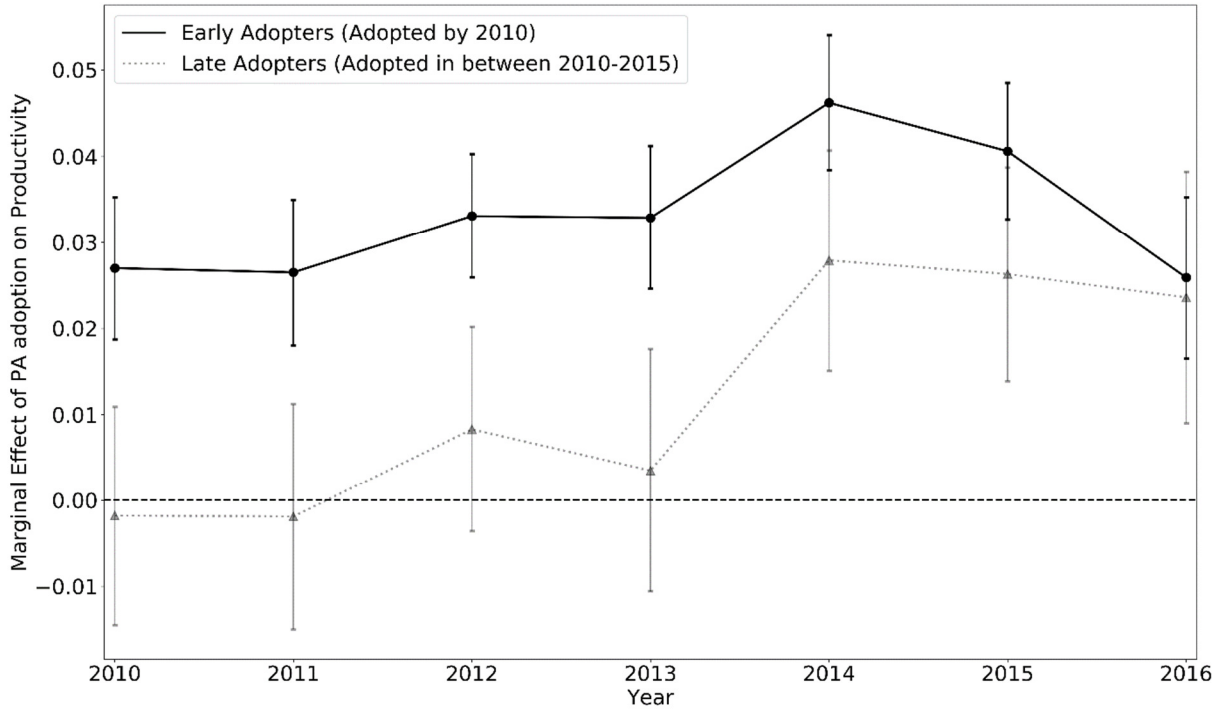
NAICS 3	Industry Definition	NAICS 3	Industry Definition
311	Food Manufacturing	326	Plastics and Rubber Products
312	Beverage and Tobacco Product	327	Nonmetallic Mineral Product
313	Textile Mills	331	Primary Metal
314	Textile Product Mills	332	Fabricated Metal Product
315	Apparel Manufacturing	333	Machinery
316	Leather and Applied Product	334	Computer and Electronic Product
321	Wood Product	335	Electrical Equipment, Appliance, and Component
322	Paper	336	Transportation Equipment
323	Printing and Related Support Activities	337	Furniture and Related Product
324	Petroleum and Coal Products	339	Miscellaneous Manufacturing
325	Chemical		

Figure 3. Conditional Correlations between Predictive Analytics and Potential Complements



