Management for the Analytics Age[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

January 17, 2023

In-Progress….Some results pending disclosure avoidance review!

Abstract

Despite rapid and widespread adoption of data and predictive analytics by firms, performance gains have arisen unevenly. Working with the U.S. Census Bureau, we surveyed over 30,000 manufacturing establishments about their capacity to leverage data-driven insights, focusing on both diverse management practices and distinct use cases for data analysis. We find strong complementarities between a workplace's use of predictive analytics and its managerial capacity to leverage objective data. However, while production process monitoring and data-driven decision making reinforce the use of and returns to analytics, we find that high levels of other structured management practices erase analytics-derived productivity gains, likely through the use of high-powered people-management practices. In contrast, approaches that emphasize "soft" inputs from frontline workers and grant them greater workplace voice augment the returns to analytics. In keeping with these gains from maintaining "humans in the loop," we further find no evidence that either frontline workers or managers are replaced by this form of digital automation. Finally, applications of analytics to supply chain management and demand forecasting are both more-prevalently adopted and systematically productive than product-design applications. However, OLS estimates appear downward-biased, particularly in the innovation-focused use cases: when we instrument for predictive analytics using government mandates to collect data, all three use cases demonstrate positive, significant, and causal productivity advantages. Taken together, these findings provide novel and nuanced guidance for both managers and workers navigating a rapidly-evolving and increasingly digital landscape.

Keywords: digitization, data, prediction, predictive analytics, productivity, complementarities, voice, supply chain, forecasting, innovation

---

[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.

1. Introduction

The saying "what gets measured gets managed" is a persistent catchphrase in management. Misattributed most often to Peter Drucker or E. Edwards Deming, this saying conveys the insight that improved measurement can unlock new opportunities for managers to better understand their businesses and, in many cases, improve decision making and outcomes (McAfee and Brynjolfsson 2012). Indeed, evidence is mounting that increasingly-available digital information, data-driven management practices, and analytics are boosting performance across a wide range of organizations and activities (Brynjolfsson and McElheran 2019; Jin and Sun 2019; Wu, Lou, and Hitt 2019, 2020; Berman and Israeli 2022).

However, critics of an over-zealous focus on objective measurement – notably including both Deming and Drucker – have emphasized the pitfalls associated with this approach. Deming put "management by use only of visible figures, with little consideration of [those] that are unknown or unknowable" on his list of seven "deadly management diseases" (Deming 1986). Management-by-the-numbers is argued to cause "surrogation," whereby employees replace abstract strategic goals with tangible metrics (Harris and Taylor 2019), harming "the purpose of the organization" (Caulkin 2008) and "ceas[ing] to function…because people start to game [them]" (Biagioli 2016). Criticisms frequently center on the use of objective measures in knowledge work (Drucker 2007).[2] Are such cautions becoming obsolete? Or, can they co-exist with a growing role for data and analytics in an increasingly digital age?

To answer this, we explore in detail the management practices that may shape the productivity gains derived from predictive analytics. We triangulate on this core question, first, by examining how adoption and performance are influenced by "hard" structured management practices. Next, we investigate the "softer" side of management, seeking to understand if this distinct approach complements or substitutes for more measurement-focused ones. Third, we disentangle different use cases for predictive analytics to uncover what applications may yield greater benefits – and in which organizational contexts.

[2] Many of the objections center on measurement problems in knowledge work. Drucker (2007) wrote, "because knowledge work cannot be measured the way manual work can, one cannot tell a knowledge worker in a few simple words whether he is doing the right job and how well he is doing it."

Our study relies on rich and representative U.S. Census Bureau microdata that was developed for this purpose, yet has been unexplored along certain key dimensions. Specifically, we leverage the data for 2010 and 2015 collected by the Management and Organizational Practices Survey (MOPS), which grants insight into the execution and organizational design of over 30,000 establishments across the U.S. manufacturing sector (Buffington et al. 2017). Our approach further leverages rich administrative data to estimate multi-factor productivity, and instruments for selection into predictive analytics use to unpack causal relationships, where novel.[3] As such, this study provides the most-nuanced – and managerially relevant – insights to date into key correlates and contingencies shaping the economics of predictive analytics use.

We assemble our collage of evidence in steps. We first confront the often-implicit assumption that what is measured *may* be managed. In particular, a range of management practices have been found to complement investments in IT, data, and analytics (Bloom et al. 2012; Aral et al. 2012; Tambe et al. 2012; Bloom et al. 2014; Tambe 2014; Brynjolfsson and McElheran 2016, 2019; Wu et al. 2019; Brynjolfsson, Jin, et al. 2021). However, the presence of appropriate managerial capacity to act on data-driven insights is not a foregone conclusion, and details of the practices that may support –or hinder –gains from predictive analytics remains incompletely understood.

Here, we find that robust monitoring of the production process – specifically, relying on more than a handful of key performance indicators (KPIs)—strongly complements predictive analytics. This holds not only for predicting its use, but also strongly gates productivity benefits. Simply put, only workplaces with the managerial capacity to *leverage* analytics-driven insights demonstrate significant productivity advantages.

Other practices focused specifically on the availability and use of data in decision making (Brynjolfsson and McElheran 2016, 2019) show mixed evidence of complementarity. They contribute strongly to adoption, one measure of this reinforcing relationship based on rising diffusion of the understanding of their increasing returns (Brynjolfsson and Milgrom 2013). However, they demonstrate a noisier interaction with predictive analytics in the performance estimation, the other standard test for complementarity (Tambe et al. 2012).

---

[3] Brynjolfsson, Jin, and McElheran (2021) estimate the causal relationship between generic use of predictive analytics and productivity using a similar approach.

Finally, top-quartile use of management practices that take a "structured" approach to both operations and people management show surprising interactions. These practices, such as setting "stretch" targets, dismissing underperforming workers, or providing individual-level performance-based bonuses, have been associated with superior firm performance in a number of studies (Scur et al. 2021). While they correlate strongly with both predictive analytics use and overall multi-factor productivity, they interact negatively with predictive analytics in the production function. Having one *or* the other in place is associated with significantly greater output, controlling for a broad range of inputs; however, their combined returns are not statistically different from zero. Given the strong correlation among measurement-focused *operations* management practices found elsewhere (Scur et al. 2021), and based on preliminary investigation, we anticipate that the structured "*people* management" practices are key drivers of this pattern (*results pending disclosure avoidance review*).

This naturally raises the question of what role humans play in the face of digitally-enhanced automation (Autor 2015, Acemoglu and Restrepo 2019, Bessen et al. 2019). Further, it leaves unaddressed the Deming-esque critiques that many essential dimensions of firm management are resistant to objective measurement. "Known unknowns" may reduce incompletely to objective metrics, and "unknown unknowns," definitionally, cannot be tracked *ex ante*. A growing body of research suggests that such concerns may yield to managerial intervention, but through distinct practices that emphasize trusting, eliciting, and leveraging the insights of frontline workers (e.g., Gibbons and Henderson 2012), as well as creating a workplace where employees have substantial "voice" at work (Freeman 1980, Freeman and Lazear 1995, Harju et al. 2021).

We contribute to understanding the "soft" practices that may complement analytics – and management, more generally (Deming 2017, 2021; Hansen et al. 2021) – by leveraging purpose-designed measures in the MOPS that capture the extent to which workplaces rely on uniquely human inputs to their production. In particular, we observe whether a given plant solicits input from production workers, and (in multi-divisional firms) local managers in determining and providing input to decision making. We find, overall, that reliance on production worker input, in particular, has increased significantly over time in the U.S. Manufacturing sector. Preliminary results *(pending disclosure avoidance review*) further point to strong complementarities between "soft" management practices that both garner insights from local workers and that give them greater voice at work. Human managers also persist in these increasingly digital manufacturing plants, in contrast to findings focusing on physical automation, i.e., robots (Dixon et al. 2021). Specifically, we find no reduction in managerial headcount at plants relying on predictive analytics. Net-net, there seems to be a robust role for "humans in the loop" when it comes to enhancing prediction (Agrawal et al. 2019) in this fashion.

