

MIT Open Access Articles

The power of prediction: predictive analytics, workplace complements, and business performance

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Brynjolfsson, Erik, Jin, Wang and McElheran, Kristina. 2021. "The power of prediction: predictive analytics, workplace complements, and business performance."

Published Version: <https://doi.org/10.1057/s11369-021-00224-5>

Publisher: Palgrave Macmillan UK

Permanent Link: <https://hdl.handle.net/1721.1/137842>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: <http://creativecommons.org/licenses/by-nc-sa/4.0/>



The power of prediction: predictive analytics, workplace complements, and business performance

Cite this article as: Erik Brynjolfsson, Wang Jin and Kristina McElheran, The power of prediction: predictive analytics, workplace complements, and business performance, Business Economics <https://doi.org/10.1057/s11369-021-00224-5>

This Author Accepted Manuscript is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication but has not been copyedited or corrected. The official version of record that is published in the journal is kept up to date and so may therefore differ from this version.

Terms of use and reuse: academic research for non-commercial purposes, see here for full terms. <https://www.palgrave.com/aam-terms-v1>

Author accepted manuscript

The Power of Prediction: Predictive Analytics, Workplace Complements, and Business 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 suggesting that the use of predictive analytics boosts firm performance. However, large-scale representative data on this phenomenon have been lacking. Working with the Census Bureau, we surveyed over 30,000 American manufacturing establishments on their use of predictive analytics and detailed workplace characteristics. We find that productivity is significantly higher among plants that use predictive analytics—up to \$918,000 higher sales compared to similar competitors. Furthermore, both instrumental variables estimates and timing of gains suggest a causal relationship. However, we find that the productivity pay-off only occurs when predictive analytics are combined with at least one of three workplace complements: significant accumulation of IT capital, educated workers, or workplaces designed for high flow-efficiency production. Our findings support claims that predictive analytics can substantially boost performance, while also explaining why some firms see no benefits at all.

Keywords: digitization, data, predictive analytics, productivity, complementarities

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. The authors thank the Stanford Digital Economy Lab and the MIT Initiative on the Digital Economy, the Social Science and Humanities Research Council of Canada (SSHRC), and the Kauffman Foundation for generous funding, as well as Chris Forman, Shane Greenstein, Megan MacGarvie, Ali Tafti, Lynn Wu and participants at the Annual Conference of the National Association of Business Economics, the Conference on Information Systems and Technology (CIST), the Workshop on Information Systems and Economics (WISE), the NBER Productivity Lunch Seminar, the MIT Initiative Lunch Seminar, and the Wharton Technology Conference for valuable comments.

“As data piles up, we have ourselves a genuine gold rush. But data isn’t the gold...The gold is what’s discovered therein.”

- Eric Siegel, *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*

1. Introduction

According to Ransbotham *et al.* (2016), “the hype around data and analytics has reached a fever pitch.” Greater volume, variety, and timeliness of digital information, produced by sources ranging from social media to the Internet of Things (IoT), have transformed what is measurable within firms and markets (McAfee and Brynjolfsson 2012). Exponential growth in computing power, declining costs of information technology (IT), and the rise of new computational methods have fueled efforts to extract more value from data (Tambe 2014; Bughin 2016; Wu *et al.* 2020). Worldwide revenue for “big data” and business analytics solutions is forecasted to reach \$274.3 billion by 2022 (IDC 2019). However, these investments have yet to yield productivity gains in the aggregate (Syverson 2017; Brynjolfsson *et al.* 2021). At the firm level, managers struggle to close the gap between the promise of predictive analytics and its performance (Ransbotham *et al.* 2015, 2017; Wu *et al.* 2019). These concerns have been difficult to tackle empirically due to the rate of technological change and, ironically, a dearth of data.

We address this challenge by providing novel and systematic evidence on the adoption, performance impacts, and enablers of predictive analytics across diverse workplace settings. Collaborating with the U.S. Census Bureau, we collect the first direct measures of predictive analytics use in a large and representative sample of U.S. manufacturing establishments,² along with critical data on tangible and intangible workplace characteristics. Linking this survey from 2010 and 2015 to an annual panel of comprehensive administrative data, we estimate the

² Prior work has addressed the measurement challenge associated with tracking analytics use in firms by triangulating on the human capital needed to adopt it, typically in smaller samples (Tambe 2014; Wu *et al.* 2019, 2020). In contrast, our data cover more than half of the U.S. manufacturing economy in the annual certainty sample. We view our approach to be complementary, with distinct advantages and challenges. See Section 2.

productivity impact of predictive analytics in over 30,000 manufacturing plants, addressing common threats to causal identification and unearthing essential contingencies.

Our novel measurement effort indicates that adoption of predictive analytics has become widespread in the American manufacturing sector. More than 70 percent of our representative sample adopted some level of predictive analytics as early as 2010, with pervasive penetration across geographies and industries, as well as plant size and age distributions. We document this prevalence of predictive analytics in recent years and present three main findings.

First, adoption of predictive analytics is associated with productivity gains that are statistically and economically significant. Workplaces adopting at least some use of predictive analytics enjoy productivity benefits of 1 to 3 percent on average. Practically, this represents roughly \$464,000 to \$918,000 greater sales for adopters versus non-adopters, even holding constant all other production inputs³ and a wide range of other factors. Benefits also increase with frequency of use suggesting a dose-response relationship.

Second, evidence points to a causal relationship. In addition to addressing a number of time-varying confounds, we use a quasi-experimental approach (instrumental variables) and the timing of the effects (performance increases only after plants adopt predictive analytics, not before) to argue that using predictive analytics causes higher performance, not the other way around.

Third, despite a significant average effect and widespread adoption, we find that productivity gains are almost entirely limited to workplaces that have high levels of accumulated IT capital, a significant share of educated employees, or high flow-efficiency manufacturing processes. The systematic need for complementary IT infrastructure and appropriately skilled workers is consistent with prior work on emerging general-purpose IT and analytics (e.g., Bresnahan *et al.* 2002, Tambe 2014), as well as recent theorizing on the “Productivity J-Curve” (Brynjolfsson *et al.* 2021). We identify a third and novel complement, as well: production process design favoring high flow efficiency. The returns to predictive analytics are much higher in efficiency-focused contexts than in flexibility-focused production environments.

Widespread heterogeneity along these dimensions points to how costly and time-consuming it

³ This is a common flexible approach to estimating revenue-based total-factor productivity (TFPR), and is possible due to establishment-level Census data on both expenditures and capital investment over time (e.g., Bloom *et al.* 2019).

is to bring key features of a workplace context into alignment with new technology, even over the five years covered by our study. The importance and durability of these three complements sheds new light on why some firms achieve high returns from their predictive analytics investments, while others see no benefit at all.

Our findings all hold when controlling for a wide range of other management practices previously associated with productivity (e.g., Bloom *et al.* 2019, Scur *et al.* 2021), including those particularly focused on using data in decision making (Brynjolfsson and McElheran 2016, 2019). Disentangling these margins is critical, as a number of practices and tools aimed at leveraging digital information have been linked to higher Tobin's *q* and profits (Brynjolfsson *et al.* 2011; Saunders and Tambe 2015) and greater productivity (Tambe *et al.* 2012; Tambe 2014; Brynjolfsson and McElheran 2019; Wu *et al.* 2020). A longstanding literature has modeled how better information can improve decision-making in firms (e.g., Blackwell 1953; Raiffa 1968; March 1994), with renewed interest in recent years linked to the rise of "big data" (Davenport 2006, Tambe 2014, Bughin 2016, Müller *et al.* 2018; Wu *et al.* 2020). Building on this trend, leveraging digital information to improve prediction is argued to further reduce the cognitive costs of decision-making, improve precision, and speed execution (e.g., Agrawal *et al.* 2019).

Thus, predictive analytics is conceptually and empirically distinct from an "evidence-based" firm culture (Pfeffer and Sutton 2006) or "data-driven decision-making" managerial practices (Brynjolfsson and McElheran 2019). Much work in this area further differentiates between descriptive and predictive uses of analytics (Berman Israeli 2020, Blum *et al.* 2015). Although definitions for emerging technology tend to shift (a significant hurdle for measurement), predictive analytics is increasingly understood to be a set of techniques—from data mining to statistical modeling, including, in some firms, machine learning and "AI"—used to analyze historical and current data in order to make predictions about future or unknown events.⁴ Despite the rising interest, causal evidence from large representative samples on its benefits remains lacking. Furthermore, while prior research explores how organizational and labor force factors explain

⁴ This definition was corroborated in our study via extensive testing of the survey instrument, led by Census experts on survey development. While of significant importance in current research and business press, the actual use of machine learning and other cognitive technologies increasingly referred to as "artificial intelligence" was still quite low in the U.S. as of 2018, including in the manufacturing sector studied here (Zolas *et al.* 2020).

variation in the returns to IT (Melville *et al.* 2004; Bapna *et al.* 2013), many slow-moving or even time-invariant organizational factors have been difficult to observe directly. This study makes progress in a number of ways.

First, our data come from a purpose-designed mandatory survey covering more than 50 percent of the U.S. manufacturing economy. We directly capture early use of a fast-changing technological advance, which is a persistent challenge for studying the digital frontier. Our measure complements prior studies of analytics (e.g., Tambe 2014; Wu *et al.* 2019, 2020) with a novel, direct measure specifically focused on prediction.

Causality has been difficult to pin down, due to the self-selection of firms into adoption (Müller *et al.* 2018; Berman and Israeli 2020). We provide two tests suggesting that predictive analytics use causally increases firm productivity. The first is instrumental variable estimation, enabled by a survey question we designed on government-mandated data collection as a promising exogenous, plant-level nudge to predictive analytics use. We also compare the timing of adoption with the timing of productivity gains, the results of which are consistent with a causal interpretation.

Finally, linking multiple Census data sets to capture tangible and intangible workplace investment yields several distinct advantages. The first is the ability to contribute to a large economics literature on business performance and productivity. Data on production inputs and outputs in this sector have been well established over time, supporting a rich exploration of productivity dispersion (e.g., Bartelsman and Doms 2000, Syverson 2011, Collard-Wexler and deLoecker 2015, White *et al.* 2018) and digitization (e.g., Doms *et al.* 1995, Black and Lynch 1996, 2001) in a well-understood setting. However, many features of the workplace context that both interact with digitization and might affect productivity have been less systematically captured. By adding new visibility to a range of tangible and intangible workplace features in this setting, we can address a large number of potential confounds and pin down essential contingencies that matter for both theory and practice.

Notably, observing these features across a five-year span shows not only that complementary investments in IT and skilled labor are indispensable to the productivity of predictive analytics, but also that they are slow to catch up with emerging technology. This has important implications for our conceptual understanding of technological diffusion and complementarity in firms—in particular, the need to be more nuanced with respect to timing. As a practical matter, this highlights

pain points for firms pursuing gains from new technologies. Finally, it reveals contingencies that may be beyond managerial control: our data on production process design indicate that time-invariant workplace characteristics strongly shape returns to predictive analytics by affecting the stationarity, availability, and value of fine-grained data. A failure to recognize constraints will invite costly technology-workplace misalignment.

These are not subtle nuances, but first-order concerns. The magnitude of each of these organizational interactions is large: in our sample, predictive analytics largely contribute to productivity *only* when combined with at least one of these complements. Adoption may be widespread, but business gains are not.

Our findings contribute to several areas of research. We build on early research in IT productivity that emphasizes heterogeneity across industries and firms (e.g., Stiroh 2002, Brynjolfsson and Hitt 1995), as well as theory arguing 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; Athey and Stern 1998; Brynjolfsson and Milgrom 2013; Brynjolfsson *et al.* 2021). A number of empirical studies have supported this theory with respect to general-purpose IT and computer use (Black and Lynch 2001; Bresnahan *et al.* 2002; Aral and Weill 2007; Bloom *et al.* 2012), specific IT applications such as electronic medical records (Dranove *et al.* 2014), and earlier waves of data-centered management practices (Aral *et al.* 2012; Tambe *et al.* 2012; Tambe 2014; Brynjolfsson and McElheran 2019). Attention has recently turned to whether similar contingencies apply to analytics deployed in specific applications, such as innovation (Wu *et al.* 2019 and 2020) and marketing (Berman and Israeli 2020). Our study is the first to specifically target prediction, documenting uneven returns to this increasingly popular type of automation. Moreover, our purpose-designed data set supports causal inference and unusual visibility into rich workplace features over time. Thus, we contribute to an increasingly rigorous empirical complementarities literature focused on firm strategy and management practices (e.g., Arora and Gambardella 1990; Hong *et al.* 2019; Choudhury *et al.* 2020).

Our goal is not only to assess the causal impacts but also to make progress on a practical roadmap that managers can follow to better leverage these new tools. Awareness of the organizational constraints can help firms allocate scarce analytics resources, targeting areas that

are most likely to yield timely returns or funding coordinated investments with better-understood timelines.

2. Empirical Setting, Prior Work, and Data

Predictive analytics leverages computer systems to investigate large data sets more quickly, and more comprehensively, than would otherwise be humanly possible. Because digital information is rapidly becoming cheaper to gather and growing in volume and complexity, leveraging these increasingly rich digital resources may generate large returns.

Anecdotal evidence suggests, however, that while many firms have benefitted from predictive analytics, others have struggled to realize returns from their investments (Schrage 2014). Unaligned workplace organization and a lack of employees with complementary skills have at times been flagged as key challenges (e.g., Ransbotham *et al.* 2015). Organizational complementarity theory suggests that firms that invest in mutually-reinforcing assets (both tangible and intangible) will perform better, though appropriate complements may take time to develop, and a mismatch may be temporarily very costly (Kandel and Lazear 1992; Milgrom and Roberts 1990, 1995; Holmstrom and Milgrom 1994; Brynjolfsson and Milgrom 2013; Brynjolfsson *et al.* 2018). Empirical studies have validated the importance of complementary investments and organizational alignment for realizing the value of IT (Bresnahan and Greenstein 1996; Black and Lynch 2001; Caroli and Van Reenen 2001; Bresnahan *et al.* 2002; Melvill *et al.* 2004; Aral and Weill 2007; Bloom *et al.* 2012; Bapna *et al.* 2013), as well as data-centered practices (Aral *et al.* 2012; Tambe *et al.* 2012; Brynjolfsson and McElheran 2019).

Building on this research, recent studies have contributed new measurement approaches and evidence of complementarities specifically related to analytics. Tambe (2014) and Wu *et al.* (2019, 2020) rely on large-scale resume data to hone in on analytics use via the complementary human skills hired to deploy it in firms. Combining this with public firm data (Compustat) or smaller-scale surveys yields insights into performance and how analytics is deployed in innovation activities. Berman and Israeli (2020) explore marketing applications of descriptive analytics, with fine-grained usage data from one analytics vendor.