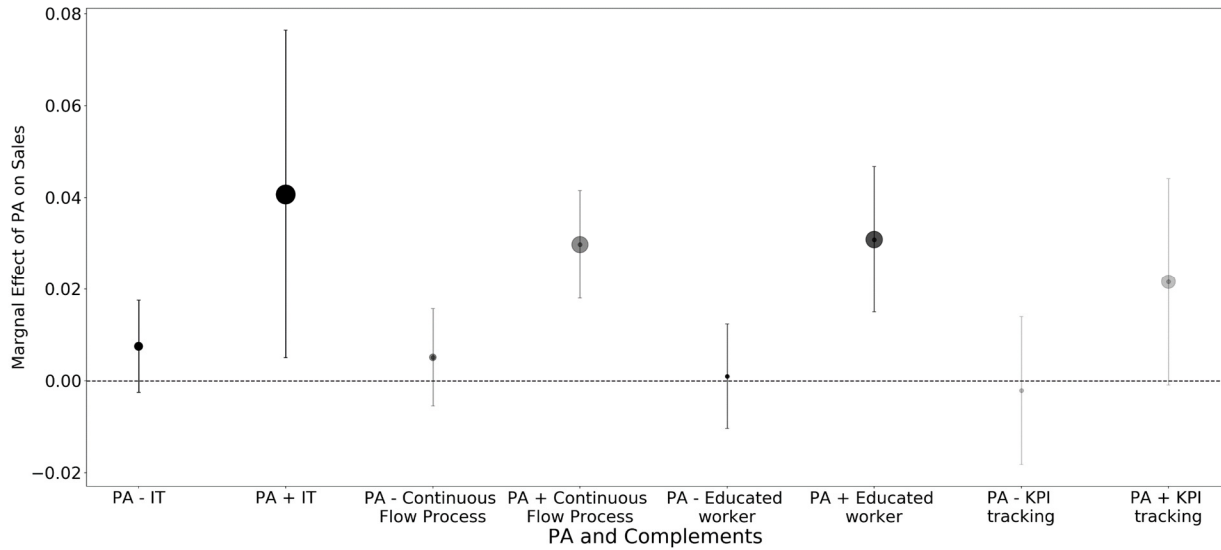
Notes: Estimates based on the baseline sample from the pooled OLS regressions controlling for plant size, plant age, and industry (6-digit NAICS) and year fixed-effects. The dependent variable is the adoption of predictive analytics. The base group is the plants in the bottom quintile of the sample based on calculated total IT capital stock for the upper left figure. For the upper right figure, the base group is the bottom quintile of the sample based on the percentage of educated workers. For the bottom figures, the based groups are plants tracking zero KPIs (left) and job shop plants (right) respectively. Histogram bars (and values on the Y-axis) represent the differences in the adoption of predictive analytics between the bottom quintile and higher quintiles (or other categories). The number of quintiles and names of the categories are labeled on the X-axis (the base group has zero value). Quintiles are used for the US Census disclosure avoidance practice and consistency across figures. Standard errors for coefficients are plotted on the histogram bars with blue lines.

Figure 4. Effects of Predictive Analytics by Early and Late Adopters over Time



Notes: Estimates based on a pooled OLS model with a specification identical to the baseline model in column 2 table 2. For this test, we construct an ASM and CMF panel, where we have annual data on most of the key inputs (except the managerial related variables from the 2015 MOPS) and sales from 2010 to 2016. We identify the groups of establishments that adopted predictive analytics by 2010, establishments that adopted predictive analytics between 2010 and 2015, and the rest of laggards (non-adopters) using the 2015 MOPS data. These indicators are then interacted with year dummies to explore the differences in sales over time (using laggards as the baseline group). Histogram bars (and values on the Y-axis) represent the marginal effect of predictive analytics adoption between 2010 to 2016. Standard errors of the coefficients are plotted on the histogram bars.

Figure 5. Effects of Predictive Analytics with and without Potential Complements on Performance



Notes: Estimates based on a pooled OLS model with a specification similar to the baseline model in column 2 table 2 using the baseline sample. For performance tests of complementarity, we add the interaction between the indicator for the adoption of predictive analytics and the potential complements including high IT capital stock, having continuous flow process, high percentage of employees with bachelors’ degree, and intensive KPI tracking respectively. The coefficients for the indicator of predictive analytics and the interaction term help us identify the differential effects of the adoption of predictive analytics on sales conditional on the presence of the complement. Histogram bars (and values on the Y-axis) present the marginal effect of PA adoption for the groups of plants with and without the presence of complement. To highlight the difference, we further use the size of the plotted dots to represent the values of the marginal effects (large sizes represents higher values of marginal effect). The error bars indicate the confidence intervals at the 95% level.

Table 1. Summary Statistics (Baseline Sample)

Variables	Definition	Mean (S.D.)
Log Sales	Plant total value of shipment in log terms (\$Thousands)	10.37 (1.52)
Log TE	Total number of employees in log term	4.56 (1.17)
PA Adoption	Indicator for plants that adopted predictive analytics (question 29 in MOPS 2015)	0.74 (0.44)
PA Adoption (Top Frequency)	Frequency of PA application based on the top value of the responses (e.g. Yearly=1, Monthly=2, Weekly=3, and/or Daily=4) ⁴⁵	1.12 (1.06)
Mandated Data Collection	Indicator for plants that are required to collect data by government regulations or agencies	0.25 (0.43)
Log IT Capital Stock	IT capital stock in log terms (\$Thousands)	5.16 (2.41)
Log Non-IT Capital Stock	Traditional capital stock on non-IT equipment and structure in log term (\$Thousands)	9.26 (1.47)
High KPI Tracking	Indicator for plants that tracked more than 10 KPIs (Top category of question 2 in MOPS questionnaire)	0.44 (0.50)
Employee Education	Percentage of employees with a bachelor's degree	0.15 (0.14)
Continuous-Flow Production	Indicator for plants with continuous flow production process	0.35 (0.48)
AGE	Plant age	24.47 (12.89)
MU	Indicator for establishments belonging to Multi-unit firms	0.73 (0.45)
HQ	Indicator for establishments reported being HQ or co-located with HQ	0.47 (0.50)
Z_MGMT	A normalized index for structured management practices using section A in the 2015 MOPS (excluding data related questions 2 and 6)	0.63 (0.17)
Number of Observations		51,000

Notes: Reported statistics are based on the baseline sample from MOPS 2015 data; Standard deviations in parentheses.

