The Power of Prediction:

Predictive Analytics, Workplace Complements, and Business Performance[1]

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

This Version: June 29, 2021

Preliminary – Do Not Circulate

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 U.S. Census Bureau, we surveyed over 30,000 manufacturing establishments on their use of predictive analytics and detailed workplace characteristics. We find that productivity is significantly higher among plants that use tools to automate prediction—up to $918,000 higher sales compared to non-adopting competitors using similar inputs. Furthermore, both instrumental variables estimates and timing of gains suggest a causal relationship. However, this productivity payoff only occurs when predictive analytics is combined with key workplace complements. Significant accumulation of IT capital, educated workers, or workplaces designed for flow-efficient production each enables non-trivial returns. Notably, managerial capacity – measured either as management practices or as manager headcount – also enables both predictive analytics adoption and performance gains. Our findings support claims that these fast-diffusing techniques can substantially improve productivity, while also explaining why some firms see no benefits at all. Further, they provide the first evidence of labor-augmenting automation in manufacturing, in contrast to the labor-substitution associated with robots.

Keywords: digitization, big data, predictive analytics, productivity, complementarities, contingencies, automation

---

[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 Sciences 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.

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

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

1. INTRODUCTION

Exponential growth in computing power, declining costs of information technology (IT), and the rise of new computational methods have created opportunities to extract more value from increasingly ubiquitous digital information (Tambe 2014; Bughin 2016; Wu et al. 2020). Worldwide revenue for "big data" and business analytics solutions is expected to reach $274.3 billion by 2022 (IDC 2019). However, these investments have yet to yield measurable productivity gains in the aggregate (Syverson 2017; Brynjolfsson et al. 2021). At the firm level, managers report numerous challenges in realizing performance benefits from these investments (Ransbotham et al. 2015 & 2017, Wu et al. 2019). This puzzling gap between voluntary firm adoption and tangible payoffs has been difficult to study due (perhaps ironically) to a dearth of large-sample data.

We shed light on the adoption-payoff gap for predictive analytics by providing the largest and richest evidence to date on the adoption, performance implications, and enablers of predictive analytics use in over 30,000 diverse workplace settings. Collaborating with the U.S. Census Bureau, we collect the first direct measures of predictive analytics use in a representative sample of U.S. manufacturing establishments,[2] also capturing critical tangible and often-intangible workplace characteristics. We further link this survey from 2010 and 2015 to an annual panel of detailed administrative data. Addressing many common threats to identification, we explore both correlational and causal relationships between predictive analytics use and productivity. Further, we unearth both "classic" and novel complementarities that overwhelmingly predict the distribution of performance gains. One is a mutually-reinforcing relationship between human managerial labor and automation in this rapidly-digitizing sector.

[2] Prior work has addressed the challenge of measuring firm use of fast-emerging analytics tools by triangulating on the human capital needed by adopters, typically in smaller samples (Tambe 2014, Wu et al. 2019 & 2020). In contrast, our data cover the majority of the U.S. manufacturing economy, including not only large public firms but also a representative swath of smaller establishments. We view our approach to be complementary, with distinct advantages and challenges. See Section 2.

In the face of considerable "hype" (Gartner 2020), our findings confirm that predictive analytics use has indeed spread throughout the American manufacturing sector. More than 70 percent of plants relied on some level of predictive analytics as early as 2010, with pervasive penetration across geographies and industries, as well as across the plant size and age distributions. More variation exists at the intensive margin of use, with a mean frequency of slightly more often than annually, increasing to between monthly and weekly at one standard deviation above the mean. In this study, we develop representative statistics describing the diffusion of predictive analytics practices in recent years and present three central findings of interest to both academic researchers and practitioners.

First, this phenomenon has important implications for the real economy. Plants adopting at least some use of predictive analytics demonstrate higher productivity of 1 to 3 percent, on average. In dollar terms, this represents $464,000 to $918,000 greater sales for adopters versus non-adopters, holding constant other inputs to production. Moreover, while benefits manifest at the extensive margin, they also increase with frequency of use, consistent with a dose-response relationship. Within the range of behavior we observe, more-frequent reliance on predictive analytics is associated with increasingly higher productivity.

Second, we find evidence for a causal relationship. In addition to addressing a number of time-varying confounds, we leverage a quasi-experimental approach (instrumental variables) and the timing of the effects (performance improves 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 benefit and voluntary adoption, productivity gains are almost entirely confined to workplaces with certain complements in place. These include accumulated IT capital and educated employees, what we call "classic complements" based on foundational studies of earlier general-purpose IT (Black and Lynch 1996, Bresnahan et al. 2002, Tambe and Hitt 2012). We identify a novel third complement: production process designs favoring flow efficiency. Predictive analytics use is much more productive in flow-efficient contexts compared to flexibility-focused ones. This quasi-fixed workplace characteristic is useful for identification but difficult for firms to adjust as technology advances.

A fourth complement warrants particular attention: human managers. Workplaces with high managerial capacity for interpreting and responding to objective information show outsized returns from predictive analytics use. This is further associated with a sustained level of managerial headcount over time. These benefits of keeping "humans in the loop" are important in light of rising apprehension about labor demand in the face of increasing automation (e.g., Autor 2015, Acemoglu and Restrepo 2018, Brynjolfsson et al. 2018, Furman and Seamans 2019) and recent evidence that humans may hinder, rather than enhance, the gains achievable with enhanced prediction (e.g., Glaeser et al. 2021). To our knowledge, ours is the first robust evidence of labor augmentation in a sector that is rapidly digitizing (Zolas et al. 2020) and has shown evidence of labor displacement associated with robotics investment (Acemoglu et al. 2020), particularly for managers (Dixon et al. 2019).

Another important feature of our study is the five-year span of time it covers. Heterogeneity in the distribution of "classic complements" is not only widespread – it persists. Evidence that the rate of organizational change has lagged the advances in technology– despite longstanding evidence about what effective digitization may demand – is informative. It helps explain why some firms achieve significant gains from predictive analytics, while others see no benefit at all. It also highlights the ongoing need for more micro-level attention to organizational adjustment and co-invention in studies of technological change and firm performance (Bresnahan and Greenstein 1996, Tambe and Hitt 2012, McElheran 2015, Raj and Seamans 2018, Brynjolfsson et al. 2021).

Also of note, our findings hold while controlling for a wide range of other practices associated with productivity (e.g., Brynjolfsson and McElheran 2016 & 2019, Bloom et al. 2019, Scur et al. 2021). Disentangling these time-varying margins is a critical advance. Like the digital frontier, many contributors to firm "quality" or intangible organizational capital are not fixed. A number of practices and tools similarly aimed at extracting insights from data have previously been linked to higher Tobin's q and profits (Brynjolfsson et al. 2011, Saunders and Tambe 2015) and greater productivity (Tambe et al. 2012, Aral et al. 2012, Tambe 2014, Brynjolfsson and McElheran 2019, Wu et al. 2020). Now, even-newer computational techniques that leverage data to improve prediction are argued to lower the cognitive costs of decision-making, improve precision, and speed execution (e.g., Agrawal et al. 2019). Yet the theory-evidence gap is only growing. A clearer and updated understanding of how distinct technological, human, and organizational "ingredients" may be combined into productive recipes is urgently needed in both research and practice.

Fulfilling this need requires additional precision with respect to both the conceptualization of the phenomenon and its measurement. As the large literature on routine-biased technical change (RBTC) robustly demonstrates, the automation of cognitive work is distinct from the automation of physical tasks (e.g., Autor et al. 2013). This distinction, while not always crisp in recent conversations, yields disparate predictions for labor market outcomes in different sectors and different occupations. Less-understood, it also likely requires different complementary combinations of technology, people, and practices. Further focusing in on the category of cognitive work, predictive analytics is usefully distinguishable from an "evidence-based" culture (Pfeffer and Sutton 2006) or "data-driven decision-making" management practices (Brynjolfsson and McElheran 2019). Even more granularly, there are widely-recognized differences between descriptive versus predictive analytics (Berman and Israeli 2020, Blum et al. 2015). These distinctions are essential for capturing the right margins of firm investment and process innovation.

Disentangling key complements also requires enhanced precision. While prior research explores how a number of organizational and human factors explain variation in returns to IT (e.g., Black and Lynch 2001, Bresnahan et al. 2002, Melville et al. 2004; Tambe 2012, Tambe et al. 2012, Tambe 2014, Bapna et al. 2013), a significant amount of unexplained variation has been attributed to an undifferentiated cluster of "intangibles" that appear increasingly important, yet remain difficult to measure (Brynjolfsson et al. 2002, Bloom et al. 2012; Saunders and Brynjolfsson 2016). Leaving too much to the "measure of our ignorance" (Abramovitz 1956) creates non-trivial barriers to leveraging research insights into actionable management practices and/or public policies.

This study innovates along a few dimensions to make progress against these concerns. First, our data come from a purpose-designed mandatory survey covering the vast majority of the U.S. manufacturing economy. In addition to being representative, our approach directly captures early use of a rising technological advance, rather than relying on digital artifacts from unrelated economic activity.