Our approach complements these works in a few ways. To begin, we use a direct survey measure that distinguishes both the extensive and intensive margin of use, which is needed to make

standard “dose-response” arguments: if some use is beneficial, more should yield greater returns up to a point.⁵ It is not clear that use scales as directly with complementary skilled labor, though these details are helpful for targeting specific techniques (Tambe 2014).

We are not immune from measurement error in our approach, which further motivates our instrumental variables estimation below. However, our approach is inclusive to a range of tools and techniques, focusing attention on prediction (which admits machine learning and “AI” in more recent years). It further distinguishes other uses of data for decision-making, as well as general managerial capacity, which might bias our inference if not appropriately accounted for. Also, our sample is both larger and more representative than prior efforts, including a more robust inclusion of smaller and non-public firms. Finally, ours is an establishment-level data set, which helps understand how localized these patterns are and sheds light on previously unobserved within-firm heterogeneity.

Managerial and Organizational Practices Survey

To generate large-scale, representative panel data on predictive analytics use and sufficiently rich workplace characteristics, we collaborated with the U.S. Census Bureau to add new, purpose-designed questions to the 2015 Management and Organizational Practice Survey (MOPS).⁶ Survey response is required by law, yielding a response rate of 70.9 percent and 30,000 complete establishment-level observations. Combined with rigorous sample stratification and data validation by Census, this obviates the standard concerns about response and selection bias that apply to most survey efforts. Our sample contains data for reference year 2015 along with recall values for 2010. Measure validity is also high: adding questions to Census surveys requires rigorous cognitive testing and validation (Buffington *et al.* 2017), essential for measuring a recent and fast-emerging technology across different industry settings.

The key question for our study 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)?” Respondents—typically a senior plant manager or accounting expert with

⁵ However, research on the value of big data shows that an increase in the amount of data available to firms has positive but diminishing impacts on prediction accuracy (Bajari *et al.* 2019).

⁶ See Bloom *et al.* (2019) and Buffington *et al.* (2017) for more details.

the help of business function or line managers—are asked to mark all that apply among *Never*, *Yearly*, *Monthly*, *Weekly*, and *Daily*, with separate columns for 2015 and recall for 2010 (See Table 1). With recall data for 2010, we have in total 51,000 observations across the two years.⁷

Table 1. Predictive Analytics Question in MOPS

Q29: How frequently does this establishment typically reply on predictive analytics (statistical models that provide forecasts in areas such as demand, production, or human resources)? *Mark all that apply*

	2010	2015
Daily		
Weekly		
Monthly		
Yearly		
Never		

Source: This table is captured from the questionnaire for the 2015 MOPS section C question 29. The PDF version of the questionnaire for 2015 MOPS can be downloaded at the U.S. Census website: https://www2.census.gov/programs-surveys/mops/technical-documentation/questionnaires/ma-10002_15_final_3-2-16.pdf.

We focus significantly on the extensive margin of analytics use, regardless of frequency, as there may be heterogeneity in inputs (such as data quality) that remain unobserved. However, we also capture variation along the intensive margin with a numeric value ranging from 0 to 4 for each frequency category in ascending order, defaulting to the highest in cases of multiple categories.⁸ We lean on this more-continuous measure in particular in our instrumental variables (IV) estimation. In so doing, we avoid potential complications due to non-linear first stage estimation, and also capture better variation among the plants' use of predictive analytics.

⁷ Note that sample counts are rounded to comply with Census disclosure avoidance requirements throughout the paper. We use the total number of observations (~51,000) as our baseline sample, but all key results are robust to restricting attention to a subsample for which respondent tenure dates back to at least one year before the recall reference year. This has been found to reduce measurement error for the other management practices measured in the MOPS (Bloom *et al.* 2019).

⁸ We also explore using a normalized score based on taking the average of multiple responses for a given establishment (see Bloom *et al.* 2019) and find results consistent to the top counted frequency measure.

Linking to Administrative Data

We merge the MOPS data with the Annual Survey of Manufactures (ASM), the Census of Manufactures (CMF), and the Longitudinal Business Database (LBD) to bring in information on detailed production inputs (including capital stocks and costs of labor, materials, and energy), outputs (total value of shipments and value-added), age, and whether the establishment belongs to a multi-establishment firm.⁹ We restrict attention to observations with complete information on sales, costs of labor, material, and energy, and employment for technical and disclosure-avoidance reasons.

Descriptive Statistics on Predictive Analytics Use

Predictive analytics has widely diffused among manufacturing plants across almost all states (Figure 1) and industries (Figure 2), as well as among plants of different sizes and ages.¹⁰ Notably, much of this diffusion took place as early as 2010, with average adoption well over 70 percent (Table 2). Among the roughly 18,000 establishments with complete data for both years,¹¹ we observe a small 1.4 percent average yearly increase.¹²

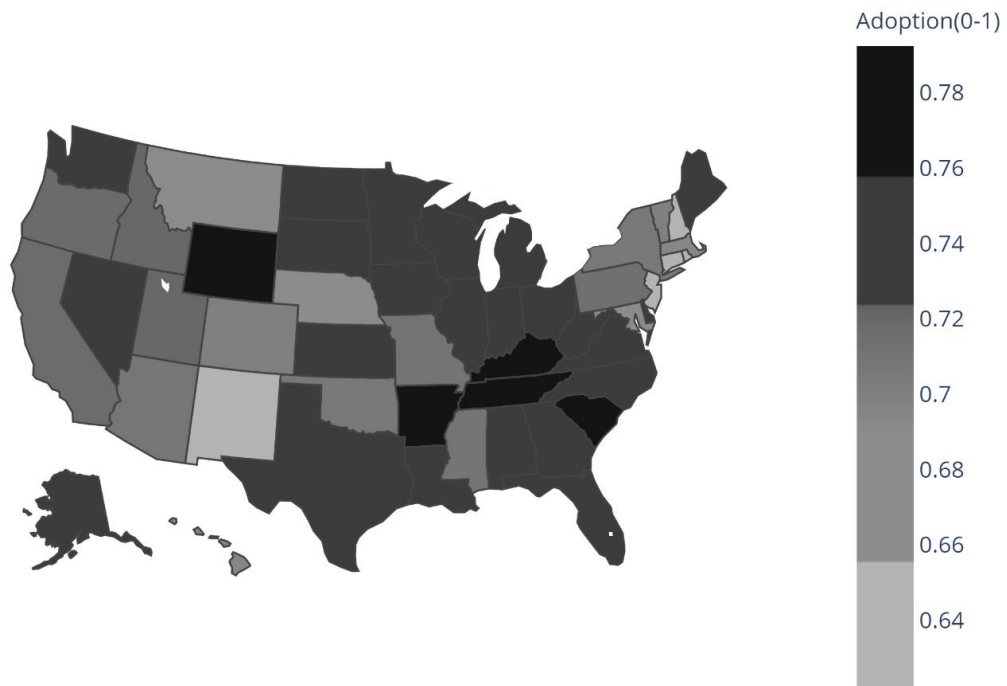
**Figure 1. Adoption of Predictive Analytics by State
(US Manufacturing in 2010)**

⁹ The ASM is conducted annually, except for years ending in 2 and 7, when it is included in the CMF. This allows us to construct a panel for all ASM/CMF variables between 2010 to 2015, which we use in our timing test to rule out reverse causality.

¹⁰ Correlations between predictive analytics and plant size and age are available upon request.

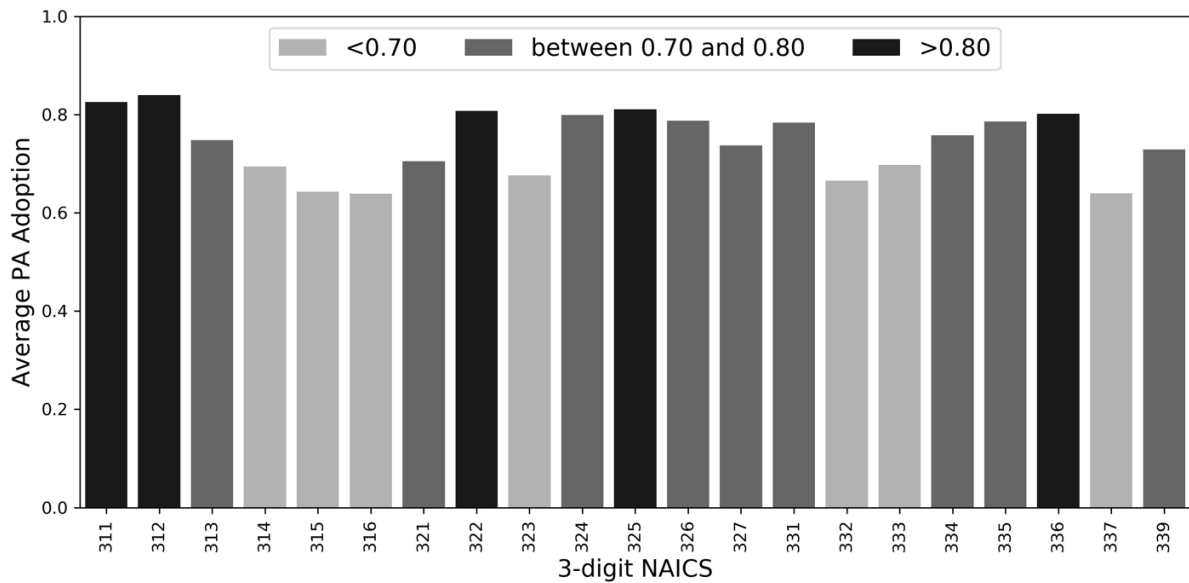
¹¹ The rotation of the ASM sample in years ending with 4 and 9 limits the number of establishments that have complete data for both reference years. However, a core “certainty sample” of larger plants covering the majority of economic activity in this sector is present for both years, conditional on survival.

¹² The adoption of predictive analytics increases from 73 percent in 2010 to 80 percent by 2015.



Notes: Reported statistics in the legend are the average adoption of predictive analytics across U.S. 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 by 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 available upon request.

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		

Table 2. Summary Statistics (Key Variables)

Variable	Definition	Mean (S.D.)	2010 (Recall)	2015
PA Use	Indicator for plants that use predictive analytics at any frequency	0.74 (0.44)	0.73 (0.44)	0.80 (0.40)
PA Use Frequency	An index for frequency of predictive analytics use based on the highest reported value (e.g. Yearly=1, Monthly=2, Weekly=3, and/or Daily=4)	1.12 (1.06)	1.09 (1.05)	1.27 (1.12)
Log Sales	Logged total value of plant shipments (\$Thousands)	10.37 (1.52)	10.68 (1.39)	10.86 (1.37)
Log L	Logged number of plant employees	4.56 (1.17)	4.79 (1.09)	4.88 (1.09)
Log (Non-IT) K	Accumulated and depreciated capital investment in non-IT equipment and structures in log terms (\$Thousands)	9.26 (1.47)	9.38 (1.58)	9.36 (1.61)
Log IT K	IT capital stock in log (\$Thousands)	5.16 (2.41)	5.58 (2.25)	5.62 (2.18)
Skilled Workers	Percentage of employees (managers and non-managers) with a bachelor's degree	0.15 (0.14)	0.15 (0.13)	0.16 (0.14)
Production Process Design (High Flow Efficiency)	Indicator for facilities with a production process designed for high flow efficiency (i.e. cellular or continuous-flow production process) as captured by question 44 of the 2015 MOPS.	0.35 (0.48)	0.38 (0.48)	0.41 (0.49)
MU	Indicator for plants belonging to multi-unit firms	0.73 (0.45)	0.78 (0.41)	0.81 (0.40)
HQ	Indicator for establishments reported as headquarters (HQ) or co-located with HQ	0.47 (0.50)	0.43 (0.50)	0.41 (0.049)
Plant age	Plant age	24.47 (12.89)	24.20 (11.31)	29.20 (11.31)
Government Mandate	Indicator that government regulations or agencies chose, at least in part, what type of data is collected at the plant	0.25 (0.43)	N/A	N/A
Structured Management	Normalized index for structured management practices using section A of the MOPS (excluding data-related questions 2 and 6)	0.63 (0.17)	0.60 (0.16)	0.68 (0.15)
DDD	Indicator for plants with high Data-Driven Decision-making practices as captured in the 2015 MOPS based on question 2 (KPI monitoring), question 6 (target setting), question 24 (data availability for decision making), and question 25 (using data in decision making); Equal to 1 for plants that set both short and long-term targets and have top 2 categories for both data availability and use.	0.27 (0.44)	0.19 (0.39)	0.38 (0.49)
Number of Observations		~51,000 (Baseline)	~18,000 (Balanced)	

Notes: Unweighted statistics based on the baseline and balanced samples from MOPS 2015 data; standard deviations in parentheses.

This high penetration and low rate of change have implications for our empirical approach (see Section 3). In particular, they hinder estimation of within-plant effects over time (a useful approach

for addressing unobserved workplace heterogeneity) for two key reasons. First, focusing on changes in the smaller subpopulation of late adopters would distort our inference. It is widely believed that later adopters of new technologies tend to be those with low anticipated returns, disproportionately high costs of adoption, and/or lagging awareness of the technology (Griliches 1957; David 1969; Bresnahan and Greenstein 1996). But we are interested in adoption and performance benefits—or the barriers thereto—for firms throughout the diffusion curve. Also, statistical power in the subsample of establishments that shift their predictive analytics use is severely limited, despite the overall size of our data set. Therefore, we rely primarily on cross-sectional variation in our analysis, addressing selection into adoption to some degree using our IV. Our primary approach to addressing workplace heterogeneity—both varying and time-invariant drivers—that could bias our inference is to directly control for an unusually rich set of workplace characteristics. See Section 3 for more on our empirical approach.

Workplace Complements

Complementarities among IT investments and organizational characteristics have long been associated with 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; Bloom *et al.* 2012; Aral *et al.* 2012; Tambe *et al.* 2012; Brynjolfsson and Milgrom 2013). More recently, heterogeneity in firm performance linked to “superstar firms” (Autor *et al.* 2020) and workplace features has attracted increasing attention.¹³ Rising concentration and inequality in workplace conditions and employee earnings are increasingly attributed to technology investment within industries and firms (Bessen 2017; Bennett 2020b; Lashkari *et al.* 2020; Barth *et al.* 2020). In addition, difficult-to-measure intangible features of firms and markets are argued to amplify these dynamics (Saunders and Brynjolfsson 2016; Haskel and Westlake 2018). Building on this prior work, we explore a few key tangible and intangible organizational characteristics that we

¹³ High cross-sectional heterogeneity in firm performance has long been established (e.g., Syverson 2004 & 2011; Hopenhayn 2014); however, a large number of 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) and is a burgeoning area of research and public policy concern.

expect to shape returns to predictive analytics: tangible IT investments, harder-to-measure employee skill, and heretofore unobserved production process design.