⁴⁵ For instance, we assigned the value for monthly to a particular plant if the respondent at the plant reported having both yearly and monthly uses of predictive analytics.

Table 2. Summary Statistics (2010 and 2015 Balanced Sample)

Variables	Definition	2010 Mean (S.D.)	2015 Mean (S.D.)
Log Sales	Plant total sales in log terms (\$Thousands)	10.68 (1.39)	10.86 (1.37)
Log TE	Total number of employees in log term	4.79 (1.09)	4.88 (1.09)
PA Adoption	Indicator for plants that adopted predictive analytics (question 29 in MOPS 2015)	0.73 (0.44)	0.80 (0.40)
Log IT Capital Stock	IT capital stock in log terms (\$Thousands)	5.58 (2.25)	5.62 (2.18)
Log Non-IT Capital Stock	Traditional capital stock on non-IT equipment and structure in log term (\$Thousands)	9.38 (1.58)	9.36 (1.61)
High KPI Tracking	Indicator for plants that tracked more than 10 KPIs (Top category of question 2 in MOPS questionnaire)	0.37 (0.48)	0.56 (0.50)
Employee Education	Percentage of employees with a bachelor's degree	0.15 (0.13)	0.16 (0.14)
Continuous Flow Production	Indicator for plants with continuous flow production process	0.38 (0.48)	0.41 (0.49)
Number of Observations Per Year		18000	

Notes: Reported statistics are based on the strongly balanced sample from MOPS 2015 data. Standard deviations in parentheses.

Table 3. Adoption of Predictive Analytics by Use Case (Correlation Test)

Models	(1)	(2)	(3)	(4)
	Linear Probability Model			
Dependent Variables	PA Adoption	PA Adoption (SCM)	PA Adoption (DF)	PA Adoption (PD)
Log IT Capital Stock	0.0034*** (0.0012)	0.0046*** (0.0012)	0.0039*** (0.0012)	0.0043*** (0.0014)
Employee Education	0.1281*** (0.0197)	0.1495*** (0.0198)	0.1393*** (0.0202)	0.1860*** (0.0210)
High KPI Tracking	0.0542*** (0.0049)	0.0610*** (0.0052)	0.0578*** (0.0051)	0.0578*** (0.0056)
Continuous Flow Production	0.0137*** (0.0053)	0.0165*** (0.0056)	0.0121*** (0.0056)	0.0098 (0.0062)
Mandated Data Collection	0.0189*** (0.0053)	0.0260*** (0.0055)	0.0222*** (0.0055)	0.0259*** (0.0062)
Log TE	0.0256*** (0.0033)	0.0304*** (0.0035)	0.0261*** (0.0035)	0.0311*** (0.0038)
AGE	-0.0009*** (0.0002)	-0.0009*** (0.0002)	-0.0008*** (0.0002)	-0.0010*** (0.0002)
MU	0.0144* (0.0082)	0.0235*** (0.0082)	0.0208** (0.0083)	0.0164* (0.0084)
HQ	-0.0245*** (0.0061)	-0.0201*** (0.0063)	-0.0124* (0.0063)	0.0153** (0.0067)
R&D focus	0.0253 (0.0202)	0.0257 (0.0206)	0.0216 (0.0205)	0.0465** (0.0208)
Industry x Year FX			Y	
Adjusted R-Squared	0.1428	0.1548	0.1365	0.1179
Number of Observations	51,000			

Notes: Estimates based on linear probability models controlling for industry (6-digit NAICS) and year fixed effects using the baseline sample. The dependent variables across columns 1-4 are the adoption of predictive analytics, and its use cases in the SCM, DF, and PD respectively. Log IT Capital Stock is accumulated and depreciated stock of IT capital expenditure at the plant in log terms. Employee Education is the percentage of employees with a bachelor's degree. High KPI Tracking is an indicator for plants that track 10 or more of KPIs based on question 2 in the 2015 MOPS survey. Continuous Flow Production is an indicator for plants whose production process is best characterized as continuous flow manufacturing. The omitted possible production categories are job shop and batch production. Mandated Data Collection is an indicator for plants that data collection that is mandated by government regulation or agencies. The unreported controls include indicators for the manager-to-non-manager employee ratio, non-IT capital stock in log terms, and average adoption of predictive analytics by other production units within the firm (excluding the focal plant). We also control an indicator for establishments having top-quartile structured management index based on variable Z_MGMT in table 1. We follow Bloom et al. (2019) to calculate index using questions in section A (excluding the data related including questions 2 and 6). Robust standard errors are clustered at the firm level. * p<0.10, ** p<0.05, *** p<0.01. Results are robust to using binary estimation models (e.g., probit in Stata 15).