Finally, we contribute to mapping out the boundaries within which predictive analytics contributes to firm productivity, at least in manufacturing production. Prior work highlights contingencies rooted in the specific activities undertaken at a particular workplace. For instance, manufacturing contexts in which flow efficiency is strategically important show much greater analytics productivity (Brynjolfsson, Jin, et al. 2021). Here, we exploit novel data on the specific use cases to which predictive analytics is applied to gain insights into which uses of predictive analytics may be more productive – and in what types of organizational settings.

We find robust evidence that most plants show significant, *causal* gains from data analytics across use cases. In fact, OLS estimates appear to be biased downward in certain applications. Looking at each, in turn, we find that analytics applied to supply chain management is large and significant, across specifications. The use of analytics to forecast demand appears largest along the intensive margin and if the plant responded to a government mandate to collect data in a way that promoted predictive analytics use. These returns are quite widespread, appearing within a large number of narrowly-defined (NAICS 6) industry codes. In sharp contrast, the use of analytics for product design appears more contingent on workplace characteristics. Adoption for this use case is significantly higher at plants that are also firm headquarters, which also tend to have a higher manager-to-employee ratio and more discretion over strategic (as opposed to operational) decisions (*pending review*). Looking across the sample, average returns to analytics use in product design is indistinguishable from zero outside the instrumental variables specification.[4] Taken together, these patterns broadly point to gains from deploying analytics in settings where objective data is more abundant and central to managerial priorities.

[4] Productivity estimation at headquarters establishments is complicated, as these establishments provide spillovers to other plants within the firm that penalize the measured outputs conditional on inputs (McElheran, Ohlmacher, and Yang 2019). Future work will explore whether contingent gains from analytics may be found at these atypical plants, caveated appropriately.

These findings contribute to a few streams of work. The impact of data and the use of analytics on firms and workers has garnered increasing attention – and no small amount of concern – in recent years. This may be observed at every level of analysis. At the macro level, a surge in digitization-focused investment over the past decade has failed to yield measurable productivity gains (e.g., Brynjolfsson, Rock, and Syverson 2021). At the firm level, managers have struggled to deliver on the promise of new tools to extract value from data (Ransbotham et al. 2016). However, gains from analytics, measured in a variety of ways, have been documented in a rising number of contexts (Brynjolfsson and McElheran 2019; Jin and Sun 2019; Wu et al. 2019, 2020; Brynjolfsson, Jin, and McElheran 2021; Berman and Israeli 2022; Galdon-Sanchez et al. 2022). Our findings add directly-surveyed measures of analytics use for prediction, a rising area of interest for digital automatino (Agrawal et al. 2019). It further adds new insights into boundary conditions, complementarities, and specific applications to this growing research conversation.

For individual workers, the value of human inputs in the face of rising automation is unclear and potentially declining (Autor 2015, Autor and Salomons 2018, Acemoglu and Restrepo 2019; Bessen et al. 2019). This mounting concern has been difficult to tackle empirically, due to the diversity of organizational interactions and the dearth of mirco-level data. Prior work points to displacement of managers in the face of physical automation at the plant level (Dixon et al. 2021), but micro-level evidence with respect to automation of cognitive work is scarce (Barth et al. 2022). Our findings point to complementarities, rather than substitution when it comes to "softer" and uniquely human inputs to production, as well as no evidence for job loss at plants that adopt predictive analytics (*details pending disclosure avoidance review*).

2. Data and Summary Statistics

While many technologies and management practices are information-enhancing, predictive analytics is distinct from and arguably more sophisticated than other approaches.[5] Predictive analytics leverages computer systems to extract regularities from much larger data sets, more rapidly, and increasingly with more predictive power than previously possible.

[5] Prior work supports a meaningful distinction between relying on predictive analytics and being "data-driven" (Brynjolfsson and McElheran 2019), as well as differences between descriptive and predictive analytics (Berman and Israeli 2022).

Empirical exploration of the impact of these new digital capabilities on performance has been hampered by the lack of large-scale data on predictive analytics use and key organizational complements. To address this, 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%, with over 30,000 complete establishment-level observations.[7] Related measures were first introduced in Brynjolfsson, Jin, and McElheran (2021), but focused only on the use of any kind of predictive analytics, whereas this study adds new measures of distinct use cases, as well as new insights into "softer" inputs into data collection and interpretation.

Predictive Analytics Use

Adding questions to Census surveys requires rigorous cognitive testing (Buffington et al. 2017), essential for measuring such a recent and fast-emerging technology across different industry settings. The first question concerning predictive analytics use 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 from: *never, yearly, monthly, weekly,* and *daily,* with separate columns for 2015 and what they recall for 2010. With the recall data in 2010, we have roughly 51,000 observations across the two years.[8]

As in Brynjolfsson, Jin, et al. (2021), we flag the extensive margin of analytics use. Variation along the intensive margin is captured by assigning a numeric value ranging from 0 to 4 for each frequency category in ascending order, defaulting to the highest in cases of multiple categories. We primarily leverage this more-continuous measure in our instrumental variables (IV) estimation (see below). Note that the distance between frequency levels is not uniform and care, therefore, is

---

[6] See Bloom et al. (2013, 2019) and Buffington et al. (2017) for more details.
[7] Additional details available in the *pending* Data Appendix.
[8] Note that the sample counts are rounded for disclosure reasons throughout the paper. We use the total number of observations (~51,000) as our baseline sample but all key results are robust to a sub-sample with a higher quality recall structured management score where respondent tenure started at least one year before the period of the recall (Bloom et al. 2019).

required in interpretation.

Predictive analytics was widely diffused among manufacturing plants across almost all states and industries as early as 2010 (Brynjolfsson, Jin, et al. 2021), with average adoption of some type of predictive analytics at well over 70% (Table 1). Among the roughly 18,000 establishments with complete data for both years,[9] we observe a small 1.4% average yearly increase.[10]

This widespread diffusion and low rate of change have implications for our empirical approach: in particular, they forestall estimation of within-plant effects over time. To begin, focusing on changes in the subpopulation of plants that adopt relatively late would yield insight on a selected sample that might not be representative of the broader population. In particular, late adopters of technological innovations tend to be those with lower expected returns and higher costs of implementation (e.g., David 1969; Bresnahan and Greenstein 1996). We address selection into adoption to some degree using our IV estimation. In addition, statistical power in the subsample of establishments that shift their predictive analytics use is severely limited, despite the overall size of our data set.

Analytics Use Cases

To gain additional insight into the applications in which predictive analytics is deployed, we added another question to the MOPS concerning where its use in three distinct business functions: supply chain management, demand forecasting, and product design. We construct the corresponding use cases indicators conditional on the plant having adopted predictive analytics and also using data analytics in the corresponding application and frequency.[11] As described in Table 1, the diffusion of the distinct use cases was similarly quite high as early as 2010, with roughly 70% uptake of supply-chain and demand analytics that increased to 77% by 2015. Use of analytics in product design was slightly lower at 63% in 2010, rising to 71% by 2015. Our objective with these new questions was primarily descriptive and in service of managerial relevance, rather than being theoretically motivated.