IT Capital Stock

The collection, storage, and communication of data inputs for predictive modeling all require tangible investments in infrastructure, such as sensors, transmission equipment, and data storage hardware. Building, training, and implementing analytics tools require corresponding data processing hardware and software. Thus, firms with existing IT capital investments that are more prepared for the industrial Internet of Things (IoT) and related “big data” innovations at the time of our study may possess fully-depreciated investments in infrastructure to collect and analyze data, as well as richer data inputs, giving them an advantage in analytics. Building and adapting such infrastructure to a particular firm setting is known to be risky and time-consuming (Bresnahan and Greenstein 1996), particularly in our manufacturing context (McElheran 2015).

To explore potential complementarities along this dimension, we calculate IT capital stocks using capital investment in computer and peripheral data processing equipment from the ASM and CMF panel dating back to 2002. We use a standard perpetual inventory approach and industry-level deflators for hardware from the Bureau of Economic Analysis (BEA), imputing values for years in which they are missing and depreciating at the rate of 35 percent per year following Bloom *et al.* (2014).¹⁴ A key advantage of this measure is that it accounts for the overall stock. If firms require time to adjust and utilize novel IT investment, we will be able to capture the lagged effect.¹⁵

Skilled Workers

Prior work has established that more-skilled and better-educated workers are key drivers of growth in both manufacturing productivity (Black and Lynch 2001; Moretti 2004) and returns to IT (Brynjolfsson and Hitt 2000; Bresnahan *et al.* 2002). With increasing digitization and growing

¹⁴ Results are quite insensitive to the choice of depreciation rate.

¹⁵ A reasonable concern here is that this measure fails 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 (Bessen and Righi 2019, 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, our findings remain consistent.

prevalence of business applications for data and predictive analytics, firms increasingly need 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 (Ransbotham *et al.* 2015). While competition for these workers may drive up wages to a point where there remain no excess returns to labor once appropriately measured, we expect that worker skill will boost predictive analytics’ contribution to productivity due to complementarities.

We leverage information from the MOPS regarding the percentages of managers and non-managers with bachelor’s degrees. 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 approach is similar to prior studies using education as a proxy for human capital (Bresnahan *et al.* 2002).

Production Process Design

Predictive analytics relies on historical and current data to predict future outcomes (e.g., Blum *et al.* 2015; Agrawal *et al.* 2018, 2019). In general, workplaces with greater automation and reduced variance will provide richer data inputs and better alignment for predictive analytics compared to other, more variable production environments (Fortuny *et al.* 2013; Martens *et al.* 2016).

In manufacturing, we expect that such environments can be identified by whether or not the plant operates primarily using a continuous-flow production process. Compared to other processes (e.g., job shops, or intermediate “batch” production), continuous flow processes are managed to maintain low variance and thus tend to be more stationary. A distinction commonly understood in operations management (e.g., Safizadeh *et al.* 1996), this type of production process is typically characterized by low product mix and high volume per product, tends to be more capital intensive, and relies on higher levels of automation. This is in contrast to establishments operating as job shops, batch-manufacturing plants, or R&D-focused facilities that are more likely to have a “jumbled flow” process; support flexible, high-mix but generally low-volume (per product) operations; and have shorter setup times between products. Lower flow-efficiency processes are central to product design and prototyping, as well as supporting a high-mix product market strategy

that may be central to a firm's product market positioning (Hayes and Wheelwright 1979).

Of relevance for our purposes, a high-volume, capital-intensive, and high flow-efficiency process tends to rely on a greater presence of instrumentation and sensors in daily production activities. The data generated are instantaneous, typically digitalized, voluminous, comparable over time, and usually transmittable to mobile or other data processing equipment. This type of data is a valuable input to predictive analytics. Combined with a more-stationary production environment, where analytics-driven insights are likely to be highly applicable, and high costs of equipment downtime (due to capital intensity of production), plants characterized by high flow-efficiency are more likely to both support and benefit from the use of predictive analytics.

We can test this hypothesis using a measure of production process design also added to the 2015 MOPS that distinguishes plants with high-flow production processes (including both cellular and continuous flow manufacturing) from more flexible and/or innovation-focused production.¹⁶ The relationships between this measure and the process design characteristics described above have been validated for the plants in our sample (McElheran *et al.* 2020).

It is worth noting that these production design characteristics are not merely “quasi-fixed,” as in prior work (e.g., Safizadeh *et al.* 1995), but are essentially time-invariant in our sample. Facility layout, production equipment, the ordering of production steps—not to mention practices and technologies to manage them—are embodied in physical infrastructure and reflected in numerous intangible and interconnected activities. This makes the transition to a different process design quite costly. In our data, the percentage of establishments transitioning into continuous-flow production process is less than 0.7 percent per year between 2010 and 2015. As a result, this feature is particularly useful for empirically identifying complementarities, as it is not subject to time-varying adjustment with respect to other, potentially unobserved factors (Athey and Stern 1998; Cassiman and Veuglers 2006; Hong *et al.* 2019).

Sample Characteristics

Table 2 presents key summary statistics for our sample. Despite the high prevalence of at least some use of predictive analytics, the intensive margin is more modest, with most plants reporting only annual and/or monthly use (mean frequency is 1.12). Although our sample represents the

¹⁶ See Kiran (2019) for a detailed description of cellular manufacturing.

majority of economic activity in the sector, very small establishments are underrepresented: sample mean annual sales and employment in log terms are 10.37 and 4.56, respectively, or about \$32 million and 96 employees, with an average age at around 24 years old. The mean plant has roughly \$175,000 of IT capital stock and slightly over 15 percent of workers with a bachelor's degree. 35 percent of plants in our sample are designed for high flow efficiency, which changes very little over the five years in our sample and is due almost entirely to shifts in sample composition between 2010 and 2015, not to changes within facility.

3. Empirical Approach

Performance Estimation

Our empirical exploration proceeds in steps. First, we estimate the average effect of predictive analytics on performance. We take a conventional approach to modeling the plant production function (Brynjolfsson and Hitt 2003; Bloom *et al.* 2012) estimating 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 sales by establishment i in industry j at time t , K denotes non-IT capital stocks at the beginning of the period. Indexed by time t : PA is an indicator (or frequency measure) for use of predictive analytics, L is labor input, M is consumption of material and energy inputs, and X is a vector of above-mentioned controls. The three potential complements are included in X in some specifications, with process design fixed in any robustness tests exploring panel models. Both w_{ijt} —the “technical productivity”—and ε_{ijt} —the “shock to productivity”—are unobservable econometrically (but w_{ijt} might be observable by establishments). Our first coefficient of interest is β_{pa} , the average relationship between predictive analytics and plant productivity, all else equal.

Assessing Causality: Instrumental Variables and Timing Tests

A standard concern with this approach is that predictive analytics use may be endogenously

determined, biasing interpretation of β_{pa} .¹⁷ Accordingly, we assess causality in two ways: with instrumental variables (IV) estimation and timing tests.

For our IV estimation, we exploit an indicator that data collection at the plant is “nudged” by government regulation or agencies. The motivation for this instrumentation strategy rests on the so-called “Porter Hypothesis” (Porter 1991; Porter and Van der Linde 1995), which argues that well-designed government regulations can stimulate firms to innovate and adopt new technology and practices. Of relevance in our setting, data collection at manufacturing facilities is often mandated by federal and local governments to demonstrate compliance with environmental and safety regulations. For instance, the Environmental Protection Agency (EPA) requires manufacturing firms (e.g., pulp and paper, petroleum, and chemical manufacturing) to install Continuous Emission Monitoring Systems (CEMS) for emissions data collection and monitoring. Leveraging this data in government-mandated reports requires that workers and managers are trained in systems and techniques for capturing, analyzing, and communicating data-driven conclusions. To the extent that data collection efforts, worker training, and data-driven monitoring practices involve sunk costs—yet may be applicable more generally—facilities exposed to such a statutory intervention will be more likely to gain infrastructure and systems for general data collection, storage, and analysis for reasons disconnected to their expected productivity benefits.¹⁸

Not all firms will be able to translate this into improved management of their production processes.¹⁹ For some, however, this external “nudge” into increased investment in and awareness of data resources may shift practices on the margin. Moreover, significant and unexpected consequences are not merely theoretical. The case of Alcoa Corporation in the late 1980s and 90s

¹⁷ This will happen if plants with higher expected returns to predictive analytics use will choose to adopt, upwardly biasing estimates of the average treatment effect. Tambe and Hitt (2012) provide a useful discussion of this common concern in the IT productivity literature, suggesting that such concerns may be overemphasized. 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), and point to quite limited upward bias due to self-selection. Unfortunately, our two-year panel lacks the longer lags typically required for this estimation approach.

¹⁸ Abundant anecdotes support the prevalence of this phenomenon. The Occupational Safety & Health Administration (OSHA) Recordkeeping rule can serve as another example: they require 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. For more details on OSHA Recordkeeping rule, see the OSHA website: <https://www.osha.gov/recordkeeping2014/records.html>

¹⁹ Note that plants already collecting and using data extensively may be less responsive to our instrument, which we discuss below.

is illustrative. When Paul O’Neil took leadership of the firm, his unexpected mandate to prioritize safety resulted in an abundance of data about accidents—but also about the performance and maintenance of infrastructure and workplace practices underlying those accidents. New data enabled new performance metrics, which were analyzed with increased frequency and linked to manager pay at the firm (*Fortune* 1991). The end result was not only improved worker safety but also improved productivity (Clark and Margolis 1991).

For this to be useful as an instrument, such oversight needs to be unrelated to the productivity of affected plants. Historically, U.S. government regulations in the manufacturing sector have fit this description. For instance, the objective of EPA CEMS requirements or OSHA’s Recordkeeping rule is restricted to public health and worker safety rather than plant performance. Although objections to such regulation have typically argued that they divert resources from other productivity-enhancing activities and investments (Gollop and Roberts 1983; Gray 1987), empirical evidence suggests that many well-designed regulations have had a limited negative impact on manufacturing competitiveness or overall performance (Jaffe *et al.* 1995; Lanoie *et al.* 2011; Ambec *et al.* 2013). Nevertheless, the standard expectation is that the direct effect will work against a positive relationship between productivity and government mandates to collect data.

Following these arguments, government-mandated data collection should satisfy both the relevance and exclusion restrictions for a valid instrument. As a practical matter, capturing this regulatory nudge at a sufficiently granular level is challenging. We addressed this by including another new question on the MOPS that captures government authority (among other decision-makers) over what type of data is collected at the plant. About 25 percent of the plants in our sample report that government regulations or agencies choose (at least in part) what type of data they collect (see Table 2).²⁰

Timing

Another threat to identifying a causal link between analytics and firm performance is the possibility that an unrelated productivity shock provides resources needed to invest in new technology or practices—not the other way around. To address this, we explore the timing of

²⁰ See question 26 in MOPS 2015 questionnaire for more detail; see Table 2 for the definition and descriptive statistics. A similar approach is used in a related study of data-driven decision-making by Brynjolfsson and McElheran (2019).

adoption vis-à-vis the timing of productivity changes. Leveraging annual data on inputs and output from the ASM and CMF, we construct a panel from 2010 to 2016 for a large subsample of our data. We exploit the recall questions to place plants in three categories: those that had adopted predictive analytics “early” by 2010, “middle” adopters (between 2010 and 2015), and “non-adopter” plants that had not adopted by 2015. Leaning on evidence that many (if not most) of the organizational practice measures in the MOPS are quasi-fixed over this period (Mundlak 1961; Bloom *et al.* 2007), we extrapolate the organizational complements outside of our core sample window and estimate comparable yearly production functions for these differently-timed groups from 2010 to 2016.

If predictive analytics causes better productivity, early adopters should have a performance premium at or near the start of our panel, compared to both middle adopters and non-adopters. In the 2010-2016 window, middle adopters should outperform non-adopters. Validating this pattern in the data would rule out reverse causality between productivity and adoption.

To test for this, we again rely on pooled OLS estimation with industry-year fixed-effects and rich organizational controls. It is worth noting here that, in addition to the limits on panel data estimation discussed above, the five-year gap in our two-period panel generates additional measurement error in this undertaking. For instance, our “middle” adopters may have adopted at any time in the 2010-2015 window; thus we anticipate estimates will be considerably noisier in this analysis.

Tests of Complementarity: Correlation and Productivity Gains

After addressing the questions of causality in the baseline performance model, we proceed with the two formal tests established for identifying complementarities: correlation in adoption and increasing returns when interacted in the performance equation (Brynjolfsson and Milgrom 2013).

First, if complementarities exist between predictive analytics and workplace features, we should observe higher adoption of predictive analytics among establishments with these investments and practices. We test for conditional correlations with IT capital stock, educated workers, and high flow efficiency in both linear probability and probit models, including a rich set of workplace controls. These include controls for structured management practices focused on operations and human resources management (Bloom *et al.* 2019, Scur *et al.* 2021), general reliance on data in decision-making (Brynjolfsson and McElheran 2019), plant age, multi-unit

status, headquarters status, and production process design.²¹ We also control for geographic differences and industry-year fixed-effects to account for any transitory industry-specific shocks.

Correlated adoption will be reinforced by mutually reinforcing returns to predictive analytics and complementary workplace features. Following the empirical strategy in Athey and Stern (1998) and Brynjolfsson and Milgrom (2013), we explore interactions in the following production function equation (2):

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

All variables in equation (2) are identical to those in equation (1) except C_{ijt} , which denotes, respectively, indicators for high IT capital stock, high percentage of educated workers, and high flow-efficiency production. A positive and significant $\beta_{interaction}$ term is indicative of such performance complementarities. The presence of significant findings in both tests is not always expected, as increased awareness of complementarity observed will lead to higher correlated adoption but potentially lower excess productivity gains. Passing both tests may be taken as strong evidence of complementarity (Brynjolfsson and Milgrom 2013).

4. Results

Predictive Analytics is Associated with Higher Performance

Table 3 explores the average conditional correlation between predictive analytics use and plant productivity. Using logged sales as the dependent variable and controlling for conventional inputs (i.e., non-IT capital, labor, materials, and energy) and the above-mentioned establishment controls, we arrive at an estimate of revenue-TFP (Foster *et al.* 2021). All columns further include industry-year controls at the narrow 6-digit NAICS level. For example, this captures the difference between

²¹ These controls are motivated by prior work associating them with technology adoption and productivity. Our management index differs from that in Bloom *et al.* (2019) by excluding the data-related MOPS questions. See Dunne (1994) and Foster *et al.* (2016) for more on the relationships between plant age, technology adoption, and performance. See Collis *et al.* (2007) for discussion of multi-unit and headquarter status. See Safizadeh *et al.* (1996) for more on manufacturing process designs. The indicator for multi-unit status equals one if the plants belong to multi-unit firms. We access the headquarter (HQ) status of a plant from the MOPS survey data where we define the HQ indicator equal to one if the plant is reported to be the HQ of a firm. Please see the definition of our measure for production process design in Table 2 (e.g. from the MOPS 2015). This set of controls is in all fully specified models for adoption and performance analysis unless stated otherwise.