Table 4. The Effect of Predictive Analytics on Plant Performance

Models	(1) OLS (Basic)	(2) OLS (Full)	(3) OLS (Frequency)	(4) IV (2SLS)
Dependent Variables	Log Sales			
PA Adoption	0.0287*** (0.0049)	0.0145*** (0.0049)		
PA Adoption Frequency (top frequency)			0.0089*** (0.0021)	0.0509*** (0.0160)
Log IT Capital Stock		0.0227*** (0.0014)	0.0227*** (0.0014)	0.0226*** (0.0014)
Employee Education		0.2068*** (0.0177)	0.2063*** (0.0177)	0.1928*** (0.0181)
High Structure Mgmt.		0.0283*** (0.0054)	0.0273*** (0.0054)	0.0165** (0.0069)
Continuous Flow Production		0.0407*** (0.0052)	0.0406*** (0.0052)	0.0385*** (0.0053)
Log TE	0.4052*** (0.0065)	0.3821*** (0.0063)	0.3820*** (0.0063)	0.3800*** (0.0063)
Log non-IT Capital Stock	0.0614*** (0.0031)	0.0616*** (0.0029)	0.0615*** (0.0029)	0.0607*** (0.0029)
MU	0.0322*** (0.0061)	0.0296*** (0.0060)	0.0295*** (0.0060)	0.0239*** (0.0063)
HQ	-0.0732*** (0.0055)	-0.0815*** (0.0055)	-0.0809*** (0.0055)	-0.0739*** (0.0059)
Mandated Data Collection (First Stage)				0.3219*** (0.0166)
Under-identification Test				287.9
Weak-identification Test				1028
Industry x Year Fixed Effects			Y	
R-Squared	0.9313	0.9327	0.9340	0.8794
Number of Observations	51,000			

Notes: Estimates based on the pooled OLS models controlling industry (6-digit NAICS) and year fixed effects using the baseline sample. The dependent variable is logged sales. Columns 1 controls for key production inputs while column 2 adds controls for IT capital stock, structure management, and percentage of employees with a bachelors' degree. Columns 3 examines the effect of predictive analytics measuring in frequency with all controls. Lastly, column 4 employs IV estimation to address the potential endogeneity of the adoption of predictive analytics. Mandated Data Collection is used as the IV for predictive analytics adoption. Unreported controls for all columns include logged cost of material and energy, plant age, and an indicator for data-driven-decision making practices. Robust standard errors clustered at the firm level. * p<0.10, ** p<0.05, *** p<0.01.

Table 5. The Effect of Predictive Analytics on Plant Performance (Robustness)

Models	(1)	(2)	(3)
	OLS	OLS PA Frequency (Average)	IV PA Frequency (Average)
Dependent Variables	Log Value Added	Log Sales	
PA Adoption	0.0223** (0.0091)		
PA Adoption Frequency (Average)		0.0023** (0.0006)	0.0167*** (0.0053)
Log IT Capital Stock	0.0491*** (0.0026)	0.0228*** (0.0014)	0.0228*** (0.0014)
Employee Education	0.4670*** (0.0351)	0.2076*** (0.0177)	0.1978*** (0.0178)
High Structure Mgmt.	0.0534*** (0.0109)	0.0269*** (0.0054)	0.0108 (0.0082)
Continuous Flow Production	0.0786*** (0.0107)	0.0408*** (0.0052)	0.0392*** (0.0052)
Log TE	0.7796*** (0.0081)	0.3820*** (0.0063)	0.3793*** (0.0063)
Log non-IT Capital Stock	0.1242*** (0.0051)	0.0615*** (0.0029)	0.0609*** (0.0028)
MU	0.0962*** (0.0112)	0.0297*** (0.0060)	0.0238*** (0.0064)
HQ	-0.1519*** (0.0107)	-0.0810*** (0.0055)	-0.0726*** (0.0061)
Mandated Data Collection (First Stage)			0.9802*** (0.0650)
Under-identification Test Statistic			186.7
Weak identification test statistic			611.3
Industry x Year Fixed Effects		Y	
R-Squared	0.7395	0.9327	0.8783
Number of Observations		51,000	

Notes: Estimates based on the pooled OLS models controlling industry (6-digit NAICS) and year fixed effects using the baseline sample. The dependent variable for column 1 is the logged value-added. The dependent variable for columns 2 and 3 is logged sales. PA Adoption Frequency is an alternative measure for the frequency of predictive analytics adoption using the average of the multiple choices in question 29 of MOPS 2015 (instead of top counted). Unreported controls for column 1 include plant age and an indicator for data-driven-decision making practices. Unreported controls for columns 2 and 3 include logged cost of material and energy, plant age, an indicator for data-driven-decision making practices. Robust standard errors clustered at the firm level. * p<0.10, ** p<0.05, *** p<0.01.