---

[9] The rotation of the ASM sample in years ending with "4" and "9" limits the number of establishments that have complete data for both years in our sample, with a core "certainty sample" of larger plants present in both.

[10] The adoption of predictive analytics rose from 73% in 2010 to 80% by 2015.

[11] Both questions are further described in the *pending* Data Appendix.

Managerial Complements and Plant Performance

Complementarities among IT investments and organizational characteristics have long been associated with differences among firms that may persist and even grow over time (Milgrom and Roberts 1990, 1995; Black and Lynch 2001; Caroli and Van Reenen 2001; Bresnahan et al. 2002; Bloom et al. 2012; Aral et al. 2012; Tambe et al. 2012; Brynjolfsson and Milgrom 2013). Rising concentration and inequality in both workplace conditions and employee earnings are increasingly attributed to technology investment within industries and firms (Bessen 2017; Bennet 2020b; Lashkari et al. 2020; Barth et al. 2022). In addition, difficult-to-measure intangible features of firms and markets are argued to amplify these dynamics (Saunders and Brynjolfsson 2016; Haskel and Westlake 2018). Prior work has found important complementarities between predictive analytics and high IT capital stock, educated workers, and production environments dedicated to flow efficiency (Brynjolfsson, Jin, et al. 2021). Building on these insights, we explore novel managerial complements to predictive analytics that might shape returns to its use.

Operations Tracking and Monitoring

Manufacturing firms have long been developing and monitoring KPIs to support a variety of operational practices such as Lean and Total Quality Management. Firms often track capacity utilization rate, manufacturing cycle time, equipment downtime, and many others. Tracking KPIs is highly correlated with the breadth and intensity of data collection and its use in decision making (Brynjolfsson and McElheran 2019). For predictive analytics, anecdotal evidence has shown that plants with well-established KPI practices will have richer inputs to feed their predictive models, potentially lowering the cost or boosting the returns to predictive analytics (Schrage and Kiron 2018). In addition, these practices reflect investments and routines that enable firms to contextualize and intepret anlatyics-driven insights, a necessary step for taking action in response to new information when execution relies on managerial intervention.

On the MOPS, respondents indicate the number of KPIs monitored at the plant, with options to select 0, 1-2, 3-9, or 10 or more. Examples given include metrics on production, cost, waste, quality, inventory, energy absenteeism, and deliveries on time. We take the top category as a proxy for establishments with high intensity of KPI monitoring to test our hypothesis that robust monitoring and tracking of workplace operations is a complement to predictive analytics. This high level of managerial capacity to choose metrics and interpret them was, on average 44% in our analysis sample. The recall data suggest it increased non-trivially from 37% in 2010 to 56% in 2015.

Data-Driven Decision Making

Following Brynjolfsson and McElheran (2016, 2019), we leverage questions in Section C of the MOPS that ask about the availability and use of data in decision-making, as well as question 6 from Section A about the use of production targets, which is an important input to interpreting data about operational metrics (as covered in the question on KPIs, above). We construct an indicator of being in the top categories of all three questions, which captured roughly 35% of the sample in 2010 and rose to 59% by 2015. While this has been shown to be distinct from predictive analytics in prior work, we anticipate that it could be important in the adoption and performance of predictive analytics by 2015.

Structured Management

Building on Bloom et al. (2013, 2019), we construct an index using other management practices reported in section A of the MOPS, excluding the data-related questions described above (specifically 2 and 6 of the MOPS). To compare magnitudes across potential complements, we also construct an indicator of "high structured management" based on being in the top quartile by industry within the sample.

"Soft" Management Practices

New questions were added to the 2015 wave of the MOPS to capture measures of "voice" and "soft information" within the workplace. Under question XYZ, which asks about the authority to decide what data collect, options include managers at this establishment, managers at headquarters or other establishment within the firm, production workers, engineers, customers, or the government. We leverage the last in our IV estimation (described in detail below). Of particular interest for understanding how managers may substitute or complement objective information with less-formal inputs, we exploit variation on whether local managers and/or production workers decide on what data to collect. There is less variation over time in the importance of local plant managers in this function. However, this climbed from 16% in 2010 to 23% 2015 and loads heavily on a principal component analysis of "soft management" (*pending disclosure review*).

We further exploit information on the frequency of reliance on formal and informal feedback from production workers, which rose from 52% in 2010 to 61% in 2015. Perhaps more strikingly, the rate of never relying on such input fell from 7% in 2010 to 3% in 2015. Again, these load heavily on a component that captures this soft information, as well as "voice" within the firm (*pending disclosure review*).

Multi-Factor Productivity

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

Key Controls

To account for plant-level heterogeneity in IT capabilities, we calculate IT capital stocks using capital expenditure on computer and peripheral data processing equipment from the ASM and CMF panel dating back to 2002, using a standard perpetual inventory approach and an industry-level deflator for hardware from the Bureau of Economic Analysis (BEA). We impute values for years in which values are missing and depreciate at the rate of 35% per year following Bloom et al. (2014). We account for variation in the skill of workers using 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 (Card 1999; Bresnahan et al. 2002). Other standard controls in most specifications include a measure of the production design at the plant (i.e., the extent to which it is organized for flow efficiency – see Brynjolfsson, Jin, et al.

2021), an indicator of whether the plant belongs to a larger, multi-unit firm, and plant age as captured in the LBD.

3. Empirical Approach

We use linear probability models to explore correlations between a wide range of firm and plant covariates with the use of predictive analytics. We address concerns regarding industry and year transitory shocks by controlling for industry-year fixed effects. Key explanatory variables include size, age, indicators for plants belonging to multi-unit firms or being the firm headquarters, and organizational complements as described above.

Next, we estimate both the average productivity return to predictive analytics, with an emphasis on and standard performance tests of complementarity. For specific use cases, we replicate approaches in Brynjolfsson, Jin, et al. (2021) that estimate both conditional correlations in the productivity function, as well as instrumental variables.

For these estimates, 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, 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 consumption of material and energy inputs, and X is a vector of abovementioned controls and the three potential complements. Both w_{ijt} - the "technical productivity" and ε_{ijt} - the "shock to productivity" are unobservable econometrically (but w_{ijt} might be observable by establishments). While the typical coefficient of interest is β_{pa}, the average relationship between predictive analytics and plant productivity, all else equal, we move directly to analysis of complementarities in specifications where we do not focus on particular use case. See Brynjolfsson et al. (2021) for estimates of the direct effects.

Following the empirical strategy in Athey and Stern (1998) and Brynjolfsson and Milgrom (2013), our empirical specification for complementarity analysis follow equation (2):

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

All variables in equation (2) are identical to those in equation (1) except C_{ijt}, which denotes, respectively, indicators for high IT tracking and monitoring (of KPIs), high data-focused management practices, high structured management, and a continuous measure of structured management, respectively A positive and significant $\beta_{interaction}$ term is indicative of such performance complementarities.

We rely primarily on pooled OLS when estimating equation (1) for several reasons. First, the 2015 MOPS only provides two snapshots for the adoption of predictive analytics and other organizational characteristics. As discussed above, the small within-plant variation for the adoption of predictive analytics and other MOPS variables would lead to significant downward bias due to measurement error in the fixed-effects model (Swaffield 2001). In particular, standard models for addressing time-invariant organizational heterogeneity through the use of firm or establishment fixed effects are of limited use in our empirical setting. Recall that the panel data methods only identify coefficients off of establishments that undergo "treatment" within the sample frame – i.e., that transition from not using predictive analytics to adopting in the 2010-2015 timeframe. Thus, they would only be applicable to a very small and highly-selected group of plants whose defining feature is being in this "laggard" group of adopters (those that had not adopted by 2010 but had finally done so by 2015). Earlier users of these practices would be differenced out of any panel data estimation, likely biasing the average productivity attributed to these practices (Brynjolfsson et al. 2021).