Folding Paperboard Box Manufacturing (NAICS 322212) and Setup Paperboard Box Manufacturing (322213).

Table 3. 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 Use	0.0287*** (0.0049)	0.0145*** (0.0049)		
PA Use Frequency (top frequency)			0.0089*** (0.0021)	0.0509*** (0.0160)
Log IT K		0.0227*** (0.0014)	0.0227*** (0.0014)	0.0226*** (0.0014)
Skilled Workers		0.2068*** (0.0177)	0.2063*** (0.0177)	0.1928*** (0.0181)
High Flow Efficiency		0.0407*** (0.0052)	0.0406*** (0.0052)	0.0385*** (0.0053)
High Structured Mgmt.		0.0283*** (0.0054)	0.0273*** (0.0054)	0.0165** (0.0069)
Log L	0.4052*** (0.0065)	0.3821*** (0.0063)	0.3820*** (0.0063)	0.3800*** (0.0063)
Log (Non-IT) K	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	0.9327	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. Column 1 controls for key production inputs while column 2 adds controls for IT capital stock, percentage of employees with a bachelors' degree, an indicator for high flow efficiency production process design, high structured management (defined as having top quartile structured management practices), and an indicator for data-driven decision-making practices. Column 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, and plant age. Robust standard errors clustered at the firm level. * p<0.10, ** p<0.05, *** p<0.01.

Column 1 indicates that the extensive margin of predictive analytics use is associated with a roughly 2.87 percent (significant at the one-percent level) higher productivity, all else equal. This magnitude is large, representing \$918,000 greater sales at the sample mean (\$32M) while holding many other factors constant. Column 2 adds a rich set of plant-level controls, including the three potential complements explored below. The coefficient drops significantly to around 1.45 percent, consistent with an important role for organizational enablers that load onto the estimated returns to predictive analytics when not directly accounted for. Note that this specification controls for top-quartile structured management practices and top use of data-driven decision-making (DDD) to address concerns that unobserved management quality or style could be affecting our estimates (Brynjolfsson and McElheran 2019; Englemaier *et al.* 2019). Indeed, they separately reduce the coefficient on predictive analytics use (likely for reasons of omitted variable bias or additional complementarities beyond the scope of this study; not shown due to space limitations). Nonetheless, the coefficient remains both statistically significant and economically non-trivial: 1.45 percent higher productivity is commensurate with \$464,000 higher sales, on average, in excess of any costs of implementing the practice.

Column 3 explores the intensive margin of predictive analytics use. The coefficient on the frequency index is positive and significant at the one-percent level and economically meaningful. Based on its mean and standard error (see Table 1), moving from yearly to monthly use of predictive analytics is associated with 0.89 percent higher productivity, equivalent to roughly \$285,000 higher sales.²²

We explore the robustness of these patterns to alternative measures for both the dependent and independent variables (see Table 4). Our findings are robust to using value added as the output measure and alternative measures for the use frequency of predictive analytics.²³

²² For easy interpretation, we treat the frequency of predictive analytics as continuous variable (Long and Freese 2006). Results from additional tests treating it as ordinal are largely consistent and available upon request.

²³ Our results are also robust to using labor productivity and estimated TFP (e.g., the conventional 4-factor TFP using cost of material, energy, labor, and capital stock following Bartelsman and Gray 1996; Foster *et al.* 2021) as alternative output measures. It is also robust to estimating a translog production function. Results are omitted due to space limitations but available upon request.

Table 4. 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 Use	0.0223** (0.0091)		
PA Use Frequency (Average)		0.0023** (0.0006)	0.0167*** (0.0053)
Log IT K	0.0491*** (0.0026)	0.0228*** (0.0014)	0.0228*** (0.0014)
Skilled Workers	0.4670*** (0.0351)	0.2076*** (0.0177)	0.1978*** (0.0178)
High Flow Efficiency	0.0786*** (0.0107)	0.0408*** (0.0052)	0.0392*** (0.0052)
High Structured Mgmt.	0.0534*** (0.0109)	0.0269*** (0.0054)	0.0108 (0.0082)
Log L	0.7796*** (0.0081)	0.3820*** (0.0063)	0.3793*** (0.0063)
Log (non-IT) K	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 Use Frequency (Average) 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.

Evidence Indicates that Predictive Analytics Causes Higher Performance

Thus far, we have explored the pooled OLS regressions without considering measurement error

or endogeneity in plant adoption of predictive analytics. Column 4 reports instrumental variables estimation using government-mandated data collection as an instrument for the predictive analytics index.²⁴

The first stage of our two-stage least square (2SLS) estimation shows that government-mandated data collection is highly correlated with the use of predictive analytics (see also Table 3). In the second stage, the effect of predictive analytics on plant productivity remains large, positive, and statistically significant. This suggests a causal relationship between predictive analytics and performance.

Despite standard concerns of upward bias due to self-selection into technology use, the IV coefficient is greater than that from OLS estimation. This larger magnitude is consistent with downward bias in OLS, possibly attributable to errors-in-variables bias arising from measurement error—something that has been found in other MOPS measures (Bloom *et al.* 2019).

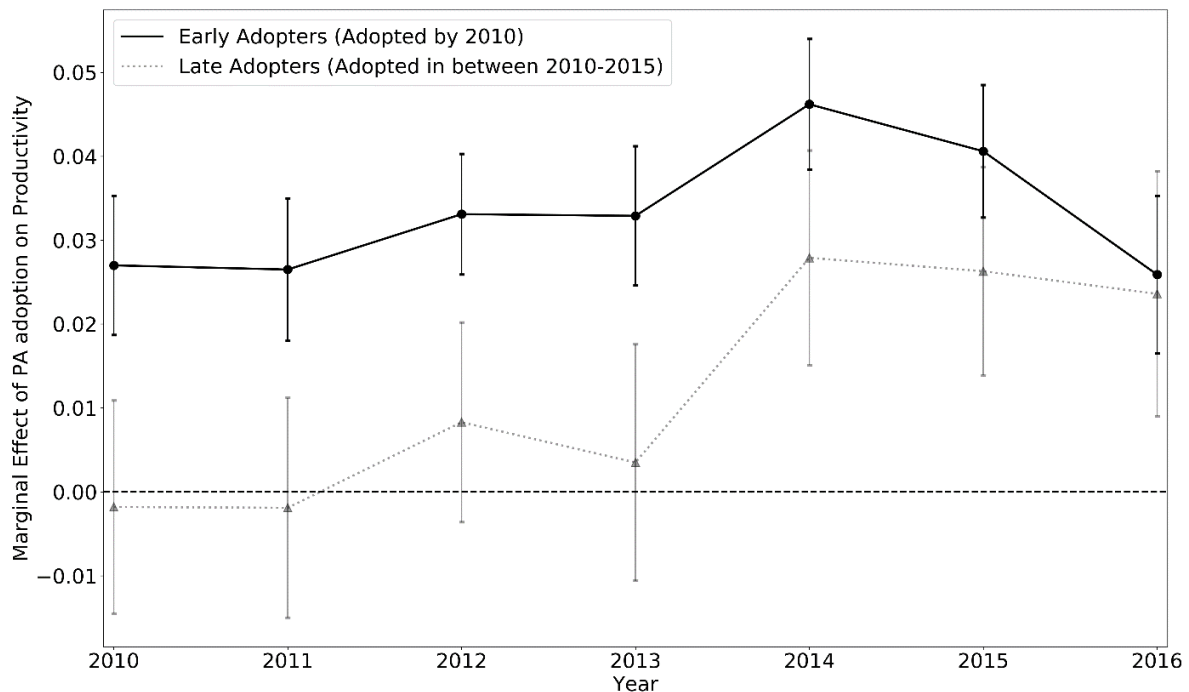
Not mutually exclusive, this pattern is also consistent with strong local treatment effects, whereby the subsample of workplaces that are the most receptive to the influence of the government mandate also experience the greatest productivity shift. This could arise if less data-savvy (and likewise less productive) plants enjoy larger indirect gains from data collection in response to requests from regulators (the “Porter Hypothesis” mechanism). In this vein, it is worth re-emphasizing that estimates in columns 2-4 control for management practices that are typically unobserved in other studies, but strongly associated with higher productivity in this sector (Bloom *et al.* 2019; Brynjolfsson and McElheran 2019). Addressing typically unobserved workplace features with the rich MOPS data significantly improves our identification of the coefficients of interest.

We further probe a causal interpretation by exploring the timing of adoption and performance in our panel (see Section 3). Figure 3 plots the coefficients for “early” and “middle” indicators in the performance model between 2010 and 2016. Consistent with our hypothesis, 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 middle adopters is not significantly different than that of non-adopters through 2013. We know that these plants adopted in the 2010-2015 window, but

²⁴ Using the index for frequency of predictive analytics for the IV estimations avoids potential complications due to non-linear first stage estimation, and also better captures variation in plant use of predictive analytics.

not precisely when. In line with steadily increasing diffusion over time (e.g., Hall 2004), performance in this group begins to rise and is significantly different from non-adopters by 2014. Notably, these middle adopters close the performance gap, becoming statistically indistinguishable from early adopters by 2016. This could be due to the rising diffusion of predictive analytics use, if competitive convergence in these practices erodes excess returns over time. However, technical productivity compared to non-adopters seems to persist. Regardless of the mechanisms at work, these overall patterns are inconsistent with reverse causality and support a causal interpretation.²⁵

Figure 3. Performance Effects for Early vs. Late Adopters of Predictive Analytics over Time



Notes: Estimates based on a pooled OLS model with a specification identical to the baseline model in column 2 table 3. 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 late adopters (non-adopters) using the 2015 MOPS data. These indicators are then interacted with year dummies to explore the differences in sales over time (using late adopters as the baseline group). Histogram bars (and values on the Y-axis) represent the marginal effect of predictive analytics adoption between 2010 and 2016. Standard errors of the coefficients are plotted on the histogram bars. Detailed regression coefficients are omitted due to space limitations but available upon request.

²⁵ Regression results for Figure 3 available upon request.

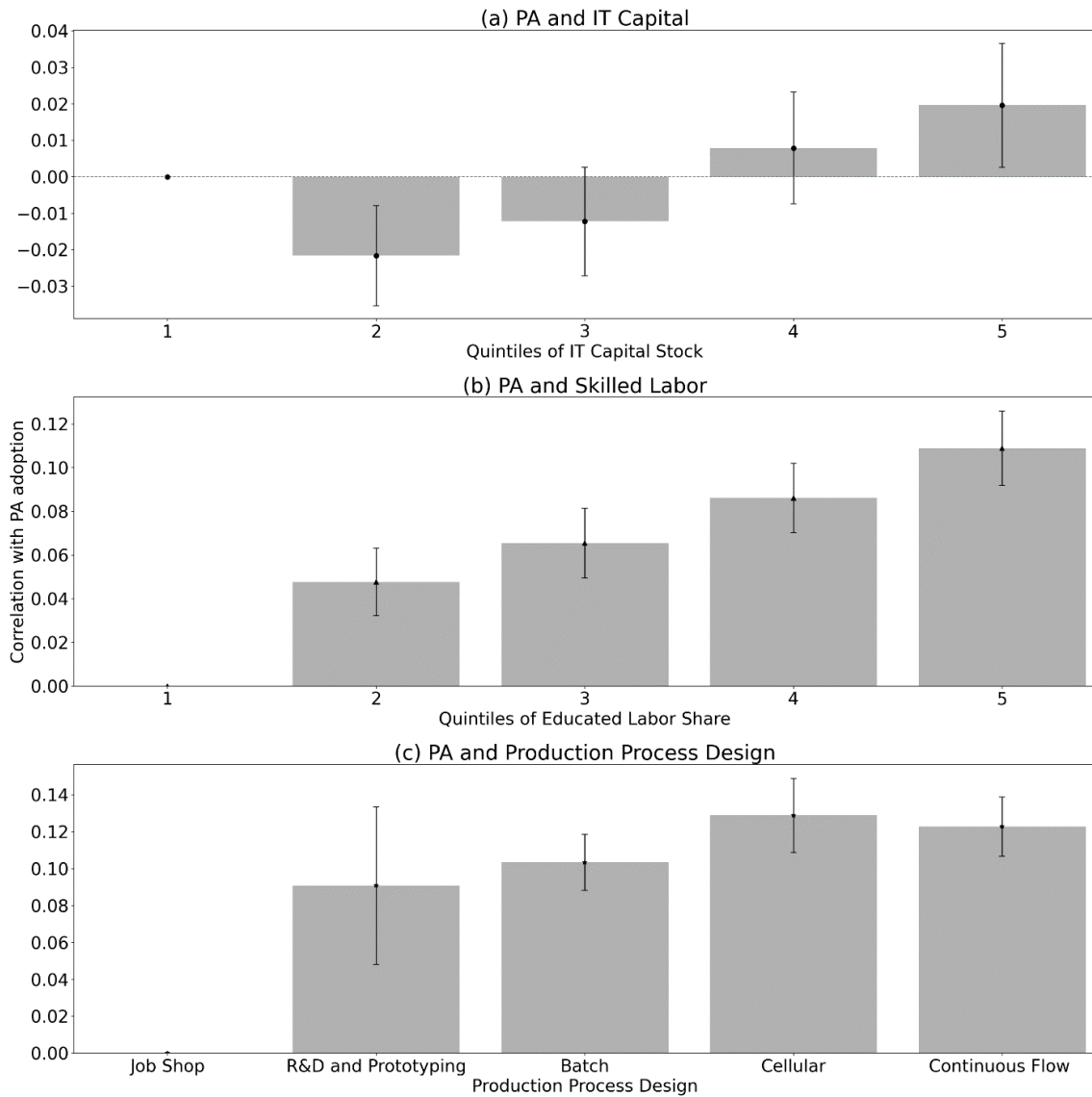
Complementary Investments are Essential

Results in Table 3 hint at the sensitivity of performance gains when key workplace features are accounted for. Here, we explore this systematically by estimating how the likelihood of predictive analytics adoption changes with the presence of key workplace features.

Results from Correlation Tests

Figure 4 depicts correlation tests for complementarity based on a linear probability model of whether or not the plant uses predictive analytics (at any frequency). This estimation includes relatively few controls, limiting them to plant size and age, as well as six-digit industry (NAICS) and year indicators. This approach is useful for understanding how adoption of predictive analytics varies across different levels of the potential complements. Panel a) of Figure 4 indicates that adoption of predictive analytics is only associated with IT capital stock accumulation above a certain threshold—only the top quintile of IT capital stock has a significant correlation with predictive analytics use. Panel b) shows that there is a more-continuous increasing relationship with skilled labor: adoption rises across all quintiles of worker education. (For ease of interpretation, we focus on above-median worker skills in the estimations to follow.) Panel c) shows that the “jumbled flow” of the job shop process design is associated with significantly lower (both statistically and empirically) lower predictive analytics use. The other, more flexible process designs have a somewhat lower and noisier correlation with predictive analytics than the efficiency-oriented designs, and robustness checks of our results support combining them to ease interpretation.