Second, causality has been notoriously difficult to pin down, due to self-selection of firms into adoption (Müller et al. 2018; Berman and Israeli 2020). Here, we exploit a novel instrumental variable (IV) approach enabled by another survey question we designed on government-mandated data collection. This new measure captures an arguably exogenous and plant-specific "nudge" to predictive analytics by shocking data availability and monitoring (see Brynjolfsson and McElheran

2019, which similarly exploits this measure).

Finally, linking multiple Census data sets allows us to assemble in one place a large number of tangible and typically intangible workplace investments. We thus make considerable headway in disentangling different dimensions of the phenomenon – some never-before explored – that matter for theory, policy, and practice. These complementarities are not subtle nuances, but rather first-order concerns. The magnitude of each of the four organizational interactions is large: predictive analytics use contributes meaningfully to productivity *only* when combined with at least one complement. Adoption may be widespread, but payoffs are not.

Our findings contribute to several streams of prior work. First, 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). Several empirical studies have found support for 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 & 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 and making progress in establishing causality.

Next, our attention to production process design recalls insights from a neglected stream of management research advocating for tighter alignment between product market strategy and production process design throughout the product lifecycle (Hayes and Wheelright 1979a & b, 1984). Reviving these insights is useful for at least two reasons. First, they identify a set of essentially fixed organizational commitments around which other workplace choices must align. This is helpful for empirically pinning down the right dependencies, rather than some omitted factor that might drive both adoption of the observed factors and outsized productivity gains (Athey and Stern 1998, Hong et al. 2019). They also make tangible a cluster of previously unobserved and accumulated organizational decisions.

Our approach, therefore, is also closely linked to growing research in economics and management that seeks to more concretely measure and estimate contingencies among understudied dimensions of firm strategy, management practices, technology adoption, and human capital (e.g., Arora and Gambardella 1990, Bloom et al. 2012, Hong et al. 2019, Blader et al. 2019, Felten et al. 2019, Choudhury et al. 2020).

Finally, we contribute to a burgeoning literature on the "future of work" for humans in the face of ever-accelerating technological change. High-profile studies have predicted large reductions in labor demand due to rising digitization and automation (e.g., Frey and Osborne 2017, Arntz et al. 2020), though primarily in the context of robot investment in manufacturing (e.g., Acemoglu and Restrepo 2018 & 2019) or, increasingly, "AI" (Agrawal et al. 2018, Brynjolfsson et al. 2018, Felten et al. 2019, Webb 2020). Micro-level evidence to date has been mixed (Bessen et al. 2019, Acemoglu et al. 2020), though managers in the manufacturing sector have been particularly displaced by increased robot investment (Dixon et al. 2019). Recent studies related to ours have looked at software or general-purpose IT (Bessen and Righi 2019, Barth et al. 2020), with limited attention to manufacturing. Also, while economics studies have dominated news headlines, attention to digitization and automation in strategic management is dramatically on the rise (e.g., Adner et al. 2019, Felton et al. 2019 & 2021, Benett 2020a & b, *inter alia*), including attention to how humans shape returns to automation at the micro level (e.g., Choudhury et a. 2020, Glaeser et al. 2021).

2. CONCEPTUAL MOTIVATION AND PRIOR WORK

2.1 Predictive Analytics and Average Performance

While definitions for emerging technology tend to shift (a significant hurdle for measurement – see Section 3), predictive analytics is increasingly understood to be a set of techniques used to analyze historical and current data in order to make predictions about future or unknown events. This may include older techniques from data mining to statistical modeling, as well as, most recently, machine learning and "AI" (Agrawal et al. 2018 & 2019).[3] 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 is often predicted to generate large returns (Davenport 2006, Bughin 2016).

Rising anecdotal evidence suggests, however, that while many firms have benefitted from predictive analytics, others have struggled to realize returns from these 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 investing in mutually-reinforcing assets (both tangible and intangible) will boost firm performance, 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. 2021). 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).

Increasingly, attention is turning to the broader competitive implications of these increasing and persistent differences among firms. Concentration and inequality in workplace conditions and employee earnings are not only receiving increased scrutiny (Song et al. 2018, Autor et al. 2020), but these trends are increasingly attributed to technology investment within industries and firms (Bessen 2017; Bennett 2020b; Lashkari et al. 2020; Barth et al. 2020).

Recent studies have moved beyond general-purpose IT to specifically studying analytics. Tambe (2014) and Wu et al. (2019, 2020) rely on large-scale resume data to home 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

---

[3] Our survey approach will tend to pick up the practices designed to take advantage of prediction, such as supply chain management and demand forecasting, leaning on older techniques and creating "use cases" for newer machine learning and "AI"-related technologies that have only recently begun to move out of the testing phase and into actual application in production in the U.S. (Zolas et al. 2020).

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 conceptually complements these studies in a few ways. To begin, because we use a direct survey measure that distinguishes both the extensive and intensive margin of use, we can make useful "dose-response" predictions: if some use is beneficial, more use along some dimension should yield greater returns up to a point.[4] Variation in intensity is difficult to measure indirectly, as it is less clear that intensity of use scales directly with demand for labor complements.

In addition, our direct measure more-accurately captures specific uses of what can often be quite general-purpose technology and skills. Our approach is inclusive to a range of analytics tools and techniques, giving up precision on specific programming inputs to focus more precisely on automation that facilitates prediction. We view this as the right tradeoff for our research question given the constraints of our survey-based approach (described below). It further disentangles prediction from other uses of data for decision-making, as well as from general managerial capacity, either or both of which might be complementary but are conceptually distinct.

Ultimately, given the widespread awareness of this phenomenon and the voluntary nature of adoption, the most straightforward prediction is that adopters of predictive analytics will enjoy a productivity payoff. Thus, we first hypothesize:

<u>*Hypothesis 1a:*</u> *The use of predictive analytics will be associated with greater workplace productivity, on average, all else equal.*

<u>*Hypothesis 1b*</u>*: More-intensive use of predictive analytics will be associated higher productivity, on average, compared to less-intensive use (all else equal).*

2.1 "Classic" Complements: IT Capital and Worker Skill

However, as touched on above, average returns typically mask critical variation in how technological change plays out in practice – often unobserved by researchers. To begin, prior work has compellingly established the importance of interactions between digital infrastructure and complementary human labor in shaping returns to technology adoption.

[4] 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).

2.1.a 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 all 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). Thus, we expect that firms with a significantly larger and embedded IT infrastructre will outperform other users of predictive analytics:

Hypothesis 2: *More accumulated IT capital stock will be complementary to predictive analytics, showing correlated adoption and mutually-reinforcing productivity gains.*

2.1.b Skilled Labor

Prior work has established that more-skilled and better-educated workers are key drivers of growth in both manufacturing productivity (e.g.,) and returns to IT (Brynjolfsson and Hitt 2000, Black and Lynch 2001, Bresnahan et al. 2002). With increasing digitization and growing prevalence of business applications that rely on data, firms increasingly need workers that know how to deploy "smart" technologies in production settings (Helper et al. 2019). They often also need workers who can translate analytical output into meaningful 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 increasing returns to worker skill will boost predictive analytics' observed contribution to productivity due to the underlying complementarities.

It is an open question whether these skills need to be widely-dispersed throughout the firm or only concentrated within a subset of employees who more-intensively interact with the technology. We explore a few approaches in the empirical section. As a first cut, we predict that firms with a higher share of educated workers will be more likely to have the requisite skills in the requisite roles to leverage more-automated prediction. Thus:

<u>Hypothesis 3:</u> *Skilled labor will be complementary to predictive analytics use, showing correlated adoption and mutually-reinforcing productivity gains.*

Before moving on to other hypotheses, it is worth noting that complementarity between IT capital and/or worker skill is not conceptually novel.[5] Our intention here is to apply prior theory in a novel setting to verify whether these intuitions hold for this latest technological advance. While we articulate our formal hypotheses in line with prior predictions, it is not obvious, *a priori*, that they should play out for this latest round of technological change. In fact, precisely because these insights have been developed and confirmed for prior generations of IT diffusion, complementarity theory predicts that firms will increasingly learn about the importance of these complements. Moreover, they should increasingly have pursued mutually-reinforcing bundles of practices and investments – possibly to the extent that they are no longer a source of differentiation (Brynjolfsson and Milgrom 2013). It thus remains an open empirical question requiring investigation – and frequent re-investigation – with data, preferably over reasonably long time periods to reveal the direction and extent of organizational adjustment. More on this, below.

2.2 Process Design Complementarities

Adjustment becomes a central concern for the next complement we investigate. Predictive analytics relies on historical and current data to predict future or currently unobserved outcomes, moving beyond descriptions of currently-available data (e.g., Blum et al. 2015, Berman and Israeli 2019). In general, workplaces with greater automation will provide richer data inputs, for instance due to instrumentation embedded in production machinery. This will lower the cost or improve the quality of inputs for a given prediction task, all else equal.