Table 6. The Effect of Predictive Analytics on Plant Performance (by Use Cases)

Models	(1)	(2)	(3)
	PA Adoption (SCM)	PA Adoption (DF)	PA Adoption (PD)
Dependent Variables	Log Sales		
PA for Supply Chain Management (SCM)	0.0116** (0.0047)		
PA for Demand Forecasting (DF)		0.0086* (0.0048)	
PA for Product Development (PD)			-0.0012 (0.0045)
Log IT Capital Stock	0.0227*** (0.0014)	0.0227*** (0.0014)	0.0227*** (0.0014)
Employee Education	0.2070*** (0.0177)	0.2077*** (0.0177)	0.2094*** (0.0177)
High Structured Mgmt.	0.0284*** (0.0054)	0.0288*** (0.0054)	0.0297*** (0.0055)
Continuous Flow Production	0.0407*** (0.0052)	0.0408*** (0.0053)	0.0410*** (0.0053)
Other inputs		Y	
Industry x Year Fixed Effects		Y	
R-Squared	0.9327	0.9327	0.9327
Number of Observations	51,000		

Notes: Estimates based on the pooled OLS models controlling industry (6-digit NAICS) and year fixed effects using the baseline sample. The dependent variable for all columns is logged sales. PA Adoption (SCM) is an indicator for plants that reported to have adopted predictive analytics and applied data analytics in supply chain management. Similarly, PA Adoption (DF) is an indicator for plants that reported to have adopted predictive analytics and applied analytics in demand forecasting. Finally, PA Adoption (PD) is an indicator for plants that reported to have predictive analytics and applied analytics in product design. Unreported controls for all columns include logged cost of material and energy, plant age, an indicator for data-driven-decision making practices, and indicators for plants belong to multi-unit firms and plants reported as HQ or co-located with HQ. Robust standard errors clustered at the firm level. * p<0.10, ** p<0.05, *** p<0.01.

**Table 7. Variation in Effects of Predictive Analytics by Use Cases
(Single Unit Firms vs Establishments from Multi-unit Firms)**

Models	(1)	(2)	(3)
	PA Adoption (SCM)	PA Adoption (DF)	PA Adoption (PD)
Dependent Variables	Log Sales		
PA Adoption (SCM)	-0.0094 (0.0074)		
PA Adoption (DF)		-0.0102 (0.0073)	
PA Adoption (PD)			-0.0113 (0.0073)
MU	0.0091 (0.0083)	0.0143* (0.0084)	0.0222*** (0.0080)
PA Adoption (SCM) X MU	0.0333*** (0.0093)		
PA Adoption (DF) X MU		0.0247*** (0.0093)	
PA Adoption (PD) X MU			0.0146 (0.0090)
Other inputs	Y		
Industry x Year Fixed Effects	Y		
R-Squared	0.9327	0.9327	0.9327
Number of Observations	51,000		

Notes: Estimates based on the pooled OLS models controlling industry (6-digit NAICS) and year fixed effects using the baseline sample. The dependent variable for all columns is logged sales. PA Adoption SCM is an indicator for plants that reported to have adopted predictive analytics and applied data analytics in supply chain management. Similarly, PA Adoption DF is an indicator for plants that reported to have adopted predictive analytics and applied analytics in demand forecasting. Finally, PA Adoption PD is an indicator for plants that reported to have predictive analytics and applied analytics in product design. MU is an indicator for establishments belonging to multi-unit firms. Unreported controls for all columns include logged total number of employees, logged non-IT and IT capital stock, logged cost of material and energy, plant age, an indicator for data-driven-decision making practices, an indicator for plants reported as HQ or co-located with HQ, and indicators for establishments with high structured management practice and establishments with continuous flow process. Robust standard errors clustered at the firm level. * p<0.10, ** p<0.05, *** p<0.01.

Table 8. Organizational Complements to Predictive Analytics (Performance Test)

Models	(1) IT Capital Stock	(2) Educated Employee	(3) Continuous Flow	(4) KPI Tracking
Dependent Variables	Log Sales			
PA Adoption	0.0075 (0.0051)	0.0051 (0.0054)	0.0010 (0.0058)	-0.0021 (0.0082)
High IT Capital Stock	0.1103*** (0.0171)			
PA × High IT Capital Stock	0.0333* (0.0185)			
High Percentage of Educated Employee		0.0371*** (0.0101)		
PA × High Percentage of Educated Employee		0.0228** (0.0114)		
Continuous Flow Production			0.0158* (0.0088)	
PA × Continuous Flow Production			0.0306*** (0.0094)	
High KPI Tracking				0.0019 (0.0090)
PA × High KPI Tracking				0.0237** (0.0101)
Joint Tests	0.0407** (0.0182)	0.0287*** (0.0106)	0.0309*** (0.0081)	0.0216** (0.0115)
Other Controls			Y	
Industry x Year Fixed Effects			Y	
R-Squared	0.9326	0.9327	0.9328	0.9328
Number of Observations	51,000			

Notes: Estimates based on pooled OLS models controlling industry (6-digit NAICS) and year fixed effects using the baseline sample. The dependent variable is logged sales. High IT Capital Stock is an indicator for plants with top ten percentile of IT capital stock. High Percentage of Educated Employee is an indicator for plants with top quartile of the percentage of employees with a bachelors' degree. High KPI Tracking is an indicator for plants that track top numbers of KPI based on question 2 in the 2015 MOPS survey. Continuous Flow Production is an indicator for plants whose production process is best characterized as continuous flow manufacturing. Columns 1-4 interact the indicator of adoption of predictive analytics with each of the potential complements while controlling for all inputs and other potential complements. Joint Tests report the calculated the coefficients of the adoption of predictive analytics with the presence of complements (using Lincom Joint Test in Stata 16). Unreported controls for all columns include logged total number of employees, log non-IT capital stock, logged cost of material and energy, plant age, indicators for data-driven-decision making practices other than KPI tracking, and indicators for plants belong to multi-unit firms and plants reported as Headquarters (HQ) or co-located with HQ. Robust standard errors clustered at the firm level. * p<0.10, ** p<0.05, *** p<0.01.