Figure 4. Conditional Correlations between Predictive Analytics and Potential Complements

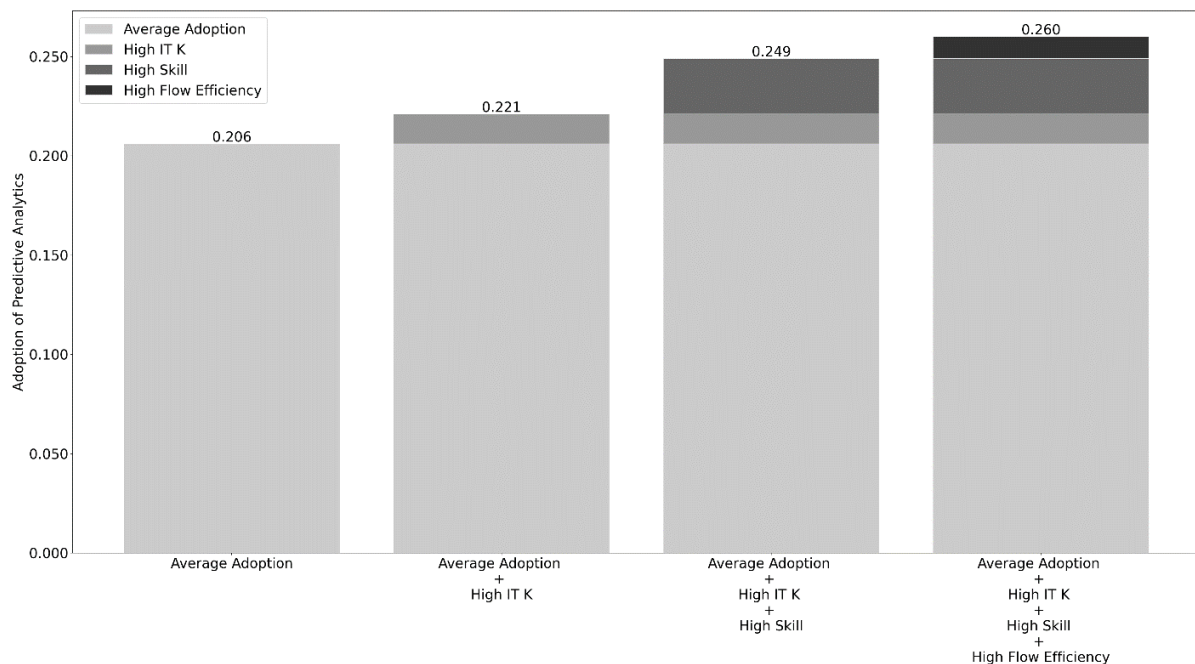


Notes: Estimates based on the baseline sample from the pooled OLS regressions controlling for plant size, plant age, and industry (6-digit NAICS) and year fixed-effects. The dependent variable is the adoption of predictive analytics. The base group is the plants in the bottom quintile of the sample based on calculated total IT capital stock for panel (a). For panel (b), the base group is the bottom quintile of the sample based on the percentage of workers with bachelor’s degrees. For panel (c), the base group is job shop facilities. Histogram bars (and values on the Y-axis) represent 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 darker lines.

In a more structural model, IT infrastructure, educated workers, and high flow efficiency are

all significantly correlated with the uses of predictive analytics at the one-percent level or higher. Figure 5 is organized to show the estimated increase in probability associated with each potential complement, building from a benchmark of 20.6 percent.²⁶ Being in the top 10 percent of the IT capital stock distribution by sample year increases the likelihood of predictive analytics use by 1.5 percentage points. Being in the top quartile for share of employees with a bachelor’s degree is associated with an additional 2.8-point increase; together they account for a 4.3 percentage-point greater likelihood of adoption. High Flow Efficiency adds 1.1 points more. A workplace with all three in place is 5.4 percentage points—or over 26 percent—more likely to use predictive analytics than a workplace without any of these reinforcing practices and investments. This is consistent with complementarity where rational managers (e.g., ones that consciously maximize profit and increase performance) will seek to adopt complementary practices together.

Figure 5. Organizational Complements to Predictive Analytics (Correlation Test)



Notes: Correlates of predictive analytics use based on linear probability estimation in the baseline pooled sample. The graph depicts estimated additive marginal contributions based on a single model that includes High IT Capital Stock, High Employee Education, and High Flow Efficiency. Average adoption represents the constant term as the benchmark adoption rate. High IT K is an indicator for plants in the top quintile of IT capital stock. High Employee Education is an indicator for plants in the top quartile for percentage of employees with a bachelors’ degree. High Flow Efficiency is an indicator for plants whose production process is best characterized as high flow manufacturing

²⁶ Based on the constant term in the linear probability model using the adoption of predictive analytics as our dependent variable, which represents the average adoption controlling for all covariates in our model. Results are omitted due to space limitations but available upon request.

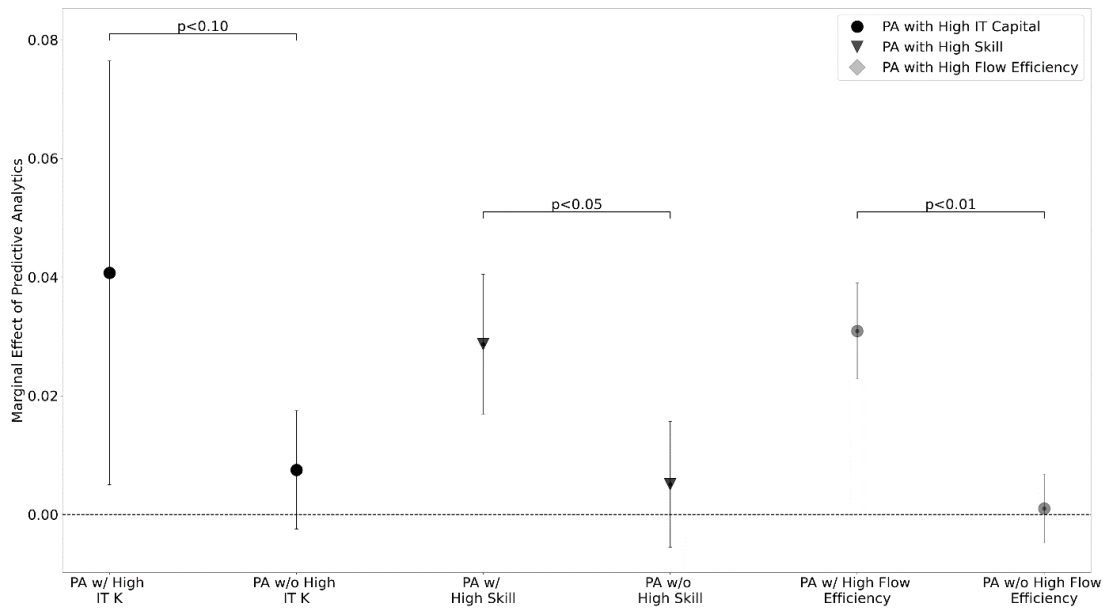
processes (i.e., continuous flow and cellular manufacturing). All three coefficients are significant at the 1% level or higher. Additional controls include industry (6-digit NAICS) and year, as well as plant-level employment in log terms; logged non-IT capital stock; structured management practices (index); having top data-driven decision-making practices (DDD); having top KPI monitoring practice; age; an indicator that government regulations or agencies chose, at least in part, what type of data is collected at the plant; multi-unit status; and headquarters status. Robust standard errors are clustered at the firm level. Findings are robust to binary (e.g., probit) estimation models. Detailed regression coefficients are omitted due to space limitations but available upon request.

Findings from Performance Tests

Following Brynjolfsson and Milgrom (2013), we explore the extent to which organizational features can help explain the heterogeneous returns of predictive analytics, as complementarity theory would predict. We find clear evidence that predictive analytics contributes far more to performance when combined with the right workplace complements.

The results are presented in Figure 6. The y-axis indicates the magnitude of the coefficients and the x-axis labels indicate the categories for each complement (e.g., predictive analytics with high IT capital stock and without high IT capital stock). Confidence intervals (at 95%) are plotted to indicate statistical significance. All three interaction terms are positive and significant, consistent with strong complementarity. Notably, with high levels of IT capital, a significant share of educated employees, or high flow-efficiency manufacturing processes, the effects of predictive analytics are at around 4.1 percent, 2.9 percent, and 3.1 percent respectively (see regression results in Table 5). Significance of the differences between the interacted terms and main effects is also reported in Figure 6. Strikingly, the marginal effects of predictive analytics are never statistically different from zero, *unless* they are combined with these other tangible and intangible workplace investments.

Figure 6. 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 3 using the baseline sample. For performance tests of complementarity, we add interactions between the indicator for the adoption of predictive analytics and each potential complement, including high IT capital stock, high percentage of employees with a bachelor's degree, and having high flow efficiency, respectively. The coefficients for the indicator of predictive analytics and the interaction term 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 predictive analytics adoption for plants with and without the presence of each complement. Additional interactions (2-way or 3-way “bundles” of complements) do not provide additional identifying variation. Error bars indicate the confidence intervals at the 95% level. Full results, including coefficients on key controls, are available upon request.

Table 5. Organizational Complements to Predictive Analytics (Performance Test)

Models	(1)	(2)	(3)
	IT Capital Stock	Skilled Workers	Continuous Flow
Dependent Variables			
PA Use	0.0075 (0.0051)	0.0051 (0.0054)	0.0010 (0.0058)
High IT K	0.1103*** (0.0171)		
PA × High IT K	0.0333* (0.0185)		
High Skill		0.0371*** (0.0101)	
PA × High Skill		0.0228** (0.0114)	
High Flow Efficiency			0.0158*

			(0.0088)
PA × High Flow Efficiency			0.0306*** (0.0094)
Joint Tests	0.0407** (0.0182)	0.0287*** (0.0106)	0.0309*** (0.0081)
Other Controls		Y	
Industry x Year Fixed Effects		Y	
R-Squared	0.9326	0.9327	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 K is an indicator for plants with the top ten percentile of IT capital stock. High Skill is an indicator for plants with the top quartile of the percentage of employees with a bachelors’ degree. High Flow Efficiency is an indicator for plants whose production process is best characterized as continuous flow manufacturing (i.e. continuous flow and cellular manufacturing). Columns 1-3 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, high structured management (defined as having top quartile structured management practices), indicators for data-driven-decision making practices other than KPI tracking, an indicator for having top KPI monitoring practice, 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.

Overall, 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.

5. Conclusion

There has been explosive growth of digital information and an accompanying increase in business expenditure on data and analytics. Although compelling anecdotal and small-sample evidence exists that predictive analytics is associated with improved performance in some settings, stories of unrealized potential also abound, and evidence outside of specific applications remains lacking. In this paper, we analyze the productivity effects of predictive analytics by analyzing over 30,000 plants while controlling a long list of potentially confounding variables. We assess causality in two ways and explore the role of complementary workplace investments in tangible and intangible infrastructure and management practices.

We were able to do this by working with the U.S. Census Bureau to field a purpose-designed survey for a representative sample of the U.S. manufacturing sector. This sector has historically

been a leading adopter of frontier technologies, 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 that also have increasingly well-understood economic dynamics (e.g., Decker *et al.* 2020).

In our sample, we find that plants have extensively adopted predictive analytics. Those plants reporting use of predictive analytics show 1 percent to 3 percent higher productivity on average, which is worth roughly \$464,000 to over \$918,000 in increased sales for the average plant. We find clear evidence that the higher performance is caused by the use of predictive analytics and not merely a spurious correlation. Specifically, our quasi-experimental evidence finds higher performance for plants where government mandates exogenously increase the use of predictive analytics, and our timing analysis finds that predictive analytics precedes performance gains but not vice-versa.

Most importantly, we identify three key complements that can explain why some firms reap large gains from predictive analytics while others see no benefit. Predictive analytics generate large payoffs when combined with IT capital investment, educated workers, or high flow-efficiency production processes but not when they are implemented without at least one of these complements. Establishments are more likely to adopt predictive analytics with the presence of these complements and enjoy significantly higher productivity post-adoption. The higher performance effects of predictive analytics depend crucially on having in place the right set of workplace complements.

Our study is not without limitations. Notably, we do not observe other benefits that may accrue to analytics use but not show up in productivity estimates, such as innovation or improved administration at headquarters, which could spill over to other establishments within the firm. This may be particularly relevant for flexibility-focused plants that undertake product and process innovation activities: these can harm multi-factor productivity—as it is traditionally measured—at the plant in question, yet support survival and performance of the parent firm. Also, we do not explore the relationship of predictive analytics survival, which if positive, would argue for a lower bound on these benefits, including for workplaces lacking the right complements. Finally, our setting is restricted to the manufacturing sector. This has desirable attributes for data availability and well-understood practices for measuring performance, but may limit generalizability to

services or other settings where automated analysis of data might require different coordinated inputs.

These caveats aside, direct evidence from over 30,000 plants indicates that predictive analytics causes significant productivity increases, but only when combined with the right complementary practices. These findings contribute to prior work concerning complementarity in the organization and provide a foundation for practical insights into the mechanisms by which predictive analytics can better provide business value.

Author accepted manuscript

References

- Akerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. Identification Properties of Recent Production Function Estimators. *Econometrica*, 83(6): 2411-2451.
- Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2018. *Prediction Machines: The Simple Economics of Artificial Intelligence*, Cambridge: Harvard Business Press.
- Agrawal, Ajay, Joshua Gans, and Avi Goldfarb. 2019. Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction. *Journal of Economic Perspectives*, 33(2): 31-50.
- Ambec, Stefan, Mark A. Cohen, Stewart Elgie, and Paul Lanoie. 2013. The Porter Hypothesis at 20: Can Environmental Regulation Enhance Innovation and Competitiveness? *Review of Environmental Economics and Policy*, 7(1): 2-22.
- Andrews, Dan, Chiara Criscuolo, and Peter N. Gal. 2015. *Frontier Firms, Technology Diffusion and Public Policy*. OECD Productivity Working Papers No. 2: OECD Publishing.
- Aral, Sinan, and Peter Weill. 2007. IT Assets, Organizational Capabilities, and Firm Performance: How Resource Allocations and Organizational Differences Explain Performance Variation. *Organization Science*, 18(5): 763-780.
- Aral, Sinan, Erik Brynjolfsson, and Lynn Wu. 2012. Three-Way Complementarities: Performance Pay, Human Resource Analytics, and Information Technology. *Management Science*, 58(5): 913-931.
- Arellano, Manuel and Stephen Bond. 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2): 277-297.
- Athey, Susan, and Scott Stern. 1998. *An Empirical Framework for Testing Theories About Complementarity in Organization Design*. NBER Working Paper No. w6600: National Bureau of Economic Research.
- Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. The Fall of the Labor Share and the Rise of Superstar Firms. *Quarterly Journal of Economics*, 135(2): 645-709.
- Bajari, Patrick, Victor Chernozhukov, Ali Hortaçsu, and Junichi Suzuki. 2019. *The Impact of Big Data on Firm Performance: An Empirical Investigation*. *AEA Papers and Proceedings* 109: 33-37.
- Bapna, Ravi, Nishtha Langer, Amit Mehra, Ram Gopal, and Alok Gupta. 2013. Human Capital Investments and Employee Performance: An Analysis of IT Services Industry. *Management Science*, 59(3): 641-658.
- Bartelsman, Eric J., and Mark Doms. 2000. Understanding Productivity: Lessons From Longitudinal Microdata. *Journal of Economic Literature*, 38(3): 569-594.
- Bartelsman, Eric J., and Wayne B. Gray. 1996. *The NBER Manufacturing Productivity Database*. NBER Working Paper No. t0205: National Bureau of Economic Research.
- Barth, Erling, James C. Davis, Richard B. Freeman, and Kristina McElheran. 2020. *Twisting the Demand Curve: Digitalization and the Older Workforce*. NBER Working Paper No. w28094: National Bureau of Economic Research.
- Bennett, Victor Manuel. 2020a. Changes in Persistence of Performance Over Time. *Strategic Management Journal*, 41(10): 1745-1769.
- . 2020b. *Automation and Market Dominance*. SSRN No. 3656713.
- Berlingieri, Giuseppe, Patrick Blanchenay, and Chiara Criscuolo. 2017. *The Great Divergence(s)*. CEP Discussion Paper No. 1488: Center for Economic Performance, London School of Economics and Political Science.