Automation is also likely to enhance returns to prediction by aligning past "states of the world" with current and future outcomes. Because automation tends to reduce variance in the production process (an oft-cited benefit of robot use in manufacturing, see Dixon et al. 2019), predictions are likely to be more accurate. This is particularly true for older, less-flexible automation, provided

[5] Though, most prior studies investigate the complementary between general-purpose IT and skilled labor, rather than how each interacts with a new technical application, which is what we emphasize, here.

the production tasks remain consistent.

This need for consistency, in turn, has led manufacturing environments with a high degree of physical automation to pursue other design choices that reduce variance. A distinction commonly understood in operations management, "continuous flow" production processes are typically characterized not only by high levels of physical automation but also by low product mix, along with high volume per product (e.g., Safizadeh *et al.* 1996). These processes tend to be more capital intensive, which requires more frequent monitoring to maintain machine uptime and ensure adequate capacity utilization (McElheran et al. 2020). Notably, this not only boosts the amount of objective data available for analysis, but it also raises the stakes with respect to anticipating and preventing equipment failure and downtime, forestalling supply chain hiccups, and maintaining a steady stream of outbound finished materials. The costs of being reactive versus proactive are much higher in these production environments.

It is conceptually useful to contrast these continuous flow processes with those operating in so-called "job shops," batch-manufacturing plants, or R&D-focused facilities. The latter are more likely to have a "jumbled flow" process, support flexible and high-mix but generally low-volume (per product) operations, and have shorter setup times between products. Novelty and experimentation in both product and process design are frequently emphasized in these settings. These lower flow-efficiency processes are central to product innovation, prototyping, and iterating on production design (Hayes and Wheelwright 1979a &b). In such settings, variance is actually "more a feature than a bug." Failures to optimize capacity utilization can still be costly, but prediction is unlikely to be as effective in avoiding these costs.

These considerations have largely been overlooked outside of operations management, excepting work by Hayes and Wheelright (1979a & b, and Hayes et al. 1984) advocating for alignment between production process design and product market strategy over the product lifecycle. Recent work leverages the quasi-fixed nature of these workplace characteristics to identify important contingencies shaping the adoption and effectiveness of many "structured management" practices (McElheran et al. 2020). Because these features of the production process are embedded in expensive and slow-moving capital investments, physical infrastructure, and process steps with multiple interdependencies, we expect that they will be slow to adjust to faster-moving technological advances. This makes them particularly useful for identifying complementarities with predictive analytics use, as one side of the interaction is quasi-fixed over relatively long time horizons (Hong et al. 2019). Our hypothesis here has more directionality (we would not expect the use of predictive analytics to have any impact on the choice of process design):

Hypothesis 4: *Production process designs focused on high flow efficiency will promote a greater likelihood of predictive analytics adoption and higher returns from predictive analytics use.*

2.3 Humans in the Loop – Managerial Capacity and Managers

We explore a few approaches to understanding the role of humans in this increasingly automated activity. We first leverage recent findings that certain managerial practices improve the capacity of firms to take advantage of more objective information (e.g., Bilicka and Scur 2019) and that objective measures can further reinforce the development of intangible managerial capacity to communicate and coordinate informally (e.g., Gibbons and Kaplan 2015), provided internal alignment is maintained (Blader et al. 2019).

Hypothesis 5: *Greater managerial capacity to interpret and respond to objective information will be complementary to predictive analytics, showing correlated adoption and mutually-reinforcing productivity gains.*

A related and open question, however, is whether increased automation of prediction will increase the productivity of individual managers. This could happen to the extent that overall demand for human labor in these tasks declines. On the other hand, rising productivity could boost labor demand across the board. Given conflicting theoretical predictions (e.g., Acemoglu and Restrepo 2019), we view the relationship between managerial headcount and predictive analytics as a fundamentally empirical question. The net answer to this question is the one of ultimate importance for making predictions about the "future of work."

3. DATA

3.1 Managerial and Organizational Practices Survey (MOPS)

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 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 here>>

Note that this approach requires first and foremost that respondents understand what is meant by "predictive analytics." While we give some examples, it also became apparent in cognitive testing of the survey instrument that this term was both well understood by 2015, but still a bit novel and "buzzword-y." A consistent challenge for measuring emerging technologies, in fact, is pinning down terminology. It is important to be specific enough to capture the phenomenon of

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

[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).

interest while being general enough to require little cognitive processing of respondents to make sense of it; measuring adoption early is essential, yet trying too early can create tremendous hurdles to measurement (McElheran 2018, Felton et al 2021).

We first explore 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. This also captures more of the variation among plants in their use of predictive analytics.

3.2 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.

3.3 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

---

[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.
[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 do not show striking patterns but are available upon request.

(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 here>>

<<Figure 2 here>>

<<Table 2 here>>

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 estimation. 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.

3.4 Measuring Workplace Complements

IT Capital Stock

We estimate IT capital stocks using data on 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

[11] The rotation of the ASM sample frame 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.

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).[13] A key advantage of this measure is that it accounts for the overall stock of IT. If firms require time to adjust and utilize novel IT investment, we will be able to capture the lagged effect.[14]

Skilled Labor

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 (e.g., Bresnahan et al. 2002).

Production Process Design

To capture this historically intangible dimension of the workplace setting, we leverage a purpose-built 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.[15] The relationships between this measure and the process design characteristics described above have been validated for the plants in this 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 virtually time-invariant in our sample. 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

---

[13] Results are quite insensitive to the choice of depreciation rate.

[14] 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.

[15] See Kiran (2019) for a detailed description of cellular manufacturing.

empirically identifying complementarities, as it is not subject to time-varying adjustment with respect to other, potentially unobserved factors (see discussions in Athey and Stern 1998, Cassiman and Veuglers 2006, and Hong et al. 2019). Moreover, it varies among establishments belonging to the same parent firm, highlighting the importance of plant-level data for our analysis.

Humans in the Loop

We explore a few approaches to measuring how human inputs may shape – and be shaped by - this particular type of automation. First, we consider two different measures of managerial capacity. The first relies on only one question from the first section of the MOPS: the number of key performance indicators (KPIs) tracked by the firm. This is a useful objective measure of how intensively the plant has engaged in identifying which aspects of its production process are critical to performance and created routines for tracking – and presumably responding to – them. One concern with this measure is that it is very narrow, ignoring information we have on a broader set of practices associated with managerial capacity to leverage objective information (e.g., Bilicka and Scur 2020). It is also subject to a greater degree of measurement error (Bloom et al. 2019). Thus, we also construct an index of all practices in the first section of the MOPS and confirm that they are distinct from the other "structured management" controls we separately include in most of our specifications.[16]

To explore complementarity with the managers themselves, we rely on measures of both the number of managers employed at the plant (question 32 on the MOPS) and the share of total plant employment represented by managers. We observe this data for both 2010 and 2015.

3.5 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 directly captures the majority of economic activity in the sector, very small establishments are under-represented: sample mean annual sales and employment in log terms are 10.37 and 4.56, respectively, or about

[16] This latter measure is our preferred approach, but these results are still pending disclosure avoidance review by Census. Thus, we present our findings using the KPIs-based measure. Our core findings are robust to either approach.

$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. In the pooled sample, 35 percent of plants are designed for high flow efficiency, which changes very little over the five years in the balanced panel.

4. EMPIRICAL STRATEGY

4.1 Production Function Estimation

Our empirical exploration proceeds in steps. First, we estimate the average effect of predictive analytics on performance. For this, we take a conventional approach to modeling the plant production function (Brynjolfsson and Hitt 2003; Bloom et al. 2012), estimating the log-transformed Cobb-Douglas production function in equation (1):

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

Y_{ijt} is sales by establishment *i* in industry *j* at time *t*. Indexed similarly by establishment *i* and time *T*, *K* denotes non-IT capital stocks at the beginning of the period, PA is an indicator (or frequency measure) for use of predictive analytics, L is labor input, M is expenditure on materials 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.

4.2 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

[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

instrumental variables (IV) estimation and timing tests.

IV Strategy

For our IV estimation, we exploit an indicator that data collection at the plant is "nudged" by government regulations 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 of such a mechanism are not merely theoretical. The case of Alcoa Corporation in the late 1980s and 90s is illustrative. When Paul O'Neil took leadership of the firm, his unexpected

---

other semi-structural estimation methods (see Arellano and Bond 1991; Blundell and Bond 2000; Levinsohn and Petrin 2003; Ackerberg 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 OHSA 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.

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 adoption vis-à-vis the timing of productivity changes. Leveraging annual data on inputs and output

[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).

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, "late" 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 (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 late adopters and non-adopters. In the 2010-2016 window, late 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 "late" adopters may have adopted at any time in the 2010-2015 window; thus we anticipate estimates will be considerably noisier in this analysis.

4.3 Formal 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 buttressed 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):

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

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 within a population of competing firms will lead to more correlated adoption but potentially lower excess productivity gains. Passing both tests may be taken as strong evidence of complementarity (Brynjolfsson and Milgrom 2013).

5. RESULTS

5.1 Conditional Correlation between Predictive Analytics and Workplace 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. 2008). All columns further include industry-

[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.

year controls at the narrow 6-digit NAICS level. For example, this distinction captures the difference between Folding Paperboard Box Manufacturing (NAICS 322212) and Setup Paperboard Box Manufacturing (322213).