Second, this measurement error issue is exaggerated as we cannot observe precise years of the adoption of predictive analytics for "late adopters", from which fixed-effects models extract information. The long 5-year gap in our two-year panel generates considerable measurement imprecision concerning this issue. Lastly, given the return of predictive analytics depends on the organizational complements (Brynjolfsson et al. 2021), the fixed-effect model is likely to substantially suppress the estimates as it would absorb most of the variation for the time-invariant organizational complements as discussed in the data section. Therefore, we focus our empirical approach primarily on the pooled OLS models for the two discrete years of data and include industry-year fixed-effects as well as a rich set of controls (many of which capture fixed or quasi-fixed organizational characteristics).

Instrumental Variable Estimation

A standard concern with this approach is that predictive analytics use may be endogenously determined, biasing interpretation of β_{pa}.[12] To address this, we explore IV estimation using an indicator of data collection driven by government regulation or agencies.

The motivation for this instrumentation strategy rests on the so-called "Porter Hypothesis" (Porter 1991; Porter and Van der Linde 1995) arguing that well-designed government regulations can stimulate firms to innovate and adopt new technology and practices. Of relevance in our setting, data collection is often mandated by federal and local governments to demonstrate compliance with 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 emission data collection and monitoring. Leveraging this data in government-mandated reports requires that workers and managers be trained in systems and techniques for capturing, analyzing, and communicating data-driven conclusions. Consistent with this mechanism, firms are more likely to build infrastructure and managerial systems for data collection, storage, and analysis when required to collect new data and devise new reports by statute.[13]

Not all firms will be able to translate this into improved management of their production processes.[14] For some, however, this external "nudge" into increased investment in and awareness

[12] 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 suggesting that such concerns may be overemphasized in IT productivity studies. System GMM and 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). However, our two-year panel lacks the longer lags typically required.

[13] Abundant anecdotes support the prevalence of this phenomenon. the Occupational Safety & Health Administration (OSHA) Recordkeeping rule can serve as another example where 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, please see the OSHA website https://www.osha.gov/recordkeeping2014/records.html

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

of data resources may shift practices on the margin among plants exposed to this additional oversight. The unexpected consequences can be striking. The case of Alcoa Corporation in the late 1980s and 90s is illustrative. When Paul O'Neil took leadership of the firm, his unexpected mandate to prioritize safety resulted in an abundance of data about accidents – but also about the performance and maintenance of infrastructure and workplace practices underlying those accidents. New data enabled new performance metrics, which were analyzed with increased frequency and linked to manager pay at the firm (*Fortune* 1991). The end result was not only improved worker safety but also improved productivity (Clark and Margolis 1991).

For this to be useful as an instrument, such oversight needs to be unrelated to the productivity of affected plants. Historically, U.S. government regulations in the manufacturing sector have fit this description. For instance, the objective of EPA CEMS requirement 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.[15]

This instrument is used to estimate the causal average of effect of any use of predictive analytics in Brynjoflsson, Jin, et al. (2021). Thus, we restrict our use in this study to the novel use-case explorations.

[15] See Data Appendix for more details. This has been used in a related study of data-driven decision making by Brynjolfsson and McElheran (2019).

4. Results

Correlation Test for Complementarities: Adoption of Predictive Analytics

Figure 1 and Table 2 describe the key contributors to predictive analytics use (of any kind, at any frequency), building from the baseline adoption of only around 41% when no potential complements are present. As found in Brynjolfsson et al. (2021) and replicated here for comparison, high IT capital stock and a higher percentage of educated workers increase the likelihood of use; combined, they raise the rate to nearly 47%. High capacity for tracking and monitoring operations has a large standalone effect on the likelihood of adoption: in column 3 of Table 2 it is shows a coefficient of 0.105, significant at the one-percent level. Column 4, however, indicates that there is some overlap among these measures. In fact, a principal component analysis reported in Table 3 suggests that there is one underlying component (the only with with an eigenvalue above the conventional threshold of 1) on which all three of these load. Of the three, this latter measure, however provides the largest-magnitude effect. Added to the others, plants with all of these factors would have a roughly 53% likelihood of using predictive analytics, all else equal.

This is only one way evaluate the managerial capacity at the plant to adopt, interpret, and leverage predictivec analytics. In fact, the next approach – using indicator of high data-focused management pratices - increases the likelihood of adoption another 11.6 percentage points (column 6, Table 2), suggesting that high-complement plants are, have a 67% adoption rate (Figure 1).

The final potential complement, structured management practices, is evaluated using the top-quartile “high” indictor, for ready comparison. Continuing in column 5, top-quartile structured management practices contributes another 8.8 percentage points to the likelihood of use. For later reference, the government mandate is also included to indicate its predictive power in the first stage. Figure 1 shows the relative importance of structured management to the overall likelihood of adoption. Keep in mind that this is only part of a standard complementarity test, which examines both correlates of adoption and interactions in a production function (Brynjolfsson and Milgrom 2013).

We further investigate whether “hard” vs. “soft” management styles contribute to adoption in column 7. *These detailed results are pending disclosure avoidance review.*

Performance Test for Complementarities

Table 4 reports on OLS estimates of equation (2). Column 1 solidifies the intuition that high managerial capacity for tracking and monitoring operations is complementary to predictive analytics. In this specification, neither the direct effect of predictive analytics nore the direct effect of this measure is statistically significant, while the interaction effect is significant at the one-percent level. The linear combination of this high levels of KPI tracking and monitoring and predictive analytics use (reported at the bottom of Table 4) is 0.022 and also significant at the one percent level.

While data-focused management practices may significantly predict the use of predictive analytics, it shows no significant interaction effect in the performance regression in column 2 of Table 4. The linear combination, however is positive and significant at the 10% level.

A puzzle arises with respect to structured management practices. A top-quartile level of structured operations and people management practices (aside from those discussed above) has a positive direct conditional correlation with performance in column 3. Predictive analytics is likewise significantly associated with productivity, both at the one-percent level. However, the interaction effect is *negative*¸ large, and statistically significant. The linear combination suggests that having both analytics and high structured management in place is not statistically distinguishable from having neither. To dig into this a bit more, we leverage the continusous z-score of the structured management practices in column 4. The interaction term is noisy, suggesting that there is important underlying heterogeneity in this relationship that requires furture attention. Based on casual inspection of the data, and *pending disclosure avoidance review*, we suspect that this is related to high levels of structured "people management" practices, which rely on high-powered incentives for workers. Additional work is required to unpack this relationship.

The Adoption of Use Case of Predictive Analytics

Figure 2 and Table 5 report on the correlates of adoption of distinct use cases for predictive analytics. Most of the covariates are consistent across use cases, with the exception of Headquarters. In panels a and b of Figure 2.

Use Cases of Predictive Analytics & Productivity

Table 6 reports on OLS estimates of equation (1), using binary indicators of use. The most clearly productive application of predictive analytics, setting aside concerns about selection into treatment, is supply chain analytics. The coefficient on analytics use for this purpose is 0.012 and significant at the 5% level (column 1). Column 2 suggests some relationship relationship between demand forecasting applications of analytics, but the coefficient is smaller and only significant at the 10% level. Analytics use in product development has no significant relationship to productivity in column 3.

Table 7 reports on categorial indices of use, which are more suitable for our IV strategy, to follow. The results are qualitatively similar to Table 6.