- Berman, Ron, and Ayelet Israeli. 2020. *The Value of Descriptive Analytics: Evidence from Online Retailers*. Harvard Business School Marketing Unit Working Paper No. 21-067: Harvard Business School.
- Bessen, James. 2020. Industry Concentration and Information Technology. *The Journal of Law and Economics* 63(3): 531-555.
- Bessen, James, and Cesare Righi. 2019. *Shocking Technology: What Happens When Firms Make Large IT Investments?* Working Paper No. 19-6: Boston University School of Law.
- Black, Sandra E., and Lisa M. Lynch. 1996. Human-Capital Investments and Productivity. *The American Economic Review*, 86(2): 263-267.
- Black, Sandra E., and Lisa M. Lynch. 2001. How to Compete: The Impact of Workplace Practices and Information Technology on Productivity. *Review of Economics and Statistics*, 83(3): 434-445.
- Blackwell, David. 1953. Equivalent Comparisons of Experiments. *Annals of Mathematical Statistics*: 265-272.
- Bloom, Nicholas, and John Van Reenen. 2007. Measuring and Explaining Management Practices Across Firms and Countries. *Quarterly Journal of Economics*, 122(4): 1351-1408.
- Bloom, Nicholas, Rafaella Sadun, and John Van Reenen. 2012. Americans Do I.T. Better: U.S. Multinationals and the Productivity Miracle. *American Economic Review*, 102(1): 167-201.
- Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron S. Jarmin, Megha Patnaik, Itay Saporta-Eksten and John Van Reenen. 2014. *IT and Management in America*. CEP Discussion Paper No. 1258: London School of Economics.
- Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron S. Jarmin, Megha Patnaik, Itay Saporta-Eksten and John Van Reenen. 2019. What Drives Differences in Management Practices? *American Economic Review* 109(5): 1648-1683.
- Blum, Bernardo, Avi Goldfarb, and Mara Lederman. 2015. The Path to Prescription: Closing the Gap Between the Promise and the Reality of Big Data. *Rotman Management Magazine*, Fall.
- Blundell, Richard, and Stephen Bond. 2000. GMM Estimation with Persistent Panel Data: An Application to Production Functions. *Econometric Reviews*, 19(3): 321-340.
- Bresnahan, Timothy, and Shane Greenstein. 1996. Technical Progress and Co-Invention in Computing and in the Uses of Computers. *Brookings Papers on Economic Activity*, 27(Micr.):1-83.
- Bresnahan, Timothy, Erik Brynjolfsson, and Lorin M. Hitt. 2002. Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *Quarterly Journal of Economics*, 117(1): 339-376.
- Brynjolfsson, Erik, and Lorin M. Hitt. 1995. Information Technology as a Factor of Production: The Role of Differences Among Firms. *Economics of Innovation and New Technology*, 3(3-4): 183-200.
- Brynjolfsson, Erik, and Lorin M. Hitt. 2000. Beyond Computation: Information Technology, Organizational Transformation and Business Performance. *Journal of Economic Perspectives*, 14(4): 23-48.
- Brynjolfsson, Erik, and Lorin M. Hitt. 2003. Computing Productivity: Firm-Level Evidence. *Review of Economics & Statistics*, 85(4): 793-808.
- Brynjolfsson, Erik, and Kristina McElheran. 2016. The Rapid Rise of Data-Driven Decision Making. *American Economic Association Papers and Proceedings* 106: 133-139.
- Brynjolfsson, Erik, and Kristina McElheran. 2019. Data in Action: Data-Driven Decision Making and

- Predictive Analytics in U.S. Manufacturing. Working Paper No. 3422397: Rotman School of Management.
- Brynjolfsson, Erik, and Paul Milgrom. 2013. Complementarity in Organizations, in *The Handbook of Organizational Economics* (Robert Gibbons and John Roberts, eds). Princeton: Princeton University Press.
- Brynjolfsson, Erik, Lorin M. Hitt, and Heekyung Hellen Kim. 2011. *Strength in Numbers: How Does Data-Driven Decision-Making Affect Firm Performance?* SSRN No. 1819486.
- Brynjolfsson, Erik, Daniel Rock, and Chad Syverson. 2021. The Productivity J-Curve: How Intangibles Complement General Purpose Technologies. *American Economic Journal: Macroeconomics*, 13(1): 333-372.
- Buffington, Catherine, Lucia Foster, Ron Jarmin, and Scott Ohlmacher. 2017. The Management and Organizational Practices Survey (MOPS): An Overview. *Journal of Economic and Social Measurement*, 42(1): 1-26.
- Bughin, Jacques. 2016. Big Data, Big Bang? *Journal of Big Data*, 3(1): 1-14.
- Caroli, Eve, and John Van Reenen. 2001. Skill-Biased Organizational Change? Evidence From a Panel of British and French Establishments. *Quarterly Journal of Economics*, 116(4): 1449-1492.
- Cassiman, Bruno, and Reinhilde Veugelers. 2006. In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition. *Management Science*, 52(1): 68-82.
- Choudhury, Prithwiraj, Evan Starr, and Rajshree Agarwal. 2020. Machine Learning and Human Capital Complementarities: Experimental Evidence on Bias Mitigation. *Strategic Management Journal* 41(8): 1381-1411.
- Clark, Kim B., and Joshua D. Margolis. 1991. *Workplace Safety at Alcoa (A)*. Case No. 692-042: Harvard Business School.
- Collard-Wexler, Allan, and Jan De Loecker. 2015. Reallocation and Technology: Evidence From the US Steel Industry. *American Economic Review*, 105(1): 131-171.
- Collis, David, David Young, and Michael Goold. 2007. The Size, Structure, and Performance of Corporate Headquarters. *Strategic Management Journal*, 28(4): 383-405.
- Davenport, Thomas H. 2006. Competing on Analytics. *Harvard Business Review*, 84(1): 98.
- David, Paul A. 1969. *A Contribution to the Theory of Diffusion*. Memorandum No. 71: Stanford University.
- Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda. 2020. Changing Business Dynamism and Productivity: Shocks Versus Responsiveness. *American Economic Review*, 110(12): 3952-3990.
- Doms, Mark, Timothy Dunne, and Mark J. Roberts. 1995. The Role of Technology Use in the Survival and Growth of Manufacturing Plants. *International Journal of Industrial Organization*, 13(4): 523-542.
- Dranove, David, Chris Forman, Avi Goldfarb, and Shane Greenstein. 2014. The Trillion Dollar Conundrum: Complementarities and Health Information Technology. *American Economic Journal: Economic Policy*, 6(4): 239-270.
- Dunne, Timothy. 1994. Plant Age and Technology Use in U.S. Manufacturing Industries. *RAND Journal of Economics*, 25(3): 488-499.
- Englemaier, Florian, Jose E. Galdon-Sanchez, Richard Gil, and Michael Kaiser. *Management Practices and Firm Performance During the Great Recession: Evidence from Spanish Survey Data*. Mimeo, LMU Munich.

- Enke, Benjamin. 2020. What You See Is All There Is. *Quarterly Journal of Economics*, 135(3): 1363-1398.
- Foster, Lucia, John Haltiwanger, and Chad Syverson. 2016. The Slow Growth of New Plants: Learning About Demand? *Economica*, 83(329): 91-129.
- Gollop, Frank M., and Mark J. Roberts. 1983. Environmental Regulations and Productivity Growth: The Case of Fossil-Fueled Electric Power Generation. *Journal of political Economy*, 91(4): 654-674.
- Gray, Wayne B. 1987. The Cost of Regulation: OSHA, EPA and the Productivity Slowdown. *The American Economic Review*, 77(5): 998-1006.
- Griliches, Zvi. 1957. Hybrid Corn: An Exploration in the Economics of Technological Change. *Econometrica*, 25(4): 501-522.
- Hall, Bronwyn H. 2004. Innovation and Diffusion, in *The Oxford Handbook of Innovation* (Jan Fagerberg and David C. Mowery, eds). Oxford: Oxford University Press.
- Haskel, Jonathan, and Stian Westlake. 2018. *Capitalism Without Capital: The Rise of the Intangible Economy*. Princeton: Princeton University Press.
- Hayes, Robert H., and Steven C. Wheelwright. 1979. Link Manufacturing Process and Product Life Cycles. *Harvard Business Review*, 57(1): 133-140.
- Helper, Susan, Raphael Martins, and Robert Seamans. 2019. *Who Profits from Industry 4.0? Theory and Evidence from the Automotive Industry*. SSRN No. 3377771
- Holmstrom, Bengt, and Paul Milgrom. 1994. The Firm as an Incentive System. *American Economic Review*, 84(4): 972-991.
- Hong, Bryan, Lorenz Kueng, and Mu-Jeung Yang. 2019. Complementarity of Performance Pay and Task Allocation. *Management Science*, 65(11): 5152-5170.
- Hopenhayn, Hugo A. 2014. Firms, Misallocation, and Aggregate Productivity: A Review. *Annual Review of Economics*, 6(1): 735-770.
- Jaffe, Adam B., Steven R. Peterson, Paul R. Portney, and Robert N. Stavins. 1995. Environmental Regulation and the Competitiveness of US Manufacturing: What Does the Evidence Tell Us? *Journal of Economic literature*, 33(1): 132-163.
- Jin, Wang, and Kristina McElheran. 2018. *Economies Before Scale: Learning, Survival and Performance of Young Plants in the Age of Cloud Computing*. SSRN Working Paper No. 3112901.
- Junqué de Fortuny, Enric, David Martens, and Foster Provost. 2013. Predictive Modeling with Big Data: Is Bigger Really Better? *Big Data*, 1(4): 215-226.
- Kandel, Eugene, and Edward P. Lazear. 1992. Peer Pressure and Partnerships. *Journal of political Economy*, 100(4): 801-817.
- Lanoie, Paul, Jérémy Laurent-Lucchetti, Nick Johnstone, and Stefan Ambec. 2011. Environmental Policy, Innovation and Performance: New Insights on the Porter Hypothesis. *Journal of Economics & Management Strategy*, 20(3): 803-842.
- Lashkari, Danial, Arthur Bauer, and Jocelyn Boussard. 2020. *Information Technology and Returns to Scale*. Working Paper No. No. G2020/14: Institut National de la Statistique et des Études Économiques.
- Levinsohn, James, and Amil Petrin. 2003. Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies*, 70(2): 317-341.
- Long, J. Scott, and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using*

Stata (Vol. 7). College Station: Stata Press.

Martens, David, Foster Provost, Jessica Clark, and Enric Junqué de Fortuny, 2016. Mining Massive Fine-Grained Behavior Data to Improve Predictive Analytics. *MIS Quarterly*, 40(4): 869-888.

McAfee, Andrew, and Erik Brynjolfsson. 2012. Big Data: The Management Revolution. *Harvard Business Review*, 90(10): 61-67.

McElheran, Kristina. 2015. Do Market Leaders Lead in Business Process Innovation? The Case(s) of E-Business Adoption. *Management Science*, 61(6): 1197-1216.

McElheran, Kristina, and Chris Forman. 2019. Firm Organization in the Digital Age: IT Use and Vertical Transactions in US Manufacturing. Available at SSRN 3396116.

McElheran, Kristina, and Wang Jin. 2020. *Strategic Fit of IT Resources in the Age of Cloud Computing*. Unpublished, University of Toronto.

McElheran, Kristina, Scott Ohlmacher, and Mu-Jeung Yang. 2020. *Strategy and Structured Management*. Unpublished, University of Toronto.

Melville, Nigel, Kenneth Kraemer, and Vijay Gurbaxani. 2004. Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value. *MIS Quarterly*, 28(2): 283-322.

Milgrom, Paul, and John Roberts. 1990. The Economics of Modern Manufacturing: Technology, Strategy, and Organization. *American Economic Review*, 80(3): 511-528.

Milgrom, Paul, and John Roberts. 1995. Complementarities and Fit Strategy, Structure, and Organizational Change in Manufacturing. *Journal of Accounting and Economics*, 19(2-3): 179-208.

Moretti, Enrico. 2004. Workers' Education, Spillovers, and Productivity: Evidence From Plant-Level Production Functions. *American Economic Review*, 94(3): 656-690.

Müller, Oliver, Maria Fay, and Jan Vom Brocke. 2018. The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics. *Journal of Management Information Systems*, 35(2): 488-509.

Nagle, Frank. 2019. Open Source Software and Firm Productivity. *Management Science*, 65(3): 1191-1215.

Pfeffer, Jeffrey, and Robert I. Sutton. 2006. Evidence-Based Management. *Harvard Business Review*, 84(1): 62.

Porter, Michael E. 1991. America's Green Strategy. *Scientific American*, 264(4): 186.

Porter, Michael E., and Claas Van der Linde. 1995. Toward a New Conception of the Environment-Competitiveness Relationship. *Journal of Economic Perspectives*, 9(4): 97-118.

Raiffa, Howard. 1968. *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*. Reading, MA: Addison-Wesley.