<<Table 3 here>>

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 potential complements explored below.[22] The coefficient drops significantly to around 1.45 percent, consistent with an important role for organizational enablers that would otherwise load onto the returns to predictive analytics when not separately accounted for. Note that this specification controls for other 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, these controls also somewhat pull down the coefficient on predictive analytics use (not shown separately due to space limitations), pointing to the risk of omitted variable bias or potential additional complementarities beyond the scope of this study. 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. This is consistent with Hypothesis 1a.

Column 3 explores the intensive margin of predictive analytics use from Hypothesis 2. 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.[23] Extending this out linearly (*pending disclosure*

---

[22] For this draft of Table 3, managerial capacity in the operations function is rolled into the "High Structured Management" measure, which also includes target-setting and Human Resources management. A model with these entering in separately is essentially unchanged and *pending disclosure avoidance review*.

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

avoidance review) would yield four times this benefit for plants at the highest frequency of use.

We explore the robustness of these patterns to alternative measures for both the dependent and independent variables (see Table A1). Our findings are robust to using value-added as the output measure and alternative measures for the use frequency of predictive analytics.[24]

<<Table 4 here>>

5.2 Exploring Causality

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 variable estimation using government-mandated data collection as an instrument for the predictive analytics index.[25]

The first stage of our two-stage least squares (2SLS) estimation shows that government-mandated data collection is highly correlated with the use of predictive analytics (see also Figure 5, below). 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 in productivity estimation 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 shifts. 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

---

[24] 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, e.g., Bartelsman and Gray 1996) as alternative output measures. It is also robust to estimating a translog production function (see Table A2 in the appendix).

[25] 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.

unobserved in other studies, but strongly associated with higher productivity in this sector (Bloom *et al.* 2019; Brynjolfsson and McElheran 2019). Addressing these "intangible" 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. Figure 3 plots the coefficients for "early" and "late" indicators of predictive analytics use (extensive margin) in the performance model between 2010 and 2016. Consistent with a causal interpretation of Hypothesis 1, 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 later adopters is not significantly different from that of non-adopters through 2013. We know that later-adopting plants did so in the 2010-2015 window, but not precisely when. In line with steadily increasing diffusion of beneficial practices or technology over time (e.g., Hall 2004), performance in this group begins to rise and is significantly different from non-adopters by 2014. Notably, these later adopters close the performance gap, becoming statistically indistinguishable from early adopters by 2016. The timing of these effects are inconsistent with reverse causality, further supporting a causal interpretation.

<<Figure 3 here>>

5.3 Complementarity Tests

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.

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 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: the likelihood of 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 a significantly lower (both statistically and empirically) likelihood of predictive analytics use than the other process design types. The other flexibility-favoring process designs have a somewhat higher (and, in the case of R&D facilities, noisier) correlation with predictive analytics. The strongest correlation is with the efficiency-oriented designs (robustness checks of our results support combining them to ease interpretation in the estimations that follow). Panel d) shows that greater managerial capacity is associated with a higher probability of using predictive analytics.

<<Figure 4 here>>

In a more saturated model, IT infrastructure, skilled workers, high flow efficiency, and managerial capacity are all significantly correlated with the use of predictive analytics at the one-percent level or higher.[26] Figure 5 is organized to show the estimated increase in probability associated with each potential complement, building from a benchmark of 20.6 percent.[27] 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 by 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 or investments. High managerial capacity has a large effect, adding another 4.6 percentage point. A workplace with all four workplace characteristics is 10 percentage points – or 30.6 percent more likely to use predictive analytics at some frequency. This is consistent with complementarity of a sufficiently advanced type, whereby managers are aware of mutually-reinforcing benefits and will seek to adopt complementary practices together.

<<Figure 5 here>>

---

[26] Regression estimates are unwieldy and omitted to conserve space but are available upon request.
[27] Based on the constant term in the unreported full linear probability regression; available upon request.

Performance Tests

We next explore the extent to which organizational features help explain the heterogeneous returns from predictive analytics, as complementarity theory would predict. We find clear evidence that predictive analytics contributes far more to performance when combined with certain 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 then 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, flow-efficient manufacturing processes, or high managerial capacity the effects of predictive analytics are around 4.1 percent, 2.9 percent, 3.1 percent, and 2.2 percent respectively (see regression results in Table 4). 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 characteristics.

<<Figure 6 here>>

<<Table 4 here>>

While our findings are visually more-intuitive using indicators for high levels of each complement, they largely hold for continuous measures, as well. As shown in Table A4, interactions between predictive analytics use (extensive margin) and continuous measures of the four complement are economically and statistically significant, excepting the IT capital measure. Consistent with the patterns in Figure 4, complementarities here are only found at quite high levels of IT capital accumulation. The interaction at the mean is not statistically different from zero (see column 1). In column 2of Table A4, a higher percentage of workers with bachelor's degrees shows a positive interaction at the mean with predictive analytics use that is significant at the 5% level. In column 3, increasing managerial capacity, measured by higher numbers of key performance indicators (KPIs) also reinforces returns to predictive analytics, even at lower levels of intensity.[28]

A natural question at this point is whether any of these complements are more important than others. The overlapping confidence intervals in Figure 4 suggest that they are not.

Not presented in the table are the results of three-way (or more) interactions between these distinct measures. This is because we do not find any statistically significant three-way interactions. One possibility is that they are actually different measures of a single underlying system of workplace practices. We investigate this through a principal component analysis and interact this principal component with predictive analytics use in the performance test.[29] The results remain consistent, suggesting that we are, indeed, picking up this integrated system of practices, the linear combination of which drives much of the variation we observe.

Overall, these findings 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.

5.4 Humans and the Future of Work

In a set of results that are still pending disclosure avoidance review by the Census Bureau, we explore interactions between predictive analytics and the jobs occupied by managers. Just because managerial capacity is mutually reinforcing with predictive analytics, it does not necessariy follow that the number of humans doing those jobs is also complementary. Specifc coefficients will be available in soon-forthcoming revision. In the meantime, qualitatively, we find that predictive anlaytics use is positive and significantly correlated with both the number of managers and the managerial share of total employment at the plant. However, when we look at change over time, we find that the *share* of mangerial labor does not change significantly in response to predictive anlaytics use. Instead, the total number of managers increases, consistent with a labor-augmentation outcome, on net. Even if individual managers are made more productive by automation (creating opportunities to reduce the total number), rising productivity leads to a net gain in overal managerial employment. This is in sharp contrast to the findings for managerial

[28] Results for ordered categories of flow efficiency and for our broader measure of managerial capacity are pending disclosure avoidance review, but the qualitative patterns are quite consistent.
[29] We extract a key principle component from the four complements and interact it with predictive analytics (without the other measures) and find a significant interaction between the two. *Results pending disclosure avoidance review*.

labor due to physical automation (Dixon et al. 2019), and suggests that fears of the "Robopacalypse" provide an incomplete view into the future of work.

6. Conclusion

There has been explosive growth in 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 explore the productivity effects of predictive analytics by analyzing its use in over 30,000 plants while controlling for 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 is 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 four 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, flow-efficient production processes, or managerial capacity for responding to data-driven insights - 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. In fact, the higher performance effects of predictive analytics depend crucially on having in place at least one element of what appears to be an integrated latent system of workplace practices.

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 to survival, which if positive, would argue for our estimates being 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 approaches to 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. Predictive analyltics use is widespread, but performance gains are not for reasons that are both well-understood and persistent and due to novel contingencies we present here. 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.

REFERENCES

Acemoglu, Daron, Claire Lelarge and Pascual Restrepo. 2020. Competing with Robots: Firm-Level Evidence from France. *American Economic Review Papers and Proceedings* 110: 383-388.

Acemoglu, Daron and Pascual Restrepo. 2019. Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives* 33(2): 3-30.

Acemoglu, Daron and Pascual Restrepo. 2018. The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment. *American Economic Review* 108(6): 1488-1542.

Ackerberg, 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.

Arntz, Melanie, Terry Gregory and UT Zierahn. 2020. Digitization and the Future of Work: Macroeconomic Consequences. *Handbook of Labor, Human Resources and Population Economics*.

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. 2015. Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of economic perspectives* 29(3): 3-30.

Autor, David H. and David Dorn. 2013. The Growth of Low-Skill Service Jobs and the Polarization of the U.S. Labor Market. *American Economic Review* 103(5): 1553-1597.

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.

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 E, Maarten Goos, Anna Salomons and Wiljan Van den Berge. 2019. Automatic Reaction-What Happens to Workers at Firms That Automate? *Boston Univ. School of Law, Law and Economics Research Paper*.

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.

Blader, S., Gartenberg, C. and Prat, A., 2020. The contingent effect of management practices. *The Review of Economic Studies*, *87*(2), pp.721-749.

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, 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, Lorin M. Hitt and Shinkyu Yang. 2002. Intangible Assets: Computers and Organizational Capital. *Brookings Papers on Economic Activity*(1): 137-198.