References

- Adner, R, Puranam P and Zhu F (2019) What is different about digital strategy? From quantitative to qualitative change. *Strategy Science*. 4(4): 253-261.
- Ambec, S., Cohen, M.A., Elgie, S. and Lanoie, P., 2013. The Porter hypothesis at 20: can environmental regulation enhance innovation and competitiveness?. *Review of environmental economics and policy*, 7(1), pp.2-22.
- Aral, S. and Weill, P., 2007. IT assets, organizational capabilities, and firm performance: How resource allocations and organizational differences explain performance variation. *Organization science*, 18(5), pp.763-780.
- 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.
- Arcidiacono, P., Bayer, P. and Hizmo, A., 2010. Beyond signaling and human capital: Education and the revelation of ability. *American Economic Journal: Applied Economics*, 2(4), pp.76-104.
- Arora, A. and Gambardella, A., 1990. Complementarity and external linkages: the strategies of the large firms in biotechnology. *The journal of industrial economics*, pp.361-379.
- Athey, S. and Stern, S., 1998. An empirical framework for testing theories about complementarity in organizational design (No. w6600). National Bureau of Economic Research.
- Babu, M. S. P. and S. H. Sastry, 2014. "Big data and predictive analytics in ERP systems for automating decision making process," *the IEEE 5th International Conference on Software Engineering and Service Science*, Beijing, 2014, pp. 259-262.
- Bennett, VM (2020) Changes in persistence of performance over time. *Strategic Management Journal*. 41(10): 1745-1769.
- Berman, E., Bound, J. and Griliches, Z., 1994. Changes in the demand for skilled labor within US manufacturing: evidence from the annual survey of manufactures. *The Quarterly Journal of Economics*, 109(2), pp.367-397.
- Black, S.E. and Lynch, L.M., 2001. How to compete: the impact of workplace practices and information technology on productivity. *Review of Economics and statistics*, 83(3), pp.434-445.
- Blackwell, D., 1953. Equivalent comparisons of experiments. *The annals of mathematical statistics*, pp.265-272.
- 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., Brynjolfsson, E., Foster, L., Jarmin, R.S., Saporta Eksten, I. and Van Reenen, J., 2013. Management in America. US Census Bureau Center for Economic Studies Paper No. CES-WP-13-01.
- Bloom, N., Sadun, R. and Van Reenen, J., 2016. Management as a Technology? (No. w22327). National Bureau of Economic Research.
- Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R., Patnaik, M., Saporta-Eksten, I. and Van Reenen, J., 2019. What drives differences in management practices?. *American Economic Review*, 109(5), pp.1648-83.
- 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. and Hitt, L., 1995. Information technology as a factor of production: The role of differences among firms. *Economics of Innovation and New technology*, 3(3-4), pp.183-200.

- Brynjolfsson, E. and Hitt, L.M., 2000. Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic perspectives*, 14(4), pp.23-48.
- Brynjolfsson, E. and Hitt, L.M., 2003. Computing productivity: Firm-level evidence. *Review of economics and statistics*, 85(4), pp.793-808.
- Brynjolfsson, E., Hitt, L.M. and Kim, H.H., 2011. Strength in numbers: How does data-driven decision-making affect firm performance?. Available at SSRN 1819486.
- Brynjolfsson, E., & McElheran, K. 2016. The Rapid Adoption of Data-Driven Decision-Making. *American Economic Review*. <https://doi.org/10.1257/aer.p20161016>
- Brynjolfsson, E., & McElheran, K. 2019. Data in Action: Data-Driven Decision Making and Predictive Analytics in U.S. Manufacturing. Rotman School of Management Working Paper No. 3422397. Available at SSRN: <https://ssrn.com/abstract=3422397>
- Brynjolfsson, E. and Mendelson, H., 1993. Information systems and the organization of modern enterprise. *Journal of Organizational Computing and Electronic Commerce*, 3(3), pp.245-255.
- 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 (No. w25148). National Bureau of Economic Research.
- Buffington, C., Foster, L., Jarmin, R. and Ohlmacher, S., 2017. The management and organizational practices survey (MOPS): An overview 1. *Journal of Economic and Social Measurement*, 42(1), pp.1-26.
- Card, D., 1999. The causal effect of education on earnings. In *Handbook of labor economics* (Vol. 3, pp. 1801-1863). Elsevier.
- Caroli, E. and Van Reenen, J., 2001. Skill-biased organizational change? Evidence from a panel of British and French establishments. *The Quarterly Journal of Economics*, 116(4), pp.1449-1492.
- Clark, K.B. and Margolis, J.D., 1991. Workplace safety at Alcoa (A).
- Côrte-Real, N., Ruivo, P. and Oliveira, T., 2014. The diffusion stages of business intelligence & analytics (BI&A): A systematic mapping study. *Procedia Technology*, 16, pp.172-179.
- Davenport, T.H., 2006. Competing on analytics. *Harvard business review*, 84(1), p.98.
- Dubey, R., Gunasekaran, A., Childe, S.J., Blome, C. and Papadopoulos, T., 2019. Big Data and Predictive Analytics and Manufacturing Performance: Integrating Institutional Theory, Resource-Based View and Big Data Culture. *British Journal of Management*, 30(2), pp.341-361.
- Duhigg, C., 2012. *The power of habit: Why we do what we do in life and business*. Random House.
- Dunne, T., 1994. Plant age and technology use in US manufacturing industries. *The RAND Journal of Economics*, pp.488-499.
- Gandhi, A., Navarro, S. and Rivers, D., 2017. How heterogeneous is productivity? A comparison of gross output and value added. *Journal of Political Economy*, 2017
- Enke, B., 2020. What you see is all there is. *The Quarterly Journal of Economics*, 135(3), pp.1363-1398.
- Forman, C., Goldfarb, A. and Greenstein, S., 2002. *Digital dispersion: An industrial and geographic census of commercial internet use* (No. w9287). National Bureau of Economic Research.
- Garicano, L., 2000. Hierarchies and the Organization of Knowledge in Production. *Journal of political economy*, 108(5), pp.874-904.
- Gorry, G.A. and Scott Morton, M.S., 1971. A framework for management information systems.
- Greenstein, S, Lerner J and Stern S (2013) Digitization, innovation, and copyright: What is the agenda?