Table 8 uses 2SLS estimation, relying on government mandated data collection as an indstrument for potentially endogenous selection into predictive analytics use. The striking observation here is that all of the coefficients are larger, postivie, and significant at the one-percent level. Given the strong first-stage (reported at the bottom of Table 8, and also hinted at in the richer specification in Table 2), this is unlikely to be the result of weak instruments. The overall pattern is consistent with downward bias in the OLS estimates, potentially due to measurement error. For an in-depth discussion of this instrumentation strategy, see McElheran, Brynjolfsson, and Yang (2023). This work suggests that a number of plants do not adopt data-based management practives endogenously, even when productivity results from a government "nudge" to adopt.

5. Conclusion – UNDER CONSTRUCTION

References

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. Journal of Economic Perspectives, 33(2), 3-30.

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Exploring the impact of artificial intelligence: Prediction versus judgment. Information Economics and Policy, 47, 1-6.

Aral, S., Brynjolfsson, E. and Wu, L., 2012. Three-way complementarities: Performance pay, human resource analytics, and information technology. Management Science, 58(5), pp.913-931.

Athey, S. and Stern, S., 1998. An empirical framework for testing theories about complimentarity in organizational design (No. w6600). National Bureau of Economic Research.

Autor, D., & Salomons, A. (2018). Is automation labor-displacing? Productivity growth, employment, and the labor share (No. w24871). National Bureau of Economic Research.

Barth, E., Davis, J. C., Freeman, R. B., & McElheran, K. (2022). Twisting the demand curve: Digitalization and the older workforce. Journal of Econometrics.

Berman, R., & Israeli, A. (2022). The Value of Descriptive Analytics: Evidence from Online Retailers. Marketing Science.

Bessen, J., Goos, M., Salomons, A., & Van den Berge, W. (2019). Automatic reaction-what happens to workers at firms that automate?.Boston University Working Paper

Biagioli 2016

Bloom, N., Sadun, R. and Van Reenen, J., 2012. Americans do IT better: US multinationals and the productivity miracle. American Economic Review, 102(1), pp.167-201.

Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R.S., Saporta Eksten, I. and Van Reenen, J., 2013. Management in America. US Census Bureau Center for Economic Studies Paper No. CES-WP-13-01.

Bloom, Nicholas and Brynjolfsson, Erik and Foster, Lucia and Jarmin, Ron S. and Patnaik, Megha and Saporta Eksten, Itay and Van Reenen, John Michael and Van Reenen, John Michael, IT and Management in America (March 2014). CEPR Discussion Paper No. DP9886, Available at SSRN: https://ssrn.com/abstract=2444907

Bloom, N., Brynjolfsson, E., Foster, L., Jarmin, R., Patnaik, M., Saporta-Eksten, I. and Van Reenen, J., 2019. What drives differences in management practices?. American Economic Review, 109(5), pp.1648-83.

Bresnahan, T., Greenstein, S. (1996). Technical progress and co-invention in computing and in the uses of computers. Brookings Papers on Economic Activity. Microeconomics, 1996, 1-83.

Bresnahan, T.F., Brynjolfsson, E. and Hitt, L.M., 2002. Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. The quarterly journal of economics, 117(1), pp.339-376.

Brynjolfsson, E. and Hitt, L., 1995. Information technology as a factor of production: The role of differences among firms. Economics of Innovation and New technology, 3(3-4), pp.183-200.

Brynjolfsson, E. and Hitt, L.M., 2000. Beyond computation: Information technology, organizational transformation and business performance. Journal of Economic perspectives, 14(4), pp.23-48.

Brynjolfsson, E. and Hitt, L.M., 2003. Computing productivity: Firm-level evidence. Review of economics and statistics, 85(4), pp.793-808.

Brynjolfsson, E., Hitt, L.M. and Kim, H.H., 2011. Strength in numbers: How does data-driven decision-making affect firm performance?. Available at SSRN 1819486.

Brynjolfsson, E., Jin, W., & McElheran, K. (2021). The power of prediction: predictive analytics, workplace complements, and business performance. Business Economics, 56(4), 217-239.

Brynjolfsson, E., & McElheran, K. 2016. The Rapid Adoption of Data-Driven Decision-Making. American Economic Review. https://doi.org/10.1257/aer.p20161016

Brynjolfsson, E., & McElheran, K. 2019. Data in Action: Data-Driven Decision Making and Predictive Analytics in U.S. Manufacturing. Rotman School of Management Working Paper No. 3422397. Available at SSRN: https://ssrn.com/abstract=3422397

Brynjolfsson, E. and Mendelson, H., 1993. Information systems and the organization of modern enterprise. Journal of Organizational Computing and Electronic Commerce, 3(3), pp.245-255.

Brynjolfsson, E. and Milgrom, P., 2013. Complementarity in organizations. The handbook of organizational economics, pp.11-55.

Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. American Economic Journal: Macroeconomics, 13(1), 333-72.

Buffington, C., Foster, L., Jarmin, R. and Ohlmacher, S., 2017. The management and organizational practices survey (MOPS): An overview 1. Journal of Economic and Social Measurement, 42(1), pp.1-26.

Card, D., 1999. The causal effect of education on earnings. In Handbook of labor economics (Vol. 3, pp. 1801-1863). Elsevier.

Caroli, E. and Van Reenen, J., 2001. Skill-biased organizational change? Evidence from a panel of British and French establishments. The Quarterly Journal of Economics, 116(4), pp.1449-1492.

Caulkin, 2008.

Clark, K.B. and Margolis, J.D., 1991. Workplace safety at Alcoa (A).

Davenport, T.H., 2006. Competing on analytics. Harvard business review, 84(1), p.98.

David, Paul A. 1969. A Contribution to the Theory of Diffusion. Memorandum No. 71: Stanford University.

Deming, W. E. (1986). *Out of the crisis.* Cambridge: MIT Press.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, *132*(4), 1593-1640.

Deming, D. J. (2021). *The growing importance of decision-making on the job* (No. w28733). National Bureau of Economic Research.

Dixon, J., Hong, B., & Wu, L. (2021). The robot revolution: Managerial and employment consequences for firms. Management Science, 67(9), 5586-5605.

Drucker, Peter F. The effective executive. Routledge, 2007.

Dubey, R., Gunasekaran, A., Childe, S.J., Blome, C. and Papadopoulos, T., 2019. Big Data and Predictive Analytics and Manufacturing Performance: Integrating Institutional Theory, Resource-Based View and Big Data Culture. British Journal of Management, 30(2), pp.341-361.

Duhigg, C., 2012. The power of habit: Why we do what we do in life and business. Random House.

Dunne, T., 1994. Plant age and technology use in US manufacturing industries. The RAND Journal of Economics, pp.488-499.

Gandhi, A., Navarro, S. and Rivers, D., 2017. How heterogeneous is productivity? A comparison of gross output and value added. Journal of Political Economy, 2017

Enke, B., 2020. What you see is all there is. The Quarterly Journal of Economics, 135(3), pp.1363-1398.

Forman, C., Goldfarb, A. and Greenstein, S., 2002. Digital dispersion: An industrial and geographic census of commerical internet use (No. w9287). National Bureau of Economic Research.

Foster, L., Haltiwanger, J. and Syverson, C., 2016. The slow growth of new plants: learning about demand?. Economica, 83(329), pp.91-129.

Freeman, Richard. 1980. “The Exit-Voice Tradeoff in the Labor Market: Unionism, Job Tenure, Quits, and Separations.” *Quarterly Journal of Economics* 94 (4):643–673.