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.

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

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.

Dixon, Jay, Bryan Hong and Lynn Wu. 2021. The Robot Revolution: Managerial and Employment Consequences for Firms. *Management Science*.

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.

Felten, Edward, Manav Raj and Robert Seamans. 2021. Occupational, Industry, and Geographic Exposure to Artificial Intelligence: A Novel Dataset and Its Potential Uses. *Strategic Management Journal*.

Felten, EW, M Raj and RC Seamans. 2019. The Variable Impact of Artificial Intelligence on Labor: The Role of Complementary Skills and Technologies. *Available at SSRN*.

Foster, Lucia, John Haltiwanger and Chad Syverson. 2008. Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability? *American Economic Review* 98(1): 394-425.

Foster, Lucia, John Haltiwanger and Chad Syverson. 2016. The Slow Growth of New Plants: Learning About Demand? *Economica,* 83(329): 91-129.

Frey, Carl Benedikt and Michael A Osborne. 2017. The Future of Employment: How Susceptible Are Jobs to Computerisation? *Technological forecasting and social change* 114: 254-280.

Furman, Jason and Robert Seamans. 2019. AI and the Economy. *Innovation Policy and the Economy* 19(1): 161-191.

Glaeser, Edward L, Andrew Hillis, Hyunjin Kim, Scott Duke Kominers and Michael Luca. 2021. Decision Authority and the Returns to Algorithms. INSEAD working paper.

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.

Gibbons, Robert and Robert S Kaplan. 2015. Formal Measures in Informal Management: Can a Balanced Scorecard Change a Culture? *American Economic Review* 105(5): 447-451.

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. 1979a. The Dynamics of Product-Process Life Cycles. *Harvard Business Review* 57(2): 127-136.

Hayes, Robert H. and Steven C. Wheelwright. 1979b. Link Manufacturing Process and Product Life Cycles. *Harvard Business Review* 57(1): 133-140.

Hayes, R.H. and Steven C Wheelwright. 1984. *Restoring Our Competitive Edge - Competing through Manufacturing*. New York, Wiley.

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.

IDC. 2019. *Worldwide Big Data and Analytics Spending Guide*. Retrieved from https://www.idc.com/tracker/showproductinfo.jsp?containerId=IDC_P33195

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.

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.

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.

Raj, Manav and Robert Seamans. 2018. Artificial Intelligence, Labor, Productivity, and the Need for Firm-Level Data. *The Economics of Artificial Intelligence: An Agenda*ed), University of Chicago Press**:** 553-565.

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

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.

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.

Syverson, Chad. 2011. What Determines Productivity? *Journal of Economic Literature,* 49(2): 326-365.

Syverson, Chad. 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.

Tambe, Prasanna, Lorin M. Hitt and Erik Brynjolfsson. 2012. The Extroverted Firm: How External Information Practices Affect Innovation and Productivity. *Management Science* 58(5): 843-859.

Van Reenen, John. 2018. *Increasing Differences Between Firms: Market Power and the Macro Economy*. CEP Discussion Paper No. 1576: London School of Economics.

Webb, Michael. 2019. The Impact of Artificial Intelligence on the Labor Market. *Available at SSRN 3482150*.

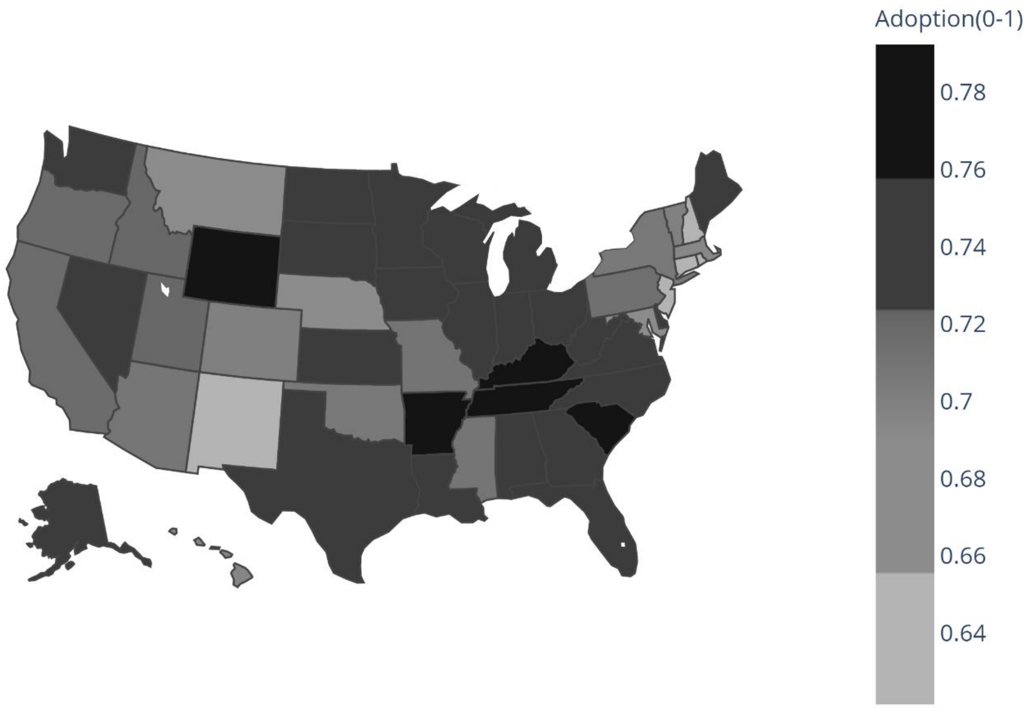
White, T. Kirk, Jerome P. Reiter and Amil Petrin. 2018. Imputation in U.S. Manufacturing Data and its Implications for Productivity Dispersion. *The Review of Economics and Statistics,* 100(3): 502-509.

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.

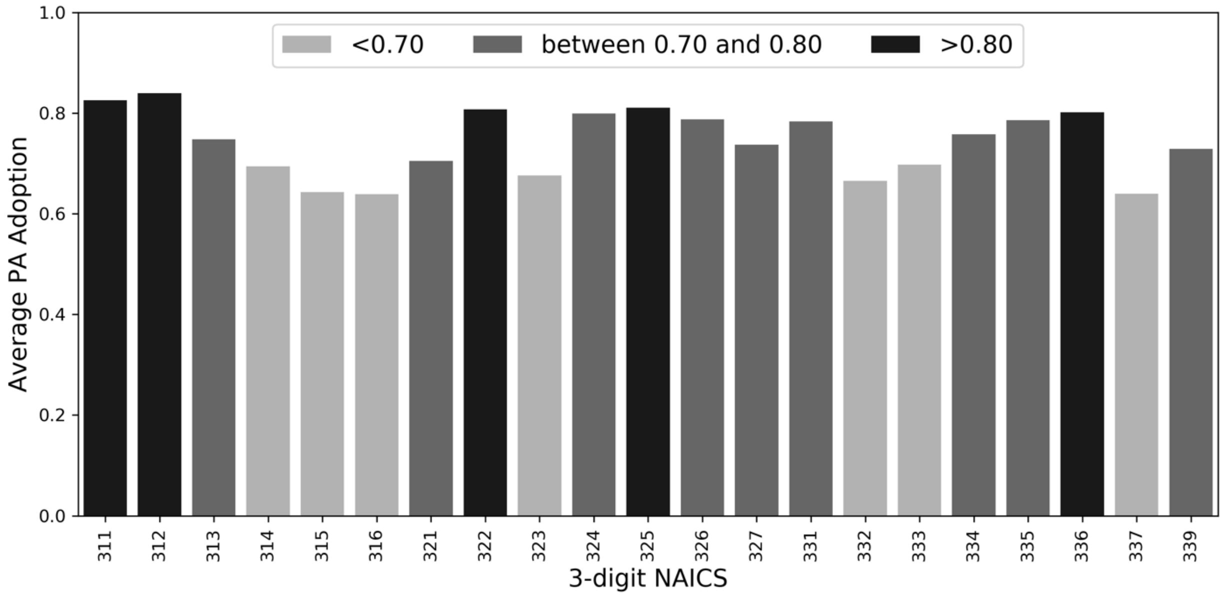
Zolas, Nikolas, Zachary Kroff, Erik Brynjolfsson, Kristina McElheran, David Beede, Catherine Buffington, Nathan Goldschlag, Lucia Foster and Emin Dinlersoz. 2020. *Advanced Technologies Adoption and Use by US Firms: Evidence From the Annual Business Survey*. NBER Working Paper No. 28290: National Bureau 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 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.