Strategic Organization. 11(1): 110-121.

Grover, V., Chiang, R.H., Liang, T.P. and Zhang, D., (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, 35(2), pp.388-423.

Hannan, T.H. and McDowell, J.M., 1984. The determinants of technology adoption: The case of the banking firm. *The RAND Journal of Economics*, pp.328-335.

Hayes, R.H. and Wheelwright, S.C., 1979. Link manufacturing processes and product lifecycles. *Harvard Business Review*. 57(1): 133-140.

Henderson, R.M. and Clark, K.B., 1990. Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, pp.9-30.

Holmstrom, B. and Milgrom, P., 1994. The firm as an incentive system. *The American economic review*, pp.972-991.

Jaffe, A.B. and Palmer, K., 1997. Environmental regulation and innovation: a panel data study. *Review of economics and statistics*, 79(4), pp.610-619.

Jaffe, A.B., Peterson, S.R., Portney, P.R. and Stavins, R.N., 1995. Environmental regulation and the competitiveness of US manufacturing: what does the evidence tell us?. *Journal of Economic literature*, 33(1), pp.132-163.

Jensen, M.C. and Heckling, W.H., 1995. Specific and general knowledge, and organizational structure. *Journal of applied corporate finance*, 8(2), pp.4-18.

Jin, W. and McElheran, K., 2018. Economies before Scale: Learning, Survival and Performance of Young Plants in the Age of Cloud Computing. SSRN: <https://papers.ssrn.com/sol3/papers.cfm>.

Jones, S., 2008. Internet goes to college: How students are living in the future with today's technology. Diane Publishing.

Kandel, E. and Lazear, E.P., 1992. Peer pressure and partnerships. *Journal of political Economy*, 100(4), pp.801-817.

Khan, A., Le, H., Do, K., Tran, T., Ghose, A., Dam, H. and Sindhgatta, R., 2018. Memory-augmented neural networks for predictive process analytics. arXiv preprint arXiv:1802.00938.

Khurana, U., Samulowitz, H. and Turaga, D., 2018, April. Feature engineering for predictive modeling using reinforcement learning. In Thirty-Second AAAI Conference on Artificial Intelligence.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. and Mullainathan, S., 2018. Human decisions and machine predictions. *The quarterly journal of economics*, 133(1), pp.237-293.

Kleinberg, J., Ludwig, J., Mullainathan, S. and Obermeyer, Z., 2015. Prediction policy problems. *American Economic Review*, 105(5), pp.491-95.

Kueng, L., Yang, M.J. and Hong, B., 2014. *Sources of firm life-cycle dynamics: differentiating size vs. age effects* (No. w20621). National Bureau of Economic Research.

Labro, E., Lang, M. and Omartian, J., 2019. Predictive Analytics and the Manufacturing Employment Relationship: Plant Level Evidence from Census Data.

Lanoie, P., Laurent-Lucchetti, J., Johnstone, N. and Ambec, S., 2011. Environmental policy, innovation and performance: new insights on the Porter hypothesis. *Journal of Economics & Management Strategy*, 20(3), pp.803-842.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S. and Kruschwitz, N., 2011. Big data, analytics and the path from insights to value. *MIT sloan management review*, 52(2), pp.21-32.

Levin, S.G., Levin, S.L. and Meisel, J.B., 1987. A dynamic analysis of the adoption of a new technology:

the case of optical scanners. *The Review of Economics and Statistics*, pp.12-17.

McAfee Andrew and Erik Brynjolfsson (October 2012). Big Data: The Management Revolution. *Harvard Business Review*. Retrieved from <https://hbr.org/2012/10/big-data-the-management-revolution>.