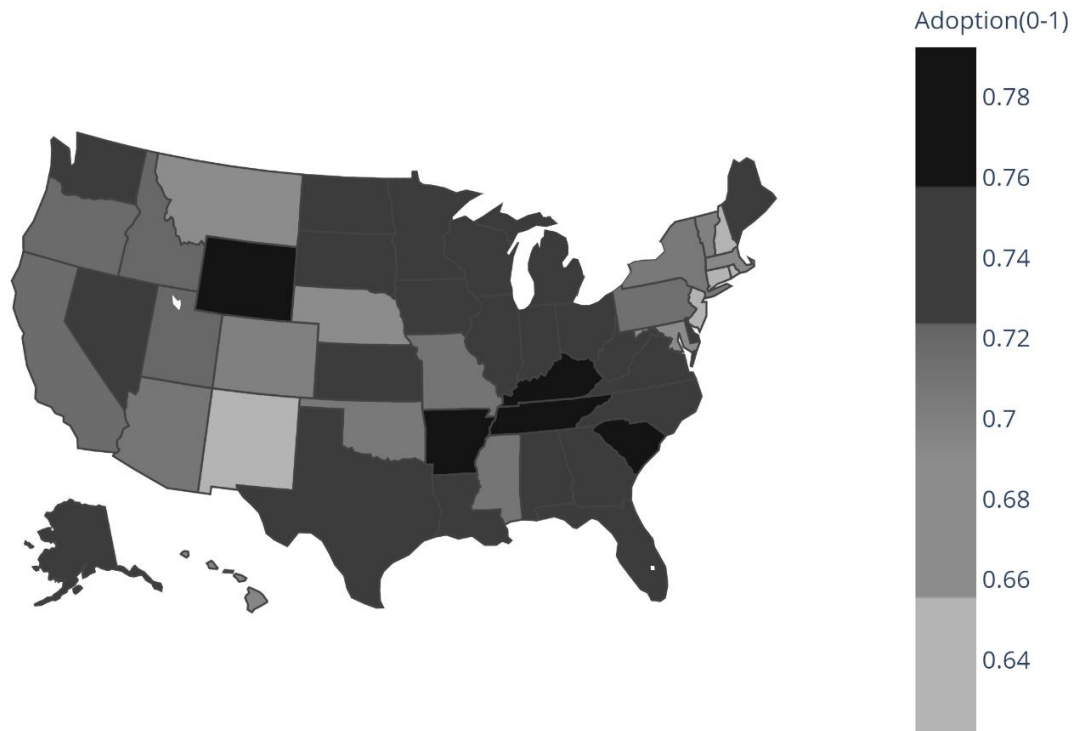
Ransbotham, Sam, David Kiron, and Pamela Kirk Prentice. 2015. Minding the Analytics Gap, *MIT Sloan Management Review*, Spring.

Ransbotham, Sam, David Kiron, and Pamela Kirk Prentice. 2016. Beyond the Hype: The Hard Work Behind Analytics Success. *MIT Sloan Management Review*, 57(3): 3-16.

Ransbotham, Sam, David Kiron, Philipp Gerbert, and Martin Reeves. 2017. Reshaping Business With Artificial Intelligence: Closing the Gap Between Ambition and Action. *MIT Sloan Management Review*, September 6.

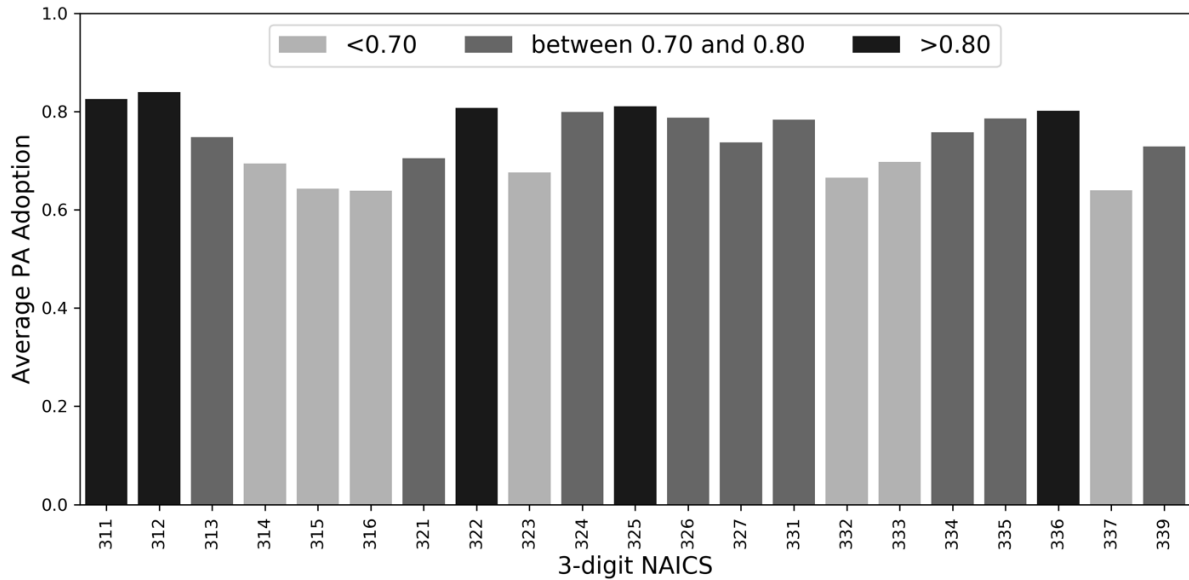
- Safizadeh, M. Hossein, Larry P. Ritzman, Deven Sharma, and Craig Wood. 1996. An Empirical Analysis of the Product-Process Matrix. *Management Science*, 42(11): 1576-1591.
- Saunders, Adam, and Erik Brynjolfsson. 2016. Valuing IT-Related Intangible Assets. *MIS Quarterly*, 40(1): 83-110.
- Saunders, Adam, and Prasanna Tambe. 2015. *Data Assets and Industry Competition: Evidence From 10-K Filings*. Available at SSRN 2537089.
- Schrage, Michael. 2014. Learn from Your Analytics Failures. *Harvard Business Review*, September.
- Scur, Daniela, Raffaella Sadun, John Van Reenen, Renata Lemos, and Nicholas Bloom. 2021. *The World Management Survey at 18: Lessons and the Way Forward*. NBER Working Paper No. w28524: National Bureau of Economic Research.
- Siegel, Eric. 2013. *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die*. Hoboken: John Wiley & Sons.
- Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter. 2019. Firming Up Inequality. *Quarterly Journal of Economics*, 134(1): 1-50.
- Stiroh, Kevin J. 2002. Information Technology and the US Productivity Revival: What Do the Industry Data Say? *American Economic Review*, 92(5): 1559-1576.
- Syverson, Chad. 2004. Product Substitutability and Productivity Dispersion. *Review of Economic Studies*, 86(2): 534-550.
- _____. 2011. What Determines Productivity? *Journal of Economic Literature*, 49(2): 326-365.
- _____. 2017. Challenges to Mismeasurement Explanations for the US Productivity Slowdown. *Journal of Economic Perspectives*, 31(2): 165-186.
- Tambe, Prasanna. 2014. Big Data Investment, Skills, and Firm Value. *Management Science*, 60(6): 1452-1469.
- Tambe, Prasanna, and Lorin M. Hitt. 2012. The Productivity of Information Technology Investments: New Evidence from IT Labor Data. *Information Systems Research*, 23(3): 599-617.
- Van Reenen, John. 2018. *Increasing Differences Between Firms: Market Power and the Macro Economy*. CEP Discussion Paper No. 1576: London School of Economics.
- White, T. Kirk, Jerome P. Reiter, and Amil Petrin. 2018. Imputation in U.S. Manufacturing Data and its Implications for Productivity Dispersion. *The Review of Economics and Statistics*, 100(3): 502-509.
- Wu, Lynn, Lorin M. Hitt, and Bowen Lou. 2020. Data Analytics, Innovation, and Firm Productivity. *Management Science*, 66(5): 2017-2039.
- Wu, Lynn, Bowen Lou, and Lorin M. Hitt. 2019. Data Analytics Supports Decentralized Innovation. *Management Science*, 65(10): 4863-4877.
- Zolas, Nikolas, Zachary Kroff, Erik Brynjolfsson, Kristina McElheran, David Beede, Catherine Buffington, Nathan Goldschlag, Lucia Foster, and Emin Dinlersoz. 2020. *Advanced Technologies Adoption and Use by US Firms: Evidence From the Annual Business Survey*. NBER Working Paper No. 28290: National Bureau of Economic Research.

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



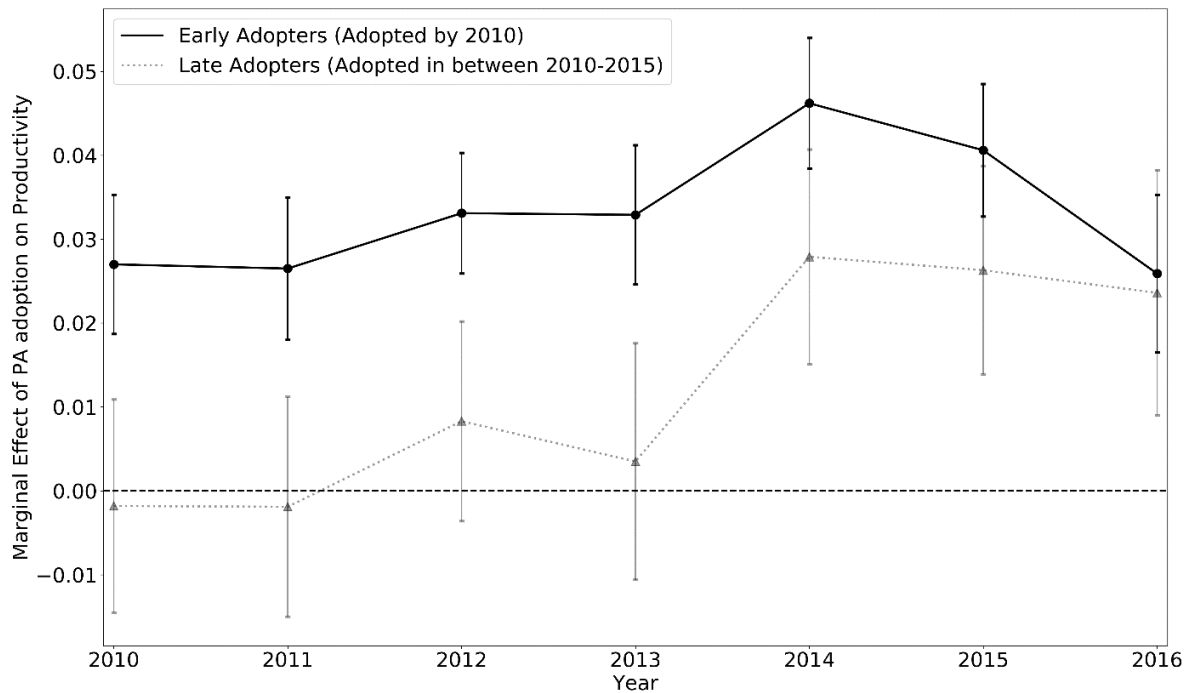
Notes: Reported statistics in the legend are the average adoption of predictive analytics across the states in the United States based on the baseline sample in 2010. This sample consists of establishments in 2015 MOPS samples (with 2010 recall) that can be merged with the Annual Survey of Manufactures (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)**



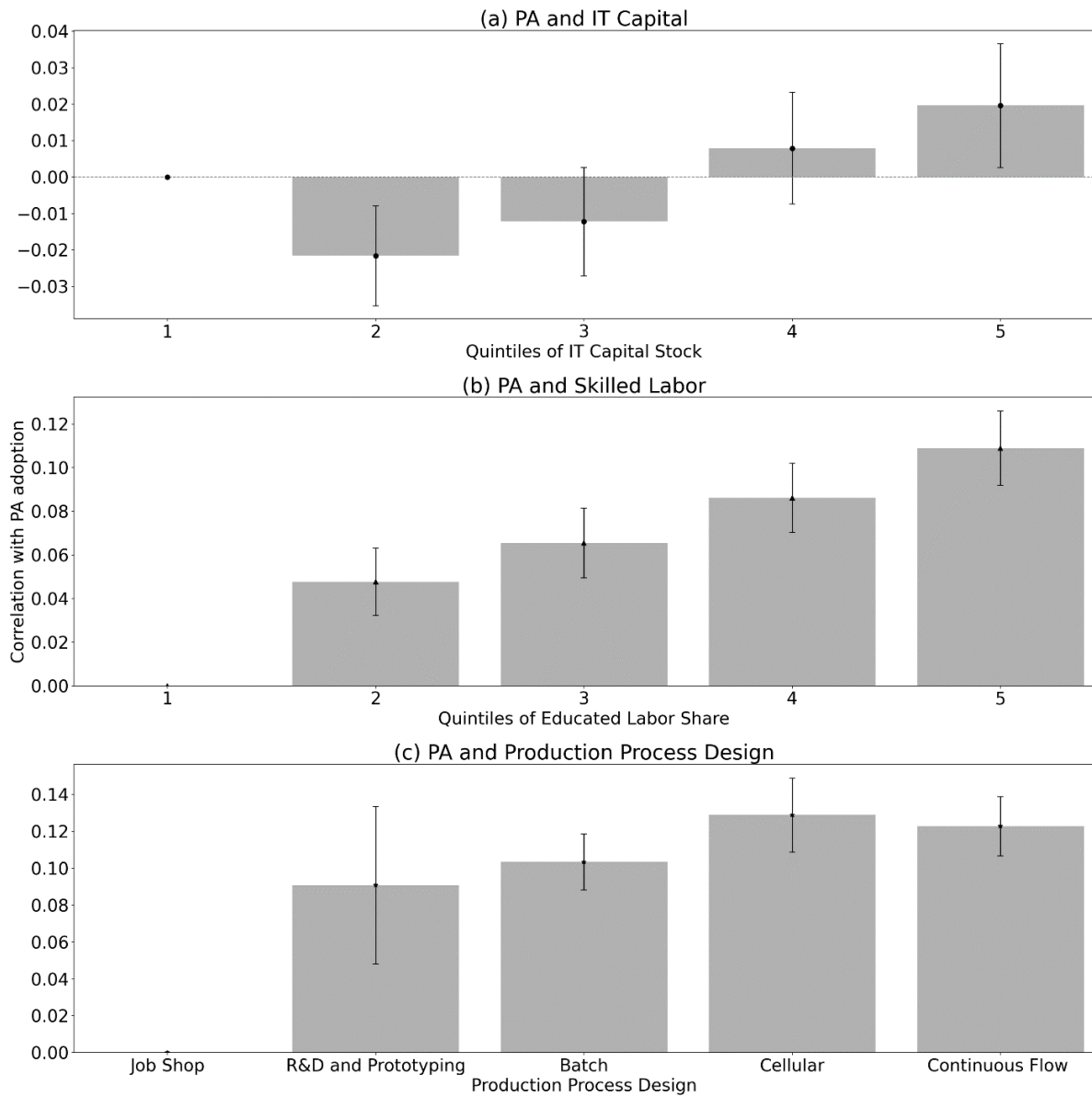
Notes: Reported statistics are based on the baseline sample for the adoption of predictive analytics in 2010. The average adoption rate is shown on the Y-axis. The 3-digit NAICS codes are shown in the X-axis and the corresponding industry definitions are listed in the table below. Darker color indicates 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 available upon request.