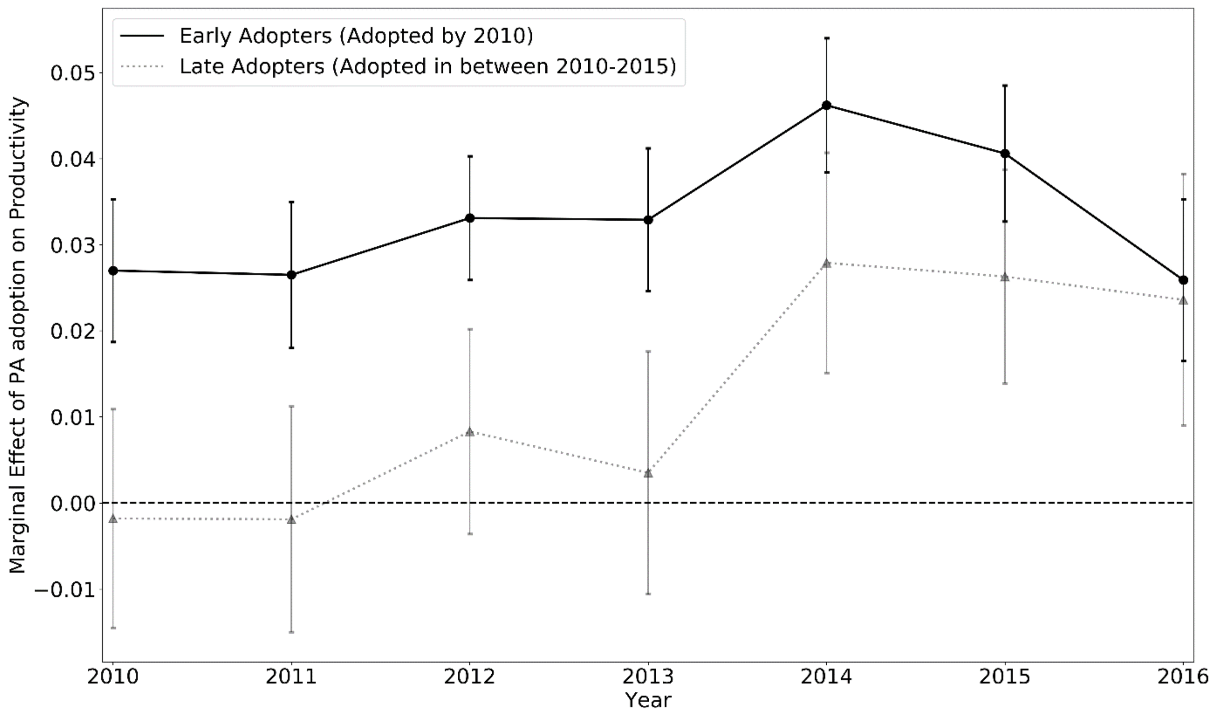
Figure 2. Adoption of Predictive Analytics by Industry
(US Manufacturing by 2010)



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.

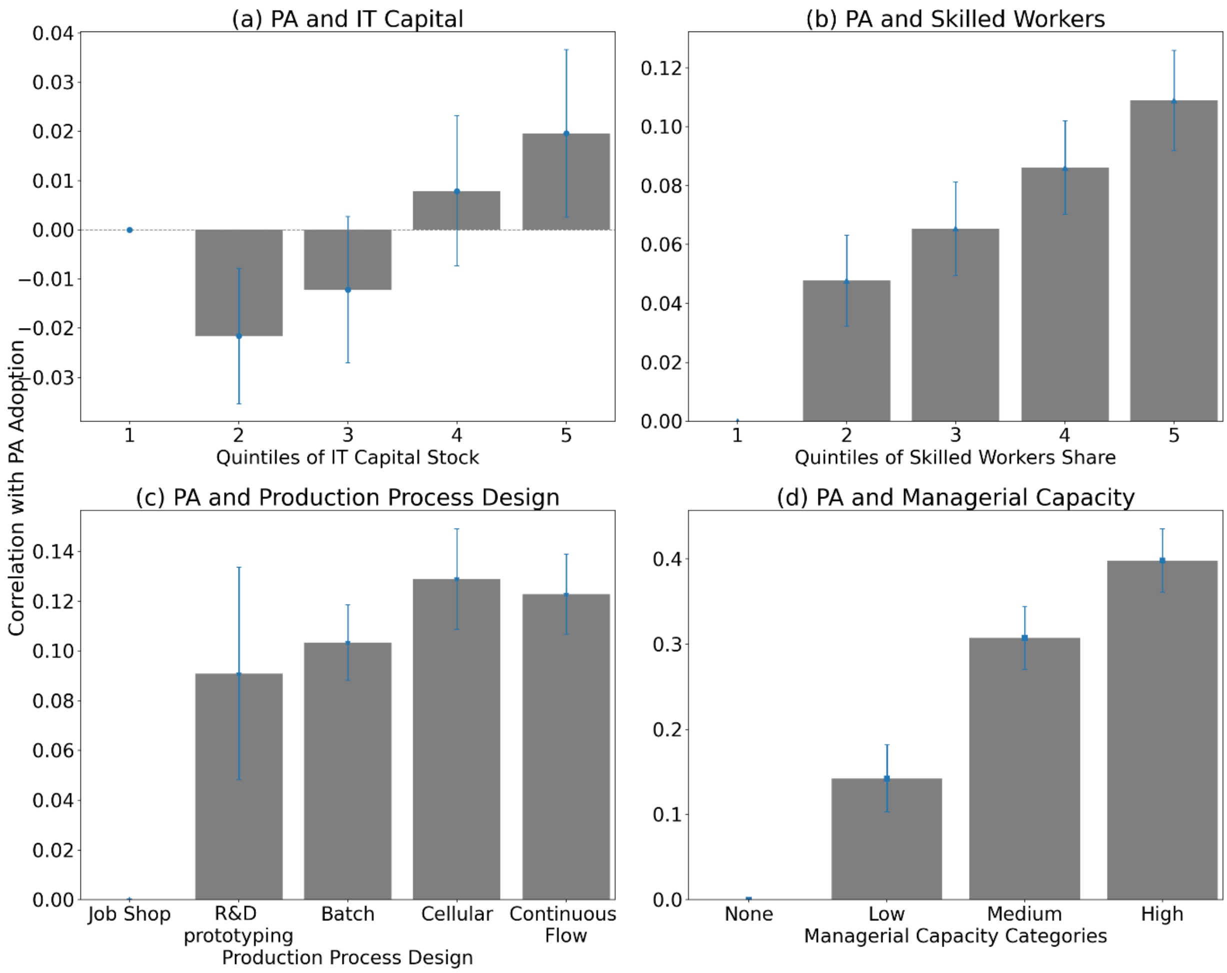
NAICS 3	Industry Definition	NAICS 3	Industry Definition
311	Food Manufacturing	326	Plastics and Rubber Products
312	Beverage and Tobacco Product	327	Nonmetallic Mineral Product
313	Textile Mills	331	Primary Metal
314	Textile Product Mills	332	Fabricated Metal Product
315	Apparel Manufacturing	333	Machinery
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		

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



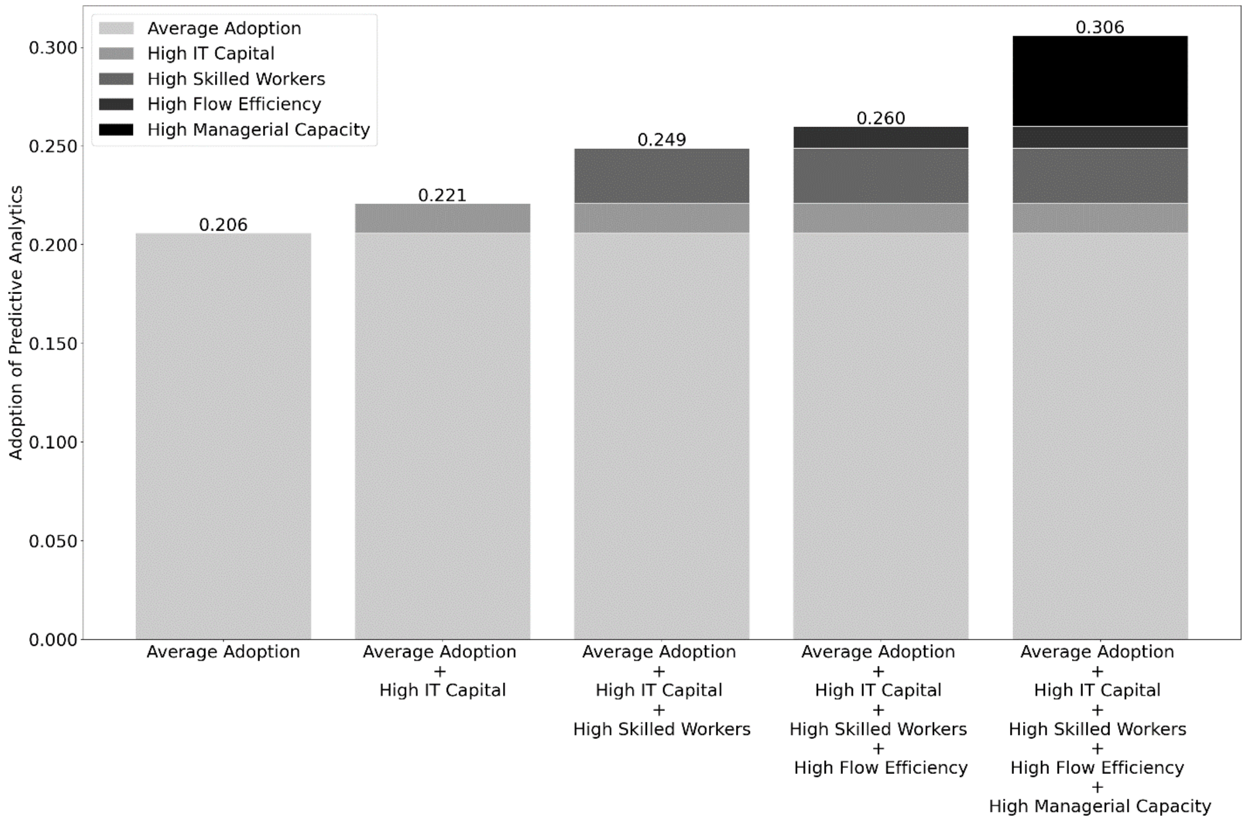
Notes: Estimates based on a pooled OLS model with a specification identical to the baseline model in column 2 Table 3. For this test, we construct an ASM and CMF panel, where we have annual data on most of the key inputs (excepting the managerial-related variables from the 2015 MOPS) and sales from 2010 to 2016. We group establishments that adopted predictive analytics by 2010 ("Early Adopters"), establishments that adopted predictive analytics between 2010 and 2015 ("Late Adopters"), and the remaining non-adopters by 2015 (the omitted category) using the 2015 MOPS data. These indicators are then interacted with year dummies to explore the differences in sales over time (using non-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