Freeman, Richard and Edward Lazear. 1995. “An Economic Analysis of Works Councils.” In *Works Councils: Consultation, Representation, Cooperation in Industrial Relations*, edited by Joel Rogers and Wolfgang Streeck. NBER Comparative Labor Markets Series.

Galdon-Sanchez, J. E., Gil, R., & Uriz Uharte, G. (2022). The Value of Information in Competitive Markets: The Impact of Big Data on Small and Medium Enterprises. Available at SSRN.

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

Hall, B.H., Innovation and Diffusion. In The Oxford Handbook of Innovation.

Hansen, S., Ramdas, T., Sadun, R., & Fuller, J. (2021). The demand for executive skills. National Bureau of Economic Research working paper # w28959.

Harju, J., Jäger, S., & Schoefer, B. (2021). Voice at work. National Bureau of Economic Research working paper # w28522.

Harris, M, and Taylor, B. 2019. Don’t Let Metrics Undermine Your Business. *Harvard Business Review*, September-October. https://hbr.org/2019/09/dont-let-metrics-undermine-your-business

Haskel, J and Westlake S (2018) Capitalism without capital: The rise of the intangible economy. Princeton University Press.

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

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

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

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

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

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

Jin, Y., & Sun, Z. (2019). Information acquisition and the return to data. Harvard University Working Paper.

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

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

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

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

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

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

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

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

Levin, S.G., Levin, S.L. and Meisel, J.B., 1987. A dynamic analysis of the adoption of a new technology: the case of optical scanners. The Review of Economics and Statistics, pp.12-17.

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

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

Milgrom, P. and Roberts, J., 1995. Complementarities and fit strategy, structure, and organizational change in manufacturing. Journal of accounting and economics, 19(2-3), pp.179-208.

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

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

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

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

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

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

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

Saunders, A. and Tambe, P., 2015. Data Assets and Industry Competition: Evidence from 10-K Filings. Available at SSRN 2537089.

Saunders, A and Brynjolfsson E., 2016. Valuing it-related intangible assets. MIS Quarterly. 40(1): 83-110

Scur, D., Sadun, R., Van Reenen, J., Lemos, R., & Bloom, N. (2021). The World Management Survey at 18: lessons and the way forward. Oxford Review of Economic Policy, 37(2), 231-258.

Schlegel, G.L., 2014. Utilizing big data and predictive analytics to manage supply chain risk. The Journal of Business Forecasting, 33(4), p.11.

Schrage Michael (September 2014). Learn from Your Analytics Failures. Harvard Business Review. Retrieved from https://hbr.org/2014/09/learn-from-your-analytics-failures

Schrage, M. and Kiron, D., 2018. Leading with next-generation key performance indicators. MIT Sloan Management Review, 16.

Tambe, P., 2014. Big data investment, skills, and firm value. Management Science, 60(6), pp.1452-1469.

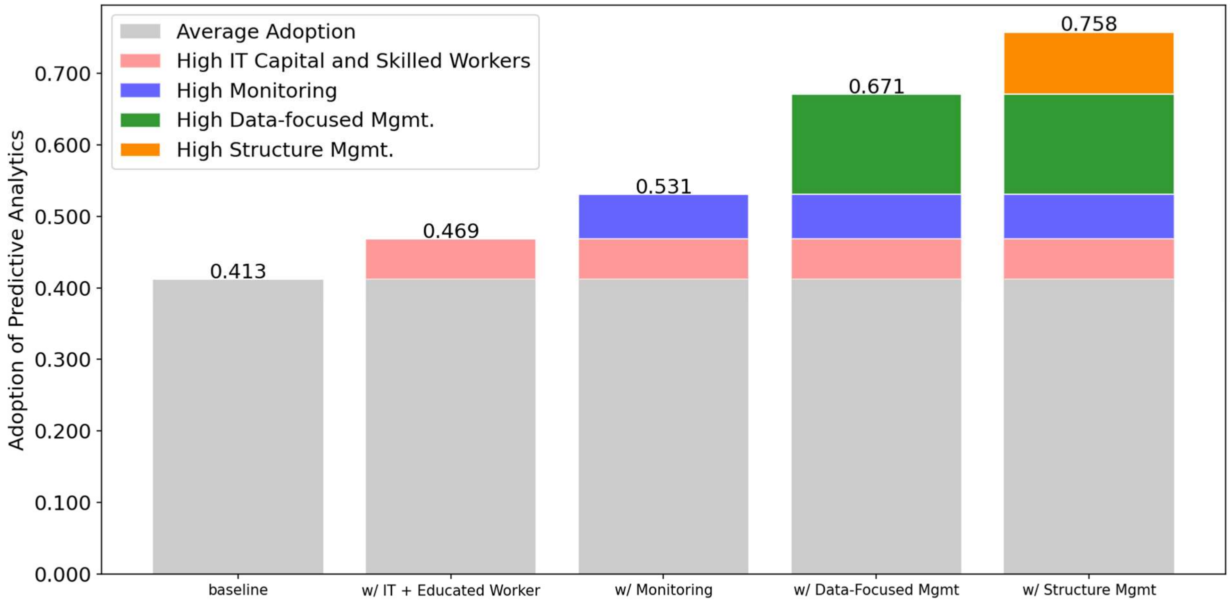
Tambe, P. and Hitt, L.M., 2012. The productivity of information technology investments: New evidence from IT labor data. Information Systems Research, 23(3-part-1), pp.599-617.

Tambe, P., Hitt, L.M. and Brynjolfsson, E., 2012. The extroverted firm: How external information practices affect innovation and productivity. Management Science, 58(5), pp.843-859.

Wu, L., Lou, B. and Hitt, L., 2019. Data Analytics Supports Decentralized Innovation. Management Science, 65(10), pp.4863-4877.

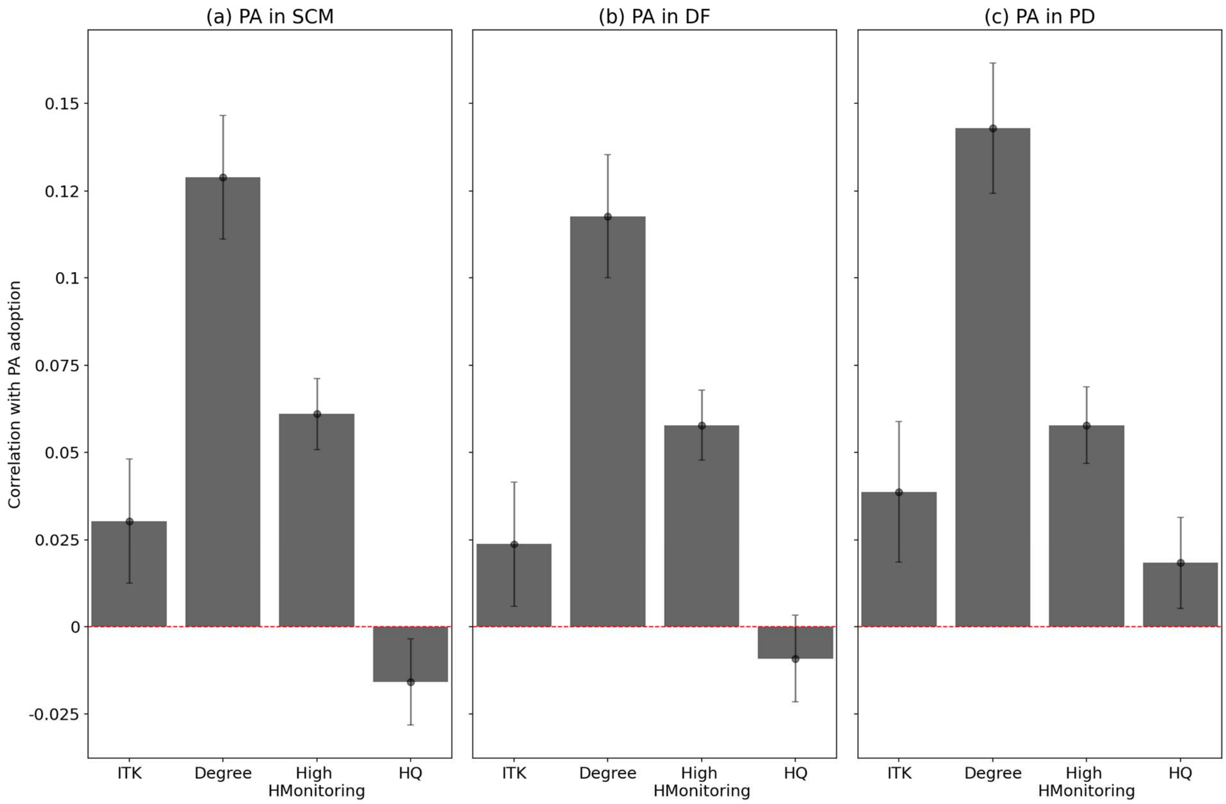
Wu, L., Hitt, L., & Lou, B. (2020). Data analytics, innovation, and firm productivity. Management Science, 66(5), 2017-2039.