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 Effects of Predictive Analytics by Early and Late Adopters Over Time

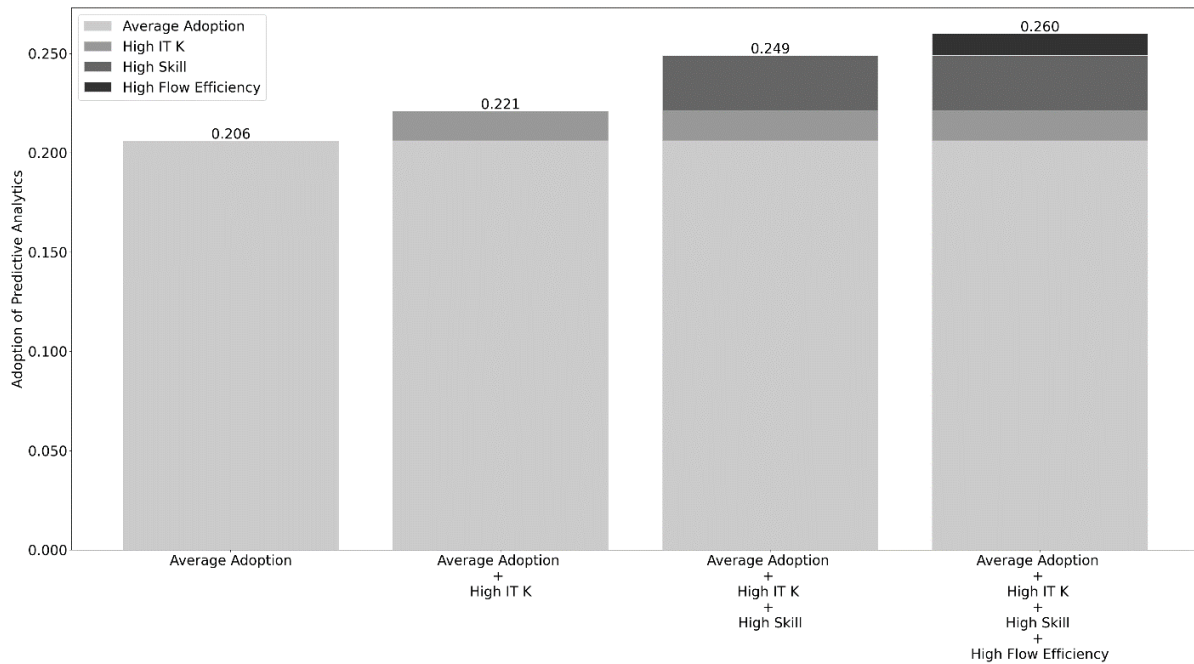
Notes: Estimates based on a pooled OLS model with a specification identical to the baseline model in column 2 table 3. For this test, we construct an ASM and CMF panel, where we have annual data on most of the key inputs (except the managerial-related variables from the 2015 MOPS) and sales from 2010 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 late adopters (non-adopters) using the 2015 MOPS data. These indicators are then interacted with year dummies to explore the differences in sales over time (using late adopters as the baseline group). Histogram bars (and values on the Y-axis) represent the marginal effect of predictive analytics adoption between 2010 and 2016. Standard errors of the coefficients are plotted on the histogram bars. Detailed regression coefficients are omitted due to space limitations but available upon request.

Figure 4 Conditional Correlations between Predictive Analytics and Potential Complements



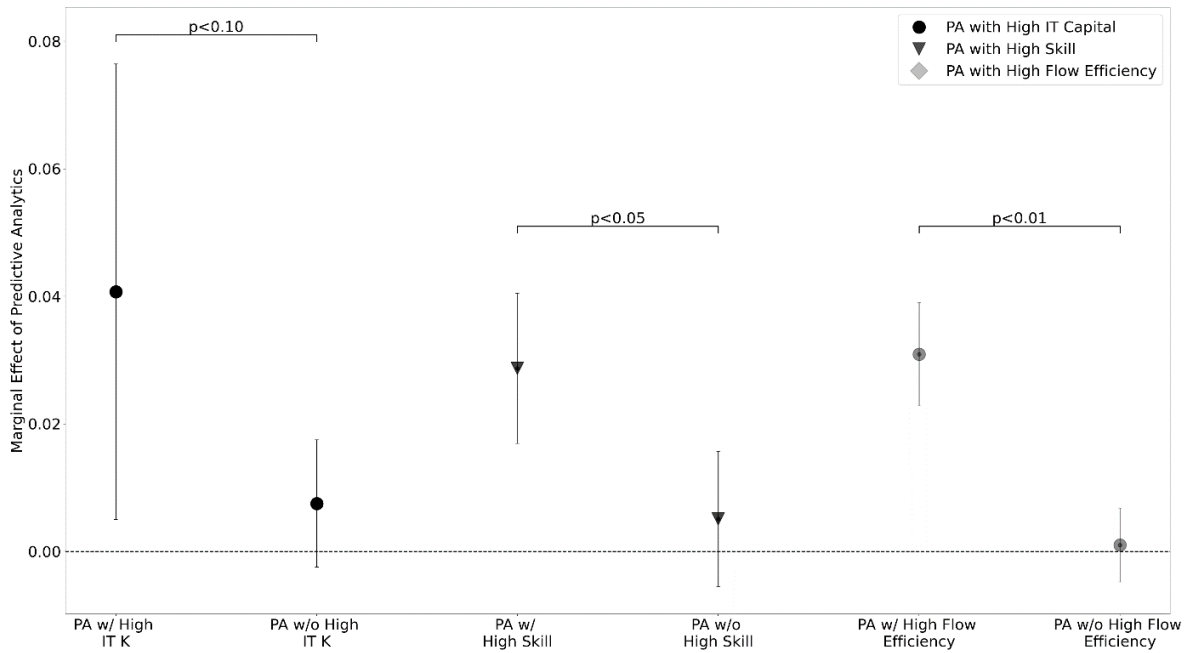
Notes: Estimates based on the baseline sample from the pooled OLS regressions controlling for plant size, plant age, and industry (6-digit NAICS) and year fixed-effects. The dependent variable is the adoption of predictive analytics. The base group is the plants in the bottom quintile of the sample based on calculated total IT capital stock for panel (a). For panel (b), the base group is the bottom quintile of the sample based on the percentage of workers with bachelor’s degrees. For panel (c), the base group is job shop facilities. Histogram bars (and values on the Y-axis) represent 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 darker lines.

Figure 5 Organizational Complements to Predictive Analytics (Correlation Test)



Notes: Correlates of predictive analytics use based on linear probability estimation in the baseline pooled sample. The graph depicts estimated additive marginal contributions based on a single model that includes High IT Capital Stock, High Employee Education, and High Flow Efficiency. Average adoption represents the constant term as the benchmark adoption rate. High IT K is an indicator for plants in the top quintile of IT capital stock. High Employee Education is an indicator for plants with top quartile of the percentage of employees with a bachelors’ degree. High Flow Efficiency is an indicator for plants whose production process is best characterized as high flow manufacturing processes (i.e. continuous flow and cellular manufacturing). All three coefficients are significant at the 1% level or higher. Additional controls include industry (6-digit NAICS) and year, as well as plant-level employment in log terms, logged non-IT capital stock, structured management practices (index), having top data-driven decision-making practices (DDD), having top KPI monitoring practice, age, an indicator that government regulations or agencies chose, at least in part, what type of data is collected at the plant, multi-unit status, and headquarters status. Robust standard errors are clustered at the firm level. Findings are robust to binary (e.g., probit) estimation models. Detailed regression coefficients are omitted due to space limitations but available upon request.

Figure 6 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 3 using the baseline sample. For performance tests of complementarity, we add interactions between the indicator for the adoption of predictive analytics and each potential complement, including high IT capital stock, high percentage of employees with bachelors’ degree, and having high flow efficiency, respectively. The coefficients for the indicator of predictive analytics and the interaction term 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 predictive analytics adoption for plants with and without the presence of each complement. Additional interactions (2-way or 3-way “bundles” of complements) do not provide additional identifying variation. Error bars indicate the confidence intervals at the 95% level. Full results, including coefficients on key controls, are available upon request.

Table 1. Predictive Analytics question in MOPS

Q29: How frequently does this establishment typically reply on predictive analytics (statistical models that provide forecasts in areas such as demand, production, or human resources)? *Mark all that apply*

	2010	2015
Daily		
Weekly		
Monthly		
Yearly		
Never		

Source: this table is captured from the questionnaire for the 2015 MOPS section C question 29. The PDF version of the questionnaire for 2015 MOPS can be downloaded at the U.S. Census website: https://www2.census.gov/programs-surveys/mops/technical-documentation/questionnaires/ma-10002_15_final_3-2-16.pdf.

Table 2. Summary Statistics (Key Variables)

Variable	Definition	Mean (S.D.)	2010 (Recall)	2015
PA Use	Indicator for plants that use predictive analytics at any frequency	0.74 (0.44)	0.73 (0.44)	0.80 (0.40)
PA Use Frequency	An index for frequency of predictive analytics use based on the highest reported value (e.g. Yearly=1, Monthly=2, Weekly=3, and/or Daily=4)	1.12 (1.06)	1.09 (1.05)	1.27 (1.12)
Log Sales	Logged total value of plant shipments (\$Thousands)	10.37 (1.52)	10.68 (1.39)	10.86 (1.37)
Log L	Logged number of plant employees	4.56 (1.17)	4.79 (1.09)	4.88 (1.09)
Log (Non-IT) K	Accumulated and depreciated capital investment in non-IT equipment and structures in log terms (\$Thousands)	9.26 (1.47)	9.38 (1.58)	9.36 (1.61)
Log IT K	IT capital stock in log (\$Thousands)	5.16 (2.41)	5.58 (2.25)	5.62 (2.18)
Skilled Workers	Percentage of employees (managers and non-managers) with a bachelor's degree	0.15 (0.14)	0.15 (0.13)	0.16 (0.14)
Production Process Design (High Flow Efficiency)	Indicator for facilities with a production process designed for high flow efficiency (i.e. cellular or continuous-flow production process) as captured by question 44 of the 2015 MOPS.	0.35 (0.48)	0.38 (0.48)	0.41 (0.49)
MU	Indicator for plants belonging to multi-unit firms	0.73 (0.45)	0.78 (0.41)	0.81 (0.40)
HQ	Indicator for establishments reported being headquarter (HQ) or co-located with HQ	0.47 (0.50)	0.43 (0.50)	0.41 (0.049)

Plant age	Plant age	24.47 (12.89)	24.20 (11.31)	29.20 (11.31)
Government Mandate	Indicator that government regulations or agencies chose, at least in part, what type of data is collected at the plant	0.25 (0.43)	N/A	N/A
Structured Management	Normalized index for structured management practices using section A of the MOPS (excluding data-related questions 2 and 6)	0.63 (0.17)	0.60 (0.16)	0.68 (0.15)
DDD	Indicator for plants with high Data-Driven Decision-making practices as captured in the 2015 MOPS based on question 2 (KPI monitoring), question 6 (target setting), question 24 (data availability for decision making), and question 25 (using data in decision making); Equal to 1 for plants that set both short and long-term targets and have top 2 categories for both data availability and use.	0.27 (0.44)	0.19 (0.39)	0.38 (0.49)
Number of Observations		~51,000 (Baseline)	~18,000 (Balanced)	

Notes: Unweighted statistics based on the baseline and balanced samples from MOPS 2015 data; Standard deviations in parentheses.

Table 3. 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 Use	0.0287*** (0.0049)	0.0145*** (0.0049)		
PA Use Frequency (top frequency)			0.0089*** (0.0021)	0.0509*** (0.0160)
Log IT K		0.0227*** (0.0014)	0.0227*** (0.0014)	0.0226*** (0.0014)
Skilled Workers		0.2068*** (0.0177)	0.2063*** (0.0177)	0.1928*** (0.0181)
High Flow Efficiency		0.0407*** (0.0052)	0.0406*** (0.0052)	0.0385*** (0.0053)
High Structured Mgmt.		0.0283*** (0.0054)	0.0273*** (0.0054)	0.0165** (0.0069)
Log L	0.4052*** (0.0065)	0.3821*** (0.0063)	0.3820*** (0.0063)	0.3800*** (0.0063)
Log (Non-IT) K	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	0.9327	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. Column 1 controls for key production inputs while column 2 adds controls for IT capital stock, percentage of employees with a bachelors' degree, an indicator for high flow efficiency production process design, high structured management (defined as having top quartile structured management practices), and an indicator for data-driven-decision making practices. Column 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, and plant age. Robust standard errors clustered at the firm level. * p<0.10, ** p<0.05, *** p<0.01.

Table 4. Organizational Complements to Predictive Analytics (Performance Test)

Models	(1)	(2)	(3)
	IT Capital Stock	Skilled Workers	Continuous Flow
Dependent Variables			
PA Use	0.0075 (0.0051)	0.0051 (0.0054)	0.0010 (0.0058)
High IT K	0.1103*** (0.0171)		
PA × High IT K	0.0333* (0.0185)		
High Skill		0.0371*** (0.0101)	
PA × High Skill		0.0228** (0.0114)	
High Flow Efficiency			0.0158* (0.0088)
PA × High Flow Efficiency			0.0306*** (0.0094)
Joint Tests	0.0407** (0.0182)	0.0287*** (0.0106)	0.0309*** (0.0081)
Other Controls		Y	
Industry x Year Fixed Effects		Y	
R-Squared	0.9326	0.9327	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 K is an indicator for plants with top ten percentile of IT capital stock. High Skill is an indicator for plants with top quartile of the percentage of employees with a bachelors' degree. High Flow Efficiency is an indicator for plants whose production process is best characterized as continuous flow manufacturing (i.e. continuous flow and cellular manufacturing). Columns 1-3 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, high structured management (defined as having top quartile structured management practices), indicators for data-driven-decision making practices other than KPI tracking, an indicator for having top KPI monitoring practice, 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.

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 Use	0.0223** (0.0091)		
PA Use Frequency (Average)		0.0023** (0.0006)	0.0167*** (0.0053)
Log IT K	0.0491*** (0.0026)	0.0228*** (0.0014)	0.0228*** (0.0014)
Skilled Workers	0.4670*** (0.0351)	0.2076*** (0.0177)	0.1978*** (0.0178)
High Flow Efficiency	0.0786*** (0.0107)	0.0408*** (0.0052)	0.0392*** (0.0052)
High Structured Mgmt.	0.0534*** (0.0109)	0.0269*** (0.0054)	0.0108 (0.0082)
Log L	0.7796*** (0.0081)	0.3820*** (0.0063)	0.3793*** (0.0063)
Log (non-IT) K	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 Use Frequency (Average) 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.