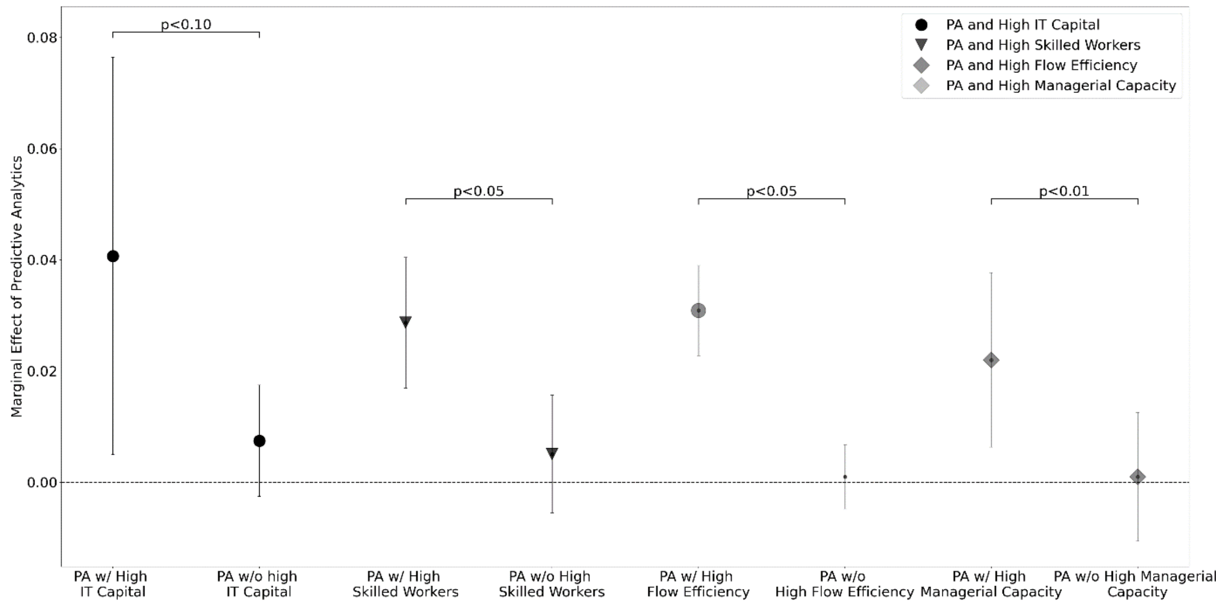
Notes: Estimates based on the baseline sample from the pooled OLS regressions controlling for plant size, plant age, and industry (6-digit NAICS) and year fixed-effects. The dependent variable is an indicator of predictive analytics use (any frequency). Histogram bars (and values on the Y-axis) represent differences in the likelihood of adopting predictive analytics compared to the first, "base" group indicated on the X-axis. For panel (d), the base group currently consists of plants that report tracking no key performance indicators (KPIs) on question 2 of the MOPS. "Low", "Medium", and "High" categories are comprised of plants that track 1-2, 3-9, and 10 or more KPIs, respectively. Categories based on a broader definition of "managerial capacity" yield a consistent pattern (*pending disclosure review).* Quintiles are used in accordance with US Census disclosure avoidance practices and for consistency across figures. Standard errors for coefficients are plotted on the histogram bars with darker lines. Detailed regression coefficients are available upon request.

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 specification that includes High IT Capital Stock, High Skilled Workers, High Flow Efficiency, and High Managerial Capacity. Average Adoption takes the constant term as the benchmark adoption rate. High IT K is an indicator for plants in the top tenth percentile of the IT capital stock distribution. High Skilled Workers is an indicator for plants in the top quartile by percentage of employees with a bachelors' degree. High Flow Efficiency is an indicator for plants whose production process is best characterized as either continuous flow or cellular manufacturing. High Managerial Capacity is currently an indicator for plants that intensively track key performance indicators at the plant (10 or more), though results are consistent for a broader measure of managerial capacity (*pending disclosure review*). All four coefficients are significant at the 1% level or higher. Additional controls include industry-year fixed effects (6-digit NAICS level), as well as plant-level employment in log terms; logged non-IT capital stock; an index of other structured management practices; having top data-driven decision-making practices (DDD); age; an indicator that government regulations or agencies influenced 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 includes High IT Capital Stock, High Skilled Workers, High Flow Efficiency, and High Managerial Capacity, respectively. Also note that results are consistent for a broader measure of managerial capacity (*pending disclosure review*). The coefficients for an indicator of predictive analytics use and the interaction term identify the differential effects of adopting predictive analytics on sales, conditional on the presence of each 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 (not shown). Error bars indicate 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	Pooled Sample	2010 (Recall)	2015 (Report)
			Mean (S.D.)	
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)
Managerial Capacity (Operations)	Indicator for monitoring 10 or more key performance indicators (KPIs) (question 2 on the MOPS 2015).	0.44 (0.50)	0.37 (0.48)	0.56 (0.50)
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	Age of establishment. Truncated for plants born before 1976.	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	Varies by specification. For most analyses, this is an index or sub-indices of the structured management practices in section A of the MOPS, excluding questions 2 and 6 (used elsewhere).	0.63 (0.17)	0.60 (0.16)	0.68 (0.15)
DDD	Indicator for plants with high Data-Driven Decision-making following Brynjolfsson and McElheran (2019) based on high levels of KPI monitoring (Q2), use of short and long-term targets (Q6), and questions 24 and 25 on the availability and use of data in decision making.	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. Average Effects 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			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	Y	Y	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.

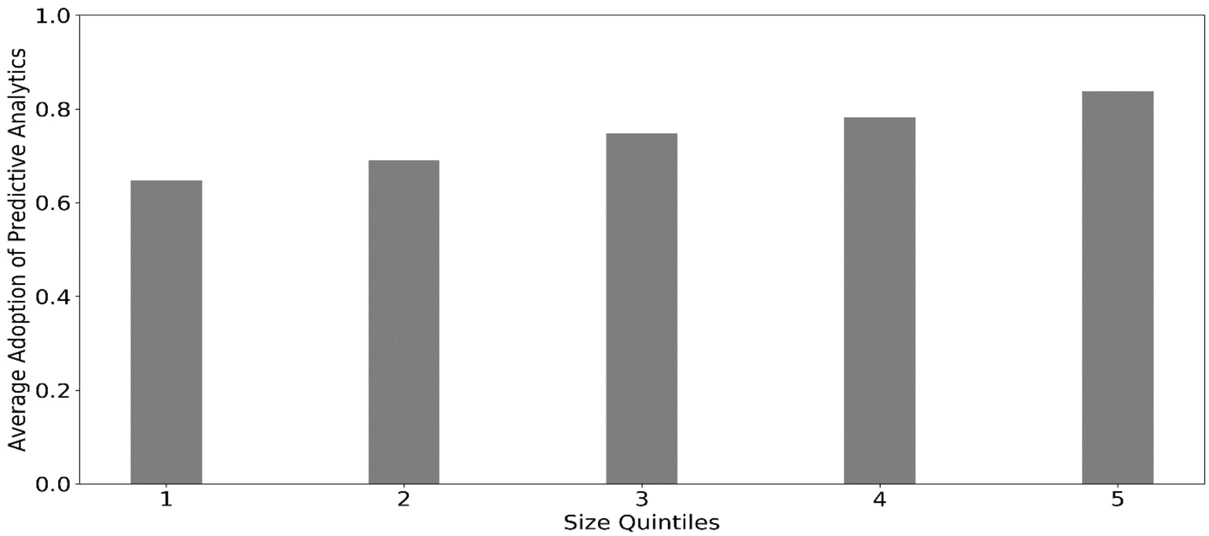
Table 4. Organizational Complements to Predictive Analytics (Performance Test)