McElheran, K, and Jin W. (2020) "Strategic Fit of IT Resources in the Age of Cloud Computing." *University of Toronto Working Paper*.

McElheran, K, S Ohlmacher, and MJ Yang (2020). "Strategy and Structured Management." *University of Toronto Working Paper*.

Mikalef, P., Boura, M., Lekakos, G. and Krogstie, J., 2019. Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research*, 98, pp.261-276

Mikalef, P., Krogstie, J., Pappas, I.O. and Pavlou, P., 2020. Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information & Management*, 57(2), p.103169.

Milgrom, P. and Roberts, J., 1990. Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica: Journal of the Econometric Society*, pp.1255-1277.

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.

Mithas, S, Tafti A and Mitchell W (2013) How a firm's competitive environment and digital strategic posture influence digital business strategy. *MIS quarterly*. 511-536.

Moretti, E., 2004. Workers' education, spillovers, and productivity: evidence from plant-level production functions. *American Economic Review*, 94(3), pp.656-690.

Müller, O., Fay, M. and vom Brocke, J., 2018. The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. *Journal of Management Information Systems*, 35(2), pp.488-509.

Novak, S. and Stern, S., 2009. Complementarity among vertical integration decisions: Evidence from automobile product development. *Management Science*, 55(2), pp.311-332.

Porter, M.E., 1991. Towards a dynamic theory of strategy. *Strategic management journal*, 12(S2), pp.95-117.

Porter, M.E. and Van der Linde, C., 1995. Toward a new conception of the environment-competitiveness relationship. *Journal of economic perspectives*, 9(4), pp.97-118.

Raghupathi, W. and Raghupathi, V., 2014. Big data analytics in healthcare: promise and potential. *Health information science and systems*, 2(1), p.3.

Raiffa, H., 1968. Decision analysis: Introductory lectures on choices under uncertainty.

Ransbotham, S., Kiron, D. and Prentice, P.K., 2015. The talent dividend. *MIT Sloan Management Review*, 56(4), p.1.

Ransbotham, S., Kiron, D. and Prentice, P.K., 2016. Beyond the hype: the hard work behind analytics success. *MIT Sloan Management Review*, 57(3).

Rivkin, J.W., 2000. Imitation of complex strategies. *Management science*, 46(6), pp.824-844.

Rogers, E.M., 2010. *Diffusion of innovations*. Simon and Schuster.

Safizadeh, MH, Ritzman LP, Sharma D and Wood C 1996. An empirical analysis of the product-process matrix. *Management Science*. 42(11): 1576-1591.

- Sandvik, J.J., Saouma, R.E., Seegert, N.T. and Stanton, C.T., 2020. Workplace knowledge flows. *The Quarterly Journal of Economics*.
- Saunders, A. and Tambe, P., 2015. Data Assets and Industry Competition: Evidence from 10-K Filings. Available at SSRN 2537089.
- Schlegel, G.L., 2014. Utilizing big data and predictive analytics to manage supply chain risk. *The Journal of Business Forecasting*, 33(4), p.11.
- Schrage Michael (September 2014). Learn from Your Analytics Failures. *Harvard Business Review*. Retrieved from <https://hbr.org/2014/09/learn-from-your-analytics-failures>
- Schrage, M. and Kiron, D., 2018. Leading with next-generation key performance indicators. *MIT Sloan Management Review*, 16.
- Siegel, E., 2013. Predictive analytics: The power to predict who will click, buy, lie, or die. John Wiley & Sons.
- Stiroh, K.J., 2002. Information technology and the US productivity revival: what do the industry data say?. *American Economic Review*, 92(5), pp.1559-1576.
- Tambe, P., 2014. Big data investment, skills, and firm value. *Management Science*, 60(6), pp.1452-1469.
- Tambe, P. and Hitt, L.M., 2012. The productivity of information technology investments: New evidence from IT labor data. *Information Systems Research*, 23(3-part-1), pp.599-617.
- 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.
- Waller, M.A. and Fawcett, S.E., 2013. Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), pp.77-84.
- Wang, G., Gunasekaran, A., Ngai, E.W. and Papadopoulos, T., 2016. Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics*, 176, pp.98-110.
- Wang, Y., Kung, L., Gupta, S., & Ozdemir, S. 2019. Leveraging Big Data Analytics to Improve Quality of Care in Healthcare Organizations : A Configurational Perspective. *British Journal OfManagement*, 30, 362–388. <https://doi.org/10.1111/1467-8551.12332>.
- Winnig, L.W., 2016. GE's big bet on data and analytics. *MIT Sloan Management Review*, 57, 5-5.
- Wu, L., Lou, B. and Hitt, L., 2019. Data Analytics Supports Decentralized Innovation. *Management Science*, 65(10), pp.4863-4877.
- Zolas, N, Kroff Z, Brynjolfsson E, McElheran K, Beede D, Buffington C, Goldschlag N, Foster L and Dinlersoz E (2020) Advanced technologies adoption and use by us firms: Evidence from the annual business survey. *NBER Working Paper*, 28290.