Figure 1. Cumulative Correlations between Predictive Analytics Use and Workplace Practices



Notes: Correlates of predictive analytics use based on linear probability estimation in the pooled 2010 & 2015 sample. The graph depicts estimated additive marginal contributions based on a single model that includes High (top-quintile) IT Capital Stock, High (top-quartile) Educated Workers, High Monitoring, Data-Driven Decision Making, and High (top-quartile) Structured Management Practices (see Table 1 and column 5, Table 2). Average adoption represents the constant term. Additional controls include controls include Log L, Log K, age, and indicators for MU, HQ, and flow-efficient process design. Robust standard errors are clustered at the firm level. Findings are robust to binary dependent-variable (e.g., probit) models. Detailed regression coefficients are are available upon request.

Figure 2. Correlates of Predictive Analytics by Use Case



Notes: Estimates based on the baseline sample from the pooled OLS regressions controlling industry (6-digit NAICS) and year fixed-effects for plots (a)-(c). The dependent variable is the indicator for the adoption of the corresponding use case of predictive analytics including supply chain analytics (panel a), demand forecasting analytics (panel b), and product design analytics (panel c). Values on the Y-axis represent the correlation of each variable (indicated by the X-axis labels) in the adoption of the corresponding use case for predictive analytics (see Table 1). 95% confidence intervals are plotted on all columns.

Table 1. Summary Plant-Level Statistics

Variable	Definition	Mean (S.D.)	2010 (Recall)	2015
PA Use	Indicator of (any) predictive analytics use	0.74 (0.44)	0.73 (0.44)	0.80 (0.40)
Supply-Chain Analytics	Predictive analytics used in supply chain management	0.71 (0.46)	0.69 (0.46)	0.77 (0.42)
Demand Analytics	Predictive analytics used in demand forecasting	0.71 (0.45)	0.70 (0.46)	0.77 (0.42)
Product Design Analytics	Predictive analytics used in product design	0.65 (0.48)	0.63 (0.48)	0.71 (0.46)
High Monitoring	Monitoring 10 or more key performance indicators (KPIs)	0.44 (0.50)	0.37 (0.48)	0.56 (0.50)
Data-Driven Decision Making (DDD)	Indicator of high levels (top two categories) of data availability and use of data in decision making, as well as relying on both short- and long-term production targets	0.47 (0.50)	0.35 (0.48)	0.59 (0.49)
Structured Management	Normalized index of structured management practices (excluding KPIs and targets questions)	0.63 (0.17)	0.60 (0.16)	0.68 (0.15)
Mandated Data Collection	Indicator that government regulations or agencies chose, at least in part, what type of data is collected	0.25 (0.43)	N/A	N/A
Discretion over Data Collection:				
Plant Managers	Managers at this establishment		**84% (37%)**	**83% (37%)**
HQ Managers	Managers at headquarters and/or other establishments		60% (49%)	65% (48%)
Production Workers	Production workers		**16% (37%)**	**23% (42%)**
Engineers	Engineers		34% (47%)	38% (49%)

Sources of Data: Frequency of Formal or Informal Feedback from Production Workers				
Daily			52% (50%)	61% (49%)
Weekly			30% (46%)	32% (47%)
Monthly			28% (45%)	27% (45%)
Yearly			17% (38%)	16% (37%)
Never			7% (25%0	3% (17%)
Log Sales	Logged value of plant shipments ($Thousands)	10.37 (1.52)	10.68 (1.39)	10.86 (1.37)
Log L	Logged number of employees	4.56 (1.17)	4.79 (1.09)	4.88 (1.09)
Log K	Accumulated and depreciated capital investment in **non-IT** equipment and structures in log terms ($Thousands)	9.26 (1.47)	9.38 (1.58)	9.36 (1.61)
Log IT_K	Accumulated and depreciated expenditure on IT capital ($Thousands)	5.16 (2.41)	5.58 (2.25)	5.62 (2.18)
Educated 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)
Flow-Efficient Process Design	Indicator of a production process designed for high flow efficiency (i.e. cellular or continuous-flow production process). Omitted categories are job shop and batch production.	0.35 (0.48)	0.38 (0.48)	0.41 (0.49)
MU	Indicator for plants belonging to multi-unit firms	0.73 (0.45)	0.78 (0.41)	0.81 (0.40)
HQ	Indicator for establishments reported being headquarters (HQ) or co-located with HQ	0.47 (0.50)	0.43 (0.50)	0.41 (0.049)
Age	Plant age	24.47 (12.89)	24.20 (11.31)	29.20 (11.31)

Table 2. Correlates of Predictive Analytics Use

	(1) IT	(2) Education	(3) Monitoring	(4) All Three	(5) Managerial Capacity	(6) Managerial Style *Pending*
High IT	0.026*** (0.007)			0.017** (0.007)	0.024*** (0.009)	
High Educated Workers		0.053** (0.006)		0.045*** (0.006)	0.032*** (0.006)	
High Monitoring			0.105*** (0.006)	0.103*** (0.005)	0.062*** (0.005)	
Data-Driven Decision Making					0.140*** (0.005)	
High Structured Management					0.088*** (0.006)	
Mandated Data Collection					0.027*** (0.006)	
“Hard” HR Management Practices						+
“Soft” People Management Practices						+
Other Controls (Log L, Log K, Age, MU, etc.)	Y	Y	Y	Y	Y	Y
Industry Fixed Effects	Y	Y	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y	Y	Y
N				51,000		

Notes: Linear probability estimates controlling for industry (6-digit NAICS) and year indicators using the pooled 2010 & 2015 sample. The dependent variable across all columns is an indicator of predictive analytics use (see Table 1). **High IT** indicates top-quartile for Log IT_K. **High Educated Workers** indicates top-quintile percentage of Educated Workers. **High Structured Management** indicates top-quartile Structured Management practices. Unreported controls include Log L, Log K, age, and indicators for MU, HQ, and flow-efficient process design. Robust standard errors clustered at the firm level. * p<0.10, ** p<0.05, *** p<0.01. Results are robust to using binary estimation models (e.g., probit in Stata 15).