Model	(1) IT Capital Stock	(2) Skilled Workers	(3) Flow-Efficient Production	(4) Managerial Capacity
Dependent Variable		Log Sales		
PA Use	0.0075 (0.0051)	0.0051 (0.0054)	0.0010 (0.0058)	0.0010 (0.0059)
High IT K	0.1103*** (0.0171)			
PA × High IT K	0.0333* (0.0185)			
High Skilled Workers		0.0371*** (0.0101)		
PA × High Skilled Workers		0.0228** (0.0114)		
High Flow Efficiency			0.0158* (0.0088)	
PA × High Flow Efficiency			0.0306*** (0.0094)	
High Managerial Capacity				0.0021 (0.0087)
PA × High Managerial Capacity				0.0210** (0.0097)
Joint Tests	0.0407** (0.0182)	0.0287*** (0.0106)	0.0309*** (0.0081)	0.0220*** (0.0080)
Other Controls	Y	Y	Y	Y
Industry x Year Controls	Y	Y	Y	Y
R-Squared	0.9326	0.9327	0.9328	0.9339
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 in the top tenth percentile of IT capital stock. High Skill is an indicator for plants in the top quartile for the percentage of employees with a bachelors' degree. High Flow Efficiency is an indicator for plants whose production process is best characterized as either continuous flow or cellular manufacturing. High Managerial Capacity is currently an indicator for plants that intensively track key performance indicators at the plant (10 or more), though results are consistent for a broader measure of managerial capacity (*pending disclosure review*). Columns 1-4 interact the indicator of predictive analytics use with each of the potential complements while controlling for other production inputs and other potential complements. Joint tests calculate the combined effect of predictive analytics use, the complement of interest, and their interacted effect (using Lincom in Stata 16). Unreported controls for all columns include logged total number of employees, logged non-IT capital stock, logged cost of material and energy, plant age, high structured management (defined as being in the top quartile of the structured management index), indicators for data-driven-decision making practices, and indicators for belonging to a multi-unit firm or being co-located with headquarters. Robust standard errors clustered at the firm level. * p<0.10, ** p<0.05, *** p<0.01.

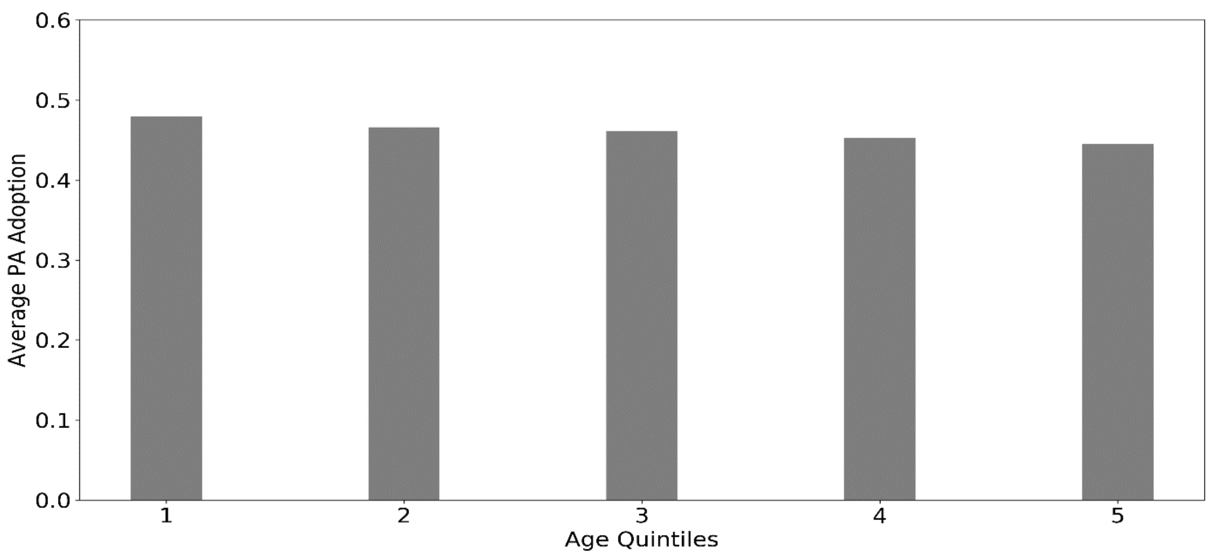
Appendix

Figure A1a Correlations between Predictive Analytics and Plant Size



Notes: Estimates based on the baseline sample from a pooled OLS regression controlling industry (6-digit NAICS) and year fixed-effects. The dependent variables are the adoption of predictive analytics. The values on the Y-axis are the calculated coefficients of each size quintile (based on logged total employment). The numbers of quintiles are presented on the horizontal axis. Quintiles are used to conform with US Census disclosure avoidance guidelines and consistency across figures.

Figure A1b Correlations between Predictive Analytics and Plant Age



Notes: Estimates based on the baseline sample from a pooled OLS regression controlling plant size, industry (6-digit NAICS) and year fixed-effects. The dependent variables are the adoption of overall predictive analytics. The values on Y-axis are the coefficients of each age quintile. Our sample covers plants from one-year-old to over 40 years old. The numbers of age groups are presented on the horizontal axis. Age quintiles are used for the US Census disclosure avoidance practice and consistency across figures.

Table A1. Average Effects of Predictive Analytics on Plant Performance (Robustness)

Models	(1) OLS	(2) OLS PA Frequency (Average)	(3) IV PA Frequency (Average)
Dependent Variables	**Log Value Added**	**Log Sales**	
PA Use	0.0223** (0.0091)		
Average PA Use Frequency		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
Other controls	Y	Y	Y
Industry x Year Fixed Effects	Y	Y	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.

Table A2. Average Effects of Predictive Analytics on Plant Performance (Additional Robustness)

Models	**(1)** **OLS**	**(2)** **OLS**	**(3)** **Translog**
Dependent Variables	**Log Sales Per Employee**	**Log TFP**	**Log Sales**
PA Use	0.0123*** (0.0047)	0.0089* (0.0048)	0.0167*** (0.0052)
High IT K	0.1366*** (0.0086)	0.1090*** (0.0085)	0.1046*** (0.0145)
High Skilled Workers	0.0530*** (0.0050)	0.0614*** (0.0051)	0.0572*** (0.0053)
High Structured Mgmt.	0.0238*** (0.0053)	0.0203*** (0.0053)	0.0290*** (0.0054)
Log L	-0.0443*** (0.0031)		0.3820*** (0.0063)
Log (non-IT) K Per Employee	0.1116*** (0.0036)		
Log (non-IT) K			0.0869*** (0.0035)
MU	0.0219*** (0.0058)	0.0136*** (0.0064)	0.0508*** (0.0081)
Other controls	Y	Y	Y
Industry x Year Fixed Effects	Y	Y	Y
R-Squared	0.8231	0.1972	0.9370
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 sales per employee. The dependent variable for columns 2 and 3 are logged total factor productivity (TFP) and logged sales respectively. High IT K is an indicator for plants with the top tenth percentile of IT capital stock. High Skilled Workers is an indicator for plants with the top quartile of the percentage of employees with a bachelors' degree. High Structured Management is an indicator for plants with top quartile structured management practices index. Unreported controls for all columns include plant age, indicators for data-driven-decision making practices other than KPI tracking, and an indicator for plants reported as Headquarters (HQ) or co-located with HQ. Additional unreported controls for column 3 include logged cost of material and energy. Robust standard errors clustered at the firm level. * p<0.10, ** p<0.05, *** p<0.01.

Table A4. Organizational Complements to Predictive Analytics (Continuous Measures)

Model	(1) IT Capital Stock	(2) Skilled Workers	(3) Managerial Capacity
Dependent Variable	**Log Sales**		
PA Use	0.0073 (0.0122)	-0.0046 (0.0067)	-0.0273** (0.0139)
PA × Log IT K	0.0001 (0.0022)		
PA × Skilled Workers		0.0869** (0.0351)	
PA × Managerial Capacity			0.0490** (0.0179)
Other controls	Y	Y	Y
Industry x Year Fixed Effects	Y	Y	Y
Adjusted R-Squared	0.9337	0.9328	0.9339
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. Log IT K is the logged IT capital stock. Educated Workers is the percentage of employees with a bachelors' degree. Managerial Capacity is currently a categorical variable based on question 2 in the 2015 MOPS survey, though results are consistent for a broader measure of managerial capacity (*pending disclosure review*). 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. Unreported controls for all columns include logged total number of employees, log non-IT capital stock, logged cost of material and energy, plant age, indicators for having top DDD-related practices (DDD) other than KPI tracking, structured management (i.e., an index calculated based on the first 16 questions of the MOPS, excluding question 2 for KPI tracking and question 6 for target setting), plant type, multi-unit status, and headquarters status. Robust standard errors clustered at the firm level. * p<0.10, ** p<0.05, *** p<0.01.