Table 3 PCA Test for IT Capital, Educated Workers, and Production Process MonitoringMonitoring

Component	**Eigenvalue**		
Comp 1	1.34		
Comp 2	0.87		
Comp 3	0.79		
Principal components (eigenvectors)	**Comp 1**	**Comp 2**	**Comp 3**
Log IT Capital	0.6131	-0.26	-0.746
Percentage of Educated Workers	0.5923	-0.4736	0.6518
Number of Key Performance Indicators (KPIs) monitored (categories = none, 1-2, 3-9, 10 or more)	0.5228	0.8415	0.1364
N	~51,000		

Note: Calculated using the **PCA** command in Stata 15.

Table 4. Performance Test of Workplace Complements to Predictive Analytics Use

	(1) Production Monitoring	(2) Data-Driven Decision Making	(3) High Structured Mgmt.	(4) Continuous Structured Mgmt.
Dependent Variables		Log Sales		
PA Use	0.001 (0.006)	0.005 (0.006)	0.016*** (0.005)	0.008 (0.017)
High Monitoring	0.002 (0.009)			
PA × High Monitoring	0.021** (0.010)			
Data-Driven Decision Making (DDD)		0.010 (0.009)		
PA x DDD		0.009 (0.010)		
High Structured Mgmt.			0.050*** (0.011)	
PA × High Structured Mgmt.			-0.031*** (0.012)	
Structured Management Index				
PA x Structured Management Index				-0.001 (0.028)
Joint Tests (linear combination)	0.022*** (0.008)	0.014* (0.008)	-0.015 (0.011)	
Other Inputs and Controls		Y		
Industry x Year Fixed Effects		Y		
R-Squared	0.934	0.934	0.934	0.934
Number of Observations		~51,000		

Notes: OLS estimates controlling for industry (6-digit NAICS)-year indicators using the pooled 2010 & 2015 sample. The dependent variable is output measured by logged sales. Unreported production inputs include: Log L, Log K, and logged material and energy expenditures; other unreported controls include age, multi-unit and headquarters indicators, and production process design. **High Monitoring** indicates plants that track 10 or more of KPIs. **DDD** indicates high availability and usage of data in decision making. **High Structured Management** is an indicator of top-quartile structured management practices based on the **Structured Management**. Joint Tests report the calculated combined coefficients for the use of adoption of predictive analytics with the presence of a given potential workplace complements (using **lincom** in Stata 16). Robust standard errors clustered at the firm level. * p<0.10, ** p<0.05, *** p<0.01.

Table 5. Correlates of Predictive Analytics by Use Case

	(1)	(2)	(3)	(4)
Dependent Variables	**Any Analytics**	**Supply Chain Analytics**	**Demand Analytics**	**Product Design Analytics**
High Monitoring	0.0542***	0.0610***	0.0578***	0.0578***
	(0.0049)	(0.0052)	(0.0051)	(0.0056)
High IT_K	0.0034***	0.0046***	0.0039***	0.0043***
	(0.0012)	(0.0012)	(0.0012)	(0.0014)
Employee Education	0.1281***	0.1495***	0.1393***	0.1860***
	(0.0197)	(0.0198)	(0.0202)	(0.0210)
Mandated Data Collection	0.0189***	0.0260***	0.0222***	0.0259***
	(0.0053)	(0.0055)	(0.0055)	(0.0062)
Log L	0.0256***	0.0304***	0.0261***	0.0311***
	(0.0033)	(0.0035)	(0.0035)	(0.0038)
Age	-0.0009***	-0.0009***	-0.0008***	-0.0010***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
HQ	-0.0245***	-0.0201***	-0.0124*	0.0153**
	(0.0061)	(0.0063)	(0.0063)	(0.0067)
	(0.0202)	(0.0206)	(0.0205)	(0.0208)
Industry x Year FX			Y	
Adjusted R-Squared	0.1428	0.1548	0.1365	0.1179
Number of Observations		51,000		

Notes: Estimates based on linear probability models controlling for industry (6-digit NAICS) and year fixed effects using the baseline sample. The dependent variables across columns 1-4 are the adoption of predictive analytics, and its use cases in the SCM, DF, and PD respectively. Log IT Capital Stock is accumulated and depreciated stock of IT capital expenditure at the plant in log terms. Employee Education is the percentage of employees with a bachelor's degree. Tracking and Monitoring is an indicator for plants that track 10 or more of KPIs based on question 2 in the 2015 MOPS survey. Mandated Data Collection is an indicator for plants that data collection that is mandated by government regulation or agencies. The unreported controls include indicators for the manager-to-non-manager employee ratio, the total number of employees, non-IT capital stock in log terms, continuous flow production design, and multi-unit status Robust standard errors are clustered at the firm level. * p<0.10, ** p<0.05, *** p<0.01. Results are robust to using binary estimation models (e.g., probit in Stata 15).

Table 6. Productivity of Predictive Analytics by Use Case (Indicators)

Models	(1) SCM	(2) Demand Forecasting	(3) Product Development
Dependent Variables		Log Sales	
Supply Chain Analytics (SCM)	0.0116** (0.0047)		
Demand Forecasting Analytics (DF)		0.0086* (0.0048)	
Product Design Analytics (PD)			-0.0012 (0.0045)
Log IT Capital Stock	0.0227*** (0.0014)	0.0227*** (0.0014)	0.0227*** (0.0014)
Employee Education	0.2070*** (0.0177)	0.2077*** (0.0177)	0.2094*** (0.0177)
Tracking and Monitoring			
High Data-focused Practices			
High Structured Mgmt.	0.0284*** (0.0054)	0.0288*** (0.0054)	0.0297*** (0.0055)
Other inputs		Y	
Industry x Year Fixed Effects		Y	
R-Squared	0.9327	0.9327	0.9327
Number of Observations		51,000	

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

Table 7. Productivity of Predictive Analytics by Use Cases (Categories)

Models	**(1)** **SCM**	**(2)** **DF**	**(3)** **PD**
Dependent Variables		**Log Sales**	
Supply Chain Analytics (SCM)	0.0105*** (0.0023)		
Demand Forecasting Analytics (DF)		0.0089*** (0.0023)	
Product Design Analytics (PD)			0.0022 (0.0028)
Log IT Capital Stock	0.0227*** (0.0014)	0.0227*** (0.0014)	0.0227*** (0.0014)
Employee Education	0.2063*** (0.0177)	0.2067*** (0.0177)	0.2085*** (0.0177)
Tracking and Monitoring			
High Data-focused Practices			
High Structured Mgmt.			
Other inputs		Y	
Industry x Year Fixed Effects		Y	
R-Squared	0.9327	0.9327	0.9327
Number of Observations		51,000	

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

Table 8. IV Estimation of Productivity of Predictive Analytics by Use Case

	(1) Supply Chain	(2) Demand	(3) Product Design
Dependent Variable		Log Sales	
Supply Chain Analytics	0.0659*** (0.0207)		
Demand Forecasting Analytics		0.0659*** (0. 0207)	
Product Design Analytics			0.0867*** (0.0274)
Log IT Capital Stock	0.0225*** (0.0014)	0.0225*** (0.0014)	0.0227*** (0.0014)
Employee Education	0.1911*** (0.0182)	0.1912*** (0.0182)	0.1846*** (0.0190)
Tracking and Monitoring			
High Data-focused Practices			
High Structured Mgmt.			
First Stage			
Mandated Data Collection	0.2485*** (0.0148)	0.2485*** (0.0147)	0.1890*** (0.0121)
Weak identification test	821.9	819.1	646.7
Under identification test	227.6	231.9	208.2
Other inputs		Y	
Industry x Year Fixed Effects		Y	
R-Squared	0.933	0.933	0.933
Number of Observations		~51,000	

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